CONF-9004365--Summ.

PUB-701

IBL-PUB--701
DE92 041207

MATHEMATICS & BIOLOGY

The Interface

Challenges and Opportunities

June 1992

Workshop Organizer and Editor:
Simon A. Levin, Cornell University

Editorial Assistant:
Colleen S. Martin, Cornell University

NSF Special Projects:
Peter Arzberger

Contributions from the participants of working groups edited by:
John Guckenheimer, Cornell University
Daniel Kersten, University of Minnesota
David T. Kingsbury, George Washington University
Marc Mangel, University of California, Davis
Michael Reed, Duke University
Wendy Silk, University of California, Davis

MASTER

This work was supported by a grant from the National Science Foundation (DMS 89-10353) to Cornell University and by the Associate Director for Health and Environmental Research, Office of Energy Research, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

CONTENTS

Preface v

Executive Summary 1

1 The Impact of Biology on Mathematics 7

1.1 Accomplishments of the Past 8

1.1.1 Statistics and Stochastic Processes 8

1.1.2 Dynamical Systems Theory 11

1.1.3 Nonlinear Partial Differential and Functional Equations 12

1.1.4 Classical Analysis 14

1.1.5 Topology and Geometry 15

1.2 Grand Challenges 15

2 The Impact of Mathematics on Cellular and Molecular Biology 19

2.1 Accomplishments of the Past 20

2.1.1 DNA Structure 20

2.1.2 Macromolecular Sequences 21

2.1.3 Genetic Mapping 22

2.1.4 Cell Motility 22

2.1.5 Structural Biology 22

2.2 Grand Challenges 24

2.2.1 Structural Analysis of Macromolecules 24

2.2.2 Molecular Dynamics Simulation 24

2.2.3 Drug Design 25

2.2.4 Nucleic Acid Sequence and Structural Analyses of Nucleic Acids 26

2.2.5 Structural Analysis of Cells 27

3 The Impact of Mathematics on Organismal Biology 29

3.1 Accomplishments of the Past 29

3.2 Grand Challenges 30

3.2.1 Complex Hierarchical Biological Systems 31
Neuroscience 31
Immunology 34
Genomic regulatory networks 35
Developmental biology 37
3.2.2 Dynamic Aspects of Structure-Function Relationships 39
4 The Impact of Mathematics on Ecology and Evolutionary Biology 41
4.1 Accomplishments of the Past 42
4.1.1 The Synthesis of Population Genetics and Evolutionary Biology 43
4.1.2 Autecology 43
4.1.3 Population Biology 44
4.1.4 Epidemiology of Infectious Diseases 45
4.1.5 Fisheries Management 46
4.1.6 Community and Ecosystem Processes 46
4.2 Grand Challenges 48
4.2.1 Global Change 48
4.2.2 Molecular Evolution 48
4.2.3 The Problem of Scale 49
5 Modes and Levels of Support 53
5.1 Research Support 53
5.2 Infrastructure 53
5.3 Training 54
5.3.1 Precollege and Undergraduate Education 54
5.3.2 Graduate and Postdoctoral Training 55
5.3.3 Senior Established Investigators 55
5.4 Human Resources 55
References 57
Appendix 1 Steering Committee 77
Appendix 2 Workshop Attendees 79
Appendix 3 Research Opportunities in Computational Biology 83
Appendix 4 Training Computational and Mathematical Biologists 87

PREFACE

THE interface between mathematics and biology has long been a rich area of research, with mutual benefit to each supporting discipline. Traditional areas of investigation, such as population genetics, ecology, neurobiology, and 3-D reconstructions, have flourished, despite a rather meager environment for the funding of such work. In the past twenty years, the kind and scope of such interactions between mathematicians and biologists have changed dramatically, reaching out to encompass areas of both biology and mathematics that previously had not benefited. At the same time, with the closer integration of theory and experiment, and the increased reliance on high-speed computation, the costs of such research grew, though not the opportunities for funding. The perception became reinforced, both within the research community and at funding agencies, that although these interactions were expanding, they were not doing so at the rate necessary to meet the opportunities and needs.

To help foster a broader understanding of this interface, and to provide an analysis of the most promising and productive areas for expanded activity, the National Science Foundation sponsored a workshop to explore the current and future trends at the interface between mathematics and biology. The workshop, which was held in Washington, DC, between April 28 and May 3, 1990, drew together a broadly based group of researchers to synthesize conclusions from a group of working papers and extended discussions. The result is the report presented here, which we hope will provide a guide and stimulus to research in mathematical and computational biology for at least the next decade. The report identifies a number of grand challenges, representing a broad consensus among the participants.

The report documents the participants' enthusiastic conclusion that mathematical and computational approaches are essential to the future of biology and that biological applications will continue to contribute to the vitality of mathematics, as they have since the days of Vito Volterra. The goal of this workshop report is to share with the scientific community our convictions about the promise of this activity and further to inform the broader community and relevant institutions about the potential for exciting growth and productivity in this interdisciplinary field and the need for nurturing it. As with other interdisciplinary efforts, it must rest on strong disciplinary foundations, but not be constrained by the narrow customs and standards of any particular discipline.

We thank the steering committee and the participants for helping to make the meeting such a success. The plans and format were developed together with this committee and

implemented with the invaluable assistance of Colleen Martin at Cornell University and Peter Arzberger and John Wooley of the National Science Foundation. In addition to the attendees acknowledged in Appendix 1, three individuals at the National Science Foundation deserve special thanks for their contributions to the development and refinement of the report: Deborah Lockhart, DeLill Nasser, and Judith Sunley. Each contributed time, attention, and constructive additions to the project and to the report. Sylvia Spengler and the Technical Information Department at Lawrence Berkeley Laboratory sheperded the manuscript to publication. The workshop and report were supported by a grant from the National Science Foundation (DMS 89-10353) to Cornell University (Simon Levin, Principal Investigator). Publication of the report was supported by the Associate Director for Health and Environmental Research, Office of Energy Research, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Simon Levin, Cornell University and Princeton University
David Kingsbury, George Washington University

June 1992

EXECUTIVE SUMMARY

MOTZ (1987) divided the history of science into three broad time periods: the era before Galileo and Newton, the period 1600–1900, and the modern period. "Why," he asked, "did [the Greeks] accomplish so little" in the first period, given their exceptional intellectual capabilities? It was because "theirs was not a systematic study of the nature of things in which experiment and theory went hand in hand but a series of unrelated speculations that stemmed from no basic principles and were never tested." For Motz, modern science began with Galileo and Newton: "Newton's contributions to science and mathematics were not independent of each other — they went hand in hand, for his scientific pursuits forced him to invent the mathematical tools that enabled him to solve the problems presented by the physics."

This is the stage in which biology finds itself today, poised for the phase transition that comes with the total integration of mathematical and empirical approaches to a subject. Many branches of biology are virtually devoid of mathematical theory, and some must remain so for years to come. In these, anecdotal information accumulates, awaiting the integration and insights that come from mathematical abstraction. In other areas, theoretical developments have run far ahead of the capability of empiricists to test ideas, spinning beautiful mathematical webs that capture few biological truths. This report eschews such areas, and instead focuses on those where the separate threads are being woven together to create brilliant tapestries that enrich both biology and mathematics.

The interface between mathematics and biology presents challenges and opportunities for both mathematicians and biologists. Unique opportunities for research have surfaced within the last ten to twenty years, both because of the explosion of biological data with the advent of new technologies and because of the availability of advanced and powerful computers that can organize the plethora of data. For biology, the possibilities range from the level of the cell and molecule to the biosphere. For mathematics, the potential is great in traditional applied areas such as statistics and differential equations, as well as in such nontraditional areas as knot theory.

This report explores some of the opportunities at the interface between biology and mathematics. To mathematicians, the report argues that the stimulation of biological application will enrich the discipline of mathematics for decades or more, as have applications from the physical sciences in the past. To biologists, the report underscores the power of mathematical approaches to provide insights available in no other way. To both communities, the report demonstrates the ferment and excitement of a rapidly evolving field.

With the advent of new types and amounts of data and with new technologies, new fields of research have appeared and existing fields have been changed beyond recognition. Over 7,000,000 nucleotides of DNA per year can be sequenced and, at least until now, such sequencing has been done around and through regions that the investigators have judged to be of biological interest. Thus, sequence comparisons often provide clues to biological function. The secondary and tertiary structure of both DNA and RNA can be analyzed, and such analysis often is conducted with the close collaboration of mathematicians. At the cellular level, recombinant technology has made it possible to ask specific questions about cell growth, cell differentiation, and pattern formation, and to interface with old and new mathematical theories. Similar excitement attends the problem of how cells communicate with each other and with their environment; the dynamics of cells, channels, and neural networks; and the behaviors of populations and aggregations of cells and organisms.

The ways in which whole fields of research are approached have changed. For example, whereas population genetics and evolutionary biology were fields historically concerned largely with inferring process from pattern, the explosion of knowledge at the cellular and molecular levels has encouraged development of a complementary approach in which one begins from knowledge of processes at the micro level. DNA sequence data make possible a different kind of analysis of patterns and processes in evolution at higher levels of organization than was possible previously from fossil data alone. Mathematical approaches allow the use of genetic data to analyze multilocus traits, which are so important, for example, to plant breeding, and have made possible a much more quantitative approach to such issues. And perhaps the greatest challenge for computational and mathematical biology will come in dealing with the problems of global change, biological diversity, and sustainable development, which will require the integration of enormous data sets across disparate scales of space, time, and organization.

As this data explosion is taking place, newer, faster, more powerful machines have become available in the form of both supercomputing centers and networked work stations. In many instances, dedicated hardware has outperformed the supercomputers and is inexpensive enough to be afforded by many scientists. One such example is a chip for sequence comparisons; other examples come from neural networks. The development of national and international networks allows immediate access to data, to software, to ideas, and to supercomputers. These changes in computation have enabled molecular geneticists to store their DNA sequence data, to search for sequence matches, and to do multiple sequence alignments. Developmental biologists can store cell lineage data and model morphogen gradients. Molecular evolutionists can reconstruct larger phylogenies. Ecologists can endeavor to relate global-level processes controlling climate and the distribution of greenhouse gases to biogenic and other mechanisms at the cell and leaf level. In all of these examples, mathematics and algorithm development are intrinsic to success.

The interaction between biology and mathematics has been a rich area of research for more than a century. Statistics and the study of stochastic processes had their origins in biological questions. Galton invented the method of correlation, motivated by questions in evolutionary biology. Fisher's work in agriculture led to the analysis of variance. The attempt to model the success (survival) over many generations of a family name led to the development of the subject of branching processes; more recently, the compilation of DNA sequence data led to Kingman's coalescence model and Ewens' sampling formula. In the area of classical applied mathematics, biological applications have stimulated the study of ordinary and partial differential equations fundamentally, especially regarding problems in chaos, pattern formation, and bifurcation theory.

Perhaps more fundamentally, mathematical approaches have long been central to biology. Before capillaries were discovered, Harvey used a mathematical model to suggest that blood circulates. Mathematical formulations are so basic to the study of ecology and evolutionary biology that they are part of the fundamental training of every biologist. Volterra's early analysis of simple models elucidated the mechanisms underlying the fluctuations of natural populations; modern work on spatial pattern is proving critical to conservation biology. Mathematical models have played a central role as well in managing the spread of infectious disease, including the development of vaccination criteria and studies of the spread of AIDS. The Luria-Delbruck fluctuation analysis, by a simple but elegant experiment based upon a mathematical concept, established that mutation was independent of selection, and mathematical arguments have been central to the analysis of the recent and potentially revolutionary suggestion that in certain situations bacteria mutate nonrandomly in response to their environment.

In molecular biology, mathematical and algorithmic developments have allowed important insights, for example, recognition of the unexpected homology between an oncogene product and a growth factor that forms the basis of the molecular theory of carcinogenesis. Statistical linkage analysis helped locate the cystic fibrosis gene. An understanding of the topology of DNA has been enhanced greatly by the close cooperation of biologists and mathematicians. Classical analysis has played a central role in image reconstruction. Radon's techniques, first developed in 1917, formed the centerpiece of computerized axial tomography that led to a Nobel prize in 1979.

At the organismal level, numerous triumphs can be cited. Mathematical modeling revealed the cause of ventricular fibrillation. Hodgkin and Huxley theorized that macroscopic current might be generated by molecular pores — ion channels that were later proved to exist. Navier-Stokes equations for flow through small bristled appendages have shown how the geometry permits the appendages of aqueous organisms to function as either paddles or rakes.

The primary purpose for encouraging biologists and mathematicians to work together is to investigate fundamental problems that cannot be approached only by biologists or only by mathematicians. If this effort is successful, future years may produce individuals with biological skills and mathematical insight and facility. Today such individuals are rare; it is clear, however, that a greater percentage of the training of future biologists must be mathematically oriented. Both disciplines can expect to gain by this effort. Mathematics is the "lens through which to view the universe" and serves to identify the important details of the biological data and suggest the next series of experiments. Mathematicians, on the other hand, can be challenged to develop new mathematics in order to perform this function.

Flexibility is essential as the funding agencies respond to the needs at this interface. Cross-disciplinary teams of researchers should be encouraged and appropriate methods for review of proposals should be developed. Methods should also be devised for selecting and training individuals at an early stage of their development, at the interface of these disciplines. And finally, meetings and workshops should be supported to explore as yet unthought of ways in which the two disciplines can serve to amplify each other.

GRAND CHALLENGES

Genomics

Attention to the human genome project and its great potential often obscures the fact that theoretical work is essential to efforts aimed at sequencing and mapping all genomes, human and nonhuman, animal and plant. Without the mathematical and statistical underpinnings and computational advances, such efforts will be severely limited; with these methods, we are poised to make dramatic advances. Intraspecific and interspecific comparative analyses of the genomes of diverse organisms can aid in identifying genes and determining their function and also increase our understanding of the natural world.

Global Change

No problem is more compelling, in terms of its importance to life as we know it, than that of global warming. Current estimates are that changes in the concentration of greenhouse gases are occurring at a rate far more rapid than anything we have experienced in the geological record, changes that could lead to equally rapid changes in climate. Furthermore, increases in pollutants of various kinds and depletion of our resource base make the analysis of these changes and their effects upon all life-forms of prime importance. We must improve our methods to describe, to predict, and to identify causes. In all of these areas, a fundamental theoretical problem involves the relationships among processes at very different spatial, temporal, and organizational scales. Closely related problems of surpassing importance are those associated with biodiversity and sustainable development.

Molecular Evolution

The understanding of the evolution of all life forms is critically dependent on our ability to analyze the historical record and to reconstruct phylogenetic relationships among species. The field of evolutionary biology currently offers few methods for this reconstruction, and only one method provides a measure of uncertainty in the final tree. The difficulty of reconstruction grows exponentially with the number of initial data points, and efforts at resolution pose challenging mathematical and computational problems. Computational and algorithmic advances can immeasurably speed up progress in this area.

Organismal Structure-Function Relationships

The relationship between the structure and function of an organism is a central theme of classical biology. Successes include the analysis of functional morphology of organisms and their parts, such as tree branches, and the analysis of fluid flow through and past organisms. The field of functional morphology is a centerpiece of modern biology, and advances in the subject offer hope not only for understanding the biological world, but also for improving the human condition. Theoretical and computational advances already have been made in analyzing

artificial heart valves. The potential is great for extending these approaches to other human and animal organ systems.

Complex Hierarchical Biological Systems

At every level of organization, biological systems are complex hierarchies in which ensembles of lower-level units become the units in higher-order ensembles. The analysis of complex hierarchical systems therefore represents one of the most important open areas in biology. At both the molecular and the cellular level, the components of biological systems are being revealed by modern experimental methodology. The organization and integration of these details into a functional biological system will require the techniques of the mathematician, as well as the data of the biologist. Problems of this sort are at the core of genetics, neurobiology, developmental biology, and immunology. Similar problems exist in understanding how individuals are organized into populations, and populations into communities.

Structural Biology

Structural biology includes the analysis of the topological and geometric structure of DNA and proteins. It also includes molecular dynamics simulation and drug design. Much basic work remains to be done on the structure and folding of crystalline and hydrated proteins. For many proteins, the structure is dictated by the sequence, so this area is closely related to genomics. Molecules are in continuous motion in nature, but NMR and X-ray crystallography necessarily produce snapshots. Mathematical and computational methods are essential to complement experimental structural biology by adding motion to molecular structures.

OPPORTUNITIES FOR MATHEMATICS

A number of fundamental mathematical issues cut across all of these challenges.

(1) How do we incorporate variation among individual units in nonlinear systems?

(2) How do we treat the interactions among phenomena that occur on a wide range of scales of space, time, and organizational complexity?

(3) What is the relation between pattern and process?

It is in the analysis of these issues that mathematics is most essential and holds the greatest potential. These challenges—aggregation of components to elucidate the behavior of ensembles, integration across scales, and inverse problems—are basic to all sciences, and a variety of techniques exist to deal with them and to begin to solve the biological problems that generate them. However, the uniqueness of biological systems, shaped by evolutionary forces, will pose new difficulties, mandate new perspectives, and lead to the development of new mathematics. The excitement of this area of science is already evident and is sure to grow in the years to come.

To achieve the great potential that is evident in this report, we make a number of specific recommendations. We encourage

- Enhanced support for individual interdisciplinary research at the interface between biology and mathematics
- Support for interdisciplinary collaboration
- Support for graduate and postdoctoral fellowships
- Support for mid-career fellowships and visiting fellowships
- Support for educational developments at the precollege and undergraduate level
- Funding for improved computer facilities, software clearinghouses, and electronic networks
- Development of minicourses
- Programs to encourage and involve underrepresented groups

Mathematical and computational biology are vital, crucial, and rapidly growing subjects that complement and guide empirical work, elucidate mechanisms, and provide model systems for study and manipulation. Indeed, such model systems can, in some circumstances, reduce the need for experimentation on living organisms or natural systems when such experimentation presents ethical, fiscal, or logistical difficulties. Mathematical and computational research is comparatively inexpensive, and great dividends can be realized from a relatively small investment of funds. Because the subject lies between traditional disciplinary areas, its support often "falls between the slats" at funding agencies. We urge that specific mechanisms be developed to recognize the unique character of the subject and to provide the support that will foster the development of work that truly can make contributions both to biology and to mathematics.

1

THE IMPACT OF BIOLOGY ON MATHEMATICS

THE application of mathematics to biology has, in turn, had considerable effect on the development of new areas of mathematics. This may seem surprising, because of the different natures of biology and mathematics. Mathematics strongly prizes rigor and precision. Mathematical fact is immutable, and successful mathematical theories have lifetimes of hundreds or thousands of years. By contrast, most of our knowledge of biological systems is recent, and most biological theories evolve rapidly. Nonetheless, the interface between mathematics and biology has initiated and fostered the development of new mathematical areas. This report highlights areas of mathematics that have been influenced greatly by biological thinking in the past, and presages future developments by identifying some areas of biology that will require the development of new mathematical tools.

Of course, many, perhaps most, applications of mathematics in biology will have little effect on core areas of mathematics. Interactions of mathematics and biology can be divided into three categories. The first involves routine application of existing mathematical techniques to biological problems. Such applications influence mathematics only when the importance to biology provokes refinements and further mathematical developments, an inherently slow process. In other cases, however, existing mathematical methods are inadequate, and new mathematics must be developed within conventional frameworks. In the final category of interactions, some fundamental issues in biology appear to require altogether new ways of thinking quantitatively or analytically. In these circumstances, creation of entirely new areas of mathematics may be necessary before it will be possible to grapple successfully with the underlying biological problems. Development of new biological technologies and the rapid accumulation of information and data will prompt the application of classical mathematics, as well as the creation of new mathematics. As in the past, some of these new mathematical theories will be quite rich and will develop lives of their own. The feedback from these applications will help mathematics retain its vitality.

The application of mathematics to biology is not new; neither is evidence of impacts on mathematics. Robert Brown, a botanist, discovered what is now called Brownian motion while watching pollen grains in water. Today, the mathematical description of such motion is central to probability theory. Similarly, catastrophe theory is a branch of mathematics stimulated to a large extent by biological theory. Inspired by Waddington's concept of an epigenetic landscape (Waddington 1957), René Thom generated interest in singularity theory and the bifurcations of

dynamical systems (Thom 1975). Although the style of modeling used by the proponents of catastrophe theory was severely criticized, the beautiful mathematics it spawned has applications that extend far beyond those originally envisaged as part of the theory itself. And perhaps most importantly, the origins of the field of statistics were intimately tied up with biology.

In other areas, the influence has been nearly as great. The theories of dynamical systems and partial differential equations represent areas of mathematics in which numerous fruitful lines of inquiry were prompted by biological questions, and in which such influences continue to be felt. In theoretical fluid mechanics, the dominant classical stream of development was toward an understanding of high-Reynolds-number (almost inviscid) flow and of compressible flows; biology has motivated a great many new developments in viscosity-dominated flows (Purcell 1977). More recently, molecular biology has stimulated advances in analysis and low-dimensional topology and geometry.

In this section, we discuss these examples in more detail, as well as genomic analysis, an area of biology that seems to demand the creation of new mathematical specialties. The section ends with a description of "grand challenges" in biological mathematics, areas that seem to demand novel mathematical and computational approaches.

1.1 ACCOMPLISHMENTS OF THE PAST

1.1.1 Statistics and Stochastic Processes

Statistics is perhaps the most widely used mathematical science. It has achieved its present position as a consequence of an intellectual development begun during the 19th century. "From the doctrine of chances to the calculus of probabilities, from least squares to regression analysis, the advances in scientific logic that took place in statistics before 1900 were to be every bit as influential as those associated with the names of Newton and Darwin" (Stigler 1986, p. 361).

What were the major influences in this development? Porter (1986) introduces his history of statistics in the 19th century as follows: "This book . . . is a study of the mathematical expression of what Ernst Mayr calls 'population thinking' " (Porter 1986, p. 6; see also Mayr 1982, 1988, pp. 350–352). He goes on to say that "the development of statistical thinking was a truly interdisciplinary phenomenon for which mathematics had no priority of position; new ideas and approaches arose as a result of the application of techniques borrowed from one or more disciplines to the very different subject matter of another" (Porter 1986, p. 8). Porter later states, "That the modern field of mathematical statistics developed out of biometry is not wholly fortuitous. The quantitative study of biological inheritance and evolution provided an outstanding context for statistical thinking, and quantitative genetics remains the best example of an area of science whose very theory is built out of the concepts of statistics. The great stimulus for modern statistics came from Galton's invention of the method of correlation, which, significantly, he first conceived not as an abstract technique of numerical analysis, but as a statistical law of heredity" (Porter 1986, p. 270). The profound problems raised by Darwin's insight have led to new fields of mathematical science. Only the surface has been scratched by these developments, and major challenges remain.

Darwin and Galton were cousins, and Darwin's ideas had a great influence on Galton (Porter 1986, p. 133 and p. 281). Likewise, problems in eugenics and plant breeding were the

motivation for R. A. Fisher's statistical work (Box 1978, Fisher 1930). The analysis of variance and the theory of experimental design were developed to interpret and plan plant-breeding experiments at the Experimental Station at Rothamsted, an institution that continues to be a major influence on statistical theory and practice. The benefits to mankind of these and later biometrical developments have been enormous. The "green revolution" in agriculture would have been quite impossible without these tools. Modern medicine and public health practice depend upon carefully designed and interpreted clinical trials and upon sophisticated studies of massive observational data sets.

Problems of the theory of evolution and genetics have had a profound influence upon probability theory as well as statistics. Galton and Watson founded the theory of branching processes in response to a problem of the extinction of human family names (Galton and Watson 1874). Yule, a student of Galton, developed the random process called the Yule process in response to a paper by Willis on the evolution of genera (Yule 1924). The same ideas appeared earlier in McKendrick (1914) and later in Furry (1937). McKendrick (1926) and Kermack and McKendrick (1927) developed their nonlinear birth and death process in response to problems in the theory of epidemics.

The influence of biology on probability theory and statistics has been equally strong in later years of this century. Feller's celebrated work on stochastic processes originated in the Volterra theory of competition and continued in response to problems in population genetics (Feller 1939, 1951; also see Kolmogorov 1959). Neyman et al. (1956) developed stochastic models in order to interpret experiments by Park on flour beetles. In these experiments, two species of beetles were pitted in competition. To Park's surprise, the outcome of a given experiment could not be predicted; but in a long series of experiments, the statistical distribution of outcomes was predictable. The flour beetle connection is still very strong (see Costantino and Desharnais 1991). The early volumes of the Berkeley Symposia contain many more examples of biological inspiration of mathematical theory (Neyman 1945 and subsequent).

Many current and future challenges for statistics and probability that are motivated by questions in molecular biology, genetics, and molecular evolution will require new techniques and theories. One such set of challenges involves the use of DNA sequence data to reconstruct phylogenetic trees, analyze genetically complex traits, and study other problems. As more and more DNA sequence data are accumulated, patterns arise, and exploratory data analysis techniques need to be developed to look through the wealth of data for these patterns. The ordering and frequency of the four nucleotides is not random (even in noncoding regions). To compare two sequences of DNA or protein (or to compare a given sequence with a databank) and to look for matches or similarities (sequence alignment) required the creation of new algorithms. New methods are needed to find regions of similarity and to assess the significance of similarities detected. Comparisons can answer both evolutionary and functional questions. Are sequences descended from a common ancestral sequence? Do they serve similar functions? One problem has been to calculate the probability of a long matching region between two DNA sequences, given that some level of matching occurs if there are overlapping regions. Strong new limit laws give rates for the longest likely matching sequences between different sequences (given mismatches) as the sequence lengths increase. Detailed distributional behavior has been obtained using the Chen-Stein method of approximation by a Poisson random variable. These new distributional results are now used as a basis for statistical tests. Arratia et al. (1990) offer a snapshot of current mathematical work on these questions.

Relevant statistical questions include the calculation of Markov-type probabilities and likelihoods over directed graphs; maximum likelihood estimation for multinomials with highly nonregular parameter spaces involving large numbers of nuisance parameters; model selection from among large numbers of hypotheses of the same dimension; and selection among small numbers of non-nested hypotheses of different dimension.

These problems would be hopelessly intractable were it not for recent and anticipated advances in computational statistics. With computing power now available, we can quickly narrow our search for promising algorithms and test their effectiveness. Other challenges involving DNA sequence data include searches of two of more pieces of data for (longest) matching subsequences. For these, new distributional results are required.

Another area of mathematical research that will be stimulated by biology is the probabilistic theory of discrete and dynamic structures. While scattered beginnings of this field have been made over the last three decades, the major developments are yet to come. Illustrative developments in the field include random graphs and random directed graphs, interacting particle systems, stochastic cellular automata, products of random matrices, and nonlinear dynamical systems with random coefficients. For example, Erdos and Renyi (1960) created the field of random graphs to model apparently random connections in neural tissue. They discovered numerous examples of "phase transitions," and many more have been discovered since (see Bollobás 1985).

Advances in computing power have revolutionized measurement techniques, which generate an abundance of biological data and a need for concomitant advances in quantitative methods of analysis. The interfaces among experimentation, mathematics, and computations are manifested at every stage of scientific investigation. A biological investigation often results in a proposal for a class of mathematical models. Such models may provide insight into the molecular processes (which need not be experimentally observable) and may also suggest new experiments.

For instance, counting process models have been developed for studying patterns of arrivals and interactions of nerve impulses from different neurons (Brillinger 1988, Tuckwell 1988). Markov processes have been used extensively in analyzing membrane channel data, in studying the kinetic behavior of ionic channels, and in understanding cell survivability and DNA damage caused by ionizing radiation (Neyman and Puri 1981; Yang and Swenberg 1991). A novel aspect of some of these studies is that both transition mechanisms and state spaces must be inferred from data. In fact, the analysis of single-channel data by Markovian models has led to new interpretations of some neural parameters, different from those offered by the Hodgkin-Huxley model (see Aldrich et al. 1983). Stochastic differential equation models have been used for investigating the depolarization of the membrane potential of spatially distributed neurons (Kallianpur and Wolpert 1987). The stochastic nature of the measurements has resulted in new developments in stochastic integration and differentiation. Neurobiology has stimulated the growth of this field.

For the corresponding problems of statistical inference, new methods and corresponding algorithms are needed for model validation and the estimation of parameters. It can happen that models appear to fit according to currently used criteria even though they have not caught the essence of the biological phenomena of interest. A relevant question to ask is, how far off can the model be and still 'fit'? In other words, subject to fitting the data, the model should be biologically interpretable. In this area of research, collaborations between neurobiologists and statisticians have been particularly successful, as evidenced by, for example, joint work on spike train pattern recognition (Brillinger and Segundo 1979), estimation of single-channel kinetic

parameters (Milne et al. 1989), temporal clustering of channels (Ball and Samson 1987), estimation of open dwell time in multichannel experiments (Yang and Swenberg, in press), and identification of kinetic states (Fredkin and Rice 1986).

Construction of confidence intervals for parameters, identifiability of models, estimation of kinetic parameters from the partially recorded current data, design of experiments to collect multivariate data as opposed to univariate data, and integration of the experimental results collected at micro and macro levels by stochastic modeling are among the important research problems. Collaborations between biologists and statisticians are essential in developing statistical modeling methods for research in biology.

A recurrent problem has been the lag between advanced theory and current practice. Many biologists now have at least an introductory course in statistics, but their understanding is generally insufficient to perform well-designed experiments or effective analysis of their data. Expert systems can help biologists make better use of their experimental resources and the data that result. The production of such expert systems offers both a theoretical challenge and the prospect of a widespread and lasting effect on the statistical practice of biologists.

1.1.2 Dynamical Systems Theory

The theory of dynamical systems has been stimulated by biological questions. For example, iterations of a single nonlinear function, based on a simple population model, capture the dynamics of an isolated population, subject to influences that regulate the population numbers exclusively through the population size. More explicitly, the population size at generation $n + 1$ is assumed to be a given nonlinear function of the population size at generation n. Models of this type were introduced in population studies a long time ago. Isolated studies of the iteration of functions were conducted near the beginning of the 20th century. Some of this work, notably that by Julia (1918) and Fatou (1919, 19920a,b) and then by Myrberg (1963) and Sarkovskii (1964), pointed to a rich mathematical structure. However, it was only in the 1970s that a widespread appreciation emerged for the depth and beauty of the mathematical phenomena involved in these mathematical problems. Population biologists, especially May, played a role in stimulating this appreciation. One can only speculate as to whether the theory of these iterations would have "taken off" as it did without this influence from population biology, but clearly, the motivation from population biology was an important part of the chain of historical events that led to very significant scientific and mathematical discoveries.

The study of simple population models provides a classic example of mutual stimulation of mathematics and biology, with resulting benefits to both. The interlocking efforts of mathematicians, biologists, and physicists formed a network of positive feedbacks that moved the subject to new levels of sophistication. Their investigations showed clearly the existence of universal sequences of bifurcations in iterations of one-dimensional maps. Libchaber provided striking confirmation of Feigenbaum's discoveries about period-doubling bifurcations in fluid convection experiments.

Substantive mathematics has grown from these beginnings. Among other developments, Lanford extended Feigenbaum's arguments with numerical analysis to give a beautiful example of a rigorous "computer" proof. The study of interval maps was generalized to encompass maps of the circle. This work on circle maps has been used by Glass, Winfree, and others for describing the phase responses of biological oscillators, particularly in cardiology. The work on

maps of the interval has also been the starting point for the work of Carleson and Benedicks on the Henon map, a two-dimensional map that is a prototype for chaotic behavior.

The mathematics described above can be evaluated both for its impact within mathematics and for its "real world" significance. On both counts, the subject appears to have lasting value. On the one hand, a rich structure is displayed by a substantial set of mathematical objects. Overstating the case slightly, one can say that all families of one-dimensional maps display the same dynamical behavior. Understanding analytically and geometrically why this is true continues to be a challenging and interesting area of research, with fascinating connections to the world of "complex dynamics" and quasi-conformal mappings. On the other hand, the theory has laid bare what appear to be the fundamental mechanisms for the creation of chaotic behavior in physical systems and for universal patterns of bifurcations that are displayed by systems otherwise unrelated to one another. Within mathematics, this sequence of events has been a success story, one in which interest in biological models provided a significant stimulus to mathematics. Feedback from the resulting mathematics to the biological sciences continues. Good mathematics often finds application in unsuspected ways.

Beyond the work involving iterations of one-dimensional mappings, many other points of contact have been established between the biological sciences and dynamical systems theory. Life itself is a dynamical process, and dynamical systems models are ubiquitous in biology. For example, the model of Hodgkin and Huxley for nerve impulses, described later in this document, is a dynamical system.

One seldom can measure all the parameter values entering dynamical models of biological phenomena, and the models themselves usually represent the behavior of aggregate quantities. Therefore, one would like to classify the possible dynamical behaviors arising from models. This challenging problem remains an important area of contact between mathematics and biology. Today, great interest is shown in the dynamics of networks of biological neurons and the dynamics of systems of coupled oscillators. In both situations, one seeks to explain details of the dynamical behavior and understand how collective behavior emerges from the coupling of individual elements. As the number of elements increases, singular perturbation methods and continuum models blend with dynamical systems theory.

Computation has played an important role in dynamical systems theory, especially in its application to specific problems. Applications in biology require the development of effective computational methods for the analysis of dynamical systems and their bifurcations. New mathematics is emerging from work in this direction.

1.1.3 Nonlinear Partial Differential and Functional Equations

Nonlinear partial differential and functional equations traditionally have been applied in the physical sciences. Several examples highlight the seminal impact of biological ideas on mathematical research in this area. Below, we focus on problems from demography, developmental biology, physiology, and population biology.

Demographic methods have been applied to the study of human and nonhuman populations for centuries. These methods, which form the basis both for population projections and for understanding population consequences of life history phenomena, have had a strong impact in mathematical theory. A snapshot of the impact of demography is provided by the history of ergodic theorems. The renewal equation, a convolution integral equation that provided the first

dynamical model for an age-dependent population, has roots in the work of Euler, Bortkiewicz, and Lotka (see Samuelson 1976). Sharpe and Lotka (1911) argued that most solutions to their renewal equation could be represented in a Fourier-type expansion. Their argument was not accepted mathematically until Feller (1941) gave a rigorous proof for asymptotic behavior under appropriate conditions. As yet, the problem of stating conditions under which the renewal equation admits a Fourier-type expansion remains partly open (see Inaba 1988).

The later demographic models of McKendrick (1926) and Gurtin and MacCamy (1974), and the epidemiological models of Kermack and McKendrick (1927) and Hoppensteadt (1974) have generated similar mathematical challenges in the realm of functional differential equations (see Jagers 1975, Cohen 1979, Metz and Diekmann 1986, Castillo-Chavez 1989). The rich interaction among demography, epidemiology, ecology, and evolutionary biology continues to be a source of new mathematical problems related to the existence, uniqueness, and characterization of the solution of nonlinear functional equations. These problems will continue to be a fertile area of mathematical research since current mathematical and numerical approaches are only partially adequate for addressing these issues.

The theory of diffusion, which describes the behavior of a population of randomly moving particles or molecules, exemplifies an area traditionally viewed within the context of chemistry or physics. However, the mathematics of nonlinear diffusion equations has received much of its impetus from biology. R. A. Fisher's (1937) interest in the problem of the spread of advantageous genes in a population stimulated his consideration of an equation that incorporates diffusion augmented by a simple ("logistic") nonlinear growth term. It was treated simultaneously by Kolmogorov et al. (1937), who proved the existence of a stable traveling wave of fixed velocity representing a wave of advance of the advantageous gene. This simple nonlinear reaction-diffusion equation was also studied by Skellam and others as a model for spatial dispersal of a population. Reaction diffusion equations were investigated by Turing (1952) to understand pattern formation and morphogenesis, fundamental problems of developmental biology. The idea that uneven distributions of chemical substances could guide cellular differentiation had preceded Turing by nearly half a century, but how such "chemical prepatterns" could arise naturally was unclear. Turing demonstrated that simple molecular diffusion, coupled with appropriate bi-molecular interactions, could spontaneously give rise to such prepatterns, because a spatially uniform solution of certain coupled parabolic equations bifurcates into a nonuniform state as certain parameters are varied.

Following the interest in Turing and Fisher equations, the study of nonlinear reaction diffusion equations has undergone a rich mathematical development. The study of standing- and traveling-wave solutions, and of characterizing the bifurcations and dynamical behavior of such equations, has spawned new and advanced mathematical techniques. Recent attention has been focused on two- and three-dimensional geometry, including target patterns, spiral, and scroll wave geometry and the like. Connections with the chemical reaction of Belousov and Zhabotinskii (see, for example, Murray 1989), with pathologies of cardiac physiology, and with uneven ("patchy") distribution of organisms in space provide new impetus and motivation for further interest in this field.

Although the equations and mathematical knowledge arising from demography and epidemiology have already found applications (e.g., in evolutionary ecology, conservation biology, and epidemiology), a strong need exists for new mathematics to address pressing new biologically motivated questions. For example, at the interface of social dynamics and epidemiology, new models describe "social mixing" (e.g., formation and dissolution of pairs)

and its role in disease dynamics. The models are novel systems of hyperbolic partial differential equations. These models may affect practical issues of public health and broader biological issues. Since current techniques are as yet in their infancy, it is likely that new mathematics will develop from these efforts.

While reaction-diffusion equations are mathematically simpler than the Navier-Stokes equations, they have presented opportunities for fertile biological and mathematical research. General techniques for studying the finite-dimensional behavior of evolution equations have found some of their first applications in reaction diffusion equations. But current theories of developmental biology provide new models that are at present barely tractable under limited circumstances. Examples include the mechanochemical models of Murray, Oster, and Odell (see Murray and Oster 1984), which incorporate traction forces exerted by cells on each other, and partial integro-differential equations that depict direct responses of cells to one another, as, for example, in neural networks. Further understanding of these models needs new mathematics.

1.1.4 Classical Analysis

Numerous examples exist of the mutual interactions of biology and classical analysis. One of the most important is in the area of digital radiography. Improved technologies for imaging biological objects have revolutionized medicine. These technologies include computerized axial tomography (CT), magnetic resonance imaging (MRI — also termed nuclear magnetic resonance imaging, or NMR), and emission tomography (PET and SPECT). Each technique has mathematical aspects to its implementation and is expected to lead to many additional problems. Regardless of technique, the wealth of digitized radiologic data has led to problems concerning their storage and transmission; solutions to these problems of data compression also require mathematical thinking.

More than 70 years ago, Radon (1917) noted that a finite Borel measure on a Euclidean space can be reconstructed in principle from its projections on one-dimensional subspaces. This was rediscovered independently in other contexts by Cramér and Wold (1936) and others. This piece of theoretical mathematics is at the heart of CT image reconstruction, for which Cormack and Hounsfield received the Nobel Prize in Physiology or Medicine in 1979. The Nobel lecture of Cormack (1980) makes clear the centrality of inversion algorithms to CT. In Hounsfield's lecture (Hounsfield 1980), he contrasts CT and NMR, which also depends on inversion algorithms for its successful application. Important early algorithms for image reconstruction were contributed by Bell Laboratories mathematicians Shepp and Logan (1974). Their work led to mathematics of interest in its own right (Logan and Shepp 1975).

Vardi et al. (1985) are responsible for a fundamental advance in positron emission tomography (PET). With emission tomography in general, a substance such as a sugar that is differentially metabolized by different tissues is tagged with an emitting molecule. In one case (PET), a positron is emitted, and in another (SPECT), a photon; with PET, each positron gives rise to two photons that move in opposite directions. In either case, individual photons are counted as they hit a detector surrounding the object (for example, a human head) being imaged. The object can be modeled as a spatially inhomogeneous Poisson process, and the mathematical task is to reconstruct the intensity function from the counts. The approach of Vardi et al. was to employ an algorithm, the EM algorithm, that was developed by Harvard statisticians Dempster, Laird, and Rubin (1977); earlier basic work on EM-like algorithms was done by the

mathematician Baum (1970) and others (see discussion of the paper by Vardi et al. [1985] for extensive references). A Bayesian approach to reconstruction in emission tomography utilizes Markov random fields that arise in statistical mechanics. Important contributions have been made by Geman and McClure (1985, 1987). Recently, Johnstone and Silverman (1990) have given minimax (in a statistical sense) rates of convergence for PET algorithms. The interface of emission tomography, mathematics, and statistics continues to be a particularly active area of research. It should be noted that PET permits quantitative measurements, in vivo, of local hemodynamics, metabolism, biochemistry, and pharmacokinetics (Fox et al. 1985), and that SPECT is best used for problems of perfusion rather than metabolism.

Data compression, i.e., storing salient aspects of pixel-by-pixel lists of binary integers, is viewed as a problem in coding. It is important to compress, in part to enable more complete medical records to be kept than is possible at present and in part to enable transmitted digital images to be utilized in real time by experts in different venues when baud rates (i.e., information transmission rates) are limited. Here, codes are of two basic types. One is lossless, in which perfect reconstruction of the original image is possible, but which seldom leads to more than a 75 percent reduction in pixel data; this is associated with Huffman, Ziv-Lempel, and other codes. The other basic type is lossy, in which perfect reconstruction is not possible, but for which it is possible to retain virtually all information contained in many images with approximately a 90 percent reduction in pixel data. Tree-structured codes of the latter type have been implemented (Chou et al. 1989).

1.1.5 Topology and Geometry

Additional areas of mathematics have recently developed interactions with biology. Three-dimensional topology and low-dimensional differential geometry are two examples. Theorems about the global topological invariants of curves and ribbons in three-space have been instrumental in studying the structural conformation of closed circular DNA. These mathematical ideas apply to supercoiling in closed DNA, topoisomerases, nucleosome winding, the free energy associated with supercoiling, and binding between proteins and DNA. These applications were carried out by experimentalists, often in collaboration with mathematicians. As collaborative work continues and our knowledge of the role of conformational changes of biological macromolecules grows, the biological problems to be solved become more complicated and the mathematical questions deepen. For example, molecular biology has renewed interest in embedding invariants for graphs (used in studying topoisomers), the study of random knots (used to study solutions of macromolecules), and the tangle calculus (used in the study of the DNA enzyme mechanism).

1.2 GRAND CHALLENGES

Two of the most influential books in the development of biological thought are Plato's *Republic* and Darwin's *Origin of Species*. Plato claimed that all the variation in observable horses, for example, is a mere shadow of an idealized abstract form of pure "horseness," not available to the senses. Plato's notion of idealized forms was the basis of scientific developments for two millennia. For example, Newton's concepts of absolute space and time are idealizations on the model of Plato's horseness. In biology, the Linnaean concept of species is an operational version

of Plato's idea: a Linnaean species is determined by a "type specimen," deposited in a museum somewhere, and all deviations between the type specimen and real members of the species are mere irrelevant accidents. For Gauss, the variation in repeated astronomical measurements led to a theory of "error" in which variation was something to be eliminated. The influence of the Platonic theory of ideal types extended far beyond science to, for example, popular notions of national or racial "types."

Darwin's theory of the origin of species gives a central place to biological variation as a necessary ingredient in explaining speciation. Because different individuals of a species vary in ways that are significant for their survival and reproduction, a given environment will select against some genotypes of a population; those that survive will produce offspring. The survival of some gene combinations and the loss of others cause a population of organisms to evolve; differences among populations may then lead to reproductive isolation and speciation. For Darwin, and for all biologists since then, the origin and consequences of variation among individuals are central to biological observation and theory.

Little more than a century has passed since Darwin's startling conceptual insight. Developments in probability theory and statistics within the last century have made a start toward developing the concepts required to fully understand variation in nature. But the mathematical concepts that will provide an integrated understanding of *nonlinear dynamics in systems with variation between individuals* have yet to be invented and analyzed. What Newton's calculus did for the ideas of Plato has yet to be done for the concepts of Darwin. Many other biological problems could be cited in which the connection between variation and nonlinear dynamics is an essential aspect of understanding the underlying phenomenon.

A second and related grand challenge recurs throughout this report: *the interaction of phenomena that happen on a wide range of scales in space, time, and organizational complexity.* In studying biological systems, one must confront an enormous range of scales. One deals with phenomena that range from molecular processes that happen in small fractions of a second, to evolutionary, ecological, and population processes that occur on geological time scales. Similar ranges exist in spatial scales, from the molecular to the biospheric, and in organizational complexity. We cannot develop the analytical or computational capability to treat this vast range of scales without encapsulating the behavior of smaller scales in models. One consequence of making such approximations is that we lose the detail that imparts confidence in models; yet we must develop ways to suppress detail and proceed to the more aggregated models that are statistically manageable.

Organisms are complex assemblies of macromolecules reacting with each other in complicated networks. Many small parts of the network have important influences upon the proper functioning of the system. Mutations, which change a single nucleotide along a strand of DNA, can affect the gross anatomy of an organism. The details of these subunits, their differences, and their interactions are important at certain levels, and we cannot yet be confident about which details become unimportant as we move to higher levels of organization. The problem may be more difficult than comparable problems in statistical physics, because the differences among subunits are greater. The distinction between these situations is analogous to the difference between assembling a large jigsaw puzzle and an orderly array of identical marbles. The complexity of biological systems is of a different order of magnitude than the problems that have been confronted successfully in mathematics, and mathematical theories are needed to develop insights into our newly accumulated store of biological knowledge.

Computation is essential for investigating mathematical problems arising in biology. The storage and retrieval of the accumulated information is an enormous task. The problems of pattern searching and matching of DNA sequences have been described above. The computer provides the critical capability to explore and study such complex situations. A useful comparison can be drawn between problems of engineering design and the structures found in biology, the products of tinkering rather than design (Jacob 1977). There is a large difference between understanding the fundamental scientific principles of mechanics and designing large buildings or automobiles. The most important aspect of a machine is its function, and design involves far more than drawing the blueprints for its manufacture. Biology confronts us continually with the inverse problem to that of engineering design. We know the basic principles of biochemistry and can laboriously determine biological structures. From these blueprints we want to infer information about biological function. The experimental tools that are available for observing functional aspects of structure are limited by the fragility of life itself. We are left with incredible puzzles to solve with literally billions of pieces and only limited clues about how they fit together. Even the problem of reconstructing the three-dimensional structure of a protein from its amino acid sequence is a major unsolved problem. Our brains are incapable of coping with the wealth of biological data without the assistance of computers. The complexity of biological problems requires that we also apply mathematical and computational approaches, and the benefits of such applications will be shared equally by the disciplines of biology and mathematics.

2

THE IMPACT OF MATHEMATICS ON CELLULAR AND MOLECULAR BIOLOGY

THE application of mathematics to cellular and molecular biology is so pervasive that it often goes unnoticed. The determination of the dynamic properties of cells and enzymes, expressed in the form of enzyme kinetic measurements or receptor-ligand binding are based on mathematical concepts that form the core of quantitative biochemistry. Molecular biology itself can trace its origins to the infusion of physical scientists into biology with the inevitable infusion of mathematical tools. The utility of the core tools of molecular biology was validated through mathematical analysis. Examples include the quantitative estimates of viral titers, measurement of recombination and mutation rates, the statistical validation of radioactive decay measurements, and the quantitative measurement of genome size and informational content based on DNA (i.e., base sequence) complexity.

Several of the "classic experiments" in microbial genetics involved mathematical insights into experimental results. For example, the Luria and Delbruck fluctuation analysis, which clearly established that mutation was independent of selection, was a mathematical argument upon which a simple but elegant experimental design was based.

These examples are cited not to document the accomplishments of mathematical biologists but to bring focus to the fact that mathematical tools are intrinsic to biological fields. The discussion that follows focuses more clearly on the more sophisticated development of new mathematical concepts and statistical models to explain the complexity of biological systems. Biological complexity derives from the fact that biological systems are multifactored and dynamic.

Quantitative research in these fields is based upon a wide variety of laboratory techniques, with gel electrophoresis and enzyme-based assays among the most common. Measurements include activity, molecular weight, diameters, and sizes in bases, and with all these an understanding of the accuracy, precision, sources of variation, calibration, etc. In short, the quality of the measurement process is of central significance.

With the greatly increased amount of data being generated by laboratory techniques, and the pressure to move to more automated analysis, it is becoming even more important to understand the statistical aspects of these laboratory procedures. Such statistical work will involve the analysis of routinely collected data, the design and analysis of special studies, the development of

new calibration and analysis techniques, and theoretical studies of the procedures in use, with an emphasis on robustness and the capacity to automate procedures. Furthermore, it will require a familiarity with the biology and the mathematical foundations of the analyses.

While the experimentalist strives to isolate single variables in order to make statistically significant measurements, many systems are not amenable to such single-factor examination. Therefore, mathematically based computational models are essential to meaningful analyses. The goal of the present discussion is to provide a framework in which ongoing research in mathematical cell and molecular biology may be logically placed, and future opportunities can be described. This framework will provide for the analysis of the resource needs for future development and carries implications for current shortfalls. One factor is that undergraduate and graduate training in biology treats mathematics too superficially, especially in light of its role as an underpinning for quantitative research.

2.1 ACCOMPLISHMENTS OF THE PAST

2.1.1 DNA Structure

Differential geometry is the branch of mathematics that applies the methods of differential calculus to study the differential invariants of manifolds. Topology is the mathematical study of shape. It defines and quantizes properties of space that remain invariant under deformation. These two fields have been used extensively to characterize many of the basic physical and chemical properties of DNA. Specific examples of particular note follow.

The recent review of Dickerson (1989) summarizes how geometric concepts of tilt, roll, shear, propeller twist, etc., have been used to describe the secondary structure of DNA (i.e., the actual helical stacking of the bases that forms a linear segment of DNA). In addition, these concepts can be used to describe the interaction of DNA with ligands such as intercalating drugs (Wang et al. 1983).

From the time that closed circular DNA was discovered, it has been clear that such DNA exhibits both physical and chemical properties that differ in fundamental ways from those of related linear (or open circular) DNA. Using differential geometry and topology, both molecular biologists and mathematicians have been able to explain many of the properties of these molecules from two basic characteristics of the linking number: first, that it is invariant under deformations, and second, that it is the sum of the two geometric quantities, twist and writhe (White 1969). Among the major applications are:

- The explanation for and extent of supercoiling in a variety of closed DNAs (Bauer 1978)
- The analysis of the enzymes that change the topology of a DNA chain (Cozzarelli 1980, Wasserman and Cozzarelli 1986)
- The estimation of the extent of winding in nucleosomes (Travers and Klug 1987)
- The determination of the free energy associated with supercoiling (Depew and Wang 1975)
- The quantitative analysis of the binding of proteins and of small ligands to DNA (Wang et al. 1983)

- The determination of the helical repeat of DNA in solution and DNA wrapped on protein surfaces (White et al. 1988)
- The determination of the average structure of supercoiled DNA in solution (Boles et al. 1990

Topology, and in particular knot and link theory of closed space curves, has been used extensively to elucidate additional intertwining of closed DNA caused by catenation of two closed duplexes or knotting of a single duplex. In particular, the recent developments in polynomial invariants for links and knots have been used to describe the structure of DNA and to characterize the action of recombinases (Wasserman and Cozzarelli 1986, White et al. 1987).

2.1.2 Macromolecular Sequences

DNA sequences are collected in the GenBank database, and protein sequences are collected in the Protein Identification Resource (PIR). When a new DNA sequence is determined, GenBank is searched for approximate similarities with the new sequence. Translations of the DNA sequence into the corresponding amino acid sequence are used to search the protein database. Sensitive search methods require time and space proportional to the product of the sequences being compared. Searching GenBank (now more than 40×10^6 bases) with a 5000 bp sequence requires time proportional to 2×10^{11} with traditional search techniques. Lipman and Pearson (1985) have developed techniques that greatly reduce the time needed. Using their techniques, one can screen the databases routinely with new sequences on IBM PCs, for example. These methods rapidly locate diagonals where possible similarities might lie and then perform more sensitive alignments. This family of programs, FASTA, FASTN, etc., are the most widely used sequence analysis programs and have accounted for many important discoveries. An example of the impact of such analysis is the unexpected homology between an oncogene and a growth factor. This discovery became the basis of the molecular theory of carcinogenesis.

More sensitive sequence analysis can be obtained by dynamic programming methods. In part they are used after the diagonals are located in the FASTN and FASTA programs. Here similar sequence elements are aligned with positive scores and dissimilar elements are aligned with negative scores. Complicating the analysis are the insertions and deletions that also receive negative scores. The challenge of the problem is to arrange two sequences into the maximum scoring alignments. Additional difficulty arises from the fact that slightly similar regions of DNA or protein sequences might lie in otherwise unrelated sequences. In spite of the complex nature of the problem, an efficient algorithm (Smith and Waterman 1981) has been devised and is in wide usage.

The problem of sequence comparison creates a related statistical problem of estimating p-values (attained significance levels) for the alignment scores. The set of possible alignment scores from two sequences are dependent random variables since they result from overlapping sequence segments. Motivated by the problems of sequence comparison, investigators have refined and extended the Chen-Stein method (Arratia et al. 1989). This method is a powerful tool for approximating the distribution of sums of dependent indicator random variables by the Poisson distribution. In addition to sequence analysis, this method is being used in regression analysis and random graphs.

2.1.3 Genetic Mapping

Genetic mapping deals with the inheritance of certain "genetic markers" within the pedigrees of families. These markers might be genes, sequences associated with genetic disease, or arbitrary probes determined to be of significance (e.g., restriction fragment length polymorphism [RFLP] probes). The sequence of such markers and probabilistic distances (measured in centimorgans) along the genome can often be determined by hybridizing each family member's genome against the predetermined probes. In essence, the genetic map most likely to produce the observed data is constructed. Only a few years ago, our knowledge of the mathematics involved and the computational complexity of algorithms based on that mathematics allowed us to analyze no more than five or six markers. As our knowledge of approximations to the formulas and likelihood estimation has improved, we have been able to produce software capable of producing maps for 60 markers or more (Lander and Botstein 1986). Progress in this area has been based on mathematical areas such as combinatorics, graph theory, and statistics.

2.1.4 Cell Motility

Cells can move, monitor changes in their environment, and respond by migrating toward more favorable regions. It is a remarkable fact that a bacterial flagellum is driven at its base by a reversible rotary motor powered by a transmembrane proton flux, and analysis of models for this device has been prolific. The study of bacterial chemotaxis (the migration of bacteria in chemical gradients) has been particularly rewarding, in part because organisms such as *Escherichia coli* are readily amenable to genetic and biochemical manipulation, and in part because their behavior is closely tied to the constraints imposed by motion at low Reynolds number and by diffusion (of both the cell and the chemoattractant). Mathematics has helped us learn how a cell moves (Brokaw 1990, Dembo 1989), how it counts molecules in its environment (Berg and Purcell 1977), and how it uses this information (Berg 1988). Mathematics has also given us a way to relate the macroscopic behavior of cell populations to the microscopic behavior of individual cells (Rivero et al. 1989).

Studies of eukaryotic cell motility (and of the motion of intracellular organelles) have been revolutionized by in vitro assays in which motor molecules (myosin, dynein, kinesin) and the polymers along which they move (actin and microtubules) are linked to glass or plastic surfaces. Following the addition of ATP, one can observe, for example, the motion of individual actin filaments over a glass slide bearing only the heads of the myosin molecules. Statistical analysis is playing an important role in determining how such assays can be extended to the study of single motor molecules (Howard et al. 1989).

2.1.5 Structural Biology

Mathematics has made perhaps its most important contribution to cellular and molecular biology in the area of structural biology. This area is at the interface of three disciplines — biology, mathematics, and physics — because its success has involved the use of sophisticated physical methods to investigate the structures of biologically important macromolecules, their assembly into specialized particles and organelles, and even higher levels of organization. A wide array of methods has been employed, but we focus on the two most powerful of these, x-ray

crystallography and nuclear magnetic resonance (NMR) spectroscopy, with a mention of other methods.

Mathematics plays three roles. First, computational methods lie at the heart of these techniques because a large amount of information about local areas or short distances is encrypted in the raw data, and it is a major computational task to deduce a structure. Second, new mathematical methods of analysis are continually being developed to improve ways of determining the structure. Third, increasingly sophisticated computer graphics have been developed in response to the need to display and interpret such structure.

In crystallography, the actual process of data collection has been enhanced by modern methods of detection (e.g., area detectors) and the use of intense synchrotron sources so that data collection per se is rarely rate limiting. Also, the use of modern techniques of recombinant DNA have greatly facilitated the isolation of material for crystallization. The rate-limiting step is often the preparation of isomorphous derivatives. As computational methods improve, fewer and sometimes no derivatives need to be analyzed.

Until the development of 2D NMR in 1978 by Richard Ernst, the use of nuclear magnetic resonance for studying the structure of biological macromolecules was limited by the need to represent too much information in a limited space. With the pioneering development of the ability to represent NMR spectra in two frequency domains, it became possible to resolve the spectra of small proteins and oligonucleotides. A key benefit was that cross peaks, resulting from magnetic interactions of nuclei close to one another, could be measured. Since these cross peaks contained spatial information, there was an immediate movement to determine the structure of these molecules at atomic resolution. The technique has been remarkably effective. The structures of a number of proteins and oligonucleotides have been determined. The use of NMR to determine structures has proved to be an important complement to x-ray crystallography because many biologically important molecules (e.g., zinc fingers by Klevit 1991, Summers 1991, and Lee et al. 1991) have resisted attempts at crystallization; these structures must be studied in solution. The success of this technique has been critically dependent on mathematics, beginning with the theoretical underpinnings developed by Ernst. The determination of structures is dependent on the mathematical technique of distance geometry that calculates all structures consistent with the distance constraints obtained from the NMR experiment. Other methods have included molecular dynamics and, more recently, the use by Altman and Jardetzky (1989) and Altman et al. (1991) of a Kalman filter to sample conformational space. There are, however, significant limitations to 2D NMR for structure determinations. First of all, the resolution obtained from NMR is less than that obtained from the best x-ray structures and is insufficient to see in detail active sites of biologically important molecules. A major mathematical challenge is to obtain such detailed structural information from structures that are basically underdetermined. One important approach is to use the structure to back-calculate the NMR data and by iteration to improve the resolution. A second limitation is that the determination of structures is limited to molecules with a weight of less than about 15,000. Better computational techniques could extend the limit.

One cannot overestimate the importance of solving structures at atomic resolution. It has led directly to an understanding of the replication of DNA and its supercoiling in chromatin; the basis of protein and nucleic acid secondary, tertiary, and quaternary structures; how proteins act as enzymes and antibodies; and how electron transfer is achieved. The medical and commercial implications of advances in structural biology are enormous.

2.2 GRAND CHALLENGES

The grand challenges at the interface between mathematics and computation and cellular and molecular biology relate to two main themes: *genomics*, which is critical, for example, to support efforts to sequence and map the human and other genomes, and *structural biology*, including structural analysis, molecular dynamic simulation, and drug design. These two areas have developed rapidly in the recent past because of the contributions of mathematics and computation, and they will continue to derive particular benefit from an enhanced interaction.

2.2.1 Structural Analysis of Macromolecules

The area of molecular geometry and its interface with visualization has been underrepresented in research to date. This research, which would benefit from the involvement of geometers and would likely contribute to new mathematics, is a major limiting area in structural biology, especially in drug design and protein folding. As noted above, new methods will enhance the use of NMR for the determination of structures. Significant advances aimed at solving the phase problem mathematically are being pursued. Important advances are also being made in the field of computer-aided drug design.

Related to the structure of crystalline and hydrated proteins is the question of how proteins fold. For many proteins, the folded structure and organelle formation (e.g., ribosomes) are dictated by the sequence. Reduction of the folding code has resisted intense efforts, but very recently important new approaches have been developed that have revealed significant new information. For example, two laboratories have shown that relatively short polypeptides can have significant secondary structure. This finding is important because it validates a piecemeal approach to protein folding, where secondary structure can be considered apart from tertiary structure. The second is the minimalist approach of DeGrado et al. (1989), in which model structures with predicted motifs are synthesized by chemical means. Experimental advances such as these, together with the explosive expansion of the available data and the development of more powerful decoding methods, mean that members of families of protein-folding codes will soon be readily identifiable. Once again this area requires mathematical innovation.

Finally, we note that microscopy is undergoing a technical revolution. Two new microscopes, the scanning tunneling and atomic force microscopes, can yield a picture of macromolecules at atomic resolution. Actually, for these computer-age microscopes, the picture is represented via a computer graphics display of digital data stored on optical media. Additionally, computational methods are the heart and soul of electron microscopic tomography. For example, using this technique, one can obtain four-dimensional information on chromatin structure (e.g., Belmont et al. 1989).

It is worth repeating that mathematical biologists are in great demand in the field of structural biology. The theoretical work also is highly important and frequently has immediate payoff.

2.2.2 Molecular Dynamics Simulation

Three-dimensional structures as determined by x-ray crystallography and NMR are static since these techniques derive a single average structure. In nature, molecules are in continuous motion; it is this motion that allows them to function (a static molecule is as functional as a static

automobile). Mathematical and computational methods have been able to complement experimental structural biology by adding the motion to molecular structure. These techniques have been able to bring molecules to life in a most realistic manner, reproducing experimental data representing a wide range of structural, energetic, and kinetic properties. Systems studied have extended from pure liquid water, through small solutes in water, to entire proteins and segments of DNA in solution.

The methods used for these calculations provide a glimpse of how simulation can be used generally in biology. Starting with a three-dimensional structure, a mathematical formulation for the forces between atoms gives the total force on each atom. These net forces are then used in Newton's second law of motion to give the accelerations, which are then integrated to give a numerical trajectory. The trajectory provides a complete description of the system, giving the position and velocity of every atom as a function of time. It is remarkable that simple forces and classical mechanics seem to give such a faithful picture of molecular motion.

At present, some of the most extensive molecular dynamics simulations have been used to study proteins and segments of DNA in solution. Such calculations involve tens of thousands of atoms and generate trajectories containing hundreds of thousands of structures changing with time; they require hundreds of hours of computer time, yet simulate periods lasting less than a nanosecond. As computer power continues to increase, it should be feasible to run simulations lasting microseconds (a billion time steps) and deal with the largest biological structures (a million atoms). In the limit of these longer time scales, there is a natural connection with analytical and stochastic theories. Indeed, such theories provide essential checks on the numerical methods used to generate trajectories. An area ripe for this combined approach involves ionic channels, where molecular dynamic simulations can provide the frictional constants used in analytical treatments. This provides a direct link to the extensively studied phenomenological equations of nerve conduction (Hodgkin-Huxley equations). The molecular dynamics method gives a fully detailed description of the system simulated; this in turn provides a unique opportunity to visualize these molecular systems at work. Such visualization is often accomplished by making a motion picture of the system as it changes with time. Numerical analysis of the trajectories is also necessary to calculate properties that relate to experimental data. Better techniques for this analysis are sorely needed.

2.2.3 Drug Design

Molecules interact strongly when they fit together well. This occurs when their three-dimensional shapes are complementary and when there are stabilizing interactions (hydrogen bonds, charged pairs, etc.). One of the most interesting and potentially useful molecular interactions concerns drugs that bind with very high affinities to protein and nucleic acid macromolecules and either block the normal function of the macromolecules or mimic other ligands for such structures as receptors and induce a normal physiological response. Inhibition can be advantageous if the protein is made in excess, or if normal cellular control of the protein's activity has been lost. Because drug binding involves spatial complementarity, and because the aim is to design a molecule that binds with the highest affinity possible, it should be possible to use the three-dimensional structure to aid design. Current work in this area has followed several directions. The most direct approach is to crystallize the protein together with the drug. Study of the structure of the complex can suggest modifications to the drug expected to enhance its

affinity for the receptor or enzyme active site. For this method to work, one needs an initial drug known to bind to the protein.

Other methods aim to circumvent this requirement by deducing the structure of the drug directly from the structure of the protein. While these methods are able to suggest completely new drug molecules, they involve a search for structures that fit a binding site. The theoretical underpinnings of such searches require further theoretical development. More specifically, they would benefit from application of better methods in global optimization and graph theory.

2.2.4 Nucleic Acid Sequence and Structural Analyses of Nucleic Acids

When a DNA sequence is determined, it is examined for a variety of sequence features known to be important: tRNA's, RNA's, protein coding regions—introns and regulatory regions, promoters, and enhancers. Since these sequence features are not identical in all organisms, it is often quite difficult to identify them. Even the widely studied bacterium *Escherichia coli* promoter sequences cannot be identified with certainty. As more and more DNA is sequenced, it becomes increasingly important to have accurate methods to identify these regions without many false positives. Statistics and mathematics should make significant contributions in this area.

As described above, pairwise alignment of sequences using dynamic programming is a well-developed area. However, alignment of more than two sequences remains a serious problem, one requiring extended computation time. Some recent advances reduce the computation time so that 10 sequences might be practical, but many problems are not approachable. Heuristic methods that align by building up pairwise alignments have been proposed, but they often fail to give good multiple alignments. Closely coupled with multiple alignment is the construction of evolutionary trees. Closely related sequences should be neighbors with few changes between them.

In the area of DNA structure, several subareas are particularly amenable to mathematical analysis: (1) A complete analysis of the packaging of DNA in chromatin. Only the first-order coiling into core nucleosomes is understood. By far the largest compaction of DNA comes from higher-order folding. (2) Presentation of the topological invariants that describe the structure of DNA and its enzymatic transformations. The goal is to be able to predict the structure of intermediates or products from enzymatic mechanisms and in turn to predict mechanisms from structure. (3) An analysis of the reciprocal interaction between secondary and higher-order structures. This includes the phenomena of bending, looping, and phasing.

This work has implications for both biology and mathematics. Mathematics will be affected in the areas of both topology and geometry. Renewed interest in the study of embedding invariants for graphs has been stimulated by the enumeration and classification of topoisomers; the study of random knots has been used to study macromolecules in dilute solution, and tangle calculus and Dehn surgery theory have been used in the study of DNA enzyme mechanisms.

In the study of kinetoplast DNA, topology and the theory of interacting particles have been brought together in a unique way. Finally, in the study of DNA-protein interactions, theorems from differential geometry and differential topology have been recast in different frameworks to solve helical periodicity problems. The determination of the configuration of closed circular DNA brings together the fields of geometry and topology and nonlinear partial differential equations, or topology and Monte Carlo techniques. These will involve extensive use of

computational techniques including the creation of new codes to use nonlinear partial differential equations to solve elasticity problems for closed circular rods.

2.2.5 Structural Analysis of Cells

Mathematical models have played, and will continue to play, an important role in cell biology. A major goal of cell biology is to understand the cascade of events that controls the response of cells to external ligands (hormones, transport proteins, antigens, etc.). The problem begins with understanding the interaction of the ligand with the cell's surface receptors. For some types of receptors, binding of the ligand to the receptor will lead to the generation of a transmembrane signal. For others, aggregation among the receptors must occur before a cell response can be triggered. The receptors themselves are under dynamic control, up or down regulating in response to external ligands, changing their rate of capture by coated pits, altering their recycling pattern, changing their rate for new receptor synthesis, changing their rate of delivery of old receptors to lysosomes for degradation, etc. The signaling pathways that are now being elucidated are equally, or more, complex. The role of mathematical models in studying these processes is to help rigorously test ideas about mechanisms and pathways, to aid in analyzing experiments, to determine parameter values, and to help in the design of new experiments. Mechanistic models for some of the stages of the receptor pathway have already been developed, e.g., aggregation of receptors on cell surfaces (Dembo and Goldstein 1978, Perelson and DeLisi 1980), capture of receptors by coated pits (Goldstein et al. 1988), receptor-ligand sorting in endosomes (Linderman and Lauffenburger 1988), and have been useful in understanding receptor dynamics. Kinetic models have been used to analyze studies of ligand binding and internalization for a variety of receptor systems. With models it should be possible to dissect the relationship between structure and function. Thus, for example, a large number of mutants of the epidermal growth factor receptor have been generated. Determining whether the induced change in structure then affects ligand binding, tyrosine kinase activity, receptor aggregation, capture of the receptor by coated pits, etc., can best be done via collaborative experimental modeling efforts. A major challenge that lies ahead is to build mathematical models of specific cell types that incorporate all the known biochemistry and that can be used to answer questions about the normal and disease states of the cell. Such an attempt is under way for the red blood cell (Yoshida and Dembo 1990), but here the effect of the biochemistry on the biomechanics of the cell is also important since the shape of the red blood cell is so critical for normal function. Predicting cell shape and the dynamic changes that occur in the cell's cytoskeleton due to interactions at the cell surface, which may lead to calcium influxes, receptor phosphorylation events, etc., are challenges for future models.

3

THE IMPACT OF MATHEMATICS ON ORGANISMAL BIOLOGY

ORGANISMAL (sometimes called organismic) biology deals with all aspects of the biology of individual animals and plants, including physiology, morphology, development, and behavior. Thus, it interfaces cellular and molecular biology at one extreme and ecology at the other. In the former, one attempts to develop integrative theories of organismal function; in the latter, one attempts to place individual behavior and function within an environmental context. Mathematical theorists have made signal contributions to organismal biology and have employed a wide range of mathematical techniques in doing so. Examples range from technological advances to theories of biological structure and function. We begin this section with a review of some of the outstanding examples.

3.1 ACCOMPLISHMENTS OF THE PAST

Image reconstruction is of importance across a range of organizational levels in biology. At the molecular level, work by the applied mathematicians Karle and Hauptman in constructing algorithms to reveal structure from x-ray data was rewarded with a Nobel Prize in 1987. As mentioned earlier, another Nobel Prize was awarded for work done at the organismic level by Cormack and Hounsfield in constructing algorithms that permit structure to be determined from tomography. PET and NMR are other areas where mathematical analysis is essential. Past achievements are impressive, but they must be supplemented by significant further advances before the difficult but vital problem of image reconstruction is minimally satisfied.

One of the most exciting areas of mathematical application has been to cardiac function. A major cause of death from malfunction of the heart is the phenomenon called ventricular fibrillation, wherein properly coordinated heart action is replaced by purposeless local oscillations of the ventricles. Mathematical modeling has revealed why this phenomenon occurs. Major experimental efforts have been suggested by the modeling. The leading figure in this line of theoretical research, Arthur Winfree, received the 1989 Einthoven Prize for his contributions to the subject. (This prize is awarded every five years to a cardiologist, usually a surgeon.)

In related work, powerful numerical algorithms and state-of-the-art computing have been applied by Peskin and others to study blood flow in the heart. Even with the use of two-

dimensional models, progress has been sufficient to enable significant input into the design of heart valves, with resulting patents and licensing agreements (see McQueen and Peskin 1983, 1986). Three-dimensional models also are under development (Peskin and McQueen 1989, McQueen and Peskin 1989).

Another major contribution of mathematics to physiology is the theory of cross-bridge dynamics in striated muscle. Introduced by A.F. Huxley (1957) and further developed by T.E. Hill, Podolsky, Lacker, and others, this theory not only has provided a satisfying explanation of the mechanical behavior of muscle, but it also has served to provide organizing principles for biochemical research on the fundamental energetic and control mechanisms of muscle contraction.

Mathematical methods for the quantitative description of morphogenesis of organs composed of nonmigrating cells (including plants, animal bone and skin, and shells) were suggested by Richards and Kavanagh (1943) and by Erickson and Sax (1956). These methods, which involve evaluation of velocity gradients from empirical data, have provided the phenomenological basis for understanding the physiology of growth (for reviews see Erickson 1976, Silk 1984, 1989).

As will be argued in detail below, theory is essential in understanding hierarchical systems phenomena in biology. A famous contribution in this area is the theoretical model made by Hodgkin and Huxley (1952) of the electrical signals in the squid axon. This Nobel-Prize-winning work incorporated the findings of a series of brilliant experiments concerning the ion permeability of the axonal membrane into a set of mathematical equations that predicted the shape and speed of the "action potential" wave that moves down the axon. Patch clamp recordings now permit investigators to relate the Hodgkin-Huxley membrane models to the opening and closing of the molecular channels that span the membrane and are responsible for its ionic conductance. Hodgkin and Huxley's inferences from macroscopic current measurements have been confirmed in basic form but have also been greatly expanded with respect to their descriptions of configurations and transition mechanisms. In recent years the work of Hodgkin and Huxley has found unexpected application in nonneural systems in which electrophysiology plays a surprising regulatory role. One example of this is the control of insulin secretion by the electrically active beta cells of the pancreas.

In developmental biology, it was hypothesized decades ago that gradients of key chemicals were responsible for triggering macroscopic events. In recent years, especially since the landmark paper of Turing (1952), the gradient idea has been greatly elaborated by theorists. In parallel, experimentalists devoted considerable efforts to find the "morphogen" chemicals whose gradients were postulated to have such importance, efforts that recently have been successful in *Drosophila*, hydra, and limb morphogenesis.

3.2 GRAND CHALLENGES

A wide variety of exciting venues exist for the application of mathematical and computational approaches to organismal biology. Among these, two stand out as having exceptional promise and importance: the study of *complex hierarchical biological systems* and of *dynamic aspects of structure-function relations.*

3.2.1 Complex Hierarchical Biological Systems

The analysis of complex hierarchical systems is one of the most important open areas in modern biology. This holds true at all levels of organization, and is a theme to which we return in the discussion of ecological and evolutionary processes. The essence of the matter is this: On several levels, the components of biological systems are being revealed by modern experimental biology. The techniques of molecular biology are most important here; other experimental advances are also of major utility. The central theoretical question is, How are the molecular details integrated into a functional unity?—a question central to at least three major fields: neurobiology, developmental biology, and immunology. We now consider each of these areas in greater depth.

Neuroscience. Mathematical modeling has made an enormous impact on neuroscience. The Hodgkin-Huxley format for describing membrane ionic currents has been extended and applied to a variety of neuronal excitable membranes. The significance of dendrites for the input-output properties of neurons was not understood before the development of Rall's cable theory (Rall 1962, 1964). Hartline and Ratliff (1972) were pioneers in developing quantitative and predictive network models. In addition, Fitzhugh's work (1960, 1969) demonstrated the value of simplified nonlinear models and of qualitative mathematical analysis. The success of these theoretical contributions, and the high degree of quantification in neurobiology, ensures continued opportunities for mathematical work.

Recent technical advances in experimentation—e.g., patch clamp recording, voltage- and ion-specific dyes, and confocal microscopy—are providing data to facilitate further theoretical development for addressing fundamental issues that range from the subcellular to cell-ensemble to whole-system levels. For thorough understanding, we must synthesize information and mechanisms across these different levels. This is perhaps the fundamental challenge facing mathematical and theoretical biology, from molecule to ecosystem. How do we relate phenomena at different levels of organization? How are small-scale processes to be integrated and related to higher-level phenomena? For example, in modeling neuronal networks, what are the crucial properties of individual cells that must be retained in order to address a particular set of questions? Most network formulations use highly idealized "neural units," which ignore much of what is known about cellular biophysics. We need to develop systematic procedures to derive, in a biophysically meaningful way, descriptions for ensemble behavior.

Correspondingly, we seek to identify low-level mechanisms from data at higher levels. The Hodgkin-Huxley theory hypothesized that macroscopic currents might be generated by molecular "pores"; only much later were these individual channels discovered. Another set of common modeling needs are methods for dealing reasonably with the wide range of time and space scales encountered in different intracellular domains and processes and in short- and long-distance interactions between cells and among different cell assemblies.

At the lowest level, improved biophysical understanding is needed of the mechanisms for ion transport through membrane channels. How does the voltage dependence of opening and closing rates arise? What accounts for ion selectivity, by which, for example, channels discriminate among ions of the same charge and similar properties? Theories at this level are beginning to involve stochastic descriptions for fluxes (Fokker-Planck equations) and simulation methods for molecular structure and dynamics. Kinetic modeling of single-channel data is being debated hotly with regard to whether a finite or infinite number of open/closed/inactivated states is appropriate.

The discovery of new channel types continues at a rapid pace (Llinas 1988). Of basic interest is how the mix of different channel types, and their nonuniform distributions over the cell surface (soma, dendrites, and axon), determine the integrative properties of neurons. Some cells fire only when stimulated, others are autonomous rhythmic pacemakers, and some fire in repetitive bursting modes. Theoretical modeling plays an important role here since channel densities cannot yet be measured directly, especially in dendritic branches. Computational models that incorporate detailed dendritic architecture, in some cases known from morphological staining, are suggesting that individual regions of dendrites can perform local processing (Fleshman et al. 1988; Holmes and Levy 1990). Differential dendritic processing has been implicated in motion detection in the visual system (Koch et al. 1986).

One of the most actively pursued goals in neuroscience research is to discover the mechanisms for plasticity and learning at the cellular/molecular level. The above techniques, together with state-of-the-art biochemical methodologies, are beginning to yield the information for feasible, detailed biophysical modeling. Dendritic spines, NMDA receptor-channels, and spatio-temporal dynamics of calcium and other intracellular second messengers are focal points for these explorations. Such studies are bringing together theoreticians, neuroscientists, and biochemists.

Although theorizing about mechanisms for synaptic plasticity is proceeding, disagreement remains about the basic mechanism of chemical synaptic transmission. Two competing hypotheses (one involving calcium alone and the other including voltage effects as well) are being explored with fervor, and mathematical modeling is a key ingredient in arguments for each case. Many additional experiments have been suggested from these debates (see Zucker and Haydon 1988 and Parnas et al. 1991).

Models of neural interactions lead to many interesting mathematical questions, for which appropriate tools must be developed. Typically, networks are modeled by (possibly stochastic) systems of differential equations. In some simplified limits, these become nonlinear integro-differential equations. The question now becomes one of proving or otherwise demonstrating that the simplified models have the desired behavior. Furthermore, one must characterize this behavior as parameters in the model vary (i.e., understand the bifurcations in the dynamics). Another important point that mathematicians must address is the extraction of the underlying geometric and analytic ideas from detailed biophysical models and simulations.

The next level of neuronal complexity beyond the single cell is the small network with tens to hundreds of neurons. Such networks have been most extensively studied in invertebrates and the sensory or motor systems of vertebrates, in which the function of small groups of neurons can be related to specific behaviors of the animal (Kandel 1984, Selverston and Moulins 1985, Lockery et al. 1989). These so-called simple systems also are attractive because one can expect to characterize their cellular and intercellular properties more completely than in vertebrates. Much research on their structural features has been based on the explicit assumption that once network structure was understood, functional understanding would follow. Recently, however, many workers have come to realize that, even with a great deal of structural information, the understanding of functional mechanisms will require the development of sound, structurally based theoretical models.

A principal challenge for modeling at this level is the development of more biologically realistic computational models and mathematical analyses that can provide insight into how these networks function. Although these networks involve relatively small numbers of neurons, their complexity will require increasingly powerful mathematical tools. At the same time, modeling at

this level is likely to be especially valuable for neurobiology. In few other neural systems is the link between neural structure and behavior more direct. Thus, it is already possible to see in the structure of the nervous system its functional correlates. Moreover, few other systems currently provide the anatomical and physiological parameters essential for realistic modeling. As models for understanding the general dynamical properties of such neural networks or for understanding the way in which feedback modifies neuronal behavior, small neural systems represent a gold mine for computational and mathematical neurobiology.

Coherent brain areas dedicated to particular functions—for example, primary sensory cortical areas—provide complex challenges for computational and mathematical models (Sereno et al. 1988). Such areas typically contain multiple types of cells, receive inputs from multiple distinct sources, and often are heavily interconnected with their links to inter-area recurrent or reentrant loops. Large bodies of anatomical and physiological data are available, but the integrative capabilities are poorly understood and modeling techniques will almost surely be needed to unravel them.

Developmental neurobiology is a source of biologically important and mathematically interesting questions. Modeling at the large-network level has played an important role in this field, with many collaborations between mathematicians and experimental biologists. Among the important questions arising in this field are the topography of connections from one part of the brain to another and how these maps might spontaneously form. Many examples exist of such maps in the central nervous system; the best characterized are in the vertebrate visual system. The earliest theoretical models and experiments concerned the "wiring" from the retina to the optic tectum. Many models have been proposed and analyzed (von der Malsberg 1973, Whitelaw and Cowan 1981, see Linsker 1990 for a review); but as new experimental results have become available, many of the models will have to be altered or eliminated. Recent investigations have led to the formulation of minimal hypotheses for the explanation of the large body of experimental manipulations (Fraser 1985). These mechanisms are ripe for mathematical formulation and analysis.

Several new technologies, such as voltage-sensitive dyes and deoxyglucose injection, have led to the discovery of beautiful regular maps in the visual cortex of mammals. The patterns include stripes of ocularity and twists and singularities of orientation preference. Models have been proposed for these patterns (Miller et al. 1989, Durbin and Mitchison 1990) involving mechanisms ranging from band-pass-filtered noise, to competitive interactions, to Hebbian rules with lateral inhibition. What must be done is to identify the common idea that underlies these models and how it might possibly be realized in the nervous system.

As we begin to understand the mechanisms of synaptic plasticity, it is natural to ask about the consequences of this for the behavior of large networks involving plastic elements. Only in this way will we understand the relation between synaptic plasticity and learning at the organismic level. This has been a major focus in the study of computational properties of large-scale neural networks across a number of disciplines, including physics, biology, psychology, and mathematics (Hopfield 1984, Rumelhart et al. 1986). Mathematical analysis promises to provide an important bridge between computational and behavioral studies and the empirical results of neurobiology (Poggio and Girosi 1990). An excellent survey is Koch and Segev (1989).

Models at the level of the complete organism provide an opportunity to make real progress on the long-sought unification of the behavioral sciences with neurobiology. Models intended to explain behavioral observations (e.g., from psychology and ethology) can be cast in terms of underlying neural mechanisms, rather than at the phenomenological or control-theory level as

before. Such models can bring about a new understanding of such phenomena as visual illusions (e.g., Treisman et al. 1990), the relation between long- and short-term memory and category formation. They will provide significant constraints on psychological explanations that have not in the past been easy to correlate with the nervous system. To carry out this analysis, one must eventually couple models of the nervous system with those of the environment in which the whole system exists (Kersten 1990).

Immunology. The immune system contains 10^{12} cells comprising at least 10^7 specificities. These cells move within the body and communicate both by cell-cell contact and via tens, maybe hundreds, of regulatory molecules. The system is capable of pattern recognition, learning, and memory expression, and thus has many features in common with the nervous system.

Theoretical ideas have played a major role in the development of the field. Controversies such as instructive vs. selective theories of antibody formation, germ-line vs. somatic-mutation models for the generation of antibody diversity, and regulatory circuits vs. idiotypic networks have dominated the intellectual development of the field and determined the direction of much experimental effort. Mathematical theories have not been nearly as important, but this appears to be changing as the field addresses more quantitative issues, such as the role of somatic mutation in the generation of antibody diversity; the role of receptor clusters in cell stimulation and desensitization signals; the effects of different concentrations of cytokines, receptor affinities, and receptor number on cell stimulation, cell proliferation, cell differentiation; and the engagement of effector functions.

Modeling the immune system requires the same type of hierarchical approach as does neurobiological modeling. At the lowest level, one must develop quantitative models of the action of single lymphocytes as they interact with antigens and cytokines. A large amount of effort involving the study of infinite systems of ordinary differential equations and branching processes has gone into the mathematical modeling of receptor cross-linking by multivalent ligands (Perelson 1984, Macken and Perelson 1985). Cell response in terms of proliferation or differentiation has been examined from an optimal control perspective (Perelson et al. 1976, 1978). The effects of the T cell growth factor IL-2 have also been incorporated into cellular models (Kevrekidis et al. 1988). At the next higher levels, small idiotypic networks containing two complementary cell populations have been modeled, as well as networks containing hundreds to thousands of B cell clones (Segel and Perelson 1989, Perelson 1989, Weisbuch et al. 1990). In the immune system, not only is the number of components large, but in distinction to the nervous system, the components turn over rapidly. The average life span of a B cell is about four days, that of serum antibody one to two weeks. Thus, on a rather rapid time scale, many immune system components may be replaced, although the system as a whole remains intact.

New ideas and mathematical representations are required to handle systems with large numbers of constantly changing components. Some promising approaches involve the formulation of models in terms of a potentially infinite dimensional "shape space," wherein emphasis is placed on determining interactions among molecules based on their shapes. In computer models, binary strings have been used to represent molecular shape, with the obvious advantage of fast algorithms to determine complementarity and the ability to represent 4×10^9 different molecular shapes with 32 bits (Farmer et al. 1986). To handle the perpetual novelty that the elimination of old components and the generation of new components introduces into the immune system, models can be formulated using "metadynamical" rules, wherein an algorithm is used to update the dynamical equations of the model, depending upon the components present in the system at the time of update (Bagley et al. 1989). One needs to understand in a mathematical

sense the dynamics of a system in which the variables of the model are in constant flux. What does it mean to have an attractor if the variables describing the attractor are eliminated from the system before a trajectory approaches the attractor? Formulation of models appropriate to unravel the observed complexity in the immune system is the first major step. Next, a massive effort is required to unravel the behavioral modes of these complex models and compare them with experiment. Here theoretical immunology merges into the mainstream of theoretical biology.

There are other areas in which we see future growth of theoretical ideas in immunology. For example, vaccine design depends on the ability to predict T cell epitopes. DeLisi and Berzofsky (1985) suggested that T cell epitopes tend to be amphipathic structures. Alternative algorithms have been suggested (e.g., Rothbard and Taylor 1988), and databases have been used to identify sequence patterns characteristic of T cell epitopes (Claverie et al. 1988). This area is clearly one in which we will see future growth and one that will rely heavily on theoretical and computational analyses.

Understanding the dynamics of HIV infection (AIDS) and its effects on the immune system is another important area for future research. Quantitative questions include these: How can the $CD4^+$ T cell population be depleted if only one in a hundred cells is infected? Why is there such a long incubation period from time of infection to the clinical symptoms of AIDS? Why is this incubation period different in children than in adults? In a seropositive patient, what does the level of serum antibody predict about the course of the disease? Can one define quantitative measures of an individual's chance of infecting a sex partner on the basis of antibody or antigen levels measured in the blood? Models also will help in determining the pathogenesis of the disease and in isolating primary effects of HIV from the secondary effects of immune dysfunction. Mathematics also can play a role in the development of optimal treatment schedules and in the design of clinical trials of multiple drug therapies for AIDS. Development of epidemiological models is currently an active area of mathematical endeavor and one that will continue at a high level as we attempt to track the course of this epidemic and develop vaccine strategies aimed at its eventual eradication.

Genomic regulatory networks. A fundamental activity over the next two decades will involve analysis of the integrated structure and behavior of the complex genetic regulatory systems underlying development in higher organisms, a massive task since the human genome encodes perhaps 100,000 genes. Its accomplishment will require uniting work in molecular and developmental genetics with new mathematical and computational tools.

In more detail, recent progress in molecular genetics in eukaryotes now is revealing the detailed composition of structural genes as well as cis-acting regulatory loci such as promoters, homeoboxes, and tissue- and stage-specific enhancer sequences, as well as trans-acting components. These genetic elements, together with their RNA and protein products, constitute the genomic regulatory network that coordinates patterns of gene expression in cell types, cell differentiation, and ontogeny from the zygote. Understanding the structure, logic, integrated dynamical behavior, and evolution of such networks is central to molecular, developmental, and evolutionary biology.

The Human Genome Initiative will provide massive sequence data from which we can eventually identify the diverse locations in the genome of each regulatory sequence, as well as the locations of many or most structural genes. These data are fundamental to understanding the "wiring diagram" of the genomic regulatory networks in eukaryotes. Analysis will require development of appropriate computer databases and development of new theory and algorithms

in the mathematical theory of directed graphs. Understanding the evolution of such genomic networks under the influence of point and chromosomal mutations that literally scramble the genomic wiring diagram will require new uses of random directed graph theory, stochastic processes, and population genetic models.

In addition to understanding the structure and evolution of genomic regulatory networks, we must understand the coordinated behavior of such systems that integrate the behavior of 100,000 molecular variables. It is here—in the effort to relate the information that we can obtain about small parts of the genomic system to the overall behavior of the integrated system—that a new marriage of mathematics and biology must be found. Without mathematical theories, we have no hope of understanding the integrated behavior of such complex systems—systems that link the "microlevel" of structure and logic with the macrolevel of behavior. Although no approach is yet clearly adequate, new avenues are available.

A first approach is via ensembles. Statistical mechanics is the paradigmatic example of a theory that links microscopic and macroscopic levels. There it is possible to explain macroscopic behaviors without knowing all the details of the microscopic dynamics. Similarly, it may be possible to build up statistical understanding of the integrated behavior of extremely complex genomic regulatory systems without knowing all the details of microscopic structure.

Molecular genetic techniques reveal small-scale features of genomic systems, such as the sequences that regulate a gene, and biases in the "rules" governing the activity of genes as a function of their molecular inputs. Using these local features, one can construct mathematically the ensemble of all genomic systems consistent with those local constraints. This ensemble constitutes the proper null hypothesis about the structure and logic of genomic systems that are random members of such an ensemble. Thus, the typical or generic behavior of ensemble members are predictions about the large-scale features of random members of the ensemble. This is a new kind of statistical mechanics, averaging over ensembles of systems (Kauffman 1969, 1974, in press, Derrida 1981). If the distributions of properties parallel those seen in genomic regulatory systems, then those properties may be explained as consequences of membership in the ensemble. Indeed, work based on this approach (Kauffman 1969, 1974, in press) has shown that many features of model genomic systems do parallel, and hence may explain, a number of features of cell differentiation, such as the numbers of cell types in an organism, the similarity of gene expression patterns in different cell types in an organism, and other statistical features. Improved ensemble models, coupled with population genetic models, offer hope of understanding how evolution can mold the structure, logic, and behavior of integrated genomic systems.

A second approach may be the development of new mathematical and experimental tools to "parse" the genomic system into structurally or functionally isolated subcircuits. Thus, clusters of genes may be regulated in overlapping hierarchical batteries, or some genes may fall to fixed steady states of activities that are common to many or all cell types, while other subsets of genes oscillate or exhibit complex patterns of temporal activity unique to different subsets of cell types. Analysis of such temporal patterns by time-series techniques, and based on temporal series of two-dimensional protein gel data, where each gel shows the synthesis patterns of up to 2000 genes at a time, may help resolve the genome into behavioral "chunks." If so, this will help block out the overall behavioral organization of the genomic system. Thereafter, analysis of detailed midsized subcircuits, with perhaps several to 100 or so genes, will require use of promoter constructs allowing activation or inhibition of arbitrary genes in arbitrary cell types at arbitrary moments, with analysis of the cascading consequences. Union with dynamical systems

theory for modestly small systems, where the "inverse problem" of guessing plausible circuitry to yield observed synthesis patterns is practical, can then be carried out.

Developmental biology. As already described, mathematics can play a crucial role in connecting different levels of organization. What biologists seek are molecular-level explanations of supramolecular phenomena. For example, embryogenesis involves the coordinated movement and differentiation of cell populations. Biologists would like to understand this in terms of chemistry and genetics. To understand organismal biology is to understand how high-level coherent organization results from mechanisms operating at the molecular level. The essence of the problem is to build from one level to another. How can we bridge this gap?

The mathematical, analytical, and numerical problems posed by the nonlinear systems of partial differential equations that arise in modeling developmental processes are extremely challenging and interesting. Reaction diffusion equations, for example, as discussed earlier, have already stimulated the creation of new mathematics to study the wide spectrum of solution behaviors exhibited by these equations. The numerical simulation techniques used to investigate solutions in three dimensions are still very difficult and need a great deal of further refinement to be practically useful. Mechanochemical models for generating pattern formation deal with more directly biological quantities (see Murray 1989 for a general survey of these and other pattern formation models); but they are more complex than, for example, the Navier-Stokes equations, which govern fluid flows, and they possess a correspondingly richer solution behavior.

Bifurcation theory, linear analysis, and singular perturbation methods already have revealed new phenomena. Numerical simulation, particularly with the mechanochemical models, is challenging even in two dimensions. Real biological applications require solutions in three-dimensional domains whose sizes increase in time. New analytical and numerical simulation techniques, as well as novel visualization methods, will have to be devised before we can explore the sophisticated solution behaviors of such models. Unfortunately, the methods developed for Navier-Stokes equations frequently are not adequate to cope with the new models that arise in biology.

Recently, several advances in experimental biology (e.g., recombinant DNA technology, computer-enhanced imaging) have created new databases so extensive and complex that mathematical and computational approaches are essential to make sense of them. For example, a network of perhaps 60 cross-regulating genes has been shown to regulate early development in *Drosophila*; similarly, cell motility, which underlies morphogenesis, is driven by the cellular cytoskeleton, whose mechanochemical regulation is controlled by a network of more than 40 regulatory molecules. These systems should catalyze new collaborations between biologists and mathematicians to deduce the macroscopic consequences of newly revealed molecular mechanisms. Below we illustrate the general case with a few specific examples.

In the past five years, recombinant DNA technology advances have produced an unprecedented molecular-level database documenting a complex network of genes that code for proteins that control the expression of other genes. Mathematics can compute the macroscopic pattern formation consequences of this molecular-level information. Indeed, mathematical analysis may be the only way to synthesize the global picture from the molecular-level parts, given the apparent complexity of genetic networks, in which each gene's expression is modulated by many other genes.

Computer graphics can be used to visualize data and the dynamical behavior of mathematical models (Odell and Segal 1987). Many instruments in the biologist's arsenal (e.g., the confocal

scanning laser microscope, gene sequencers) gather data into a computer-based graphical database. Modern computer graphics technology makes it possible to display, dynamically and pictorially, the dynamic behavior of a mathematical model in the same form in which experimental data are stored. This technology should become the common way to compare the behavior of a quantitative model with the data it purports to explain. Moreover, this same technology yields the fastest and most compelling medium of communication between mathematical modelers and biologists.

Using immunofluorescent probes from cloned gene products and scanning confocal laser microscopy on whole-mount *Drosophila* embryos, one may now obtain three-dimensional stereo reconstructions of the temporal evolution and spatial expression pattern of each of the genes that organize future morphological segmentation of the larva. Similarly, it is possible to observe intracellular and intercellular events such as cytoskeletal reorganization, calcium transients, distribution patterns in cell adhesion molecules, and putative morphogens in real time. Thus, a model of early pattern formation and/or morphogenesis (Edgar et al. 1989) in the *Drosophila* embryo, if it is correct, should produce the same output that confocal microscopy gathered as input. The intellectual challenge is to understand how the gene network, operating identically in every cell, results in a globally coherent spatial pattern as a consequence of temporal biochemical dynamics.

Theoretical models have stimulated a great deal of experimental work in developmental biology. Here we briefly describe three major classes of models that illustrate the way in which mathematics provides a framework for connecting information at the micro level to observations at the macro level.

Spatial patterns can be created according to the classical local-activation lateral-inhibition mechanism (Keller and Segel 1970, Oster and Murray 1989). A purely chemical mechanism for pattern formation (but not morphogenesis) was proposed by Turing (1952). In this model, activator and inhibitor morphogens diffuse at different rates and react with one another. Mathematical analysis shows how spatially heterogeneous patterns of morphogen concentration can arise. For pattern to emerge, it is necessary that the activator be relatively short-range relative to the inhibitor, i.e., that the activator diffusion be relatively slow. If cells can sense the morphogen level and respond, then we have a molecular mechanism for Wolpert's (1969) notion of "positional information," one of the most influential concepts in modern developmental biology. Although chemical gradients have been suspect in biological pattern formation for over 100 years, it is only recently that their existence has been unequivocally demonstrated (e.g., the bicoid protein in *Drosophila* and retinoic acid in vertebrate limb development). However, morphogenesis may not be a purely chemical phenomenon in which cells merely respond to pre-existing chemical patterns.

One possibility is generation via chemotaxis, the response to a chemical gradient. The classical example is the slime mold *Dictyolstelium*, whose cells produce the chemoattractant cAMP as well as a chemokinetic morphogen (ammonia). Starting from the view that morphogenesis is, at least proximally, a mechanical event, several modelers have shown that the same spatial patterns that arise in Turing models can be produced by biomechanical models whose variables are cellular stresses and strains. These mechanochemical models have stimulated experimental programs to address their validity (Wolpert and Hornbruch, in press).

3.2.2 Dynamic Aspects of Structure-Function Relationships

The relation between structure and function is a central theme of classical biology. Some mathematical models have already illuminated problems in this area. For instance, McMahon and Kronauer (1976) modeled the tree branch as a beam of greatest lateral extent. Another example involves the biomechanics of feeding of aqueous organisms. Solving the Navier-Stokes equations for flow through small, bristled appendages, Cheer and Koehl (1987) have shown how the geometry permits the appendage to function as either a paddle or a rake.

In temporally shifting systems, the description of structure-function relations remains especially elusive; and it is here that mathematical modeling is particularly essential. In physiology, for example, only by solving the appropriate equations of fluid mechanics and elasticity can one understand the relationships between the structure of the heart and its function of providing appropriate blood flow (and changing blood flow) in response to changing environmental conditions. Similar remarks apply to other organs; for example, the kidney. Here fluid-mechanical considerations play a role, but the details of chemical reactions are perhaps even more crucial to describe accurately. The interplay between chemistry and solid and fluid mechanics is similarly important in the description of plant growth.

Organ physiology is a natural target for mathematical and computer modeling. Such models can serve a threefold purpose: to understand the normal structure-function relationship of the organ, to study the mechanisms and impact of disease processes, and to aid in the design of artificial devices that can be used to repair, assist, or replace the organ. For plants one can add the possibility of aiding breeders by identifying structures that optimize performance.

In the case of the heart, a computational method has been introduced (Peskin and McQueen, 1989) to solve the coupled equations of motion of the muscular heart walls, the elastic heart valve leaflets, and the viscous incompressible blood that flows in the cardiac chambers. Variants of this method have been applied to other problems in bio-fluid dynamics, including platelet aggregation during blood clotting, aquatic animal locomotion, and wave propagation along the basilar membrane of the inner ear. In the heart itself, the method has been used to study the optimal timing of events of the cardiac cycle, to simulate a disease state involving prolapse of the mitral valve, and to conduct parametric studies aimed at the optimal design of prosthetic cardiac valves. At the level of mechanics, another set of challenges is to develop theories for explaining the heart's structural components: the orientation and layering of muscle fibers in the ventricles, the position and makeup of the heart valves.

Cardiac contraction is mediated by propagation of electrical activity over the three-dimensional multicellular musculature. Disturbances in this electrical system result in arrhythmias; the most severe of these is ventricular fibrillation, the principal cause of death after a heart attack. This is an active area of modeling research, with many open avenues to explore: the ionic channels underlying the cardiac signal (Noble 1962); the effects of spatial inhomogeneities, say from damaged tissue; the consequences of discreteness (finite cell size and gap-junction coupling); and the fundamental nature of synchronization and sustained propagation patterns in three dimensions (Winfree 1990).

Other organ systems that are under intense investigation, and which cannot be understood without the help of mathematics, include the kidney and pancreas. The kidney's countercurrent mechanism achieves a substantial separation of water and solutes, which determines, under the influence of antidiuretic hormone, whether a dilute or concentrated urine will be excreted. A key difficulty in this field is that the basic rules governing the transport of ions and molecules (e.g.,

Na+, Cl-, urea, and water) across the walls of renal tubules are quite different in different parts of the nephron (the fundamental unit of renal function) and are in many cases unknown. Differential equation models are leading to considerable insights in this area by illustrating the physiological consequences of different assumptions and therefore suggesting experiments critical in distinguishing the possibilities (Stephenson 1972, Weinstein and Windhager 1985, Layton 1989). The many nephrons in a kidney are spatially distributed in a particular way; modeling will be invaluable in helping us to understand the reasons.

The pancreas also plays a key role in homeostasis, the control of the body's internal environment, in which cells must operate. Although the classical view of homeostasis is based on steady-state notions, the release of insulin for metabolic regulation actually occurs in a rhythmic, pulsatile manner (period of 10 minutes or so), which appears to involve a hierarchy of oscillatory time scales. Release by cells in the islet (the functional unit of the pancreas) is correlated with their electrical activity, which exhibits a 5–10 second oscillation in response to glucose. Modeling, analogous to that for ionic currents in neurons, is helping to identify how the cellular oscillations arise, how cells are synchronized, and what the possible glucose-sensing mechanisms are (Keizer 1988, Sherman et al. 1988, Rinzel 1990). Further challenging questions have to do with coupling between electrical activity and release and with interactions among the million or so islets in the whole pancreas.

In organ morphogenesis, important challenges for future work include finite element analyses of mechanical stress fields in the cellular continuum of growing tissue; optimization models to understand the functional significance of morphologies; and hydrodynamical models for nutrient transport in plants and animals (including marine invertebrates). Another interesting class of problems involves demographic models to predict cell cycle duration, age distribution, and family trees of cells in developing tissue (Bertaud and Gandar 1986). Kinematic analyses could be used to help unravel the physiological significance of gene products recently found to be correlated with the events of the cell cycle (reviewed by Murray and Kirschner 1989).

One of the strengths of mathematics is, of course, the ability to contend with temporally varying phenomena and, in particular, to use models to deduce mechanism from kinetic data. It is a theme of modern biology, which has been reiterated several times in this report, that what was previously regarded as static has now come to be understood as dynamic. We have just cited the dynamic nature of pancreatic homeostasis. An example of similar type is the hormonal regulation of ovulation, which has been shown in the laboratory of Knobil to involve pulsatile secretion of the relevant hormones with a periodicity of about one hour. This too is an especially fertile field for mathematical investigation. The book edited by Goldbeter (1989) is a source for up-to-date references for theoretical work on this and many other dynamical problems in physiology.

4

THE IMPACT OF MATHEMATICS ON ECOLOGY AND EVOLUTIONARY BIOLOGY

ECOLOGY and evolutionary biology encompass a broad range of levels of biological organization, from the organism through the population to communities and whole ecosystems, and a tremendous range of spatial and temporal scales. Aspects of it have been discussed in the earlier chapters, from phylogenetic reconstruction to organism-environment interfaces. The grand challenges identified earlier—in particular, analysis of structure-function relations and the integration of phenomena occurring at different scales—are of particular relevance both to ecology and to evolutionary biology.

Autecology refers to the interaction of organisms with their environments, including such aspects as physiology, morphology, and behavior. Some related aspects of organismal biology have been covered in the preceding section. The need for enhanced mathematical and computational ability is most evident when one attempts to couple large numbers of individual units into highly interactive networks. Individual-based models of populations provide a case in point, as do spatially distributed analogues of simpler dynamic models. Computationally intensive areas of autecology include those linking neurobiology with behavioral models for certain tasks, such as search, and the modeling of spatial pattern formation through interacting particle systems or partial differential equations.

Population biology deals with the basic and applied aspects of ecological and evolutionary change, including links to resource management, epidemiology, and demography. The rich theoretical literature in this subject, including the work of such giants as Lotka, Volterra, and Kostitzyn in ecology; Fisher, Wright, and Haldane in genetics; and Kermack and McKendrick in epidemiology, has greatly influenced the development of fields as diverse as dynamical systems theory on the one hand and probability and statistics on the other (see Chapter 1). As already discussed, May's demonstration of how chaotic behavior could arise in simple dynamical models was a catalyst for the development of that aspect of dynamical systems theory, and interest in the dynamics of epidemics has spurred research in differential-difference equations and integro-differential equations, an area pioneered by Volterra in the classical models of mathematical ecology.

Population biology thus includes a dauntingly diverse assemblage of topics, including, for example, the construction of phylogenetic trees from data sets, the interface of game theory and population genetics, the ecology and evolution of quantitative characters, molecular evolutionary dynamics, and human population genetics.

Among the critical computational problems in population biology are those that relate to database management in the examination of risk groups for epidemiological models—for example, the classification of sexual behavior and its relationship to the spread of AIDS; categorization and analysis of information on the global environment, being collected by means of remote sensing techniques; and the manipulation of databases, such as those derived from sequence analysis, and their use in interpreting phylogenetic histories. Dynamic aspects relate to models of the spread of disease in heterogeneous populations; the interaction between evolutionary biology and neural networks, as reflected in the view of evolution as a combinatorial optimization problem in a very-high-dimensional space; more sophisticated game theoretical approaches to evolution; and quantitative genetics.

The study of communities and ecosystems includes the study of how assemblages of species are organized in space and time and how these assemblages interact with each other and the physical environment. One area that has received great interest is the analysis of the organization of trophic webs — the compilation and storage of data from hundreds of webs collected by ecologists introduces substantial problems of data storage and retrieval. Cohen's analysis of the consistent patterns exhibited by these webs (Cohen 1978) demonstrates how sophisticated mathematical analysis can lay bare patterns in the balance of nature. Biogeochemical cycles represent a complementary aspect of the dynamics of ecosystems; and the analysis of patterns in these cycles, and of how they respond to different stresses in different ecosystems, is of fundamental importance. The analysis of ecosystems—and especially of the transfer of energy and nutrients within the biota and between the biota and its physicochemical environment—involves a class of problems of considerable applied importance.

Agroecosystems, ecotoxicology (the responses of ecosystems to chemical stresses), landscape ecology, and global change represent other areas of importance. The study of agroecosystems raises problems from the characterization of rates of spread of pest species (for which the mathematical results of Kolmogorov et al. (1937) provide the mathematical underpinnings, and for which models and approaches borrowed from percolation theory and interacting particle systems allow the extension to fragmented habitats) to issues of management, as represented by dynamic programming approaches to integrated pest management, among other problems. Ecotoxicology trades heavily on diffusion-advection models of spread and on multivariate statistical methods for the analysis of the fate, transport, and spread of chemicals.

4.1 ACCOMPLISHMENTS OF THE PAST

For interdisciplinary work, such as theoretical and computational biology, a success occurs in one or more of three ways. First, new mathematics can develop from the biological problem. Second, the theory can affect in a fundamental way the world view of biologists, most of whom are not theoreticians. Third, the theoretical contribution can lead to modifications of practice. Ecology and evolutionary biology have had numerous instances of each kind of success.

The application of mathematical methods in this area is a very old enterprise; as already discussed, it spans a range of topics from the very basic to the very applied (Roughgarden 1979,

May 1981, Hallam and Levin 1986, Levin et al. 1989). Demographic methods have been applied to the study of human and nonhuman populations for centuries (see, for example, Keyfitz 1977) and form the basis both for population projections and for the understanding of the population consequences of life history phenomena (Cole 1954). The interface with population genetics, and more recent game theoretical approaches, has produced a rich mathematical literature that forms the basis for our understanding of the evolution of the living world. At the other extreme, mathematical models have been fundamental in describing the fate and transport of pollutants in the environment (Levin et al. 1989), the spread of agricultural pests, the dynamics and control of epidemics, the management of renewable and nonrenewable resources, and the response of ecological systems to such stresses as toxicants, acid deposition, and global climate change.

4.1.1 The Synthesis of Population Genetics and Evolutionary Biology

A major role of mathematical biology, and of biology in general, must be to improve our understanding of the evolution of the living world. The theory of evolution by natural selection, and the associated extensions that include the neutral theory, relate to the central organizing principle of modern biology. A key aspect of the elaboration of that theory lay in the mathematical contributions of Fisher, Haldane, and Wright, already discussed, and in relating evolutionary change to the underlying genetic mechanisms (see Provine 1971).

The suggestion that most molecular genetic variation within a species and between species is selectively neutral (i.e., has no adaptive or functional significance) stimulated a great deal of mathematical work on random changes in allele frequencies due to sampling effects in finite populations. Diffusion approximations to finite population models have been employed successfully to understand the amount and pattern of genetic variation in populations, including sampling properties (work by Kimura 1983, Ewens 1972, Watterson 1977, Griffiths 1979). The mathematical analyses of these models had an enormous impact on the biological view of molecular genetic variation and led to the development of statistical tests and estimation procedures useful in the analysis of enzyme polymorphism and sequence variation (see Nei 1987 for examples). This theoretical and empirical work also stimulated important work on models with random temporal and spatial variation of selection coefficients by J. Gillespie (1978).

Modern topics of fundamental interest that involve considerable mathematical content include punctuated equilibrium, coevolution, and sociobiology. Quantitative methods have been involved intimately in the development and logical structure of sociobiology, broadly construed to encompass all interactions among individuals that affect reproductive success. Quantitative theory has been instrumental both in establishing the hypothesis itself within an evolutionary framework (Hamilton 1964, Cavalli-Sforza and Feldman 1981) and in testing and revising the fundamental theory (Hamilton 1964, Uyenoyama and Feldman 1980).

4.1.2 Autecology

Classic studies in heat balance in leaves and plant parts (Raschke 1960, Gates 1965) and animals (Porter and Tracy 1973) were used to predict "climate space," the set of microclimate variables (exposure to sunlight, wind, etc.) consistent with maintaining body temperature within nonlethal limits, and to predict activity times of animals and whole-plant water and gas exchange. Cowan (1965) used electrical circuit analogues of flow of water from roots to leaves and out through

stomatal pores to predict the onset of wilting. More recently, plant physiologists have developed models to represent photosynthesis and carbon allocation at scales ranging from biogeochemical (Farquhar et al. 1980) to global. These models draw from studies of physiology, biophysics, and adaptation, and are important tools in theoretical and applied studies of plant biology. Similarly, a range of models exists for transpiration, many based on the Penman-Monteith formulation for surface-energy balance (Monteith 1973), but with many versions including more sophisticated biology. Models include relationships between carbon assimilation and water use based on optimization principles (Cowan and Farquhar 1977) or on isotope discrimination during carbon assimilation. These models can be used in applications ranging from crop production, through evolutionary studies of plant adaptation, to examination of the role of vegetation in global climate change.

Other work of considerable importance in this area, focusing on the relationship between the structure of an organism and its ability to function in its environment (see for example, McMahon and Kronauer 1976, Wainwright et al. 1976, Cheer and Koehl 1987, Vogel 1988), has already been discussed in Chapter 3.

4.1.3 Population Biology

Population modeling and population projection have been important parts of demography and ecology since the pioneering contributions of John Graunt (1662). Demographic methods have been applied to the study of human and nonhuman populations for centuries (see, for example, Keyfitz 1977), and they form the basis for population projections and for the understanding of the population consequences of life history phenomena (Cole 1954). These mathematical methods provide organizing principles for collecting and analyzing data on the rates of fertility and mortality. Such analyses are now commonplace in many areas of population biology and are applied to numerous species, ranging from humans to insects of economic importance (Keyfitz 1977, Carey, in press). The theory of age-structured populations, and the theories built on Leslie's matrix and the Perron-Frobenius operator theory, are among the most elegant and important advances in mathematical biology. Recent advances treat other aspects of population structure (e.g., Nisbet and Gurney 1982) and open population systems (e.g., Roughgarden et al. 1985).

The seminal work of Volterra and Lotka on predator-prey mechanisms showed how simple assumptions could lead to sustained oscillations of predator and prey populations. The predator-prey models of Volterra and Lotka are rarely taken literally. Yet they have formed the cornerstone of the subject, being the point of departure for more sophisticated models and stimulating both experimental studies of individual behavior and further mathematical studies of the bifurcation properties of systems of continuous time-differential equations.

Related closely to these predator-prey models are complementary models of competing organisms; again, the original models assume a simple quadratic form but have stimulated more sophisticated approaches. The theory of the ecological niche (see, for example, Whittaker and Levin 1975) and the associated theory of competitive exclusion, among the most influential concepts in community theory, derive in large part from the mathematical approaches. Other work dependent upon that theory has examined the limits to similarity and niche width of coexisting species (MacArthur 1972), studies of coevolution and character displacement

(Roughgarden 1979, Slatkin 1980, Fenchel and Christiansen 1977), and stochastic models of competition and predation.

One of the greatest successes of mathematical theory has been the application of diffusion models and their extensions to the spread of populations. The methods have been available, of course, for over a century (Skellam 1951), and early successes in the theory of epidemics occurred shortly after the turn of the century (Brownlee 1911). But the first major advances came from population genetics, especially the work of Haldane (1937) and Fisher (1937) and later work by Malécot (1969) and (in a discrete setting) Kimura and Weiss (1964), Maruyama (1977), and others.

Fisher modeled the spread of an advantageous gene through the use of diffusion-reaction equations, hypothesizing that, in the generic case, allelic spatial distributions would relax to ones characterized by fronts, spreading at the rate of twice the product of square root of the diffusion coefficient and the maximal selection coefficient. This remarkable insight—confirmed in simultaneous mathematical analyses by Kolmogorov et al. (1937)—has been a stimulus to much modern mathematical work (e.g., Aronson and Weinberger 1978, Bramson 1983). In ecology, there are direct analogues (Skellam 1951, Okubo 1980), and such models have been applied to study the rate of advance of invading species (Lubina and Levin 1988, Andow et al. 1990). Kareiva (1983), stimulated by the mathematical theory, examined the link between these population level descriptions and the individual movements of foraging insects.

Closely related to this work, and building upon it, has been the development of models to explain patchiness in the distribution of organisms (e.g., Segel and Jackson 1972, Steele 1978). This has stimulated research into critical patch size (Skellam 1951, Kierstead and Slobodkin 1953, Okubo 1980) and other mechanisms for generating and maintaining nonuniform spatial distributions (Levin 1979).

Evolutionary approaches to ecological problems have had a tremendous growth and influence over the past two decades. Maynard Smith (1982) applied theoretical approaches to evolutionary problems. Earlier, optimal foraging theory (Emlen 1966, MacArthur and Pianka 1966) linked behavior and optimization by the assumption that certain behaviors had been optimized by natural selection. Optimal foraging theory stimulated considerable biological research, including more than 100 empirical tests of the theory (through 1986, reviewed in Stephens and Krebs 1986). The most recent conceptual advance in this field involves the use of stochastic dynamic programming and computational methods to derive biological insights (Mangel and Clark 1988). This latter work shows one of the first instances in ecology (although common in physics and chemistry) of gaining biological insight through numerical computation.

Life history theory (Cole 1954) has been a fundamental and active area of research, providing a link between demographic and evolutionary theories. Problems of interest include senescence (evolution of the mortality schedule, Hamilton 1966), the timing of reproduction and tradeoffs with respect to mortality (Cole 1954, Caswell 1982), dispersal and dormancy (Cohen 1966, Cohen and Levin 1987), and density-dependent selection on equilibrium population sizes (Roughgarden 1979).

4.1.4 Epidemiology of Infectious Diseases

The mathematical theory of infectious diseases, pioneered by Ross, MacDonald, Kermack and McKendrick, and others, has been an important applied tool, especially for the establishment of

vaccination strategies. (See various papers in Levin et al. 1989.) Recently, Anderson and May (1979) and May and Anderson (1979) stimulated a renaissance of activity in this area, especially involving viral diseases such as influenza (Liu and Levin 1989, Castillo-Chavez et al. 1988, 1989a); rubella (Hethcote 1989a); myxoma (Dwyer et al. 1990); and AIDS (Anderson and May 1987, Castillo-Chavez 1989, Castillo-Chavez et al. 1989b).

Models of gonorrhea transmission were used to evaluate the effectiveness of strategies to combat the rapid rise in gonorrhea incidence in the United States in the 1960s. The initial step was the formulation and analysis of a simple model (Cooke and Yorke 1973), which was later extended to incorporate a "core" group of highly sexually active individuals. Tracing and treating the sexual contacts of members of the core group was shown to be a more cost-effective control than random screening of asymptomatic women (Hethcote and Yorke 1984). The work of Hethcote and Yorke has been one of the success stories of the application of mathematical models in epidemiology to influence management practice.

4.1.5 Fisheries Management

Fisheries management has proved a fertile area for the interaction of mathematics and biology. Fisheries managers recognized early that the problems involved were not only difficult, but could benefit considerably from a quantitative approach. The biological side has contributed concepts of nonlinear maps, such as the Ricker map. Many mathematical methods of optimal control and adaptive management (Clark 1985, Walters 1986) have been developed to solve problems in fisheries management. The recent work on nonclassical control problems by Clark (1985) was directly motivated by the problems of irreversible investment in fisheries. The methods developed by Clark, Walters, Ludwig, and their students and colleagues are currently applied worldwide to manage renewable resources.

A strong link also exists between fisheries management and evolutionary ecology. Although allozyme variation has been used for about 20 years in the study of evolutionary processes, in the last 10 years such variation also has been used to provide genetic "markers" that can be used to assess the composition of populations. This method, called Genetic Stock Identification, currently is used in Washington and California to determine the composition of oceanic mixtures of salmon in terms of the contributing source stocks. Because of the complexities of the analysis, the teams working on this problem always include biologists and mathematicians. The calculations are done by use of the EM algorithm (Dempster et al. 1977).

4.1.6 Community and Ecosystem Processes

Historically, the applications of mathematics to community- and ecosystem-level processes have been of two types: the simplistic dynamic approaches patterned after the Lotka-Volterra theory and the descriptive multivariate methods, of which Whittaker (1975) was the most important practitioner. Recently, however, a number of directions that blend theory and data have proved promising.

Understanding the causes of vegetation change has been an important long-term goal of ecology. A recent class of models has linked individual-based simulations of populations to models of detritus composition and nutrient release. Because species differ in the chemistry of their detritus, and because this difference influences decomposition and nutrient release, this

class of models exhibits a rich behavior that mimics real systems. It is becoming apparent that these models exhibit a rich array of dynamical behaviors, including deterministic chaos and multiple stable states. These models provide important information on plant community processes, constraints over selection and biogeochemistry. The development of succession/production/decomposition models is continuing with applications to paleobiology and global change.

One of the most important advances in community theory in the past decades has been the recognition of the patchy nature of most systems and the importance of spatially localized disturbances in maintaining diversity. The seminal paper here was Watt (1947), but its influence was negligible for a quarter of a century. More recent work in the marine intertidal (Levin and Paine 1974, Paine and Levin 1981), in forests (Pickett and White 1985), and in other systems has made this one of the most active areas of research in ecology.

A number of important studies in biogeochemistry have relied heavily on simulation models. Dynamic watershed models simulate water movement and biogeochemical reactions affecting soil and lake water chemistry, and have been central to integrated assessment of aquatic effects of acid deposition. They have been used as heuristic tools to improve understanding of watershed dynamics and as bases for projecting regional responses of watersheds to changes in acidic deposition. The comparison of the mathematical basis of these models; their calibration and application to watersheds that differ in size, slope, and geology; and the experiments that these models have stimulated have been significant components of the national integrated assessment of acid deposition effects. Similarly, the models of Parton et al. (1988), Pastor and Post (1988) and Schimel et al. (1990) have been used to analyze the effects of climate change on carbon and nitrogen biogeochemistry.

The early developments in ecosystem analysis also dealt with problems concerning the transfer of energy and materials among biota and their physico-chemical environment. The relevant models were composed of linear differential equations and, with the availability of computers, led to development of a suite of mathematical and simulation tools based on thermodynamics (Odum 1960), compartmental analysis (Patten 1971), and systems analysis (Watt 1966). The transfer of energy and nutrients among the biotic and abiotic components of ecosystems is one of the classic areas of application of mathematical models in ecology. Perhaps the most fruitful applications have been in nonlinear simulation models at levels from individuals (Botkin et al. 1972, Shugart and West 1977) to spatially explicit long-term ecosystem succession (Costanza et al. 1990).

Trophic webs describe the flow of energy among biological components in an ecological community and are of applied importance because they help predict, for example, how environmental toxins propagate through living species and which predators may help regulate weed species or pests. From the first monographs on food webs (Cohen 1978, Pimm 1982) have followed collections of hundreds of food webs (catalogued in machine readable form) from different habitats. These catalogues have led to the discovery of several new quantitative regularities, previously unsuspected, in the structure of food webs. These regularities, in turn, have led to the development and analysis of new mathematical models based on random directed graphs, which have made new and testable predictions about food web structure. A current general reference, the result of collaboration between an aquatic ecologist, a population biologist, and a mathematician, is Cohen et al. (1990).

4.2 GRAND CHALLENGES

In this section, we identify two grand challenges, among the many confronting mathematical ecology and evolutionary biology. The first, *global change*, includes relations to biodiversity and sustainable development of the biosphere (see, for example, Lubchenco et al. 1991), as well as global changes in the carbon cycle, climate, and the distribution of greenhouse gases. The second, *molecular evolution*, builds bridges between population biology and the problems of cellular and molecular biology, as discussed in Chapter 1.

4.2.1 Global Change

Global change, with its great implications for the future of our biosphere, presents one of the grandest challenges to computational biology. The proliferation of information from remote sensing, as well as more traditional ground surveys, introduces the need for geographical information systems that provide a framework for classifying information, spatial statistics for analyzing patterns, and dynamic simulation models that allow the integration of information across multiple spatial, temporal, and organizational scales. Multigrid techniques, parallel processing, and other advances will be essential tools in interfacing general circulation models with ecological models and will require substantive partnerships among physical scientists, biological scientists, and computational scientists.

The deficiencies of our knowledge about the patterns and processes of individuals, populations, and communities are serious enough even for static climatic conditions. But these shortcomings are magnified in any attempt to deal with long-term changes in global climate. Historical measures of production contain information on the variations in the climate, but the global increase in "greenhouse" gases portends a trend of unknown magnitude in climatic change. We are challenged to predict how such global changes will be reflected in the genetic structure of organisms, in biodiversity, in the behavior of individuals, in the recruitment and growth of populations, and in the behavior of communities, and we are challenged to develop strategies for mitigation and sustainable development. Understanding and dealing with the biological implications of global climate change, from every perspective, requires a significant new initiative. One of the central challenges, as discussed many times in this report for other problems and again in more detail below, is the development of approaches for dealing with and relating phenomena across disparate scales of space, time, and organizational complexity.

4.2.2 Molecular Evolution

Many challenging and important problems remain to be solved in the application of population genetic theory to molecular evolution. The existing methods of population genetics, such as the neutral theory, which were developed to describe variation at single loci, require restructuring to address questions that arise in the analysis of DNA sequence data. For example, the implications of tight but incomplete linkage among nucleotide sites within loci present a serious challenge. Molecular evolution is an area of rapid growth in the acquisition of sequence data as well as in theoretical development; it also has enormously important economic and political implications, ranging from the environmental release of genetically engineered organisms to improvements in biotechnology.

Although several methods are available for reconstructing phylogenies from sequence data (Cavalli-Sforza and Edwards 1967, Felsenstein 1981, Nei 1987), robust methods for assessing the reliability of the inferred phylogenies are not available. Realistic models of the evolutionary process that can form the basis for statistical inference are needed. Rapidly accumulating sequence data raise questions far ahead of current statistical methods. Progress in this area will be important for understanding the evolutionary relationships of virtually all organisms that lack a detailed fossil record, including bacteria and plants, as well as recently diverged human populations.

Molecular variation within populations and divergence between species contain information about the relative importance of evolutionary forces, including mutation, recombination, natural selection, migration, transposition, and gene conversion (e.g., Hughes and Nei 1988, Hudson 1990). Efficient methods of extracting this information and for testing alternative models are needed, particularly since large amounts of DNA sequence data are becoming available. The task of characterizing the properties of sequence variation expected under neutral models is under way, but alternative models with various forms of natural selection interacting with genetic drift are only beginning to be developed and explored. Gillespie (1989) has begun the analysis of the evolutionary process in a highly structured molecular landscape. Takahata and Nei (1990) have obtained some results for a model with many alleles maintained by overdominant selection and frequency-dependent selection. These studies indicate the possibility of progress, but much more effort in these directions is needed.

A promising area for further research, and one in which important progress already has been made, is in understanding molecular variation in populations by consideration of gene genealogies. This research was initiated by analysis of the coalescent process (Griffiths 1980, Kingman 1982) and extended by Tavare (1984), Watterson (1984), Kaplan et al. (1988), Tajima (1983), Takahata (1988), Slatkin (1989), and others. The analysis of measure-valued diffusions (Fleming and Viot 1979) represents another powerful approach for the study of multidimensional population genetic processes.

4.2.3 The Problem of Scale

An important factor motivating new developments in ecology is the expanding temporal and spatial scale of many critical environmental problems. Within a decade we have moved from forest and lake studies on the scale of tens of hectares, to acid precipitation and air pollutants operating on entire regions, to carbon dioxide problems on a global scale. The mathematical challenge will be to develop a theory of scale that can (1) guide the aggregation and extrapolation of fine-scale understanding to larger scales and (2) suggest hypotheses and methods for the direct investigation of large-scale phenomena.

Fundamentally new approaches to studies in population biology will be made possible by an understanding of phenomena that occur at different spatial and temporal scales. For example, genes express their effect at the individual level, but the effect of individual variation on population dynamics is poorly understood. Some recent successes in this area include the expression of genes in individuals and the role of individual behavior and variation in population dynamics (Mangel and Clark 1988). New approaches (i.e., theory, models and data) are needed to link subpopulations that are intermittently connected by stochastic events mediated by fluid flow (e.g., water, wind) and even plate tectonics. The key problems are to (1) determine the

"characteristic scales" for various ecological processes, (2) formulate the corresponding models that capture scale-dependent effects, and (3) test these models at the appropriate spatial and temporal scales.

Different dynamical characteristics are displayed by epidemiological systems, depending on the level of spatial aggregation of observations. At the individual level, stochastic effects are very important. In a small group, a disease may enter and quickly disappear. However, in cities and counties as a whole, persistence is more likely and the patterns of incidence appear more regular. For larger aggregations, deterministic models have proved to be useful. An understanding of the appropriate ways to link small-scale and large-scale epidemic behavior is important for understanding the impact of disease. Greater access to powerful computers will make it possible to study the relationships between different scales.

The development of new models and innovative mathematical and statistical methods for addressing the interaction of social dynamics and epidemiology—at distinct biological and sociological levels of aggregation and at distinct temporal and spatial scales—is a rapidly expanding area of research. Models and methods that follow the dynamics of pairs (or groups) of individuals in different "sociological spaces" are now being extensively studied.

Population biology and ecosystem ecology long have been disjunct subdisciplines. Challenges posed by environmental problems, including global change, are causing these two areas to pull together. Paleobiology, process studies, and theoretical examinations show that biogeochemical cycling imposes important resource constraints on populations. In return, patterns of resource use specific to populations, such as type of gaseous product, carbon element ratios and organic compounds, produce feedback to local and global element cycles. This linkage is central to our current understanding of plant populations dynamics, dynamics of species invasions, and marine biogeochemistry.

Linking population biology to biogeochemistry involves some major challenges to mathematical representation. For example, species of phytoplankton that have highly contagious distributions affect global air chemistry, possibly influencing global climate. Soil carbon levels change over time scales of hundreds to thousands of years, yet control soil nutrients that regulate plant growth and competition over short intervals. In addition, the interactions take place in a spatial context, which requires large input data sets for realistic simulation. Better theory, more powerful computations, and large-scale field studies are all required to achieve the coupling of these subdisciplines. The requirement to predict the effects of human use of ecosystems and global climate change makes this coupling essential.

Global problems ultimately must be studied at global scales. This is especially true when linking spatial and temporal scales in the study of oceanic processes and global climate change. For example, zooplankton respond to the spatial and temporal distribution of their food resources (phytoplankton) and their predators (planktivores), while the planktivores respond to the temporal and spatial distribution of the zooplankton and their predators (piscivores). In order to predict the patterns of these organisms, one must deal with spatial scales that range from millimeters or less (phytoplankton), centimeters (zooplankton size), meters (zooplankton aggregation size), tens of meters (planktivore school size), to kilometers (planktivore school group size, piscivore group size and movement scale). Each of these spatial scales has its own temporal scale (Okubo 1980).

The concept of self-similarity, derived from fractal geometry (Mandelbrot 1977), implies that extrapolation of information across scales is possible as long as the underlying process remains unchanged. However, ecological processes (e.g., energy flow, nitrogen exchange) are not always

self-similar at all scales, because processes often change abruptly between locations. Relatively uniform areas might be measured with a few samples and extrapolated to large regions with little error, while heterogeneous regions with complex gradients of soils, light, and moisture might produce major differences within a single watershed (i.e., not self-similar) and thus be difficult to extrapolate. The principal questions are: (1) How does one identify self-similar processes? (2) How can situations that are not self-similar be anticipated? (3) Can extrapolation methods be developed for these situations?

Many ecological processes occur in spatially patterned environments. Plant succession, biodiversity, foraging patterns, predator-prey interactions, dispersal, nutrient dynamics, and the spread of disturbance all have important spatial components. Many theoretical studies (e.g., Levin and Paine 1974, Steele 1974, Clark et al. 1978) have demonstrated the significance of spatial considerations in processes such as energy flow, nutrient cycling, and population growth rates. However, the difficulty of analyzing these processes often has caused the spatial dynamics to be ignored.

Models based on percolation theory (Stauffer 1985) have recently been used to relate the spatial distribution of resources to the propagation of disturbance (Turner et al. 1989) and the dynamics of species dispersal and habitat utilization (Gardner et al. in press, O'Neill et al. 1988). Other ideas from the theory of interacting particle systems are being applied to ecosystem problems. For instance, models that simulate the change in critical thresholds of disturbance propagation as a result of climatic change (i.e., drier forests), previous disturbance history, and the effects of human intervention will be useful for unravelling the issues associated with global change. Studies of the percolation "backbone" (a connected series of sites that transports material, energy, or organisms through a spatial system) may provide an objective view of critical habitats for the design and management of conservation areas.

Spatially explicit models can be very useful in addressing the problem of linking scales. The spatial resolution (grain) can be manipulated and changed in modeling studies. Evaluation of model predictions against spatial data available in geographical information systems allows the uncertainties of model predictions to be evaluated and key processes and parameters to be identified. It is expected that measurement of these parameters will significantly improve the accuracy and reliability of spatial predictions.

5

MODES AND LEVELS OF SUPPORT

THE potential for interactions between mathematics and biology can be developed only by careful nurturing. Some of the isolated interactions between the two disciplines have been discussed in a limited fashion in the Executive Summary. The interactions occur because the questions exist and because efforts are being made by individuals to promote interdisciplinary collaborations. But this interaction can be strengthened through attention to the modes and levels of support that will encourage such interactions. Specific recommendations follow.

5.1 RESEARCH SUPPORT

It is recommended that funds for the support of interdisciplinary research between biological and computational scientists and mathematicians be dramatically increased. Projects could take several forms, including the following.

- Projects involving interdisciplinary research by a single investigator
- Interdisciplinary groups of mathematicians, biologists, and computational scientists of sizes ranging from two individual investigators to networks of individuals from the different disciplines at different universities

Related to these projects is the further recommendation that specific guidelines for the review process of such proposals be considered. Databases containing the names of reviewers who have biological, mathematical, and computational expertise, or any combination of those skills, should be developed and made available to administrators at all involved federal agencies.

5.2 INFRASTRUCTURE

It is extremely important that adequate computer facilities and support of such facilities be provided. It is recommended that:

- Funding for computer facilities and support of those facilities be considered to be intrinsic to all awards made

- Funding for clearinghouses for software development, maintenance and distribution be made available
- Support be provided for networks for database access and network collaborations

5.3 TRAINING

Training in this interdisciplinary field must be regarded as a lifelong exercise. It must start early and continue for the professional lifetime of the scientist. Thus, recommendations for such continuing education are made for several levels of training.

5.3.1 Precollege and Undergraduate Education

Children, both in primary and secondary school, have a natural interest in biology. In the past, this has been limited largely to field biology or experiments that are chosen for minimal cost rather than the long-range view of building a base for exciting the child into consideration of quantitative biology as a career. While this remains perhaps the most natural avenue for arousing the curiosity of a child about the nature of biology, quantification must be introduced at an earlier stage. This is a natural area to show students projects and group efforts consistent with new mathematics curricula (see National Council of Teachers of Mathematics, Commission on Standards for School Mathematics 1989, Mathematical Sciences Education Board, National Research Council 1990). Few of these applications appear in textbooks, and most are absent entirely from the preparation of teachers. Our recommendations are to:

- Develop curriculum materials in mathematical biology for grades K through 12 and commit special teacher enhancement funds to introduce these materials to the nation's cadre of teachers
- Establish a program of summer internships for high school students and/or undergraduate students in which the students would spend two months working with mathematicians or biologists
- Support faculty at undergraduate institutions for training and research experiences that will further their knowledge and interest in the cross-discipline
- Support summer workshops developed for high school or undergraduate level faculty or/and students that focus on biological subdisciplines in which mathematics and computation play a large role. Instructors should be recruited from both disciplines. A paradigm might be the Computational Neurobiology Course at Woods Hole, MA
- Support workshops and conferences to develop methods for introducing significant quantitative tools to be introduced naturally into precollege and undergraduate biology curricula
- Support graduate students in applied mathematics and/or biology with an interest in the other discipline to work with high school and/or undergraduate students as teaching assistants or in other more imaginative ways to develop the younger students' interest

5.3.2 Graduate and Postdoctoral Training

The recommendations in this category are designed to improve the quantitative knowledge of biologists, to improve the biological knowledge of mathematicians, to facilitate ongoing collaborations, and to encourage new collaborations.

- Specifically target a substantial number of graduate fellowships in the biological sciences to individuals with undergraduate degrees in the mathematical or computer sciences and vice versa
- Support special cross-disciplinary postdoctoral fellowships that will allow Ph.D.s in one field to work in the other field
- Hold mini-courses, lasting four to eight weeks, in areas where both biological insight and mathematical or computational expertise are needed. Levels would be appropriate for graduate or postdoctoral students in one of the disciplines with more basic information in the cross-discipline

5.3.3 Senior Established Investigators

It is recognized that established scientists with expertise in biology, mathematics, and computational sciences are rare. It thus becomes important, at least initially, to encourage and facilitate efforts by scientists in one discipline to cross over into the other discipline to answer significant questions. The recommendations are to:

- Establish special mid-career fellowships for mathematical or computer scientists to join biological teams or individuals to enhance their biological insight and for biologists to work with mathematicians for varying lengths of time
- Support special visiting arrangements, both short- and long-term, be supported for scientists from one discipline to work with scientists from the other disciplines to encourage greater insight into the use of mathematics in biology

5.4 HUMAN RESOURCES

Several federal funding agencies already have a number of programs that encourage and seek out underrepresented groups (women, minorities, and persons with disabilities) in the sciences. This effort should continue that emphasis. All of the disciplines considered in this initiative have underrepresentation of minorities and people with disabilities. A significant number of biologists are women, but the number of female mathematicians decreases as the level of the degree increases. It is hoped that, as mathematical biology develops as a field, this statistic will change.

REFERENCES

Aldrich, R.W., D.P. Corey, and C.F. Stevens. 1983. A reinterpretation of mammalian sodium channel gating based on single channel recording. Nature 306:436–441.

Altman, R.B. and O. Jardetzky. 1989. The heuristic refinement method for the determination of the solution structure of proteins from NMR data. pp. 218–246 In Nuclear Magnetic Resonance, Part B: Structure and Mechanisms (N.J. Oppenheimer and T.L. James, eds.). Methods in Enzymology, Vol. 177. Academic Press, New York.

Altman, R.B., C. Arrowsmith, R. Pachter, and O. Jardetzky. 1991. Determination of large protein structures from NMR data: definition of the solution structure of the trp repressor. pp. 363–374 In Computational Aspects of the Study of Biological Macromolecules by NMR Spectroscopy (J.C. Hoch, F.M. Poulsen, and C. Redfield, eds.). Plenum, New York.

Anderson, R.M. and R.M. May. 1979. Population biology of infectious diseases. Part I. Nature 280:361–367.

Anderson, R.M. and R.M. May. 1987. Transmission dynamics of HIV infection. Nature 326:137–142.

Andow, D.A., P.M. Kareiva, S.A. Levin, and A. Okubo. 1990. Spread of invading organisms. Landscape Ecol. 4(2/3):177–188.

Aronson, D.G. and H.F. Weinberger. 1978. Multidimensional nonlinear diffusion arising in population genetics. Adv. in Math. 30(1):33–76.

Arratia, R., L. Goldstein, and L. Gordon. 1989. Two moments suffice for Poisson approximations: the Chen-Stein method. Ann. Probab. 17:9–25.

Arratia, R., L. Goldstein, and L. Gordon. 1990. Poisson approximations and the Chen-Stein method. Statistical Science 5:403–434.

Bagley, R.J., J.D. Farmer, S.A.. Kauffman, N.H. Packard, A.S. Perelson, and I.M. Stadnyk. 1989. Modeling adaptive biological systems. Biosystems 23:113–137.

Ball, F.G. and M.S.P. Samson. 1987. Temporal clustering of ion channel openings incorporating time interval omission. IMA J. Math. Appl. Med. & Biol. 4:333–361.

Bauer, W.R. 1978. Structure and reactions of closed duplex data. Ann. Rev. Biophys. Bioeng. 7:287–313.

Baum, L.E., T. Petrie, G. Soules, and N. Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Statist. 41:164–171.

Belmont, A.S., M.B. Braunfeld, J.W. Sedat, and D.A. Agard. 1989. Large-scale chromatin structural domains within mitotic and interphase chromasomes in vivo and in vitro. Chromasoma 98(2):129–143.

Berg, H.C. 1988. A physicist looks at bacterial chemotaxis. Cold Spring Harbor Symp. Quant. Biol. 53:1–9.

Berg, H.C. and E.M. Purcell. 1977. Physics of chemoreception. Biophys. J. 20:193–219.

Bertaud, D.S. and P.W. Gandar. 1986. A simulation model for cell proliferation in root apices. II. Patterns of cell proliferation. Ann. Bot. 58:303–320.

Boles, T.C., J.H. White, and N.R. Cozzarelli. 1990. Structure of plectonemically supercoiled DNA. J. Mol. Biol. 213:931–951.

Bollobás, B. 1985. Random Graphs. Academic Press, London, Orlando.

Botkin, D.B., J.F. Janak, and J.R. Wallis. 1972. Rationale, limitations, and assumptions of a northeastern forest growth simulator. IBM J. Res. Develop. 16:101–116.

Box, J.F. 1978. R.A. Fisher, The Life of a Scientist. Wiley, New York.

Bramson, M. 1983. Convergence of solutions of the Kolmogorov equation to travelling waves. Mem. Am. Math. 44(285):1–190.

Brillinger, D. R. 1988. Some statistical methods for random process data from seismology and neurophysiology. Ann. Statist. 16(1):1–54.

Brillinger, D.R. and J.P. Segundo. 1979. Empirical examination of the threshold model of neuron firing. Biol. Cybern. 35:213–220.

Brokaw, C.J. 1990. Descriptive and mechanistic models of flagellar motility. pp. 128–144 In Biological Motion (W. Alt and G. Hoffmann, eds.). Lect. Notes Biomath. 89. Springer-Verlag, Berlin.

Brownlee, J. 1911. The mathematical theory of random migration and epidemic distribution. Proc. R. Soc. Edinburgh 31:262–289.

Carey, J.R. (in press). Insect Population Demography. McGraw Hill, New York.

Castillo-Chavez, C. (ed.) 1989. Mathematical and Statistical Approaches to AIDS Epidemiology. Lect. Notes Biomath. 83, Springer-Verlag, Berlin.

Castillo-Chavez, C., H. Hethcote, V. Andreasen, S.A. Levin, and W-m. Liu. 1988. Cross-immunity in the dynamics of homogeneous and heterogeneous populations. pp. 303–316 In Mathematcal Ecology, Proc. of the Autumn Course Research Seminars, Trieste 1986 (T. Hallam, L. Gross, and S.A. Levin, eds.). World Scientific, Singapore.

Castillo-Chavez, C., H. Hethcote, V. Andreasen, S.A. Levin, and W-m. Liu. 1989a. Epidemiological models with age structure, proportionate mixing, and cross-immunity. J. Math. Biol. 27:233–258.

Castillo-Chavez, C. K. Cooke, W. Huang, and S.A. Levin. 1989b. On the role of long periods of infectiousness in the dynamics of acquired immunodeficiency syndrome (AIDS). Part 1. Single population models. J. Math. Biol. 27:373–398.

Caswell, H. 1982. Optimal life histories and the maximization of reproductive value: a general theorem for complex life cycles. Ecology 63:1218–1222.

Cavalli-Sforza, L.L. and A.W.F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Am. J. Hum. Genet. 19:233–257.

Cavalli-Sforza, L.L. and M.W. Feldman. 1981. Cultural Transmission and Evolution: A Quantitative Approach. Monographs in Population Biology 16. Princeton University Press, Princeton, NJ.

Cheer, A.Y.L. and M.A.R. Koehl. 1987. Paddles and rakes: fluid flow through bristled appendages of small organisms. J. Theor. Biol. 129(1):17–39.

Chou, P.A., T. Lookabaugh, and R.M. Gray. 1989. Optimal pruning with applications to tree-structured source coding and modeling. IEEE Trans. Inf. Theory 35:299–315.

Clark, C.W. 1985. Bioeconomic Modelling and Fisheries Management. Wiley, New York.

Clark, W.C., D.D. Jones, and C.S. Holling. 1978. Patches, movements, and population dynamics in ecological systems: a terrestrial perspective. pp. 385–432 In Spatial Pattern in Plankton Communities (J. H. Steele, ed.). Plenum Press, New York.

Claverie, J.-M., P. Kourilsky, P. Langlade-Demoyen, A. Chalfour-Prohnicka, G. Dadaglio, F. Tekais, F. Plata, and L. Bougueleret. 1988. T-immunogenic peptides are constituted of rare sequence patterns. Use in the identification of T epitopes in the human immunodeficiency virus gag protein. Eur. J. Immunol. 18:1547–1553.

Cohen, D. 1966. Optimizing reproduction in a randomly varying environment. J. Theor. Biol. 16:267–282.

Cohen, D. and S.A. Levin. 1987. The interaction between dispersal and dormancy strategies in varying and heterogeneous environments. pp. 110–122 In Mathematical Topics in Population Biology, Morphogenesis and Neurosciences, Proc. Kyoto 1985 (E. Teramoto and M. Yamaguti, eds.). Springer-Verlag, Berlin.

Cohen, J. E. 1978. Food Webs and Niche Space. Princeton University Press, Princeton, NJ.

Cohen, J. E. 1979. Ergodic theorems in demography. Bull. Am. Math. Soc. 1(2):275–295.

Cohen, J.E., F. Briand, and C.M. Newman. 1990. Community Food Webs — Data and Theory. Biomathematics 20. Springer-Verlag, Berlin.

Cole, L.C. 1954. The population consequences of life history phenomena. Q. Rev. Biol. 20:103–137.

Cooke, K. and J.A. Yorke. 1973. Some equations modeling growth processes and gonorrhea epidemics. Math. Biosci. 16:75–101.

Cormack, A.M. 1980. Early two-dimensional reconstruction and recent topics stemming from it. Science 209:1482–1486.

Costantino, R.F. and R.A. Desharnais. 1991. Population Dynamics and the Tribolium Model: Genetics and Demography. Monographs on Theoretical and Applied Genetics 13. Springer-Verlag, New York.

Costanza, R., F.H. Sklar, and M.L. White. 1990. Modeling coastal landscape dynamics. BioScience 40:91–107.

Cozzarelli, N.R. 1980. DNA gyrase and the supercoiling of DNA. Science 207:953–960.

Cowan, I.R.. 1965. Transport of water in the soil-plant-atmosphere system. J. Appl. Ecol. 2:221–239.

Cowan, I.R. and G.D. Farquhar. 1977. Stomatal function in relation to leaf metabolism and environment. Symp. Soc. Exp. Biol. 31:471–505.

Cramér, H. and H. Wold. 1936. Some theorems on distribution functions. J. London Math. Soc. 11:290–294.

DeGrado, W.F., Z.R. Wasserman, and J.D. Lear. 1989. Protein design, a minimalist approach. Science 243:622–628.

DeLisi, C. and J.A. Berzofsky. 1985. T-cell antigenic sites tend to be amphipathic structures. Proc. Nat. Acad. Sci. USA 82:7048–7052.

Dembo, M. 1989. Field theories of the cytoplasm. Comments, Theor. Biol. 1(3):159–177.

Dembo, M. and B. Goldstein. 1978. Theory of equilibrium binding of symmetric bivalent haptens to cell surface antibody: application to histamine release from basophils. J. Immunol. 121:345–353.

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Statist. Soc. Ser. B 39:1–38.

Depew, R.E. and J.C. Wang. 1975. Conformational fluctuations of DNA helix. Proc. Nat. Acad. Sci. USA 72:4275–4279.

Derrida, B. 1981. Random energy model: an exactly solvable model of disordered systems. Phys. Rev. B 24:2613.

Dickerson, R.E. 1989. Definitions and nomenclature of nucleic acid structure components. Nucleic Acids Research 17(5):1797–1803.

Durbin, R. and G. Mitchison. 1990. A dimension reduction framework for understanding cortical maps. Nature 343:644–647.

Dwyer, G., S.A. Levin, and L. Buttel. 1990. A simulation model of the population dynamics and evolution of myxomatosis. Ecological Monographs 60:423–447.

Edgar, B.A., G.M. Odell, and G. Schubiger. 1989. A genetic switch, based on negative regulation, sharpens stripes in Drosophila embryos. Devel. Genetics 10:124–142.

Emlen, J. 1966. The role of time and energy in food preference. Am. Nat. 100:611–617.

Erdos, P. and A. Renyi 1960. On the evolution of random graphs. Publ. Math. Inst., Hungar. Acad. Sci. 5:17–61.

Erickson, R.O. 1976. Modeling of plant growth. Ann. Rev. Plant Physiol. 27:407–434.

Erickson, R.O. and K.B. Sax. 1956. Elemental growth rate of the primary root of Zea mays. Proc. Amer. Philos. Soc. 100:487–498.

Ewens, W. 1972. The sample theory of selectively neutral alleles. Theor. Popul. Biol. 3:87–112

Farmer, J.D., N.H. Packard, and A.S. Perelson. 1986. The immune system, adaptation, and machine learning. Physica D 22:187–204.

Farquhar, G.D., S. von Caemmerer, and J.A. Berry. 1980. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. Planta 149:78–90.

Fatou, M.P. 1919. Sur les équations fonctionelles. Bull. Soc. Math. France 47:161–271.

Fatou, M.P. 1920a. Sur les équations fonctionelles. Deuxiéme memoire (Chapitre IV). Bull. Soc. Math. France 48:33–94.

Fatou, M.P. sur les équations fonctionelles. Troisiéme memoire (Chapitre IV). Bull. Soc. Math. France 48:208–314.

Feller, W. 1939. Die Grundlagen der Volterraschen Theorie des Kampfes ums dasein in wahrscheinlichkeitstheoretischer Behandlungen. Acta Biotheor. 5:11–40.

Feller, W. 1941. On the integral equation of renewal theory. Ann. Math. Stat. 12:243–267.

Feller, W. 1951. Diffusion processes in genetics. pp. 227–246 In Proc. 2nd Berkeley Symp. Math. Stat. Probab. Univ. of California Press, Berkeley.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368–376.

Fenchel, T. and F. Christiansen. 1977. Selection and interspecific competition. pp. 477–498 In Measuring Selection in Natural Populations (F.B. Christiansen and T.M. Fenchel, eds.). Lect. Notes Biomath. 19. Springer Verlag, Berlin.

Fisher, R.A. 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Fisher, R.A. 1937. The wave of advance of advantageous genes. Ann. Eugen. (London) 7:355–369.

Fitzhugh, R. 1960. Threshold "and plateaus in the Hodgkin-Huxley nerve equations. J. Gen. Physiol. 43:867–896.

Fitzhugh, R. 1969. Mathematical models of excitation and propagation in nerve. pp. 1–85 In Biological Engineering (H.P. Schwan, ed.). McGraw-Hill, New York.

Fleming, W.H. and M. Viot. 1979. Some measure-valued Markov processes in population genetics theory. Indiana Univ. Math. J. 28:817–843.

Fleshman, J.W., I. Segev, and R.E. Burke. 1988. Electrotonic architecture of type-identified alpha-motorneurons in the cat spinal cord. J. Neurophysiol. 60:60–85.

Fox, P.T., J.S. Perlmutter, and M.E. Raichle. 1985. Stereotactic localization for positron emission tomography. J. Cereb. Blood Flow Metab. 5 Suppl. 1:S611–S612.

Fraser, S.E. 1985. Cell interactions involved in neuronal patterning: an experimental and theoretical approach. pp. 481–507 In Molecular Bases of Neural Development (G.M. Edelman, W.E. Gall, and W.M. Cowan, eds.). Wiley, New York.

Fredkin, D.R. and J.A. Rice. 1986. On aggregated Markov chains. J. Appl. Probab. 23:208–214.

Furry, W.H. 1937. On fluctuation phenomena in the passage of high energy electrons through lead. Phys. Rev. 52:569–581.

Galton, F. and H.W. Watson. 1874. On the probability of extinction of families. J. R. Anthropol. Inst. 4:138–144.

Gardner, R.H., M.G. Turner, R.V. O'Neill, and S. Lavorel. (in press). Simulation of the scale-dependent effects of landscape boundaries on species persistence and dispersal. In The Role of Landscape Boundaries in the Management and Restoration of Changing Environments (M.M. Holland, P.G. Risser and R.J. Naiman, eds.). Chapman and Hall, New York.

Gates, D. 1965. Energy, plants and ecology. Ecology 46:1–13.

Geman, S. and D.E. McClure. 1985. Bayesian image analysis: an application to single photon emission tomography. pp. 12–18 In Proc. of the Am. Statistical Assoc. Statistical Computing Section.

Geman, S. and D.E. McClure. 1987. Statistical methods for tomographic image reconstruction. Bull. Int. Stat. Inst. 52(4):5–21.

Gillespie, J.H. 1978. A general model to account for enzyme variation in natural populations. V. The SAS-CFF model. Theor. Popul. Biol. 14:1–45.

Gillespie, J.H. 1989. Molecular evolution and polymorphism: SAS-CFF meets the mutational landscape. Am. Nat. 134:638–658.

Goldbeter, A. (ed.) 1989. Cell to Cell Signalling: From Experiments to Theoretical Models. Academic Press, New York.

Goldstein, B., C. Wofsy, and H. Echavarria-Heras. 1988. Effect of membrane flow on the capture of receptors by coated pits. Biophys. J. 53:405–414.

Graunt, J. 1662. Natural and political observations mentioned in a following index, and made upon the bills of mortality, with reference to the government, religion, trade, growth, ayre, diseases, and the several changes of the said city (London). Journal of the Institute of Actuaries 90, Part 1 (384):4–61.

Griffiths, R.C. 1979. Exact sampling distributions form the infinite neutral alleles model. Adv. Appl. Probab. 11:326–354.

Griffiths, R.C. 1980. Lines of descent in the diffusion approximation of neutral Wright-Fisher models. Theor. Popul. Biol. 17:37–50.

Gurtin, M.E. and R.C. MacCamy. 1974. Non-linear age-dependent population dynamics. Archive Rational Mech. Anal. 54:281–300.

Haldane, J.B.S. 1937. The effect of variation on fitness. Am. Nat. 71(735):337–349.

Hallam, T.G. and S.A. Levin (eds.) 1986. Mathematical Ecology: An Introduction. Biomathematics 17. Springer-Verlag, Berlin, Heidelberg.

Hamilton, W.D. 1964. The genetical evolution of social behaviour. I & II. J. Theor. Biol. 7:1-16, 17–52.

Hamilton, W. D. 1966. The modeling of senescence by natural selection. J. Theor. Biol. 12:12–45.

Hartline, H.K. and F. Ratliff. 1972. Inhibitory interaction in the retina of Limulus. pp. 381–448 In Handbook of Sensory Physiology, Vol. VII/2 (M.G.F. Fuortes, ed.), Springer-Verlag, Berlin.

Hethcote, H.W. 1989. Rubella. pp. 212–234 In Applied Mathematical Ecology (S.A. Levin, T.G. Hallam, and L.J. Gross, eds.). Biomathematics 18. Springer-Verlag, Berlin.

Hethcote, H.W. and J.A. Yorke. 1984. Gonorrhea Transmission Dynamics and Control. Lect. Notes Biomath. 56. Springer Verlag, New York.

Hodgkin, A.L. and A.F. Huxley. 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. 117:500–544.

Holmes, W. R. and W.B. Levy. 1990. Insights into associative long term potentiation from computational models of NMDA receptor-mediated calcium influx and intracellular calcium concentration changes. J. Neurophysiol. 63:1148–1168.

Hopfield, J.J. 1984. Neurons with graded response have collective computational properties like those of two-state neurons. Proc. Nat. Acad. Sci. U.S.A., 81:3088–3092.

Hoppensteadt, F. 1974. An age-dependent epidemic model. J. Franklin Inst. 297:325–333.

Hounsfield, G. 1980. Computed medical imaging. Nobel lecture, December 8, 1979. J. Comput. Assisted Tomogr. 4(5):665–674.

Howard, J., A.J. Hudspeth, and R.D. Vale. 1989. Movement of microtubules by single kinesin molecules. Nature 342:154–158.

Hudson, R.R. 1990. Gene genealogies and the coalescent process. In Oxford Surveys in Evolutionary Biology. Oxford University Press, Oxford (England).

Hughes, A.L. and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 355:167–170.

Huxley, A.F. 1957. Muscle structure and theories of contraction. Prog. Biophysics 7:235–318.

Inaba, H. 1988. A semigroup approach to the strong ergodic theorem of the multistate stable population process. Math. Popul. Stud. 1(1):49–77.

Jacob, F. 1977. Evolution and tinkering. Science 196:1161–1166.

Jagers, P. 1975. Branching Processes with Biological Applications. Wiley, New York.

Johnstone, I.M. and B.W. Silverman. 1990. Speed of emission in positron emission tomography and related inverse problems. Ann. Statist. 18:251–280.

Julia, C. 1918. Memoire sur l'iteration des fonctions rationelles. J. Math. Pures Appl. 4:47–245.

Kallianpur, G. and R. Wolpert. 1987. Weak convergence of stochastic neuronal models. pp. 116–145 In Stochastic Methods in Biology (M. Kimura, G. Kallianpur and R. Hida, eds.). Lect. Notes Biomath. 70, Springer-Verlag, Heidelberg.

Kandel, E.R. 1984. Steps toward a molecular grammar for learning: explorations into the nature of memory. pp. 555–604 In Medicine, Science and Society. Symposia Celebrating the Harvard Medical School Bicentennial (K.J. Isselbacher, ed.). Wiley, New York.

Kaplan, N.L., T. Darden, and R.R. Hudson. 1988. The coalescent process in models with selection. Genetics 120:819–829.

Kareiva, P.M. 1983. Local movement in herbivorous insects: applying a passive diffusion model to mark-recapture field experiments. Oecologia 57:322–327.

Kauffman, S.A. 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. 22:437–467.

Kauffman, S.A. 1974. The large scale structure and dynamics of gene control circuits: an ensemble approach. J. Theor. Biol. 44:167–190.

Kauffman, S.A. (in press). The Origins of Order: Self Organization and Selection in Evolution. Oxford University Press, Oxford (England).

Keizer, J. 1988. Electrical activity and insulin release in pancreatic beta cells. Math. Biosci. 90:127–138.

Keller, E.F., and L.A. Segel. 1970. The initiation of slime mold aggregation viewed as an instability. J. Theor. Biol. 26:399–415.

Kermack, W.O. and A.G. McKendrick. 1927. Contributions to the mathematical theory of epidemics. Proc. R. Soc. Lond. Ser A 115:700–721.

Kersten, D. 1990. Statistical limits to image understanding. In Vision: Coding and Efficiency (C. Blakemore, ed.). Cambridge University Press, Cambridge (England).

Kevrekidis, I.G., A.D. Zecha, and A.S. Perelson. 1988. Modeling dynamical aspects of the immune response. I. T cell proliferation and the effect of IL–2. pp. 167–197 In Theoretical Immunology, Part One, SFI Studies in the Sciences of Complexity (A.S. Perelson, ed.). Addison-Wesley, Redwood City, CA.

Keyfitz, N. 1977. Introduction to the Mathematics of Populations (with revisions). Addison Wesley, Reading, MA.

Kierstead, H. and L.B. Slobodkin. 1953. The size of water masses containing plankton bloom. J. Mar. Res. 12:141–147.

Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge (England).

Kimura, M. and G.H. Weis. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576.

Kingman, J.F.C. 1982. On the genealogy of large populations. J. Appl. Probab. 19A:27–43.

Klevit, R.E. 1991. Recognition of DNA by Cys2, His2 zinc fingers. Science 253(5026):1367, 1393.

Koch, C. and I. Segev (eds.). 1989. Methods in Neuronal Modeling: From Synapses to Networks. MIT Press, Cambridge, MA.

Koch, C., V. Torre, and T. Poggio. 1986. Computations in the vertebrate retina: motion discrimination gain enhancement and differentiation. Trends Neurosci. 9(5):204–211.

Kolmogorov, A.M. 1959. The transition of branching processes into diffusion processes and associated problems of genetics. Teor. Veroyatn. i Primen. 4:233–236.

Kolmogorov, A., I. Petrovsky, and N. Piscounov. 1937. Etude de l'equation de la diffusion avec croissance de la quantite de matiere et son application a un probleme biologique. Moscow Univ. Bull. Ser. Internat. Sec. A 1:1–25.

Lander, E.S. and D. Botstein. 1986. Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. Proc. Nat. Acad. Sci. 83:7353–7357.

Layton, H.E. 1989. Concentrating urine in the inner medula of the kidney. Comments, Theor. Biol 1:179–196.

Lee, M.S., J.M. Gottesfeld, and P.E. Wright. 1991. Zinc is required for folding and binding of a single zinc finger to DNA. FEBS Letters 279(2):289–294.

Levin, S.A. 1979. Non-uniform stable solutions to reaction-diffusion equations: applications to ecological pattern formation. pp. 210–222 In Pattern Formation by Dynamic Systems and Pattern Recognition (H. Haken, ed.). Springer-Verlag, Berlin.

Levin, S.A. and R.T. Paine. 1974. Disturbance, patch formation, and community structure. Proc. Nat. Acad. Sci. 71:2744–2747.

Levin, S.A., T.G. Hallam, and L.J. Gross (eds.). 1989. Applied Mathematical Ecology. Biomathematics 18.

Levin, S.A., M.A. Harwell, J.R. Kelly, and K.D. Kimball (eds.). 1989. Ecotoxicology: Problems and Approaches. Springer-Verlag, New York.

Linderman, J.J. and D.A. Lauffenburger. 1988. Analysis of intracellular receptor/ligand sorting in endosomes. J. Theor. Biol. 132:203–245.

Linsker, R. 1990. Perceptual neural organization: some approaches based on network models and information theory. Annu. Rev. Neurosci. 13:257–281.

Lipman, D.J., and W.R. Pearson. 1985. Rapid and sensitive protein similarity searches. Science 227:1435–1441.

Liu, W.-m. and S.A. Levin. 1989. Influenza and some related mathematical models. pp. 235–252 In Applied Mathematical Ecology (S.A. Levin, T.G. Hallam and L.J. Gross, eds.). Biomathematics 18. Springer-Verlag, Heidelberg.

Llinas, R.R. 1988. The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. Science 242:1654–1664.

Lockery, S.K., G. Wittenberg, W.B.J. Kristan, and G.W. Cottrell. 1989. Function of identified interneurons in the leech elucidated using neural network trained by back-propagation. Nature 340:468–471.

Logan, B.F. and L.A. Shepp. 1975. Optimal reconstruction of a function from its projections. Duke Math. J. 42:645–659.

Lubchenco, J., A.M. Olson, L.B. Brubaker, S.R. Carpenter, M.M. Holland, S.P. Hubbell, S.A. Levin, J.A. MacMahon, P.A. Matson, J.M. Melillo, H.A. Mooney, C.H. Peterson, H.R. Pulliam, L.A. Real, P.J. Regal, and P.J. Risser. 1991. The sustainable biosphere initiative: an ecological research agenda. Ecology 72(2):317–412.

Lubina, J. and Levin, S.A. 1988. The spread of a reinvading organism: range expansion of the California sea otter. Am. Nat. 131(4):526–543.

MacArthur, R.H. 1972. Geographical Ecology — Patterns in the Distribution of Species. Harper and Row, New York.

MacArthur, R.H. and E. Pianka. 1966. On the optimal use of a patchy environment. Am. Nat. 100:603–609.

Macken, C.A. and A.S. Perelson. 1985. Branching Processes Applied to Cell Surface Aggregation Phenomena. Lect. Notes Biomath. 58. Springer-Verlag, Berlin.

McKendrick, A.G. 1914. Studies on the theory of continous probabilities, with special reference to its bearing on natural phenomena of a progressive nature. Proc. London Math. Soc. 2 13:401–416.

McKendrick, A.G. 1926. Application of mathematics to medical problems. Proc. Edinburgh Math. Soc. 44:98–130.

McMahon, T.A. and R.E. Kronauer. 1976. Tree structures: deducing the principle of mechanical design. J. Theor. Biol. 59:443–466.

McQueen, D.M. and C.S. Peskin. 1983. Computer-assisted design of pivoting-disc prosethetic mitral valves. J. Thorac. Cardiovasc. Surg. 86:126–135.

McQueen, D.M. and C.S. Peskin. 1986. Computer-assisted design of butterfly bileaflet valves for the mitral position. Scand. J. Thorac. Cardiovasc. Surg. 19:139–148.

McQueen, D.M. and C.S. Peskin. 1989. A three-dimensional computational model for blood flow in the heart. II. Contractile fibers. J. Comput. Phys. 82:289–297.

Malécot, G. 1969. The Mathematics of Heredity. W.H. Freeman, San Francisco, CA.

Mandelbrot, B.B. 1977. Fractals, Form, Chance, and Dimension. W.H. Freeman, San Francisco, CA.

Mangel, M. and C.W. Clark. 1988. Dynamic Modeling in Behavioral Ecology. Princeton University Press, Princeton, NJ.

Mathematical Sciences Education Board, National Research Council. 1990. Reshaping School Mathematics: A Philosophy and Framework for Curriculum. National Academy Press, Washington, D.C.

Maruyama, M. 1977. Stochastic Problems in Population Genetics. Lect. Notes Biomath. 17. Springer-Verlag, Berlin.

May, R.M. (ed.). 1981. Theoretical Ecology, Principles and Applications. Second Edition. Blackwell Scientific, Oxford (England).

May, R.M. and R.M. Anderson. 1979. Population biology of infectious diseases. Part II. Nature 280:455–461.

Maynard Smith, J. 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge (England).

Mayr, E. 1982. The Growth of Biological Thought. Belknap Press of Harvard University, Cambridge, MA.

Mayr, E. 1988. Toward a New Philosophy of Biology: Observations of an Evolutionist. Belnap Press of Harvard University Press, Cambridge, MA.

Metz, J.A.J. and O. Diekmann (eds.). 1986. Dynamics of Physiologically Structured Populations. Lect. Notes Biomath. 68. Springer-Verlag, Heidelberg.

Miller, K.D., J.B. Keller, and M.P. Stryker. 1989. Ocular dominance column development: analysis and simulations. Science 245:605–615.

Milne, R. K., G.F. Yeo, B.W. Madsen, and R.O. Edenson. 1989. Estimation of single channel kinetic parameters from data subject to limited time resolution. Biophys. J. 55:673–676.

Monteith, J.L. 1973. Principles of Environmental Physics. American Elsevier, New York.

Motz, L. 1987. Introduction. pp. 17–27 In The World of Physics (J.H. Weaver, ed.). Simon and Schuster, New York.

Murray, A.W. and M.W. Kirschner. 1989. Dominoes and clocks: the union of two views of the cell cycle. Science 246:614–621.

Murray, J.D. 1989. Mathematical Biology. Biomathematics 19. Springer-Verlag, Berlin.

Murray, J.D. and G.F. Oster. 1984. Cell traction models for generating pattern and form in morphogenesis. J. Math. Biol. 19:265–279.

Myrberg, P.J. 1963. Iteration der reellen polynome zweiten grades III. Ann. Acad. Sci. Fenn. 336/3:1–18.

National Council of Teachers of Mathematics, Commission on Standards for School Mathematics. 1989. Curriculum and Evaluation Standards for School Mathematics. NCTM, Reston, VA.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Neyman, J. 1945 and subsequent. Proc. Berkeley Symp. Math. Stat. Probab. Univ. of California Press, Berkeley.

Neyman, J. and P.S. Puri. 1981. A hypothetical stochastic mechanism of radiation effects in single cells. Proc. R. Soc. Lond. 213:139–160.

Neyman, J., T. Park, and E.L. Scott. 1956. Struggle for existence. The Tribolium model: biological and statistical aspects. Proc. 3rd Berkeley Symp. Math. Stat. Probab., Vol IV:41–79.

Nisbet, R.M. and W.S.C. Gurney. 1982. Modelling Fluctuating Populations. Wiley, Chichester (England).

Noble, D. 1962. A modification of the Hodgkin-Huxley equation applicable to Purkinje fibre action and pacemaker potentials. J. Physiol. 160:316–352.

Odell, G.M. and L.A. Segel. 1987. BIOGRAPH: A Graphical Simulation Package with Exercises. To accompany Lee A. Segel's Modelling Dynamic Phenomena in Molecular and Cellular Biology. Cambridge University Press, Cambridge (England).

Odum, H.T. 1960. Ecological potential and analogue circuits for the ecosystem. Am. Sci. 48:1–8.

Okubo, A. 1980. Diffusion and Ecological Problems: Mathematical Models. Springer Verlag, New York.

O'Neill, R.V., B.T. Milne, M.G. Turner, and R.H. Gardner. 1988. Resource utilization scales and landscape pattern. Landscape Ecol. 2:63–69.

Oster, G.F. and J.D. Murray. 1989. Pattern formation models and developmental constraints. J. Exp. Zool. 251:186–202.

Paine, R.T. and S.A. Levin. 1981. Intertidal landscapes: disturbance and the dynamics of pattern. Ecol. Monogr. 51:145–178.

Parnas, H., I. Parnas, and L.A. Segel. 1991. On the contribution of mathematical models to the understanding of neurotransmitter release. Int. Rev. Neurobiol. 32:1–50.

Parton, W.J., J.W.B. Stewart, and C.V. Cole. 1988. Dynamics of C, N, P, and S in grassland soils: a model. Biogeochem. 5:109–131.

Pastor, J. and W.M. Post. 1988. Response of northern forests to CO2-induced climate change. Nature 334:55–58.

Patten, B.C. (ed.) 1971. Systems Analysis and Simulation in Ecology. Vol. 1. Academic Press, New York.

Perelson, A.S. 1984. Some mathematical models of receptor clustering by multivalent ligands. pp. 223–276 In Cell Surface Dynamics: Concepts and Models (A.S. Perelson, C. DeLisi and F.W. Wiegel, eds.). Marcel Dekker, New York.

Perelson, A.S. 1989. Immune network theory. Immunol. Rev. 5:36.

Perelson, A.S. and C. DeLisi. 1980. Receptor clustering on a cell surface. I. Theory of receptor cross-linking by ligands bearing two chemically identical functional groups. Math. Biosci. 48:71–110.

Perelson, A.S., M. Mirmirani, and G.F. Oster. 1976. Optimal strategies in immunology. I. B-cell differentiation and proliferation. J. Math. Biol. 3:325–367.

Perelson, A.S., M. Mirmirani, and G.F. Oster. 1978. Optimal strategies in immunology. II. B-memory cell production. J. Math. Biol. 5:213–256.

Peskin, C.S. and D.M. McQueen. 1989. A three-dimensional computational model for blood flow in the heart. I. Immersed elastic fibers in a viscous incompressible fluid. J. Comput. Phys. 81:372–402.

Pickett, S.T.A. and P.S. White. 1985. The Ecology of Natural Disturbance and Patch Dynamics. Academic Press, Orlando, FL.

Pimm, S. 1982. Food Webs. Chapman and Hall, New York.

Poggio, T. and F. Girosi. 1990. Regularization algorithms for learning that are equivalent to multilayer networks. Science 247:978–982.

Porter, T.M. 1986. The Rise of Statistical Thinking 1820–1900. Princeton University Press, Princeton, NJ.

Porter, W.P. and C.R. Tracy. 1973. Biophysical analyses of energetics, time-space utilization, and distributional limits. pp. 55–81 In Lizard Ecology: Studies of a Model Organism (R.B. Huey, E. Pianka, and T.H. Schoener, eds.). Harvard University Press, Cambridge, MA.

Provine, W.B. 1971. The Origins of Theoretical Population Genetics. The University of Chicago Press, Chicago.

Purcell, E.M. 1977. Life at low Reynolds number. Am. J. Physics. 45(1):3–11.

Radon, J. 1917. Uber die Bestimmung von Funktionen durch ihre integralwerte langs gewisser Mannigfaltigkeiten. Berichte Sachsische Akademie der Wissenschaften, Leipzig, Math.-Phys. Kl. 69:262–267.

Rall, W. 1962. Theory of physiological properties of dendrites. Ann. N.Y. Acad. Sci. 96:1071–1092.

Rall, W. 1964. Theoretical significance of dendritic trees for neuronal input-output relations. pp. 73–97 In Neural Theory and Modeling (R.F. Reiss, ed.). Stanford University Press, Stanford, CA.

Raschke, K. 1960. Heat transfer between the plant and the environment. Ann. Rev. Plant Physiol. 11:111–126.

Richards, O.W. and A.J. Kavanagh. 1943. The analysis of the relative growth gradients and changing form of growing organisms: illustrated by the tobacco leaf. Am. Nat. 77:385–399.

Rinzel, J. 1990. Discussion: electrical excitability of cells, theory and experiment: review of the Hodgkin-Huxley foundation and an update. In Classics of Theoretical Biology (Part One) (M. Mangel, guest editor), Special Issue, Bull. Math. Biol. 52(1/2):5–23.

Rivero, M.A., R.J. Tranquillo, H.M. Buettner, and D.A. Lauffenburger. 1989. Transport models of chemotactic cell populations based on individual cell behavior. Chem. Eng. Sci. 44:2881–2897.

Rothbard, J.B. and W.R. Taylor. 1988. A sequence pattern common to T cell epitopes. EMBO J. 7:93–100.

Roughgarden, J. 1979. Theory of Population Genetics and Evolutionary Ecology: An Introduction. Macmillan, New York.

Roughgarden, J., Y. Iwasa, and C. Baxter. 1985. Demographic theory for an open marine population with space-limited recruitment. Ecology 66:54–67.

Rumelhart, D.E., J.L. McClelland, and the PDP Group. 1986. Parallel Distributed Processing — Explorations in the Microstructure of Cognition. MIT Press, Cambridge, MA.

Samuelson, P.A. 1976. Resolving a historical confusion in population analysis. Hum. Biol. 48:559–580.

Sarkovskii, A.N. 1964. Coexistence of cycles of a continuous map of a line into itself. Ukr. Mat. Zh. 16:61–71.

Schimel, D.S., W.J. Parton, T.G.F. Kittel, D.S. Ojima, and C.V. Cole. 1990. Grassland biogeochemistry: links to atmospheric processes. Climate Change 17:13–25.

Segel, L.A. and J.L Jackson. 1972. Dissipative structure: an explanation and an ecological example. J. Theor. Biol. 37:545–559.

Segel, L.A. and A.S. Perelson. 1989. Shape space: an approach to the evaluation of cross-reactivity effects, stability and controllability in the immune system. Immunol. Lett. 22:91–99.

Selverston, A.I. and M. Moulins. 1985. Oscillating neural networks. Ann. Rev. Neurophysiol. 47:29–48.

Sereno, M.E., D.J. Kersten, and J.A. Anderson. 1988. A neural network model of an aspect of motion perception. Consortium for Scientific Computing, Science at the John von Neumann National Supercomputer Center, pp. 173–178.

Sharpe, F.R. and A.J. Lotka. 1911. A problem in age-distribution. Philos. Mag. Ser. 6, 21:243–267.

Shepp, L.A. and B.F. Logan. 1974. The Fourier reconstruction of a head section. IEEE Trans. Nucl. Sci. NS-21:21–43.

Sherman, A., J. Rinzel, and J. Keizer. 1988. Emergence of organized bursting in clusters of pancreatic B-cells by channel sharing. Biophys. J. 54:411–425.

Shugart, H.H. and D.C. West. 1977. Development of an Appalachian deciduous forest succession model and its application to assessment of the impact of the chestnut blight. J. Environ. Managemt. 5:161–179.

Silk, W.K. 1984. Quantitative descriptions of development. Ann. Rev. Plant Physiol. 35:479–518.

Silk, W.K. 1989. Growth rate patterns which maintain a helical tissue tube. J. Theor. Biol. 138:311–327.

Skellam, J.G. 1951. Random dispersal in theoretical populations. Biometrika 38:196–218.

Slatkin, M. 1980. Ecological character displacement. Ecology 61:163–178.

Slatkin, M. 1989. Detecting small amounts of gene flow from phylogenies of alleles. Genetics 121:609–612.

Smith, T.F. and M.S. Waterman. 1981. The identification of common molecular subsequences. J. Molec. Biol. 147:195–197.

Stauffer, D. 1985. Introduction to Percolation Theory. Taylor and Francis, London.

Steele, J.H. 1974. Spatial heterogeneity and population stability. Nature 248(5443):83.

Steele, J.H. (ed.). 1978. Spatial Pattern in Plankton Communities. NATO Conference Series, Series IV: Marine Sciences, Vol. 3. Plenum, New York.

Stephens, D.W. and J.R. Krebs. 1986. Foraging Theory. Princeton University Press, Princeton, NJ.

Stephenson, J. L. 1972. Central core model of the renal counterflow system. Kidney Int. 2:85–94.

Stigler, S.M. 1986. The History of Statistics. The Measurement of Uncertainty Before 1900. Belknap Press of Harvard University Press, Cambridge, MA.

Summers, M.F. 1991. Zinc finger motif for single-stranded nucleic acids? Investigations by nuclear magnetic resonance. J. Cell. Biochem. 45(1):41–48.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

Takahata, N. 1988. The coalescent in two partially isolated diffusion populations. Genetics Research 52:213–222.

Takahata, N. and M. Nei. 1990. Allelic genealogy under overdominant and frequency dependent selection and polymorphism of major histocompatability complex loci. Genetics 124:967–968.

Tavare, S. 1984. Line of descent and genealogical processes, and their applications in population genetic models. Theor. Popul. Biol. 26:119-164.

Thom, R. 1975. Structural Stability and Morphogenesis. W.A. Benjamin, Reading, MA.

Travers, A.A. and A. Klug. 1987. The bending of DNA in nucleosomes and its wider implications. Philos. Trans. R. Soc. Lond. B 317:537–561.

Treisman, A., P. Cavanagh, B. Fischer, V.S. Ramachandran, and R. von der Heydt. 1990. Form perception and attention. pp. 273–316 In Visual Perception: The Neurophysiological Foundations (L. Spillman and J.S. Werner, eds.). Academic Press, San Diego, CA.

Tuckwell, H. 1988. Stochastic processes in the neurosciences. CBMS-NSF Regional Conference Ser. N.56, Society for Industrial and Applied Mathematics, Philadephia, PA.

Turing, A.M. 1952. The chemical basis of morphogenesis. Philos. Trans. R. Soc. Lond. B 237:37–72.

Turner, M.G., R.H. Gardner, V.H. Dale, and R.V. O'Neill. 1989. Predicting the spread of disturbance across heterogeneous landscapes. Oikos 55:121–129.

Uyenoyama, M. and M.W. Feldman. 1980. Theories of kin and group selection: a population genetics perspective. Theor. Popul. Biol. 17:380–414.

Vardi, Y., L.A. Shepp, and L. Kaufman. 1985. A statistical model for positron emission tomography (with discussion). J. Am. Statist. Assoc. 80:8–37.

Vogel, S. 1988. Life's Devices — The Physical World of Animals and Plants. Princeton University Press, Princeton, NJ.

von der Malsburg, C. 1973. Self-organization of orientation sensitive cells in the striate cortex. Kybernetik 14:85–100.

Waddington, C.H. 1957. The Strategy of the Genes. Allen and Unwin, New York.

Wainwright, S.A., W.D. Biggs, J.D. Currey, and J.M. Gosline. 1976. Mechanical Design in Organisms. Princeton University Press, Princeton, NJ.

Walters, C. 1986. Adaptive Management of Renewable Resources. Macmillan, New York.

Wang, J.C., L.F. Peck, and K. Becherer. 1983. DNA supercoiling and its effects on DNA structure and function. Quant. Biol. 47:85–91.

Wasserman, S.A. and N.R. Cozzarelli. 1986. Biochemical topology: applications to DNA recombination and replication. Science 232:951–960.

Watt, A.S. 1947. Pattern and process in the plant community. J. Ecol. 35:1-22.

Watt, K.E.F. (ed.). 1966. Systems Analysis in Ecology. Academic Press, New York.

Watterson, G.A. 1977. Heterosis or neutrality? Genetics 85:789–814.

Watterson, G.A. 1984. Lines of descent and the coalescent process. Theor. Popul. Biol. 26:77–92.

Weinstein, A.M. and E.E. Windhager. 1985. Sodium transport along the proximal tubule. pp. 903–932 In The Kidney: Physiology and Pathophysiology (D. Seldin and G. Giebisch, eds.). Raven Press, New York.

Weisbuch, G., R.J. De Boer, and A.S. Perelson. 1990. Localized memories in idiotypic networks. J. Theor. Biol. 146:483–499.

White, J.H. 1969. Self-linking and the Gauss integral in higher dimensions. Am. J. Math 91:693–728.

White, J.H., N.R. Cozzarelli, and W.R. Bauer. 1988. Helical repeat and linking number of surface–wrapped DNA. Science 241:323–327.

White, J.H., K.C. Millett, and N.R. Cozzarelli. 1987. Description of the topological entanglement of DNA catenanes and knots by a powerful method involving strand passage and recombination. J. Mol. Biol. 197:585–603.

Whitelaw, V.A. and J.D. Cowan. 1981. Specificity and plasticity of retinotectal connections: a computational model. J. Neurosci. 1:1369–1387.

Whittaker, R.H. 1975. Communities and Ecosystems. Macmillan, New York.

Whittaker, R.H. and S.A. Levin (eds.). 1975. Niche: Theory and Application. Benchmark Papers in Ecology/3. Dowden, Hutchinson & Ross, Stroudsburg, PA.

Winfree, A.T. 1990. Stable particle-like solutions to the nonlinear wave equations of three-dimensional excitable media. SIAM Rev. 32:1–53.

Wolpert, L. 1969. Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol. 25:1–47.

Wolpert, L. and A. Hornbruch. (in press). Double anterior chick limb buds and models for cartilage rudiment specification. Development.

Yang, G.L. and C.E. Swenberg. 1991. Stochastic modeling of dose-response for single cells in radiation experiments. Math. Scientist 16:46–65.

Yang, G.L. and C.E. Swenberg. (in press). Estimation of open dwell time and problems of identifiability in channel experiments. J. Statist. Plann. Inference.

Yoshida, T. and M. Dembo. 1990. A thermodynamic model of hemoglobin suitable for physiological applications. Am. J. Phys. 258:C563–C577.

Yule, G.U. 1924. A mathematical theory of evolution based upon the conclusions of Dr J.C. Willis, FRS. Philos. Trans. R. Soc. Lond. B 213:21–87.

Zucker, R.S. and P.G. Haydon. 1988. Membrane potential has no direct role in evoking neurotransmitter release. Nature 335:360–362.

Appendix 1

STEERING COMMITTEE

Peter Arzberger (NSF Liaison)
Statistics and Probability
National Science Foundation
1800 G Street, N.W.
Washington, DC 20550

Nicholas Cozzarelli
Dept. of Molecular Biology
229 Stanley Hall
University of California, Berkeley
Berkeley, CA 94720

Robert Eisenberg
Dept. of Physiology
Rush Medical College
1750 West Harrison St.
Chicago, IL 60612

Marcus W. Feldman
Dept. of Biological Sciences
Stanford University
Herrin Hall
Stanford, CA 94305

Daniel Kersten
Dept. of Psychology
Elliott Hall
University of Minnesota
75 E. River Rd.
Minneapolis, MN 55455

Simon A. Levin (Chairman)
Section of Ecology and Systematics
347 Corson Hall
Cornell University
Ithaca, NY 14853

Deborah Lockhart (NSF Liaison)
Special Projects Program
National Science Foundation
1800 G Street, N.W.
Washington, DC 20550

Andre Manitius
Electrical and Computer Engineering 4400
University Drive
George Mason University
Fairfax, VA 22030

Colleen Martin (editorial assistant)
Ecology and Systematics
347 Corson Hall
Cornell University
Ithaca, NY 14853

Alan S. Perelson
Los Alamos National Laboratory
Mail Stop K710
Los Alamos, NM 87545

Charles Peskin
Courant Inst. of Mathematical Sciences
251 Mercer St.
New York, NY 10012

Michael Reed
Dept. of Mathematics
136 Physics Bldg.
Duke University
Durham, NC 27706

Wendy Silk
Dept. of Land, Air, and
Water Resources
Hoagland Hall
University of California, Davis
Davis, CA 95616

John Wooley (NSF Liason)
Biological Instrumentation and Resources
National Science Foundation
1800 G Street, NW
Washington, DC 20550

Appendix 2

WORKSHOP ATTENDEES

Joan Aron
Dept. of Population Dynamics
Johns Hopkins School of Hygiene
and Public Health
615 N. Wolfe St.
Baltimore, MD 21205

Peter Arzberger (NSF Liaison)
Statistics and Probability
National Science Foundation
1800 G Street, N.W.
Washington, DC 20550

Howard Berg
Dept. of Cellular and
Developmental Biology
Harvard University
16 Divinity Ave.
Cambridge, MA 02138

Carlos Castillo-Chavez
Dept. of Plant Breeding
and Biometry
341 Warren Hall
Cornell University
Ithaca, NY 14853

Joel E. Cohen
Rockefeller University
1230 York Ave., Box 20
New York, NY 10021

Robert Costanza
Chesapeake Biology Laboratory
University of Maryland
Box 38
Solomons, MD 20688-0038

Nicholas Cozzarelli
Dept. of Molecular Biology
229 Stanley Hall
University of California, Berkeley
Berkeley, CA 94720

Leah Edelstein-Keshet
Dept. of Mathematics
#121-1984 Mathematics Road
Vancouver, B.C. V6T 1Z2 Canada

Robert Eisenberg
Dept. of Physiology
Rush Medical College
1750 West Harrison St.
Chicago, IL 60612

Bard Ermentrout
Dept. of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15260

Robert H. Gardner
Oak Ridge National Laboratory
Bldg. 1505 MS 6036
P.O. Box 2008
Oak Ridge, TN 37831-6036

John Guckenheimer
Center for Applied Mathematics
305 Sage Hall
Cornell University
Ithaca, NY 14853

Richard Hudson
Dept. of Ecology
and Evolutionary Biology
University of California, Irvine
Irvine, CA 92717

Fern Hunt
Dept. of Mathematics
Howard University
Washington, DC 20059

Stuart Kauffman
Dept. of Biochemistry and Biophysics
School of Medicine G3
University of Pennsylvania
Philadelphia, PA 19174

Daniel Kersten
Dept. of Psychology
Elliott Hall
University of Minnesota
75 E. River Rd.
Minneapolis, MN 55455

David Kingsbury
Dept. of Microbiology
George Washington University Medical Center
2300 Eye St., N.W.
Washington, DC 20037

Nancy Kopell
Dept. of Mathematics
Boston University
111 Cummington Street
Boston, MA 02215

Eric S. Lander
Dept. of Mathematics
Harvard University
151 Bishop Richard Allen Dr.
Cambridge, MA 02139

Simon A. Levin
Ecology and Systematics
347 Corson Hall
Cornell University
Ithaca, NY 14853

Michael Levitt
Dept. of Cellular Biology
Stanford University Medical Center
Palo Alto, CA 94305

Deborah Lockhart (NSF Liaison)
Special Projects Program
National Science Foundation
1800 G Street, N.W.
Washington, DC 20550

Don Ludwig
Dept. of Mathematics
121-1984 Mathematics Rd.
University of British Columbia
Vancouver, BC V6T 1Z2 Canada

Marc Mangel
Dept. of Zoology
University of California, Davis
Davis, CA 95616

Andre Manitius
Electrical and Computer Engineering Dept.
4400 University Drive
George Mason University
Fairfax, VA 22030

James D. Murray
Centre for Mathematical Biology
Mathematical Institute
University of Oxford
24-29 St. Giles'
Oxford OX1 3LB, U.K.

Garrett Odell
Dept. of Zoology, NJ-15
University of Washington
Seattle, WA 98915

Richard Olshen
Stanford University
School of Medicine
HRP Bldg., Rm. 110A
Stanford, CA 94305

George F. Oster
Division of Entomology
201 Wellman Hall
University of California, Berkeley
Berkeley, CA 94720

Alan S. Perelson
Los Alamos National Laboratory
Mail Stop K710
Los Alamos, NM 87545

Michael Reed
Dept. of Mathematics
136 Physics Bldg.
Duke University
Durham, NC 27706

John Rinzel
National Institutes of Health
Public Health Service
Dept. of Health and Human
Services
Bldg. 31, Room 4B-54
Bethesda, MD 20892

Jonathan Roughgarden
Dept. of Biological Sciences
Stanford University
Stanford, CA 94305

David Rumelhart
Dept. of Psychology
Stanford University
Stanford, CA 94305

David Schimel
Natural Resource Ecology
Laboratory
Colorado State University
Fort Collins, CO 80523

Lee A. Segel
Dept. of Applied Mathematics
The Weizmann Institute of Science
Rehovot 76100 Israel

Wendy Silk
Dept. of Land, Air, and Water Resources
Hoagland Hall
University of California, Davis
Davis, CA 95616

Terry Speed
Dept. of Statistics
University of California, Berkeley
Berkeley, CA 94720

Sylvia S. Spengler
Human Genome Center
459 Donner Laboratory
Lawrence Berkeley Laboratory
University of California, Berkeley
Berkeley, CA 94720

Marcy K. Uyenoyama
Dept. of Zoology
Duke University
Durham, NC 27706

Michael S. Waterman
Dept. of Mathematics
University of Southern California
University Park
Los Angeles, CA 90089

James H. White
Dept. of Mathematics
University of California,
Los Angeles
Los Angeles, CA 90024

Grace Yang
Dept. of Mathematics
University of Maryland
College Park, MD 20742

Appendix 3

RESEARCH OPPORTUNITIES IN COMPUTATIONAL BIOLOGY

EXECUTIVE SUMMARY[1]

Computational Biology is emerging as a current analog to the development of molecular biology as a discipline in its own right. In the late 1950s and early 1960s a group of scientists began to apply the tools of several disciplines, genetics, microbiology, physics, biochemistry, and biophysics, to analyze biological problems in a new way. The power of this approach was so great that it emerged as a discipline itself and is now known as molecular biology. As described in this document, the application of mathematical and computational tools to all areas of biology is producing equally exciting results, is providing insights into biological problems too complex for traditional analysis, and is emerging as a new discipline within the biological sciences.

There is a consensus among all observers that biology, regardless of the sub-speciality, is overwhelmed with a large amount of very complex data. However, what sets biology apart from other data rich fields is the *complexity* rather than the sheer volume of the data produced. In contrast to other data rich fields, biology remains a scientific "cottage industry," with the data generation done in a highly distributed mode, with no standard format or syntax.

Thus, all areas of the biological sciences have urgent needs for the organized and accessible storage of biological data. Generally this is referred to as biological database development, however, this terminology infers traditional database technology such as transaction oriented relational database systems. Unfortunately, relational database technology is inadequate to serve many areas of the biological sciences due to the complexity of biological data and the absence of a standardized data structure. It is clear that collaboration between computer scientists and biologists will be necessary to design information platforms which accommodate the needs for variation in the representation of biological data, the distributed nature of the data acquisition system, the variable demands placed on different data sets, and the absence of adequate algorithms for data comparison, which forms the basis of biological science.

[1] Executive Summary from "Opportunities in Computational Biology" (D. Kingsbury, ed.), a companion publication to this document. It summarizes the current status of computational biology as a field and focuses on the future development of computational approaches to investigating and modeling biological problems. Following publication in mid-1992, it will be available from the National Science Foundation and the Department of Energy.

biological data, the distributed nature of the data acquisition system, the variable demands placed on different data sets, and the absence of adequate algorithms for data comparison, which forms the basis of biological science.

There have been dramatic advances in commercially available hardware over the past few years and it has had an effect at both the high and low ends of the spectrum. In the past this general purpose hardware was inadequate to address the most computationally intense problems in the biological sciences. These computational problems were best handled by special purpose equipment designed by teams of biologists and chip and circuit designers. This condition has been dramatically altered in the past two years as high performance general purpose instruments have become more widely available. Not only hardware limitations have affected the productivity of the computational biologist. There is a continuing need for new algorithm development to cover many tasks, especially comparisons between objects and images. Imaging technology is central to almost all of biology and data representation though image construction remains an elusive but astoundingly powerful tool. The full utilization of modern CAD tools in computational biology will advance image analysis, but will require intense software and hardware development because of the complexity of biological data

During the last decade there were dramatic advances in instrumentation and related methodologies for both light and electron microscopy. The advances lie not simply in higher resolution, but rather in a broader size range of structures that can be analyzed, and more powerful methods for putting together the pieces of three-dimensional puzzles of cell form, and the addition of dynamic details of biological form and function, ranging from the subcellular to the physiological level. The new approaches are computationally demanding. Extant computational resources, which were typically set up for entirely different processing needs, not surprisingly, are proving inadequate for dealing with the massive data flow. An effort to develop new computational approaches is under way in a few laboratories around the world. However, it is important that new software be developed within the context of the experimental research driving the needs; that is, there must be close collaboration between those developing the software and the groups carrying out research on static and dynamic structures. Furthermore, augmentation of the experimental environment, particularly image processing equipment and other specialized equipment, is needed. Positions for sophisticated programmers are even more important. A prime example of the need for such a laboratory-based specialized programming effort is the development of workstations for interactive visualization and interpretation of 3-D data. The development will proceed in pace with experimental research only if it is done in an environment "open" in the terms used by the computer science world, where new applications are developed free from proprietary restraints and distributed as source code to other laboratories facing the same experimental needs. Commercial interests or specialized production groups will be required finally, to add value to the base line development, producing highly reliable ("bullet proof") production line products.

X-ray crystallography and NMR are the major experimental methods for deducing macromolecular structures at atomic resolution. NMR and X-ray crystallography both produce extremely large amounts of data and are entirely dependent upon the availability of powerful computers and sophisticated processing algorithms for the interpretation of raw data. In addition, there are fundamental scientific problems in both areas that require major computational advances. In addition, substantial opportunities exist for combining

structural information from several experimental techniques. This may provide the basis for a structural solution where only partial data are available from any single technique. With improved computational tools, combining physical data from a variety of sources may become commonplace. These developments will allow solutions to be obtained for structural problems which would otherwise be intractable. Analysis of errors in structures based upon experimental data from several sources also represents a new computational challenge.

Advances in X-ray and NMR data analysis will lead directly to rapid developments in the field of protein folding which will be synergistic with developments in other areas of biology itself, and especially computational biology. Common problems of data representation, search strategy, pattern recognition and data visualization appear in many fields. There is a particularly exciting synergistic relationship between the protein folding field and those of structure determination by X-ray crystallography and 2-D NMR. Each field will benefit from rapid advances in the other disciplines. Improved folding algorithms provide a new way to attack the phase problem in crystallography, and new, more carefully refined protein structures provide rich new insights into protein folding.

Various initiatives in computational neurobiology give us the hope of interpreting the mass of anatomical and physiological information about the nervous system that is now available in functional terms. Better interpretation of these data will permit neurobiology to make contact with other fields such as psychology and artificial intelligence. This work will make specific, testable predictions in the areas of sensory perception (visual, olfactory, and auditory), memory, learning, and motor control. Above all, it will lead to the integration of all these aspects to provide an eventual understanding of the total functioning of the nervous system. Such integration can be expected to provide new insights that will lead to improvements in the treatment of diseases of the nervous system at all levels, from neuropharmacology to psychotherapy. In addition, studies of this kind may be expected to contribute to major advances in artificial intelligence and practical robotics.

In the area of genome analysis significant progress has been made over the past few years, including the use of molecular tools such as Restriction Fragment Length Polymorphism (RFLP) analysis. However, considerable effort is still required to make genetic linkage maps effective tools for genetic research. To be useful in common situations, more markers must be identified and mapped to produce higher-resolution maps. In many cases marker analysis requires the ability to analyze small families and consider quantitative traits. To be fully useful in a meaningful quantitative sense this analysis will require powerful computer simulation and modeling. Common to all of the problem areas examined is the need for good visualization of data. Visualization is necessary because the sequence analysis phase for a molecular biologist is equivalent to exploratory analysis for a statistician. It is at this point that the experimentalist gains the feeling for, and understanding of, a sequence which may then guide many months of experimental work. The complexity inherent in biological systems is so great that very sophisticated methods of analysis are required. These are the tools which must be readily accessible to molecular and cellular biologists untrained in computer technology.

Ecology and evolutionary biology encompass a broad range of levels of biological organization, from the organism through the population to communities and whole ecosystems. This complexity demands computational solutions. The need for enhanced

computational ability is most evident when one attempts to couple large numbers of individual units into highly interactive and largely parallel networks, whether at the tissue, community or ecosystem level of organization. The proliferation of information from remote sensing introduces the need for geographical information systems that provide a framework for classifying information, spatial statistics for analyzing patterns, and dynamic simulation models that allow the integration of information across multiple spatial, temporal, and organizational scales. Today, in these fields application software is mostly nonexistent except in a few special special cases such as image processing and remote sensing. As more researchers begin to use computational techniques, we can expect to see a wider sharing of applications developed by an individual or small group. This will require additional resources to take research codes and make them "bullet-proof" enough for community use and to add adequate documentation. In order to take advantage of all these new capabilities, we need to increase training modalities. This can take a wide variety of forms, from on-line self training techniques to special sessions at universities, national centers, or workshops.

It is recommended that Federal granting agencies place greater emphasis on the area of Computational Biology through a number of mechanisms. This support must be developed over a period of several years with *a particular emphasis on infrastructure and training*. Many of the necessary changes may be instituted immediately while others will require a longer time in order to generate budgetary resources to build in new areas. The current focus on biological databases is a good beginning, however, the need is so great that the initiative needs considerable additional resources. These resources should be directed in three areas. First, the enhancement of current databases which are in wide use but need concerted effort at standardization of data structures and broadened access. Second, a continued examination of new databases which will incorporate important information needed by many investigators, but also explore new database ideas and representations. Third, research on the representation of objects and images which will be searchable and comparable within database structures. For example, there is a great need to be able to search a database of enzyme or antibody active site configurations to test for binding of newly developed ligands. Database development remains the highest priority item since this area is common to all fields of biology. A second area of high priority is the development of more powerful visualization tools for data interpretation. This area too is a need shared by almost all fields of biology. Funding agencies could immediately respond to some of the needs of the research community by recognizing the need for professional programmers and hardware and software facilities on grants in this area. Agencies must break out of the habit of immediately removing these items from budget requests in order to reduce the overall cost of an award, since these items are critical not only to doing the proposed work but also to making the results of the work (in the form of usable source code) available to the rest of the research community.

Appendix 4

TRAINING COMPUTATIONAL AND MATHEMATICAL BIOLOGISTS[1]

SUMMARY

1. Introduction

It has been estimated that in mid-1990, there were approximately 4000 professional level scientists identifiable as computational or mathematical biologists. These scientists were found in a wide variety of institutions and in a wide range of positions within those institutions.

The pattern of distribution of these individuals among and within different institutions appears to be related to their academic training. For example, mathematicians and computer scientists who have primarily followed an interest in the biological sciences generally work as biologists and find themselves in nonacademic research positions in industry, government or private research institutes, or quasi-academic research centers (e.g., supercomputer centers). A small minority are in biology departments. In contrast, mathematicians who have continued to pursue research activities in mathematics, choosing biologically related problems or examples, or collaborating with biologists, tend to remain in departments of mathematics or applied mathematics in academic institutions. Computer scientists follow a similar pattern. Statisticians may be found in statistics departments, biostatistics groups or departments, or even in biological sciences departments, depending on the extent of their involvement with biological problems, and the local structure of the institutions within which they work.

Biologists who rely on computational and mathematical tools in their research activities are found in many institutions. A large number have moved into industry where they play a role in the analysis of macro-molecules in biotechnology and pharmaceutical companies. Another major source of employment is in government and private research institutes, which tend to focus on problem-oriented research and directly utilize their computational biology skills. In the academic environment, computational biologists pursuing accepted biological problems are found in a variety of departments of biology

[1] Final report of the NSF-sponsored Workshop on Training Computational and Mathematical Biologists, held at the Banbury Center of the Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, December 9–11, 1990.

pursuing accepted biological problems are found in a variety of departments of biology (including departments of related name such as genetics, ecology and evolutionary biology, molecular biology, and microbiology), chemistry, and biochemistry.

The character of the institutional acceptance of these interdisciplinary activities depends on two factors: the need of the institution for problem-oriented work, and the traditional academic expectations on the performance of the individual. For example, biology departments place their emphasis on disciplinary achievements, and computational and mathematical approaches are secondary to the disciplinary results. Therefore, the infusion of mathematical and computational tools is dependent on the confidence of the researcher that they can afford to invest the time and effort to enable them to use this approach, let alone develop new tools. Thus in many cases, computational and mathematical biology makes a backdoor entrance into the academic world. In contrast, these approaches are embraced more directly by industry and research institutes whose problem-oriented programs utilize a broader range of approaches, including direct application of mathematical and computational techniques.

The workshop participants' assessment is that in the immediate future, this situation will not undergo a substantial change. Therefore, scientists expecting to enter the academic research world will continue to need a strong disciplinary grounding for their cross disciplinary work. Employment opportunities in industry and research institutes appear to be stable, or growing slowly. Such centers will continue to be major sites for the development of computational techniques and applications in biology.

Because of their frequently strong mathematical and computational environments, and the less frequent presence of rigid departmental structures, one possible source of future growth for computational biology is the four-year college. Mathematical and computational approaches fit well within the research environments found in these institutions, and they are likely to find effective implementation in the teaching programs. In this context, faculty in these institutions may be expected to employ mathematical and computational techniques in both research and the development of teaching aids that will eventually find their way into research institutions. However, here again, strong disciplinary training will be essential as the basis for the research approach.

2. Profiles of Computational and Mathematical Biologists

In the past, most of the migration of scientists into computational biology has been from disciplines outside of biology (e.g., math, physics, chemistry, computer science, etc.). Physicists become biologists, but not the reverse. This migration and its asymmetry has been prompted by successful applications of domain-specific technology to solving biological problems.

Many early successes in computational biology were obtained by scientists who were primarily biologists with marginal skills in computer science and mathematics (programming skills and some algorithmics), while many others were the result of work by scientists with extensive mathematical and computational backgrounds. However, as the problems under investigation become more complex, training which provides great depth in quantitative analysis will be essential.

Current interest and excitement in computational and mathematical biology is driven in large part by neurobiology, global change, and genomics. In all of these areas, vast amounts of information are accumulating at a rate that precludes human absorption and, hence, understanding. Biology needs tools for manipulating and analyzing information. In order for training environments to be maximally effective there must be a clear understanding of which professional profiles are suitable for current and future researchers in computational and mathematical biology.

The profiles which follow are dependent upon the nature of the position. Academicians tend to reside within traditional departmental units; whereas, in industrial settings and research institutes there is a wider range in the mixtures of disciplines in working groups. The following lists of specialities within computer science, mathematics and biology are those in which there is substantial research activity today and where there is likely to remain some research focus in the future.

Computer Scientists:

Most computer scientists retain their primary professional identification with computer science. They tend to view biological applications as a source of computer science problems. Biological applications are new to computer scientists, and the traditions across the interface are developing at a moderate pace. The tendency is to cross the line as a senior scientist by developing collaborations. There are some successful scientists in this field whose first exposure to biology was at the graduate level. Examples of the areas of computer science in which such collaborations take place are:

Artificial Neural Networks (AI)

Algorithmics

Database design and theory

Visualization (Graphics)

Biologists:

Biologists working on computational problems come from a plethora of backgrounds: computer science, mathematics, statistics, engineering, physics and chemistry as well as biological disciplines. The biological sciences are themselves diverse and different areas of biology draw upon very different quantitative skills. Those biologists who have crossed the boundaries between biology and other disciplines have often done so to address specific biological problems. Their acceptance by the biological community has been out of necessity since many biological problems require technology that has been driven by insight and intuition from other disciplines. This report is motivated by the assumption that this trend will accelerate in the near future in areas such genomics, neurobiology, imaging, structural biology and issues of global climate change. Many of these developments have been initiated by scientists whose initial training was outside biology (e.g., mathematics, chemistry and physics). The current technological advances will require a new range of quantitative skills beyond the norm of current curricula in the

biological sciences. Biological Sciences that currently draw substantially from the computational and mathematical sciences include:

Population Biology, including Ecology and Genetics

Molecular Biology

Molecular Genetics

Cellular Biology

Neurobiology

Biophysics and Structural Biology

Ecosystem Ecology

Epidemiology

Physiology

Mathematicians:

There is a long tradition of mathematicians and statisticians working on biological problems. Indeed, the field of statistics grew largely out of biological origins, and there is a substantial portion of the statistics community working on problems of biometry and biostatistics. There is also a small but stable community of mathematical biologists working within departments of pure and applied mathematics. Some members of this community migrate to biological departments during the course of their careers while others remain in mathematical science departments. Those who do remain within mathematical science departments either establish a career based upon collaborations with biologists, or focus upon mathematical questions driven by biological problems. In some cases, threads of mathematical research initiated by biological problems take on a life of their own as interesting areas of mathematics per se. Areas of mathematics making substantial contributions to biology include:

Applied Mathematics (Differential Equation Models, Image Processing and Analysis)

Probability (Sequence Analysis, Interacting Particle Systems)

Statistics

Discrete Mathematics

Topology and Differential Geometry

2.1. Summary of the Current Status

With regard to the current panorama of activity, we perceive that several difficulties exist. First, computer scientists are not sufficiently involved in computational biology. Their work is frequently on problems so abstracted from the application as to make them less

than fully effective as collaborators. Another limitation is that biologists tend to view the work of computational scientists as service, and not original research, which tends to alienate this community. Mathematicians are caught between mathematical peers who evaluate their work on the basis of its mathematical depth and elegance, and biologists who have little appreciation for theory that does not have a direct bearing on the interpretation of experimental data. Finally, those biologists who have invested in cross-training are frequently misunderstood and undervalued by their colleagues, most of whom do not understand how to evaluate their work.

Computer science is a new discipline that is rapidly maturing. As the field develops, a tradition of interdisciplinary work will evolve much as it has for mathematics, especially statistics. This will, in part, alleviate the problem of computer scientists' involvement. A greater emphasis on the early grounding in scientific disciplines while at the undergraduate level should also help to cultivate computer scientists with a stronger interdisciplinary focus. As the need for computation in the various areas described above becomes clearer, the biological community must become increasingly more tolerant and accepting of computational biologists within their midst. As a result of this and other factors, such as heavy dependence on physical measurement, the training of biologists at all levels must become increasingly more quantitative in nature.

3. Encouraging Interactions

The most effective way to encourage interactions between mathematicians and computer scientists on the one hand, and biologists on the other, is through direct co-involvement with a particular problem. This applies at all levels from undergraduate through senior scientist. The ways in which this interaction may be encouraged depend on the level and direction of movement (math/CS to biol or biol to math/cs). At present, the pattern is generally unidirectional, with movement from mathematics or computer science into biology as the dominant paradigm. Significant changes in this state of affairs are likely to require substantial curricular changes based upon effective means of overcoming the apprehension of most biology students towards mathematics.

Interaction can be improved through a strengthening of mechanisms that already exist. However, one area deserves much greater emphasis than is now the case, and that is support of small research groups with a genuine interdisciplinary focus: within this, substantial support is needed for post-doctoral scientists. Support of small group research will develop critical mass in important areas, will help to foster and sustain collaborative research, and provide a crucial home for individuals who are in the early stages of (what is now) a cross-disciplinary research career.

The most effective mechanisms for stimulating these fields vary by the level of a scientist's career stage as outlined below.

(a) Senior researchers (tenured and above)

(CS, Math -> Biol) Support for sabbaticals and, later, research in biology.

(Biol -> Math/CS) Support for visits to math research groups to learn/update new technical areas.

(b) Pre-tenure

Most mathematics and statistics PhD students will start in untenured positions. Changing fields (or, at least becoming more interdisciplinary) at such an early stage is a very risky career move, particularly by individuals approaching a tenure decision. One way to ameliorate this situation is through a new focus on PYI-level type support (National Science Foundation Presidential Young Investigator) for promising people (prestigious competitive awards).

(c) Postdoctoral

Support for postdoctoral training within existing grants is essential. Postdocs are an important educational component of existing research groups, and are very scientifically profitable in the short term. These should support a given individual for multiple years, and not be specifically tied to a particular investigator within the group. This mechanism allows quick response to changing areas of interest, while providing enough time for a postdoctoral fellow to develop a useful independent research focus.

Another aid to young investigators is the computational research associates program at the NSF sponsored Supercomputing Centers. This program is of great value to the biological sciences and the field would benefit from its continued existence. However, to be maximally effective these investigators must be part of an active and focused research program and not "generalists" in applied computer science.

The concepts behind these training programs are not based on the assumption that all people passing through them will eventually obtain tenure track positions in universities.

(d) Graduate students

An important source of mathematical biologists comes from mathematically trained undergraduates who change fields early in their postgraduate education. Such students are then main-stream biologists, with the requisite quantitative background to enter the fields of mathematical or computational biology. The educational challenge for students with this background is the continuation of the quantitative approach to biology in a supportive environment. This requires an appropriate mentor and an appropriate departmental or graduate group environment so that the student's background is valued and prior training reinforced. Given the many opportunities available to an undergraduate with computer science or mathematical training, it is essential that graduate student support be provided to entice these students to forego the immediate gratification of lucrative employment for the longer term prospects of graduate training and research careers in biology. To this end the continued and renewed support of training grants or traineeships (for example in the research groups described above) are of central and continuing importance.

Furthermore, educational institutions must be encouraged to recognize the need for training students in these areas as a means of dealing with the future of biological research. To this end institutional and departmental support of fellowships and RA (Research Assistant) positions are of supreme significance. Cross-training students at the

graduate level will lengthen an educational process that already can be inordinately long. Freeing a student from the demands of a teaching assistantship or a research assistantship with responsibilities to further the work of a principal investigator will help make such programs educationally feasible. It would be especially appealing to find a mechanism to support mathematical or computational biologists within the structure of departments of mathematics or computer science.

One of the most significant factors in the training of graduate students is the role model of the major professor. This mentorship plays a greater role in the ultimate aspirations of a student than is generally acknowledged. The successes, failures, and frustrations of a student's mentor plays a profound role in the expectations and aspirations of a student. In this context the small group research environment is a highly significant environment in which to train students for the future of the biological sciences.

(e) Undergraduate

In most institutions it is very common for the top biology students, especially those interested in eventual graduate study, to participate in undergraduate research projects, especially in their Junior and Senior years. this opportunity should not be confined to biology students, but should be expanded wherever possible to include interested students from mathematics and computer sciences whenever possible. The proper environment is essential to the nurturing of a student that might wish to commit to a career in the biological sciences, using this valuable undergraduate training. To this end the National Science Foundation REU (Research Experiences for Undergraduates) program provides an extraordinary opportunity in the Math/Biol area.

One area of extreme importance for the future development of a cadre of computational and mathematical biologists, and for the continued recruitment of students into biophysics and related disciplines is the development of better course materials devoted to the quantitative approach to biology. The workshop participants valued very highly the concept of "enculturation of quantitative thought" through the introduction of quantitative approaches in biology courses

(f) Pre-college

While there was considerable discussion during the workshop regarding the state of pre-college science education, no specific recommendations were developed. Many private and government agencies have focused great attention on this problem, and it remains a top national priority. There was general agreement that two issues posed particular concern to the participants. First, the need to involve more fully parents in the educational process. This is particularly important in groups which do not have a cultural history of educational achievement. The second concern was the current selection of the "ultimate underachiever" as the folk hero of the nation's children. We believe that this message is alarmingly inappropriate in the current context of rapid technological change and global competition. The participants hope that the leadership of the Education and Human Resources Directorate of the National Sciences Foundation will use its influence and insight to find a mechanism to reverse this trend.

(g) Summary Principles

(1) If time is limited for education, spend it in mathematics, not computer science.

(2) What we want is an attitude/consciousness change, so that people are aware of the input of the "other" type of science in their own area.

(3) While collaboration will enhance the science of the current generation, we are seeking to change the way that biology is done by changing the way biologists are educated for the next forty years.

4. Fundamental Educational Principles

(a) Undergraduate Education

(i) General Course Content

The cross-disciplinary aspects of modern science must be emphasized in all undergraduate science and mathematics courses. The role of computer science and mathematics, as well as technologies from physics and chemistry, need to be presented in biology courses. In contrast, the research areas that have used various tools of computer science and mathematics in the experimental sciences should be identified throughout mathematics and cs courses.

(ii) Mathematics/Computer Science Majors

All mathematics and computer science majors should have required experimental science courses. We recommend a minimum of two years that can be concentrated in one area or spread over the basic sciences. The purpose of this is to provide the student with an understanding of the vocabulary and concepts and an experience of the ways in which mathematics or computer science have contributed to other disciplines.

(iii) Biological Sciences

In order to produce biological scientists who will be qualified to do modern research, we strongly recommend that the science curricula require four years of mathematics and/or computer science. Representative courses might include programming, theory of algorithms, probability and statistics, linear algebra, calculus, discrete mathematics, and numerical analysis.

(b) Consequences

Failure to implement these recommendations at a minimal level will foreclose the future for many undergraduates majoring in biological sciences. This originates in the types of problems that are coming into existence and that are consistently more and more dependent on quantitative skills for their solution. Secondarily, lack of training in these

quantitative areas will limit the questions that can be asked by an investigator, and may come to threaten an individual's levels of funding. We must remember that we are addressing the education of persons who will be in the pool for the next forty years. If education changes are not implemented, much of biology will fail to thrive.

The broad education that we are proposing also permits people to change their minds and acquire additional course work in another field, even late in their studies, without having to start from the beginning.

Our recommendations should not be construed to support any concept that presupposes a gender-specific bias in the ability to perform. It may be that a type of math/cs anxiety will become apparent if our recommendations are instigated. In order to counter this, we propose that support groups, personal tutorials, study circles and other tools of encouragement and enhanced performance/esteem be supported so that they are readily available.

5. Additional Recommendations

Part of the difficulty in implementing the course recommendations may be the prevalence of pre-med education as a major component of biology curricula. Although there will be a number of additional consequences, it would be well worth considering the restructuring of the undergraduate major so that pre-meds follow a separate track and their presence does not determine the future of an academic discipline.

It is incumbent upon those who practice cross-disciplinary science and mathematics/CS to become both role models and mentors for others. It is particularly important for representatives of under-represented groups to make an effort to encourage others.

Several members of the group have suggested that a new type of biology course should be developed. It would cover the elements of modern biology, but highlighting the contributions of other disciplines. The hope is that someone will be inspired to write a founding text, one that will change the field.

GRADUATE EDUCATION

Continue to create opportunities for cross-disciplinary work. NIH programs in molecular biophysics and the NSF research training groups are examples of attempts to encourage this type of interaction.

One-on-one mentor/student relationships are not sufficient to maintain cross-disciplinary development. Direct support for cross-disciplinary efforts would help to break down the interdepartmental barriers that frequently exist. Seminar groups or other frequent interactions should be encouraged.

New graduate students (and postdocs) might acquire an elementary grounding in a new field through summer institutes or some other "crash course." The courses would be taught by highly interactive, expert, senior level researchers. For example, a course in basic molecular biological concepts could include molecular biology, biochemistry, and molecular biophysics. Emphasis would be on the vocabulary and point of view, that is, how the science is done and what are its assumptions. For a course on computation in

genetics, this material might include basic computer science concepts, e.g., files, databases, algorithms and their use, graphics and statistics. The benefits of such a course could also be made available to more senior investigators.

6. Women And Other Under-Represented Groups

In high school, women represent a reasonable proportion, approximately 30-40%, of those students who are interested in the physical sciences and mathematics. Partitioning begins in college and is nearly finished by graduate school. Some disciplines within the biological sciences do have equivalent or even over-balanced representation by women. Increasing the level of course work in mathematics and computer science may be threatening to some of these women. In order to prevent this, specific actions may well be necessary. Similarly, for some students from other under-represented groups, it may be necessary to have additional courses available at the undergraduate level to improve the level of computational competence of entering students.

END

DATE
FILMED
12 / 15 / 92