MATHEMATICS AND 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 Associate: 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

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.

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) divides 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 Ñ 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 non-traditional 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 and the cellular and molecular levels have permitted complementation of that approach with that 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 at higher levels of organization than was possible previously from fossil data alone. Mathematical approaches allow the use of genetic data to analyze multi-locus 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 stochastic processes have 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 fundamentally integrated into the training of every scientist. 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 non-randomly 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 have occurred. Mathematical modeling revealed the cause of ventricular fibrillation. Hodgkin and Huxley theorized that macroscopic current might be generated by molecular pores Ñ ion channels that were proven later to exist. Navier-Stokes equations for flow through small bristled appendages have shown how the geometry permits the appendages of aqueous organisms to function either as 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. At this time 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 by the funding agencies to the needs at this interface is essential. Cross-disciplinary teams of researchers should be encouraged and appropriate methods for review of proposals should be developed. Methods of selection and training of interdisciplinary individuals at an early stage of their development in the interface of these disciplines should be devised. Meetings and workshops to explore as yet unthought of ways in which the two disciplines can serve to amplify each other should be supported.

Grand Challenges

Genomics

Attention to the human genome project and its great potential often obscures the fact that theoretical work is essential to efforts at sequencing and mapping all genomes, human and non-human, animal and plant. Without the mathematical and statistical underpinnings and computational advances, efforts directed to sequencing and mapping 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 finding solutions, and also increase our understanding of the natural world.

Global Change

No problem is more compelling, from the viewpoint of 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, a fundamental theoretical problem involves the relationships between 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 current status of the field 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 speed up immeasurably the development of the subject.

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 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. Basic work must be done related to 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 continual motion in nature, but NMR and X-ray crystallography necessarily involves snapshots. Mathematical and computational methods are essential to complement experimental structural biology by allowing the addition of motion to molecular structures.

Mathematical Theories

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 under-represented groups.

Mathematical and computational biology is a vital, crucial, and rapidly growing subject that complements and guides empirical work, elucidates mechanisms, and provides model systems for study and manipulation. Such model systems indeed, in some circumstances, can 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.

.