

Modeling of Biological Systems

A Workshop at the National Science Foundation

March 14 and 15, 1996

Panel Members:

Peter Kollman, University of California, San Francisco, Chair
Simon Levin, Princeton University, Co-Chair
Alberto Apostolico, University of Padova
Marjorie Asmussen, University of Georgia
Bruce L. Bush, Merck Research Labs
Carlos Castillo-Chavez, Cornell University
Robert Eisenberg, Rush Medical College
Bard Ermentrout, University of Pittsburgh
Christopher Fields, Santa Fe Institute
John Guckenheimer, Cornell University
Alan Hastings, University of California, Davis
Michael Hines, Yale University
Barry Honig, Columbia University
Lynn Jelinski, Cornell University
Nancy Kopell, Boston University
Don Ludwig, University of British Columbia
Terry Lybrand, University of Washington
George Oster, University of California, Berkeley
Alan Perelson, Los Alamos National labs
Charles Peskin, Courant Institute of Mathematical Sciences
Greg Petsko, Brandeis University
John Rinzel, National Institutes of Health
Robert Silver, Marine Biological Laboratory
Sylvia Spengler, Lawrence Berkeley Labs
DeWitt Sumners, Florida State University
Carla Wofsy, University of New Mexico

***Sponsored by the National Science Foundation
(MCB 96-29868 to the University of California, San Francisco)***

TABLE OF CONTENTS

I. [Summary](#)

II. [Introduction](#)

III. [Molecular and Cellular Biology](#)

IV. [Organismal Biology](#)

V. [Ecology and Evolution](#)

VI. [Cross-Cutting Issues](#)

VII. [Educational Issues](#)

VIII. [References](#)

I. SUMMARY

The common theme of this report is the tremendous potential of mathematical and computational approaches in leading to fundamental insights and important practical benefits in research on biological systems. Mathematical and computational approaches have long been appreciated in physics and in the last twenty years have played an ever-increasing role in chemistry. In our opinion, they are just coming into their own in biology.

The goals of these mathematical and computational approaches are to elucidate mechanisms for seeming disparate phenomena. For example, how does the atomic level structure of an enzyme lead to its functional, enzyme catalysis? To understand this structure/function relationship requires fundamental quantum mechanical and molecular dynamical calculations, but successful simulations may lead to understanding of disease and drug therapy. Knowing the three dimensional structure of the muscle protein kinesin may lead to understanding of muscle action as well as other cellular motors. Simulations of the embryonic and fetal heart at different stages of development are helping to elucidate the role of fluid forces in shaping the developing heart. The structure and dynamics of earth's ecosystems are critical elements in how they function and mathematical/computational methods play a critical role in understanding their function.

In these examples and the many others in the body of this report (sections III-V), mathematical/computational methods, based either on fundamental physical laws (e.g. quantum mechanics), empirical data, or a combination of both, are providing a key element in biological research. These methods can provide hypotheses that let one go beyond the empirical data and can be constantly tested for their range of validity.

Our report also highlights (section VI) computational issues that are common across biology, from the molecular to the ecosystem. Computers are getting more powerful at a prodigious rate and, in parallel, the potential for computational methods to ever more complex systems is also increasing. Thus, it is essential that the next generation of biological scientists have a strong training in mathematics and computation from kindergarten through graduate school. We discuss educational issues in section VII of our report.

A purpose of this report is to increase the awareness among biological scientists of the ever-increasing utility of mathematical and computational approaches in biology. Sometimes newly emerging areas and interdisciplinary areas are in danger of falling between the cracks at funding agencies. Specifically, we hope that this report will raise the level of awareness at the National Science Foundation and other funding agencies on nurturing computational and mathematical research in the biological sciences.

II. INTRODUCTION

Characterization of biological systems has reached an unparalleled level of detail. To organize this detail and arrive at a better fundamental understanding of life processes, it is imperative that powerful conceptual tools from mathematics and the physical sciences be applied to the frontier problems in biology. Modeling of biological systems is evolving into an important partner of experimental work. All facets of biology, environmental, organismic, cellular and molecular biology are becoming more accessible to chemical, physical and mathematical approaches. This area of opportunity was highlighted in a 1992 report, supported by the National Science Foundation, entitled "Mathematics and Biology, the Interface, Challenges and Opportunities." (MBICO)

A workshop was held at the National Science Foundation (NSF) on March 14 and 15, 1996 that built on the findings of MBICO in order to critically evaluate its findings and to suggest which areas were the most promising as foci for further research. This workshop brought together 25 scientists, with expertise ranging from the molecular to the cellular to the organism to the ecosystem level, all of whom have an interest in applications of mathematical/computational approaches to biological systems. The goal of the workshop was to identify important research areas where theoretical/computational studies could be of most use in giving insight and in aiding related experimental work. This is done below. Because of the small size of our group, the limited time we had, and our not unlimited vision, one must view the areas of research opportunities presented below as representative, not exhaustive. Hopefully, our report can provide some guidance and an historical marker as to the state-of-the-art in Modeling of Biological Systems, ca. 1996.

Our report is divided into five sections. We follow the organization of the NSF in dividing our description of research opportunities into three areas: Molecular and Cellular Biology, Organismal Biology and Ecology and Evolution. These three sections are followed by a section focussing on issues that cross the boundaries between these areas and a final section on educational issues.

III. MOLECULAR AND CELLULAR BIOLOGY

OVERVIEW

A central organizing theme in Molecular and Cellular Biology is the relationship between structure of molecules and high level complexes of molecules and their function, both in normal and aberrant biological contexts. The connection between structure and function was most clearly illustrated in the paper that began Molecular Biology, the elucidation of the structure of DNA by Watson and Crick.

This study immediately illustrated how DNA can replicate and retain the original information stored in it. Thus, the structure showed how this molecule functions. But this example also shows the important role of mathematics, chemistry and physics in elucidating structure/function relationships in biology. Both the information contained in DNA duplexes and their higher order structures have been usefully analyzed by mathematics, as the sections below on the GENOME and MOLECULAR HISTOLOGY illustrate, and important questions have been answered and many still remain unanswered in these areas.

The developments in physics and chemistry have played fundamental roles in enabling structure determination of the essential molecules in biology - proteins, nucleic acids, membranes and saccharides - and in that fashion, helping one to understand their function. Some aspects of these efforts are described below in the sections on PROTEIN STRUCTURE and NUCLEIC ACIDS. The use of the simulation methodologies first developed in the physics and chemistry communities to simulate molecules of biological interest is described in SIMULATIONS. Evolution has occurred on a molecular as well as a macroscopic scale and some of the molecules and their properties that have evolved are quite astonishing. The section on BIOINSPIRED MATERIALS points out the possibilities inherent in making use of some of the materials that have evolved in

the process of molecular evolution.

Although much progress has been made in understanding structures of molecules of biological interest and using this to infer function, a tremendous amount remains to be done. Some of the key questions include: What is the structure of the DNA in the nucleus and how does this structure govern DNA transcription? Given the DNA sequence, what determines the RNA and protein structures that the DNA codes for? Given the protein structure, what is its function? How did this function evolve and is it optimized? How can one use this function to design pharmaceuticals that will have a really impact on disease without upsetting the rest of the delicately balanced biological system? What can we learn from other organisms, some that grow under extreme conditions of temperature and pressure, about the nature and limits of living cells and the molecules that make them up?

The above are just some of the key questions, but it is clear from their nature that mathematical and physical/chemical methods will be essential in answering these questions. These methods provide the tools and language of molecular structure from the smallest to the largest molecules and the fundamental laws to explain how molecules interact and form their three dimensional shape. It is this three dimensional shape which determines the molecular function. We have reached an incredibly exciting time of the determination of protein structure, with over 200 different types of globular protein structures known and an estimate of the order of 10^3 expected to exist in all of biology. Thus, we may soon have examples of every type of globular protein structure, as well as insight into the nature of the gene which determines it.

It is clear that the nature of biological signaling pathways is very complex and involves many feedback loops and fail safe mechanisms. The tools of mathematics are essential to understanding these. These signaling pathways are just one example where there is a connection between the material presented in Molecular and Cellular and in Organismal Biology. How do these molecular signals ultimately get transmitted into neural signals and how can we understand possible defects at every level of these pathways --are defects due to mutations in the proteins, subtle changes in concentration of normal molecules or some external influence? These are exciting and extremely important questions that involve understanding the connections from the molecular to the cellular to the organismal level.

GENOME

In the six years since the MBICO report, genomic sequence information has continued its exponential growth. Sequencing technology is being applied directly to sequence diversity analysis and gene expression analysis via high throughput, chip-based, automated assay systems. This influx has changed both the questions that are asked, as well as the range of the interactions considered.

For example, high throughput expression data are now both tissue-specific and specific to stages of development. Over 300,000 human expressed sequence tags are now available in public databases, representing at least 40,000 human genes. Moreover within the next 5 years, as many as 50 complete genomes will be sequenced. Indeed the complete genomes of number of simple organisms have already been sequenced (see e.g. (Fleischman, 1995)), the sequence of the yeast genome has recently been published (see e.g. Williams, 1996), and *C. elegans* is reported to be a year or two away. No one is sure as how best to exploit genomic data but it is clear that there will soon be an explosion of biological information on an unprecedented scale.

It will become increasingly important to carry out comparisons of entire genomes rather than just single genes, with a concomitant expansion in the time to compute. Multiple comparisons remain even more problematic. A similar expansion of queries from local regions of interest (say 50,000 bp) to long range patterns of sequence or expression is now necessary, with synthetic regions on the order of 25 Mb considered a reasonable length for consideration.

Biological and biochemical research is producing exponentially-growing data sets. In addition to the examples cited above of DNA sequences (currently doubling about every 6 months) and gene expression data (chips supporting 1000s of assays per day), combinatorial library screens (10,000s of compounds against 1000s of targets) are producing vast quantities of systematic data on function. Technological developments will increase

these data acquisition rates by an order of magnitude or more in the next few years.

Significant work is required to develop data management systems to make these data not just retrievable, but usable as input to computations and amenable to complex, ad hoc queries across multiple data types. Significant work is also required on techniques for integrating data obtained for multiple observables, at different scales, with different uncertainties (data fusion) and for formulating meaningful queries against such heterogeneous data (data mining).

For example, it should be possible in the future to ask what differences to expect in the kinetic efficiencies of a signal-transduction pathway across multiple individuals, given the differences in the sequences of the proteins involved in the pathway. Answering such queries will require improvements in data models, heterogeneous database management systems, multivariate correlation analysis, molecular structure prediction, constrained-network modeling, and uncertainty management.

PROTEIN STRUCTURE AND FUNCTION

As the amount of genomic data grows, three dimensional structure will provide an increasingly important means for exploiting and organizing this information. Structure provides a unique yet largely unexplored vehicle for deducing gene function from sequence data. Structure also links genomic information to biological assays and serves as a basis for rational development of bioactive compounds, including drugs and vaccines.

Research opportunities in this area can be divided into four distinct categories: experimental structure determination, structure prediction, structure exploitation of globular proteins, and modeling of membrane proteins, where the determination of high resolution structures is much more difficult.

Structure Determination

During the past decade, advances in protein crystal growth, diffraction data collection and experimental phase determination have led to an explosion of structural information. (Ringe and Petsko, 1996) Despite this rapid growth, the demand for new structural data remains high. Areas where mathematical and computational approaches are still needed to increase the throughput further to include direct phase determination, improved structure solution by molecular replacement, and automated electron density map interpretation.

1. Direct phasing: The phases of diffracted x-rays cannot be observed; they must be deduced experimentally by indirect methods. Despite recent advances in experimental phase determination by techniques such as MADD phasing (Leahy *et al*, 1992), this step is often a bottleneck in structure solution. Direct solution to the phase problem for macromolecule crystal structures would revolutionize structural biology.
2. Improved molecular replacement methods: As the database of solved structures grows, molecular replacement methods, in which a homologous model of the unknown protein structure is used to phase the measured diffraction pattern (Arnold and Rossmann, 1992), will become increasingly important. This method often fails, especially when sequence identity between the unknown protein and the homologous structure is low. Better approaches to molecular replacement are needed as are better methods for model building of homologous structures.
3. Automated electron density map interpretation: Once the measured diffraction amplitudes have been assigned phases, an electron density map is calculated by Fourier summation. This map must then be interpreted in terms of an atomic model. Present computer graphics methods for fitting models to electron density are tedious, labor intensive and not always accurate. Computational approaches to automating a large part of this process are thus needed. These can take several forms: recognition of overall protein folding types from low to medium resolution maps (with impact on electron microscopy and electron tomography as well); automated identification of secondary structure; automated chain tracing, and alignment of the amino acid sequence with the electron density.

NMR: Solution NMR is now providing high resolution protein, DNA and RNA structures that rival those from x-ray crystallography. The limit to the size of molecules whose structures can be solved by NMR is dictated by chemical complexity, solubility, redundancy and molecular tumbling time. For proteins, the current upper size limit is currently about 200 KDa. Mathematical and modeling issues that pertain to NMR structures include developing methods to account for the effects of molecular dynamics, to specify the reliability of the structures, and to specify regions of molecular disorder and/or under-determined constraints. Data reduction for NMR should be automated. Current research along this line includes developing data management tools for assigning resonances and keeping track of cross-peaks from multi-dimensional experiments (Johnson and Blevins, 1994).

Structure prediction

The most effective methods of structure prediction currently available involve constructing models of proteins with unknown structures based on templates derived from protein structures that have been determined (see e.g. Bowie *et al*, 1991). There has been remarkable progress in the development of these "fold recognition" methods in the past few years and they offer new opportunities in structure prediction that simply did not exist a few years ago (see e.g. the November 1995 issue of *Proteins* (Asilomar, 1995)).

Fold recognition methods can be used to predict the structures of proteins that have not yet been determined experimentally and to find homology relationships between proteins that cannot be detected with traditional sequence alignment methods. The challenges that now arise offer research opportunities in a number of areas. These include the integration of structural information in sequence alignment methods, the development of improved scoring functions for the association of a given sequence with a given structure (see e.g. Bryant and Lawrence, 1993), and the identification of folding templates that focus on key structural elements to be matched to sequence fragments (Orengo *et al*, 1995). These problems will all require the development of new computational methods that allow the analysis and integration of large quantities of structural and sequence data and new simplified physical models that are designed to the requirements of this emerging field.

Once an overall structural template has been derived, there is a need for methods to predict three dimensional structure at the atomic level. There has been significant progress in the past few years in the building of site chain conformations onto backbone templates (see e.g. Lee and Subbiah, 1991) but faster and more accurate solutions to this problem would be extremely useful. Assuming the conserved structural framework regions are known there is also a need for new methods which model the structures of loops onto fixed structural framework regions (see e.g. Levitt, 1993) a problem which is of unique importance for membrane proteins. These can benefit from fast minimization and conformational search procedures and from improved physical models which relate structure to free energy (see e.g. Smith and Honig, 1994)).

Structure exploitation

The growing body of structural information provides a new way of organizing biological data, with applications including the prediction of function given a structure, the discovery of new principles of protein-protein interactions, and the discovery of new evolutionary relationships that were not evident from sequence alone. Structure determination is usually done to address fundamental problems in cell biology, biochemistry or pharmacology. The specific questions raised by a structure include: where on the protein surface are the binding sites? What are the chemical groups that prefer to bind to these sites? How do the protein and ligand structures change in response to binding? What are the roles of protein and cofactor groups in catalysis? How do protein dynamic properties influence protein function?

The construction of a new class of protein structure/function databases offers a possible approach to these problems. For example, the characterization of different protein binding sites in terms of physical and geometric properties will be useful in predicting the function of new proteins whose structures have been determined, and more, generally, provides a new way of organizing and interpreting biological data. This area offers research opportunities in problems including the construction of new methods to represent three dimensional objects and their incorporation into databases, the merging of these databases with sequence and function databases, and the

development of new physical models to characterize functionally active regions in proteins.

Structure-based drug design requires locating all usable binding sites followed by the design of small molecules that bind tightly and specifically to them (Guida, 1994). Existing computational methods often fail because they do not adequately account for solvent effects (see e.g. Eisenberg and McLachlan, 1986) nor for the possibility of conformational adjustment (Kearsley *et al*, 1994) Better procedures are urgently needed.

Studies of enzyme catalysis ultimately require simulation of entire reaction pathways including all bond breaking and bond-making steps as well as the random motion of the enzyme substrate system. Existing methods of combining quantum mechanical and molecular mechanical potential functions to carry out such simulations are still rather inaccurate. This is particularly true for the interactions of metal ions and clusters which are found in a high percentage of enzymes. Improved mathematical and computational methods are needed in all of these areas and it is an area of much active research (Gao, 1996).

One new experimental area that is certain to have major impact on the exploitation of structural information is combinatorial chemistry. (Gordon *et al*, 1994) New techniques for high-speed parallel synthesis of novel organic compounds are generating libraries of literally hundreds of thousands of molecules, many of which bind to important biological targets. Methods must be developed for organizing, correlating and interpreting the plethora of structure/activity data produced by screening such libraries. The union of combinatorial chemistry and structural biology offers the possibility of deducing the rules for molecular recognition, which may ultimately allow us to build accurate models of multiprotein complexes from the structures of their components. The merging of small molecule and structural databases offers unique and important challenges in this regard.

Membrane Proteins

Study of membrane proteins presents special challenges, but also promises to yield exciting and important information. Greater understanding of membrane protein structure and function will enhance dramatically our understanding of basic biochemical processes such as signal transduction, and make possible significant advances in biotechnology (e.g., receptor-based biosensors) and biomedical sciences (e.g., structure-aided drug design). Technical problems make it difficult or impossible to determine high-resolution structures for most membrane proteins at present. However, a great deal of experimental data is available for many membrane proteins, and this information can often be used in concert with computational tools to generate reasonable three-dimensional models (Findlay, 1996). The models in turn are beneficial in formulation of hypotheses and design of future experiments (Kontoyianni and Lybrand, 1993). A number of developmental issues must be addressed to enhance modeling capabilities for study of proteins in general, and membrane proteins in particular. For example, it is not well understood at present how much "constraint" information is needed to permit construction of a reasonable three-dimensional model structure, or even which types of experimental information are most useful in model building exercises. Additional methodological developments are also needed for improved representation and treatment of lipid bilayers (e.g., efficient treatment of long-range electrostatic interactions, modified Hamiltonians for representation of anisotropic pressure tensors, etc.) and lipid-protein interactions. A number of prokaryotic membrane proteins are now quite well characterized (e.g., bacterial chemotaxis receptors (Bourret *et al*, 1991) and porins (Kreusch and Schulz, 1994), and can serve as useful models for more complex membrane proteins from higher organisms. These systems are ideal test cases for evaluation of new procedures for membrane protein modeling.

Rapid progress in the understanding of membrane protein structure and function has been hindered by the lack of a large number of high-resolution structures. Structures from x-ray crystallography are limited to those complexes that crystallize, whereas those from high-resolution solution NMR are limited to cases where the assemblies have sufficiently short correlation times to produce narrow lines. Techniques from solid state NMR, including rotational resonance (RR) and rotational echo double resonance (REDOR) and EPR spectroscopy (Steinhoff *et al*, 1994), offer special opportunities for obtaining highly specific distance constraints for membrane proteins. A promising avenue of research is to delineate the minimum amount of distance information needed to specify a structure, and to predict in what order one could perform the least number of specific NMR or EPR experiments to arrive at a structure.

NUCLEIC ACIDS

The problem of RNA structure prediction and DNA and RNA interactions with proteins is of central biological interest. There is a need here for improved physical models to describe the interactions of nucleic acids which differ from most proteins in that they induce large local electric fields. Recently, methods have been developed for treating highly charged macromolecules which are surrounded by concentrated ion atmospheres (see e.g. (Misra *et al*, 1993; York *et al*, 1995). These and related methods open up a variety of opportunities for simulating important biological phenomena involving nucleic acids at atomic level resolution.

The explosive growth of information about RNA structure and function offers new opportunities that were nonexistent a few years ago. Requirements in this area range from computational and mathematical techniques to describe the interaction of large fragments (see e.g. (Easterwood *et al*, 1994) which are treated as rigid structural units to accurate atomic-level representations. Similarly, methods must be developed to integrate experimental and phylogenetic data into modeling studies (Jaeger *et al*, 1994).

SIMULATIONS

Simulations of molecules of biological interest use computational representations that range from simple lattice models to full quantum mechanical wave functions of nuclei and electrons. If one has access to a macromolecular structure derived from NMR or X-ray crystallography, then one can begin with a full atom representation and fruitfully examine "small changes" in the system such as ligand binding or site specific mutation. Again, the goal is to reproduce and predict structure, dynamics and thermodynamics. In fact, simulations can provide the connecting link between structure (X-ray and NMR) and function (experimental measurements of thermodynamic properties).

In the last 10 years, because of increased computer power, molecular dynamics calculations have progressed from the short-time simulation a macromolecule without explicit solvent to full representations of solvent and counterions carried out over a few nanoseconds (Berendsen, 1996). Developments in both hardware and software for parallel computing have played a major role. However, the longest time simulations that have been carried out are still 9 orders of magnitude away from the typical time scale for experimental protein folding. Simplified but realistic models, for example using a continuum treatment of the solvent (Gilson *et al*, 1995), could increase the time scale by 1-2 orders of magnitude. Continuum representations may more readily incorporated into Monte Carlo methods and thus allow large movements of the molecule during simulation (Senderowitz *et al*, 1996). In some cases, the use of Langevin and Brownian dynamics and multiple time step algorithms (Humphreys *et al*, 1994) may be warranted. The simulation of biological molecules at the molecular level has generated much excitement and these approaches have become an increasingly important partner with experimental studies of these complex systems.

Electrostatic interactions are a crucial component in the structure and function of biological macromolecules. In the last few years electrostatic models based on numerical solutions to the Poisson-Boltzmann (PB) equation have been used extensively as a basis for interpreting experimental observations on proteins and nucleic acids (Honig and Nicholls, 1995) including for example the prediction of the pKa's of ionizable groups (see e.g. Bashford and Karplus, 1990). Electrostatic potential plays a special role in membrane phenomena: the energies involved are large and the experimental effects of potential changes are also large, often dominant. The extension of PB methods to membranes and channels is an area of great interest.

BIO-INSPIRED MATERIALS

Bio-inspired materials represent a special area of opportunity for developing new high-performance engineering materials based on ideas inferred from Nature (Tirrell *et al*, 1994). For example, the proteins derived from spider silk serve as the inspiration for high-strength fibers (Simmons *et al*, 1996); the adhesives from barnacles suggest how to produce glues that cure and function underwater; and the complex protein-inorganic interactions in

mollusk shells supply ideas for producing ceramics that are less brittle than current ones. It is likely that ultimate bio-inspired materials will be chimeric, that is, they will be produced as a hybrid between biological and synthetic components. Consequently, these materials represent a special class of the protein folding problem and of polymer physics. In addition to the molecular level interactions, the ultimate mechanical properties of such materials derive also from long-range interactions, orientation and crystallite size. Models from polymer science and from protein folding must be combined and adapted to predict how mechanical properties such as modulus, strength and elasticity depend on these physical parameters. Once such models are also able to explain the mechanical properties of wild-type biomaterials, they can be used in a predictive sense to guide the production of chimeric materials.

MOLECULAR HISTOLOGY

Understanding the spatial conformation of biological macromolecules (DNA, RNA, protein) and functional changes in conformation provides an ongoing challenge to mathematics. Analytical and computational models based in geometry and topology continue to be very successful in providing a theoretical and computational framework for the analysis of enzyme mechanism and macromolecular conformation (Rybenkov, 1993; Schlick and Olson, 1992; White, 1992; Sumners *et al*, 1995; Lander and Waterman, 1995).

New experimental modalities, such as cryo-electron microscopy, (Stasiak *et al*, 1996), optical tweezers (Smith *et al*, 1996), provide spatial and structural data of ever-increasing resolution. This new spectrum of high-resolution data will require correspondingly high-resolution mathematical models to aid in the design and interpretation of experiments. Refinement of existing models will provide a starting point, but new ideas and new combinations of old ideas are needed. One particularly important need is the development of efficient descriptors of spatial conformation of macromolecules; descriptors that will afford efficient database entry and retrieval of information, while encoding biologically significant structural information.

IV. ORGANISMAL BIOLOGY

OVERVIEW

The central organizing theme for Cells and Cell Systems is how behavior and function at one level of organization emerges from the structure and interactions of components at lower levels. In the set of topics described in this section the lower level of organization is subcellular or cellular. Though some of the subcellular components that play a role in these models are molecular, the focus is not on the structure of those molecules, but on the part that they play in cellular and multicellular function. The section on CELL SIGNALING deals with the role of specific molecules in regulation of processes such as cell division, cellular communication, and gene expression. In the MECHANICS AND EMBRYOLOGY section, the focus is on how mechanochemical processes at the molecular level can drive the processes that lead to macroscopic changes in shapes of tissues and organs. The problems discussed in BIOFLUID DYNAMICS again start at the level of individual (bacterial) cells, with substructures (flagella) interacting at tiny scales with the hydrodynamics to produce macroscopic behavior (swimming).

The sections on IMMUNOLOGY AND VIROLOGY and NEUROSCIENCES focus on scientific problems that involve larger multicellular systems. Understanding the immune system requires insights about how classes of molecules found on the cell surface generate the complex signals which lead to a normal immune response; this response, which includes a memory of previous interactions with antigens, is a property of the entire immune system, not of individual cells. Similarly, the nervous system can be studied at the level of individual cells, to understand how the biophysical properties of cellular membranes contribute to the responses of individual cells; but an understanding of the functioning of the nervous system also requires a study of the behavior of large scale networks of neurons.

CELL SIGNALING

Control of cellular processes, mediated by interactions of signaling molecules and their cell surface receptors, is a central and unifying theme in current experimental cell biology. Within the past five years, techniques of molecular biology have revealed many of the kinases, phosphatases and other molecules involved in signal transduction pathways, as well as molecular sub-domains and sequence motifs that determine distinct functions. New techniques for measuring phosphorylation, calcium fluxes, and other early biochemical responses to receptor interactions are being applied to study many cell signaling systems (e.g., chemotactic bacteria, neurons and lymphocytes). Genetically engineered experimental systems consisting of homogeneous cell lines, transfected with homogeneous populations of wild type and mutant receptors and effector molecules, have facilitated acquisition of much of the new information about the intracellular molecules that mediate signal transduction. Improved measurement and experimental design make mathematical modeling an increasingly feasible tool for testing ideas about the interactions of these molecules.

Modeling has contributed to our understanding of key cell surface interactions (e.g., ligand-induced receptor aggregation, cell-cell interactions, and cell adhesion). Modeling has also clarified the nature and effects of cellular responses (e.g., internalization and secretion of proteins, cell division and differentiation, and cell motility). Recent combinations of modeling and experiment have brought a deeper understanding of the role of calcium in the regulation of cell division, neuronal communication, regulation of muscle contraction, pollination, and other cellular processes. (Silver, 1996) Representative descriptions of collaborative work applying mathematics to problems in experimental cell biology are found in Alt *et al*, 1996; Goldstein and Wofsy, 1994 and Lauffenburger and Linderman, 1993. Other recent examples of the productive application of theory to cell signaling and cell motility include Alon *et al*, 1995; Bray, 1995; Jafri and Keizer, 1994; Naranja *et al*, 1994; Tranquillo and Alt, 1996 and Tyson *et al*, 1996. Over the next few years, we can expect mathematical modeling to play a central role in the design and interpretation of experiments aimed at understanding in detail the biochemical reactions leading from receptor interactions to changes in gene expression, cell division, and other functional responses.

MECHANICS AND EMBRYOLOGY

Recent advances in instrumentation have made it possible to measure motions and mechanical forces at the molecular scale (Svoboda and Block, 1994). Concomitant with these new mechanical measurements are crystallographic and x-ray diffraction techniques that have revealed the atomic structure and molecular geometry of mechanochemical enzymes to angstrom resolutions (Rayment and Holden, 1994). Together, these techniques have begun to supply data that has revived interest in cellular mechanics, and reinvigorated the view of enzymes as mechanochemical devices. It is now possible to make realistic models of molecular mechanochemical processes that can be related directly to experimentally observable, and controllable, parameters (Peskin and Oster, 1995). These advances in experimental technology have initiated a renaissance in theoretical efforts to readdress the central question: how do protein machines work? More precisely, how is chemical energy transduced into directed mechanical forces that drive so many cellular events?

Embryology has also moved beyond descriptive observation to encompass genetic control of development and localization of protein effectors. The stress and strain measurements that are now possible at the cellular scale promise to unite the genetics, biochemistry and biomechanics of development (Oliver *et al*, 1995). By characterizing the mechanical properties of embryonic cells and tissues, mathematical models can be used to discriminate between various possible mechanisms for driving morphogenesis (Davidson *et al*, 1995).

Examples encompass all phenomena that involve the coordinated movement of macromolecules, cells or tissues. How do embryonic cells crawl and bacteria swim (Dembo, 1989; Berg, 1995; Mogilner and Oster, 1996)? How are proteins shuttled about the cell (Scholey, 1994)? What drives the grand progression of cell division (Murray and Hunt, 1993)? What drives the shaping of tissues and organs during embryonic development (Murray and Oster, 1984; Brodland, 1994) and the reshaping of organs after injury (Tranquillo and Murray, 1993; Olsen *et al*, 1995)?

BIOFLUID DYNAMICS

Because of the ongoing revolution in computer technology, we can now solve fluid dynamics problems in the three spatial dimensions and time (Ellington and Pedley, 1995). This opens up biological opportunities on many different scales of size. On the organ scale, for example, one can now perform fluid dynamics simulations of the embryonic and fetal heart at different stages of development. Such models will help to elucidate the role of fluid forces in shaping the developing heart. The swimming mechanics of microorganisms are also accessible to computer simulation. A particularly challenging problem in this field concerns the intense hydrodynamic interaction among the different flagella of the same bacterium: When the flagella are spinning so that their helical waves propagate away from the cell body, they wrap around each other to form a kind of superflagellum that propels the bacterium steadily along; when their motors are reversed and the flagella spin the other way, the superflagellum unravels and the bacterium tumbles in place. Because of the difficulty of measuring microscopic fluid flows, hydrodynamics within cells is a much neglected aspect of cellular and intracellular biomechanics. Indeed, computation provides our only window onto this important aspect of cellular physiology. The incompressibility and viscosity of water have the effect of coupling motions along different axes, and between objects quite distant from one another; biomolecular processes are also modulated by the necessity of moving water out of the way. A new feature in this realm of micro and nano hydrodynamics is the importance of Brownian motion and the related significance of osmotic mechanics (including sol-gel transformations) for controlling fluid motions.

Progress in this field will depend on access to large-scale scientific computing. It is important that the best technology be made available to scientists on a scale sufficient to sustain this kind of research. This will also necessitate supporting people with the expertise to make effective use of these powerful machines. At universities, such people are often in non-faculty, non-tenured research positions. We need support to sustain their crucial role.

IMMUNOLOGY AND VIROLOGY

During the last two years mathematical modeling has had a major impact on research in immunology and virology. Serious collaborations between theorists and experiment provided breakthroughs by viewing experiments in which AIDS patients were given potent anti-retroviral drugs as perturbations of a dynamical system. Mathematical modeling combined with analysis of data obtained during drug clinical trials established for the first time that HIV is rapidly cleared from the body and that approximately 10 billion virus particles are produced daily (Ho *et al*, 1995). This work had tremendous impact on the AIDS community and has, for the first time, given them a quantitative picture of the disease process. The impact of this type of analysis has extended beyond AIDS, and opportunities exist for developing realistic and useful models of many viral diseases. Challenges remain in studying drug therapy as a nonlinear control problem, and the issue of how rapidly viruses mutate and become drug resistant under different therapeutic regimes needs to be considered. Such issues also apply to the development of antibiotic resistance in bacterial disease.

Opportunities exist for substantial advances in immunology by the use of modeling techniques. Molecular modeling is providing insights into the structure and function of the cell surface molecules crucial for the operation of the immune system: immunoglobulin, the T cell receptor, and molecules coded for by the major histocompatibility complex genes, as well as molecules being recognized by the immune system. The biochemical sequelae of molecular recognition involve the generation of complex biochemical and enzymatic signals, whose net effect are changes in gene expression followed in many cases by cell proliferation, cell differentiation and cell movement. How these changes are orchestrated to produce an immune response remain to be elucidated. However, modeling can give us insights into how cells interact by direct contact and via secreted molecules, cytokines, to produce the coordinated behavior necessary to meet immune system challenges.

NEUROSCIENCES

The fundamental challenge in neuroscience is to understand how behavior emerges from properties of neurons and networks of neurons. Advances in experimental methodologies are providing detailed information on ionic channels, their distribution over the dendritic and axonal membranes of cells, their regulation by modulatory

agents, and the kinetics of synaptic interactions. The development of fast computing, sophisticated simulation tools, and improved numerical algorithms has enabled the development of detailed biophysically-based computational models that reproduce the complex dynamic firing properties of neurons and networks. Such computations provide a two-fold opportunity for advancing our knowledge: (1) they both explain and drive new experiments, (2) they provide the basis for new mathematical theories that enable one to obtain reduced models that retain the quantitative essence of the detailed models. These reduced models, which allow the bridging of multiple spatial and temporal scales, are the building blocks for higher level models.

Modeling tools and mathematical analysis allow us to address the central question: What are the cellular bases for neural computations and tasks such as sensory processing, motor behavior and cognition? (Koch and Segev, 1989; Bower, 1992) More specifically, how do intrinsic properties of neurons combine in networks with synaptic properties, connectivity, and the cable properties of dendrites to produce our interaction with the world? Neural modulators affect both the intrinsic currents and the synaptic interactions between neurons. (Harris-Warrick *et al*, 1992) The effects of these changes at the network level are difficult to work out even for small networks. The largest challenge in this area is to understand how systems with enormous numbers of degrees of freedom and large numbers of different modulators combine to produce flexible but stable behavior. The geometry and electrical cable properties of the branching dendrites of neurons also affect network activity. (Stuart and Sakmann, 1994) Mathematical analysis is needed to interpret the results of massive computations, and to incorporate the insights into network models.

The dynamics of neural networks (Golomb *et al*, 1996; Kopell and LeMasson, 1994) affect both cognitive and sensory-motor behavior. To understand motor behavior, one must construct models that illuminate the role of feedback between neural and mechanical subsystems. For sensory systems, one of the most important problems is to understand how the brain controls the data that it receives, including understanding more rigorously the quantitative parameterization/description of natural stimuli. A current active area of inquiry is the characterization of codes used in information processing in the nervous system. (Softky and Koch, 1993; Shadlen and Newsome, 1995; Softky, 1995) Among the issues raised by this question is how the complex dynamics of the cortex can help shape responses to stimuli, including selecting pathways that lead to different behavior.

Modeling has become an accepted and central tool in neurobiology. The current scientific goals listed above create specific challenges in modeling. Some of these concern the handling and interpretation of the far greater volume of data that is now, or potentially, available, e.g. through multiunit recording techniques. With very large and complex models (Whittington *et al*, 1995), techniques for systematically choosing parameters are important, as are methods for comparing models and understanding their differences. Both computers and mathematical analysis will play major roles in dealing with the technical problems; mathematical analysis remains the fundamental tool for providing a deep understanding of how models differ in their predictions.

V. ECOLOGY AND EVOLUTIONARY BIOLOGY

OVERVIEW

Evolution is the central organizing theme in biology (e.g. Roughgarden, 1979), and its manifestation in the relationships among types of organisms spans levels of organization, and reaches out from biology to earth and social sciences. Thus, the core problems in ecology and evolution run the gamut from those that address fundamental biological issues to those that address the role of science in human affairs. Fundamental challenges facing ecologists and evolutionary biologists relate to the threats of the loss of biological diversity, global change, and the search for a sustainable future, as well as to the continued search for an understanding of the biological world and how it came to assume its present form. To what extent is the organization of the biological world the predictable and unique playing out of the fundamental rules governing its evolution, and to what extent has it been constrained by historical accident? How are the interactions among species, ranging from the tight interdependence of host and parasite to the more diffuse connections among plant species in a forest, manifested in their coevolutionary patterns and life history evolution? What are the evolutionary relationships among closely related species, in terms of their shared phylogenetic histories? How do human influences, such

as the use of antibiotics and pesticides, exploitation of fisheries and land, and accelerated patterns of global change, influence the evolutionary dynamics of species and patterns of invasion? To what extent can an evolutionary perspective help us to prepare for the future, in terms of understanding what species might be best suited to new environments? The latter is important both in terms of natural patterns of change, and deliberate manipulations through breeding and species introductions.

Among the central issues are those relating to biodiversity (Tilman, 1994) How it is maintained, how it supports ecosystem services, likely patterns of change, and steps to preserve it. This leads to a fundamental set of core issues, both in terms of their importance, and in terms of their ripeness for success:

Conservation biology, and the preservation of biodiversity

What factors maintain biodiversity? How can new approaches to phylogenetic analyses, in clarifying the evolutionary relationships within and among species, help us to understand how we should measure biodiversity? How are ecosystems organized into functional groups, ecologically and evolutionarily, and how does that organization translate into the maintenance of critical ecosystem processes, such as productivity and biogeochemical cycles, as well as climate mediation, sequestering of toxicants, and other issues of importance to human life on earth.

Global change

What are the connections between the physical and biological parts of the global biosphere, and the multiple scales of space, time and organizational complexity on which critical processes are played out? (Bolker *et al*, 1995) In particular, how are individual plants influenced by changes in atmospheric patterns; and, more difficult, how do those effects on individual plants feed back to influence regional and global patterns of climate and biological diversity? How do effects on phytoplankton and zooplankton relate to each other, and to the broader patterns that may be observed?

Emerging disease

How do patterns of population growth and resource use, as well as the profligate use of antibiotics, contribute to the emergence and reemergence of deadly new diseases, many of them antibiotic resistant? (Ewald, 1995) Are there approaches to management of the diversity of those diseases, guided by both an evolutionary and an ecosystem perspective, that can reduce the threat and provide new strategies for mitigation?

Resource management

The history of the management of our sources of food and fiber is not one of unmitigated successes, and many of these crucial resources are threatened to a level that they will be unable to support the needs of humanity in the coming decades. The prospect of large-scale alterations of the earth's physical and biological systems creates a potential conflict between human needs, desires and capabilities. (Walters and Parma, 1996; Walters and Maguire, 1996) This situation is further complicated by the limitations of our understanding and ability to control complex biological systems. We must develop methods for decision-making and management that are appropriate for an uncertain future. (Hilborn *et al*, 1995)

In all of these issues, there are a variety of cross-cutting themes, some biological, some methodological or conceptual. From a biological point of view, the essential point is that all that we see has been shaped by evolutionary processes; from an ecological point of view, it is that organisms do not exist in isolation, but have existed within the context of other species and an abiotic environment, making essential an ecosystem perspective on issues ranging from the management of diseases to the management of our global surroundings. Indeed, a central challenge is to understand how the properties even of ecosystems, those loose assemblages of species in particular habitats, can be understood in terms of the diffuse coevolution of the components within

very open systems.

From a modeling point of view, fundamental issues remain how to deal with variation within as well as variation among units, for example in the importance of heterogeneity in evolutionary processes or infectious transfer. The interplay among processes operating on very different scales also pervade these questions, from evolution through global change. And finally, techniques for simplification, and for relating behaviors at the level of individuals to macroscopic descriptions, provide the tools for making the essential connections.

Progress in all of these research areas will derive from the application of a suite of approaches, ranging from explicit spatial and stochastic simulations to more compact (Durrett and Levin, 1994) mathematical descriptions that allow analysis and simplification. Recent advances in computer technology have opened up the possibility of including much more detail than ever before in simulation approaches, yielding the possibility of including much more biological detail. This detail comes at a cost, however. The ability to generate information does not equal understanding, and the mathematical challenge is to develop techniques which can include the essential details driving the complex models, while allowing an understanding of the features driving the biological behavior at a deeper level that will allow generalization. This will require both close attention to the underlying biological details and fundamental mathematical progress in taking appropriate limits and achieving manageable simplification of complex, spatially explicit, stochastic models.

Below, we focus on modeling opportunities in some of the specific subfields in the general areas of ecology and evolution.

POPULATION GENETICS

While evolution is the great unifying principle underlying all of biology, evolutionary genetics forms the foundation of evolution. Challenging mathematical and computational applications in this critical area range from the development of theoretical frameworks from which to infer the operation of evolutionary mechanisms such as natural selection at the molecular level through the organismal level, to understanding the genetic basis of interactions among species.

One critical area, still in its infancy, concerns the identification and genetic analysis (Coyne *et al*, 1991) of genes that play key roles in species and environmental interactions. The mapping of such quantitative trait loci consists of three interrelated inference problems: detecting the effects of these loci, determining the number of major loci affecting a trait, and locating them relative to genomic markers. A complete solution thus involves problems of testing, model selection, and estimation. Once ecological and genetic analysis of traits limiting adaptive responses is complete, it will be possible to address crucial evolutionary questions such as the relative importance of gene flow, genetic trade-offs, and genetic constraints.

A second exciting area concerns life history evolution, which often focuses upon the timing of life history events or the allocation of organismal resources and time among conflicting demands such as longevity and fecundity. Evolution of these traits can be studied from quantitative genetic descriptions in which transient dynamics are explored (Tuljapulkar and Wiener, 1995), while the selective environment is reduced to a selection gradient. Alternatively, the nature of the environment's selective effect on a trait can be explored through optimization approaches. There is a pressing need for more complex formulations such as models bridging the gap between problems of allocation and timing, models explicitly (Charlesworth, 1994) incorporating how genes act at different ages and over time, for models at the interface between life history evolution and behavior (Charlesworth, 1994), and for models examining how life histories (Tuljapulkar, 1994) are influenced by temporal and spatial variation in the environment.

Beyond the species level, the coevolutionary dynamics of the quantitative traits that are often involved in species interactions pose many challenges and opportunities to theoretical, computational, and mathematical biologists that cut across all areas of ecology and evolution. For example, the study of the evolution of virulence (Frank, 1993, 1994) in insect-parasitoid-host systems and fungi-virus interactions in plants and the study of mechanisms of specialization and the analysis of hybrid zones are part of the cutting-edge research being conducted at the

interface of biology and the mathematical sciences.

With the rapid accumulation of sequence data for entire genomes, we are now poised to analyze the set of genes, their order and organization, codon usage, etc. across taxa (Griffiths and Tavare, 1996) and how and perhaps why this has evolved over time. (Thorne *et al*, 1992) This requires an increased ability to model how information is represented and acted upon in biological systems (Griffiths and Tavare, 1996) based on tools from such fields as discrete mathematics, combinatorics, and formal languages. Novel, perhaps ad-hoc formulations are needed to form the mathematical basis of genomic analyses because classical quantitative formulations of notions such as information, similarity, and classification - all inextricably related to biology - are inadequate. Correspondingly, methods for organizing vast sequence data into data structures and databases suited for the most efficient data storage and access are needed, along with improved algorithms for sequence analysis and the identification of homologies among sequences.

Population genetic surveys of the genetic structure of natural populations are a critical tool from which to deduce the evolutionary history of, and evolutionary forces at work in, natural populations. Current population genetic theory and data analysis methods are largely based upon single or a few genetic loci, each with two alternate forms (alleles). Current data, however, typically includes the genetic makeup at a large number of genetic markers which, with the advent of new molecular techniques such as the polymerase chain reaction, are increasingly hypervariable with a large number of alternate forms segregating at each. New theoretical frameworks and statistical methods are needed to extract and utilize the full evolutionary information contained in these complex data sets.

CONSERVATION BIOLOGY

Virtually all important questions in conservation biology require making predictions, so theory and mathematical methods have played and will continue to play a central role. Although many of the underlying scientific issues have been defined during the past decade, many questions remain to be resolved. What species would be lost in the wake of an invasion, and what are the effects on ecosystem function? For example, what are the consequences of the replacement of native fish species by introduced species? Substantial progress is likely (and needed) in the near future in understanding the dynamics of invading exotic species, determining more carefully the role genetics plays in the dynamics of rare or endangered species, and in the ecological dynamics of threatened species.

Theoretical studies have focused on the population size or characteristics needed to allow species to maintain the genetic diversity necessary to allow long term persistence. (Lande, 1993, 1994) These answers have shown that an effective population size is required, but further work is needed to understand how effective population size is related to actual population size and structure and life history characteristics -- what can actually be observed. These lead to interesting mathematical challenges dealing with structured populations, and with integrating ecological and genetical models.

The impact of invading exotic species on existing native ecological communities and species is perhaps the most important conservation issue today (OTA, 1993). There has been almost no development of theories predicting rates of spread of species within the context of even simple communities, and the related mathematical problems of coupled reaction diffusion equations are challenging as well. Although the basic mathematical models of spatial spread can be traced at least as far back as Fisher (1937), recent work has shown that the situation is far more complex, as rates of spread can vary by at least an order of magnitude as model assumptions are changed. (e.g. Lewis & Kareiva, 1993; Zadoeks and Van Den Bosch, 1994) Further work will be able to lead to robust quantitative predictions of rates of spread.

MANAGEMENT OF NATURAL SYSTEMS

In recent years there has been an abrupt shift in management philosophy. (Hiborn *et al*, 1995) The old goal of managing individual species in order to reach and maintain optimal conditions has been replaced by a new goal

of maintaining ecosystem function and adapting to new conditions or changes in the system. This shift reflects a more mature attitude towards nature that recognizes the limitations of our knowledge and capabilities, the importance of interactions between species and an appreciation of the dangers of a command and control mode of operation.

This new approach to management makes it possible to apply elements of the scientific method in a new and significant context: we may design experimental management schemes to provide information that is required to improve the management process and adapt to changes, even unforeseen changes. This new approach challenges our mathematical and statistical skills. Successful adaptation requires effective and timely organization of data through estimation of parameters that affect system dynamics, including the dynamics of our learning. That information then must be translated into an assessment of the likely consequences of management strategies and actions.

The major challenges facing the human species cannot be met by a reductionist or piecemeal approach. Instead we must muster all of our ingenuity and resources to learn about the behavior of intact natural systems under stress and perturbation, and adapt our human institutions to a finite and vulnerable world.

GLOBAL CHANGE AND BIODIVERSITY

Climate change and associated changes in greenhouse gases have made imperative the examination of the potential impacts on natural systems, and associated feedbacks. Advances in computational capabilities have made possible the construction of detailed individual-based models that take account of the responses of individual trees to changes in environmental conditions, and their mutual effects. Yet such models are tremendously data-hungry, and have great potential for error propagation. To make their predictions robust, and to allow those predictions to be interfaced with the much broader scale predictions of climate models, and the masses of broad scale information that are becoming available from remote sensing, we must find ways to reduce dimensionality and simplify those overly detailed models. Similar comments apply to models of other systems, such as the aggregation of social organisms from cellular slime molds to marine and terrestrial invertebrates and vertebrates. Methods such as moment closure and hydrodynamic limits, borrowed from other disciplines, are proving remarkably promising, especially when coupled with experimental approaches (Levin and Pacala, 1996).

This represents one of the most challenging and important issues in ecosystem science. At the same time, masses of data are becoming available from global observation systems, and critical experiments are providing understanding of the linkages between ecosystem structure and function, and in particular the role of biodiversity in maintaining system processes. The next 5-10 years hold remarkable potential for integrated theoretical, empirical and computational approaches to elucidate profound and important issues (Field, 1992; Bolker, 1995).

THE DYNAMICS OF INFECTIOUS DISEASES

The subject of infectious disease dynamics has been one of the oldest and most successful in mathematical biology for a century, and has seen powerful advances in recent years in mathematical theory, and in the application of that theory to management strategies (see, for example, Anderson and May, 1991). Much of the literature has assumed homogeneous mixing, so that every individual is equally likely to infect every other individual; but such models are inadequate to describe the central qualitative features of most diseases, especially those that are sexually transmitted, or for which spatial or socioeconomic structure localizes interactions. The classical work of Hethcote and Yorke (1984) on core-group dynamics highlighted the importance of such effects, and formed the basis upon which much recent work rests. Such work, involving spatial structure, frequency and density dependence, and behavioral factors have not only forced us to revise old paradigms, but have reenergized the interplay among nonlinear dynamics, ecology and epidemiology.

VI. CROSS-CUTTING ISSUES

MATHEMATICAL AND COMPUTATIONAL ISSUES SPANNING ALL DOMAINS - - RELATIONSHIP BETWEEN SIMULATION AND MATHEMATICS

The revolution in computer technology enables us to perform complex simulations only dreamed of a decade ago. Effective use of this technology requires substantial use of mathematics throughout all stages of the simulation process: the quantitative (or qualitative) formulation of models, the design of appropriate data types and algorithms, translation of models into efficient computer implementations, estimation of parameter values, visualization of the output, and comparison of simulation results with results of further experimentation. Mathematics is also essential in the critical step of developing algorithms that compute important properties of models without recourse to numerical simulation.

Furthermore, mathematics can significantly enhance our understanding of processes that are studied through simulation. For example, theories of dynamical systems describe patterns that are widespread, so much so that they have been called "universal." The elucidation of such recurring patterns is a central part of mathematics. Mathematics ponders a common language, a context that gives meaning to simulation results and a firm foundation for the algorithmic infrastructure of simulation. Such a foundation ensures that simulation methods are generalizable and capable of generating predictions. Moreover, theory can serve as a basis for reducing models without loss of information, thereby improving the efficiency of large-scale simulations.

MAJOR CHALLENGING ISSUES THAT SPAN ALL AREAS OF MODELING SYSTEMS

A. Integrating data and developing models of complex systems across multiple spatial and temporal scales.

- scale relations and coupling
- temporal complexity and coding
- parameter estimation and treatment of uncertainty
- statistical analysis and data mining
- simulation modeling and prediction.

B. Structure-function relationships

- large and small nucleic acids
- proteins
- membrane systems
- general macromolecular assemblies
- cellular, tissue, organismal systems
- ecological and evolutionary systems.

C. Image analysis and visualization

- image interpretation and data fusion
- inverse problems
- 2, 3, and higher-dimensional visualization and virtual reality

D. Basic mathematical issues

- formalisms for spatial and temporal encoding
- complex geometry
- relationships between network architecture and dynamics
- combinatorial complexity
- theory for systems that combine stochastic and nonlinear effects, often in partially distributed systems.

E. Data management

- data modeling and data structure design
- query algorithms, especially across heterogeneous data types
- data server communication, especially peer-to-peer replication
- distributed memory management and process management.

VII. EDUCATIONAL ISSUES

As noted above, mathematical analysis and computer modeling have become indispensable tools in biology in recent years. These techniques have had a major impact in areas ranging from ecology and population biology to neurosciences to gene and protein sequence analysis and three-dimensional molecular modeling. Mathematical and modeling techniques make it possible to analyze and interpret enormous amounts of data, yielding information and revealing patterns and relationships that would otherwise remain hidden.

Given the essential role that mathematical and modeling techniques play in so many diverse areas of biology, there is a clear need for appropriate training opportunities in computational, mathematical, and theoretical biology. Suitable and practical mechanisms to encourage and nurture training in computational biology might include 1) graduate training grant programs that involve faculty engaged in both computational and experimental approaches, 2) postdoctoral fellowships to encourage mathematicians and computational scientists to pursue research training in biology, and to enable biologists to acquire computational and modeling skills, and 3) summer workshops and short courses to help practicing biologists, mathematicians, and computational scientists to begin to bridge the gap between these rather diverse disciplines.

In addition to training of computational biology specialists, there is a clear and dramatic need for enhanced training in mathematics and computational methods for biological science students or others who might enter the workforce in any scientific discipline. A systematic approach, beginning at the K-12 level, that emphasizes the importance of mathematics and modeling in biology activities (as outlined in the National Science Standards) would help insure that students are better prepared to utilize mathematical approaches in undergraduate biology curricula, and less likely to avoid mathematically rigorous courses in undergraduate programs because of weak mathematics backgrounds or "math phobia". Improved mathematics training at the earliest levels will also likely increase the number of students interested in pursuing graduate study in interdisciplinary areas of mathematical and computational biology. Greater emphasis on mathematics and computational studies at the K-12 and undergraduate levels can also be coupled effectively with programs to encourage women and underrepresented minorities to pursue careers in science, especially in interdisciplinary areas that bridge the biological, mathematical, and computational sciences.

Finally, it should be recognized that computer simulations and mathematical modeling tools can be effective teaching aids in the biological sciences. Topics like protein structure-function relationships benefit greatly from interactive, three-dimensional graphics demonstrations. Computer simulations and animations based on mathematical models can be an extremely effective way to illustrate the behavior and properties of complex systems, ranging from protein-ligand interactions to migration behavior of large animal populations. Therefore, inclusion of mathematical and computational course work as a logical and sequential theme articulated in K-12 and undergraduate curricula will likely have far-reaching benefits for biology education.

REFERENCES

Alon, R., Hammer, D. A., and Springer, T. A. (1995). Lifetime of the P-selectin -- Carbohydrate Bond and Its Response to Tensile Force in Hydrodynamic Flow. *Nature* 374, 539-542.

Alt, W., Deutsch, A., and Dunn, G., eds., (1996). *Mechanisms of Cell and Tissue Motion*. Birkhaeuser Verlag, Basel.

Anderson, R.M. and May, R.M. (1991) Infectious Diseases of Humans. Oxford Univ. Press.

Arnold, B. and Rossmann, M.G. (1986). Effect of Errors, Redundancy, and Solvent Content in the Molecular Replacement Procedure for the Structure Determination of Biological Macromolecules. Proc. Natl. Acad. Sci. USA 83, 5489-5493.

Asilomar (1995). Proteins 23, 295-460.

Bashford, D. and Karplus, M. (1990). pKa's of Ionizable Groups in Proteins - Atomic Detail from a Continuum Electrostatic Model. J. Mol. Biol. 29, 10219-10225.

Berendsen, H. J. C. (1996). Bio-molecular Dynamics Comes of Age. Science 271, 954-955.

Berg, H. (1995). Torque Generation by the Flagella Rotary Motor. Biophys. 68 (4 Suppl), 163S-166S.

Bolker, B.M., Pacala, S.W., Bazzaz, F.A. and Canham, C.D. (1995) Species Diversity and Ecosystem Response to Carbon Dioxide Fertilization -Conclusions from a Temperate Forest Model. Global Change Biology 1, 373-381.

Bolker, B. M., Pacala, S.W., Canham, C., Bazzaz, F. and Levin, S.A. (1995) Species Diversity and Ecosystem Response to Carbon Dioxide Fertilization: Conclusions from a Temperate Forest Model. Global Change Biology 1, 373-381.

Bourret, R. B., Borkovitch, K. A. and Simon, M. I. (1991). Signal Transduction Pathways Involving Protein Phosphorylation in Prokaryotes. Ann. Rev. Biochem. 60, 401-441.

Bower, J.M., Guest Editor (1992). Special Issue: Modeling the Nervous System. Trends In Neuroscience 15, #11.

Bowie, J. U., Luthy, R. and Eisenberg, D. (1991). A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure. Science 253, 164-170.

Brodland, G. (1994). Finite Element Methods for Developmental Biology. In International Review of Cytology, 150, Academic Press, Inc. pp. 95-118.

Bray, D. (1995). Protein Molecules as Computational Elements in Living Cells. Nature 376, 307-312.

Bryant, S. H. and Lawrence, C. E. (1993). An Empirical Energy Function for Threading Protein Sequence Through the Folding Motif. Proteins 16, 92-112.

Charlesworth, B. (1994). Evolution in Age-Structured Populations. Cambridge University Press, 2nd edition.

Coyne, J.A., Aulard, S. and Berry, A. (1991). Lack of Underdominance in a Naturally Occurring Pericentric Inversion in *Drosophila-Melanogaster* and Its Implications for Chromosome Evolution. Genetics 129, 791-802.

Coyne, J.A., Charlesworth, B. and Orr, H.A. (1991). Haldane's Rule Revisited. Evolution 45, 1710-1714.

Davidson, L., Koehl, M., Keller, A. and Oster, G. (1995). How Do Sea Urchins Gastrulate? Distinguishing Between Mechanism of Primary Invagination Using Biomechanics. Development 121, 2005-2018.

Dembo, M. (1989). Mechanics and Control of the Cytoskeleton in *Amoeba proteus*. Biophys. J. 55, 1053-1080.

Doering, C., Ermentrout, B. and Oster, G. (1995). Rotary DNA Motors. Biophys. J. 69, 2256-2267.

- Durrett, R.** and Levin, S.A. (1994) Stochastic Spatial Models: A User's Guide to Ecological Applications. Phil. Trans. Soc. Lond. B. 343, 329-350.
- Easterwood, T. R.,** Major, F., Malhotra, A. and Harvey, S. C. (1994). Orientations of Transfer RNA in the Ribosomal A and P Sites. Nucl. Acid. Res. 22, 3779-3789.
- Ewald, P.W.** (1995). The Evolution Of Virulence - A Unifying Link Between Parasitology and Ecology. J. Parasitology 81, 659-669.
- Eisenberg, D.** and McLachlan, A. D. (1986). Solvation Energy in Protein Folding and Binding. Nature 319, 199-203.
- Ellington, C.P.** and Pedley, T.J., eds. (1995). Biological Fluid Dynamics. Company of Biologists Limited, Cambridge UK
- Field, C.F.,** Chapin III, F. S., Matson, P. A. and Mooney, H. A. (1992) Responses of Terrestrial Ecosystems to the Changing Atmosphere: A Resource-Based Approach. Ann. Rev Ecol. Syst. 23, 201-235.
- Findlay, J. B. C.** (1996). Membrane Protein Models. BIOS, Oxford
- Fisher, R.A.** (1937). The Wave of Advance of Advantageous Genes. Ann. Eugen. (Lond.) 7, 355-369.
- Fleischmann, R.D.,** Adams, M.D., White, O., Clayton, R.A. (1995) Whole-Genome Random Sequencing and Assembly of Haemophilus influenzae Rd. Science 269, 496-512.
- Frank, S.A.** (1993). Evolution of Host-Parasite Diversity. Evolution 47, 1721-1732.
- Frank, S.A.** (1994). Coevolutionary Genetics of Host and Parasites with Quantitative Inheritance. Evolutionary Ecology 8, 74-94.
- Gao, J.** (1996). Hybrid Quantum and Molecular Mechanical Simulations - An Alternative Avenue to Solvent Effects in Organic Chemistry. Acc. Chem. Res. 29, 298-305.
- Gilson, M. K.,** McCammon, J. A. and Madura, J. D. (1995). Molecular Dynamics Simulation with a Continuum Electrostatic Model of the Solvent. J. Comput. Chem. 16, 1081-1095.
- Goldstein, B.,** and Wofsy, C., eds. (1994). Lectures on Mathematics in the Life Sciences 24: Cell Biology. American Mathematical Society, Providence, RI.
- Golomb, D.,** Wang, X-J, and Rinzel, J. (1996) Propagation of Spindle Waves in a Thalamic Slice Model. J. Neurophys 75, 750-769.
- Griffiths, R.C** and Tavaré, S. (1996). Computational Methods for the Coalescent. IMA volume, P. Donnelly and S. Tavaré, eds. In press.
- Guida, W. C.** (1994). Software for Structure-Based Drug Design. Curr. Opin. Struc. Biol. 4, 777-781.
- Harris-Warrick, R.,** Marder, E., Selverston, A. and Moulins, M., eds. (1992). Dynamic Biological Networks: The Stomatogastric Nervous System, MIT Press
- Hethcote, H.W.** and Yorke, J.A. (1984) Gonorrhea: Transmission Dynamics and Control. Lect. Notes in Biomath. 56, 1-105.
- Hilborn, R.,** Walters, C.J. and Ludwig, D. (1995) Sustainable Exploitation of Renewable Resources. Annual

Review Of Ecology And Systematics 26, 45-67.

Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M. and Markowitz, M. (1995). Rapid Turnover of Plasma Virions and CD4 Lymphocytes in HIV-1 Infection. *Nature* 373, 123-126.

Honig, B. and Nicholls, A. (1995). Classical Electrostatics in Biology and Chemistry. *Science* 268, 1144-1149.

Humphreys, D. D., Freisner, R. A. and Berne, B. J. (1994). A Multiple Time-Step Molecular Dynamics Algorithm for Macromolecules. *J. Phys. Chem.* 98, 6884-6892.

Jaeger, L., Michel, F., and Westhof, E. (1994). Involvement of a GRNA Tetraloop in Long-Range Tertiary Interactions. *J. Mol. Biol.* 236, 1271-1276.

Jafri, S. M., and Keizer, J. (1994) Diffusion of Inositol 1,4,5-Trisphosphate But Not Ca^{2+} Is Necessary for a Class of Inositol 1,4,5-Trisphosphate-Induced Ca^{2+} Waves. *Proc. Natl. Acad. Sci.* 91, 9485-9489.

Johnson, B. A. and Blevins, R. A. (1994). NMRView: A Computer Program for the Visualization and Analysis of NMR Data. *J. Biomolec. NMR* 4, 603-614.

Kearsley, S.K., Underwood, D.J., Sheridan, R.P. and Miller, M.D. (1994). Flexibases - A Way to Enhance the Use of Molecular Docking Methods. *J. Comp. Assist. Mol. Design* 8, 565-582.

Koch, C. and Segev, I., eds. (1989). *Methods in Neuronal Modeling: From Synapses to Networks*, MIT Press, Cambridge MA. 2nd Edition, in press 1996.

Kontoyianni, M. and Lybrand, T. P. (1993). Three Dimensional Models for Integral Membrane Proteins: Possibilities and Pitfalls. *Perspect. Drug Disc. Design* 1, 291-300.

Kopell, N. and LeMasson, G. (1994) Rhythmogenesis, Amplitude Modulation, and Multiplexing in a Cortical Architecture, *Proc. Natl. Acad. Sci. USA* 91, 10586-10590.

Kreusch, A. and Schulz, G. E. (1994) Refined Structure of the Porin from *Rhodopseudomonas blastica*. Comparison with the Porin from *Rhodobacter capsulatus*. *J. Mol. Biol.* 243, 891-905.

Lande, R. (1993). Risks of Population Extinction from Demographic and Environmental Stochasticity. *Am. Nat.* 142, 011-927.

Lande, R. (1994). Risks of Population Extinction from New Deleterious Mutations. *Evolution* 48, 1460-1469.

Lander, E.S. and Waterman, M.S. (1995). *Calculating the Secrets of Life*, National Academy Press, Washington, D.C.

Lauffenburger, D. A. and J. J. Linderman (1993). *Receptors: Models for Binding, Trafficking, and Signaling*. Oxford University Press, Oxford.

Leahy, D.J., Hendrickson, W.A., Aukhil, I. and Erickson, H.P. (1992). Structure of a Fibronectin Type III Domain from Tenascin Phased by MAD Analysis of the Selenomethionyl Protein. *Science* 158, 987-991.

Lee, C. and Subbiah, S. (1991). Prediction of Protein Side-Chain Conformation by Packing Optimization. *J. Mol. Biol.* 217, 373-388.

Levitt, M. (1993). Accurate Modeling of Protein Conformation by Automatic Segment Matching. *J. Mol. Biol.* 226, 507-533.

- Levin, S.A.** and Pacala, S.W. (1996). Theories of Simplification and Scaling of Spatially Distributed Processes. In press, 1997. In *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*. D. Tilman and P. Kareiva, eds, Princeton University Press, Princeton NJ.
- Lewis, M.A.**, Kareiva, P. (1993). Allee Dynamics and the Spread of Invading Organisms. *Theoretical Population Biology* 43, 141-158.
- Lyubchenko, Y.** Shlyakhtenko, L., Harrington, R., Oden, P. and Lindsay, S. (1993). Atomic Force Microscopy of Long DNA: Imaging in Air and Under Water. *Proc. Natl. Acad. Sci. USA* 90, 2137-2140.
- Misra, V.K.**, Hecht, J.L., Sharp, K.A., Friedman, R.A. and Honig, B. (1993). Salt Effects on Protein-DNA Interactions - The Lambda-CI Repressor and EcoRI Endonuclease. *J. Mol. Biol.* 238, 264-280.
- Mogilner, A.** and Oster, G. (1996). Cell Motility Driven by Actin Polymerization. *Biophys. J.*, in press.
- Murray, A.** and Hunt, T. (1993). *The Cell Cycle: An Introduction*. New York, W.H. Freeman.
- Murray, J.** and Oster, G. (1984). Cell Traction Models for Generating Pattern and Form in Morphogenesis. *J. Math. Biol* 19, 265-80.
- Naranjo, D.**, Latorre, R., Cherbavaz, D., McGill, P., and Schumaker, M. F. (1994). A Simple Model for Surface Charges on Ion Channels. *Biophysical J.* 66, 59-70.
- OTA (Office of Technology Assessment)**, (Sept, 1993). Harmful Non-Indigenous Species in the United States. OTA-F-565. US Govt. Printing Office, Washington D.C.
- Oliver, T.**, Dembo, M. and Jacobson, K. (1995). Traction Forces in Locomoting Cells. *Cell Motil. Cytoskel.* 31, 225-240.
- Olsen, L.**, Sherratt, J. and Maini, P. (1995). A Mechanochemical Model for Adult Dermal Wound Contraction and the Permanence of the Contracted Tissue Displacement Profile. *J. Theor. Biol.* 177(2), 113-128.
- Orengo, C. A.**, Swindell, M. B., Michie, A. D., Zvelebil, M. J., Driscoll, P. C., Waterfield, M. D. and Thornton, J. M. (1995). Structural Similarity between Pleckstrin Homology Domain and Verotoxin: The Problem of Measuring and Evaluating Structural Similarity. *Prot. Sci.* 4, 1977-1983.
- Peskin, C.** and Oster, G. (1995). Coordinated Hydrolysis Explains the Mechanical Behavior of Kinesin. *Biophys. J.* 68(4), 202s-210s.
- Rayment, I.** and H. Holden (1994). The Three-Dimensional Structure of a Molecular Motor. *TIBS* 19, 129-134.
- Ringe, D.** and Petsko, G.A. (1996). A User's Guide to Protein Crystallography. In *Protein Engineering and Design*, P.R. Carey, ed. Academic Press, San Diego.
- Roughgarden, J.** (1979). *Theory of Populations Genetics and Evolutionary Ecology: An Introduction*. Macmillan, New York.
- Rybenkov, V.V.**, Cozzarelli, N.R. and Vologodskii, A.V. (1993). Probability of DNA Knotting and the Effective Diameter of the DNA Double Helix, *Proc. Nat. Acad. Sci.* 90, 5307-5311.
- Schlick, T.** and Olson, W.K. (1992). Supercoiled DNA Energetics and Dynamics by Computer Simulation, *J. Mol. Biol.* 223, 1089-1119.
- Scholey, J.** (1994). Kinesin-Based Organelle Transport. In *Modern Cell Biology: Microtubules*. J. S. Hyams and

C. W. Lloyd, eds. New York, Wiley-Liss. 13: pp. 343-365.

Senderowitz, H., Guanieri, F. and Still, W. C. (1996), A Smart Monte Carlo Technique for Free Energy Simulations of Multiconformational Molecules, Direct Calculation of the Conformational Populations of Organic Molecules. J. Amer. Chem. Soc. 117, 8211-8219.

Shadlen, M. and Newsome, W. (1994), Noise, Neural Codes and Cortical Organization. Curr. Opin. Neurobiol. 4, 569-579.

Silver, R.B. Calcium, BOBs, Microdomains and a Cellular Decision: Control of Mitotic Cell Division in Sand Dollar Blastomeres, Cell (in press).

Simmons, A. H., Michal, C. A. and Jelinski, L. W. (1996). Molecular Orientation and Two-Component Crystalline Fraction of Spider Dragline Silk, Science 271, 84-87.

Smith, K. C. and Honig, B. (1994). Evaluation of the Conformational Free Energies of Loops in Proteins. Proteins: Structure, Function, and Genetics 18, 119-132.

Smith, S.B., Cui, Y. and Bustamante, C. (1996). Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules, Science 271, 795-799.

Softky, W.R. (1995). Simple Codes Versus Efficient Codes. (Commentary) Curr. Opin. Neurobiol. 5, 239-247.

Softky, W.R. and Koch, C. (1993). The Highly Irregular Firing of Cortical Cells is Inconsistent with Temporal Integration of Random EPSPs. J. Neuroscience 13, 334-350.

Stasiak, A., et al. (1996). Determination of DNA Helical Repeat and of the Structure of Supercoiled DNA by Cryo-Electron Microscopy. In Mathematical Approaches to Biomolecular Structure and Dynamics, IMA Proceedings 82, Springer Verlag, New York, p. 117.

Steinhoff, H. J., Mollaaghabada, R., Altenbach, C., Khorana, H. G. and Hubbell, W. L. (1994). Site-Directed Spin Labeling Studies of Structure and Dynamics in Bacteriorhodopsin. Biophys. Chem. 56, 89-94.

Stuart, G.J. and Sakmann, B. (1994). Active Propagation of Somatic Action Potentials into Neocortical Pyramidal Cell Dendrites. Nature 367, 69-72.

Sumners, D.W., Ernst, C., Spengler, S.J. and Cozzarelli, N.R. (1995). Analysis of the Mechanism of DNA Recombination Using Tangles, Quarterly Reviews of Biophysics 28, 253-313.

Svoboda, K. and S. Block (1994). Force and Velocity Measured for Single Kinesin Molecules. Cell 77, 773-84.

Thorne, J.S., Kishino, H. and Febenstein, J. (1992). Inching Toward Reality: An Improved Likelihood Model of Sequence Evolution. J. Mol. Evolution 34, 3-16.

Tilman, D. (1994) Competition and Biodiversity in Spatially Structured Habitats. Ecology 75, 2-16.

Tirrell, J. G., Fournier, M. J., Mason, T. L. and Tirrell, D. A. (1994). Biomolecular Materials. Chem. Eng. News, December 19, 40-51.

Tranquillo, R. T., and Alt, W. (1996). Stochastic Model of Receptor-Mediated Cytomechanics and Dynamic Morphology of Leukocytes. J. Math. Biol. 34, 361-412.

Tranquillo, R. and J. D. Murray (1993). Mechanistic Model of Wound Contraction. J. Surg. Res 55, 233-47.

- Tuljapurkar, S.** and Wiener, P. (1994). Migration in Variable: Exploring Life History Evolution Using Structured Population Models. *J. Theor. Biol.* 166 75-90.
- Tuljapurkar, S.** (1994). Stochastic Demography and Life Histories. In *Frontiers in Mathematical Biology*, S.A. Levin, ed. Springer-Verlag, Berlin, pp. 254-262.
- Tyson, J. J.,** Novak, B., Odell, G. M., Chen, K., and Thron, C. D. (1996). Chemical Kinetic Theory: Understanding Cell-Cycle Regulation. *Trends in Biochemical Sciences* 21, 89-96.
- Walters, C.** and Maguire, J.J. (1996). Lessons For Stock Assessment from the Northern Cod Collapse. *Reviews In Fish Biology And Fisheries* 6, 125-137.
- Walters, C.** and Parma, R.M. (1996). Fixed Exploitation Rate Strategies for Coping with Effects of Climate Change. *Canadian Journal Of Fisheries And Aquatic Sciences* 53, 148-158.
- Williams, N.** (1996). Yeast Genome Sequence Ferments New Research, *Science* 272, 481-481.
- White, J.H.,** (1992). Geometry and Topology. In *Proceedings of Symposia in Applied Mathematics* 45, American Mathematical Society, Providence, R.I., 17.
- Whittington, M.A.,** Traub, R.D., and Jefferys, J.G.R. (1995). Synchronized Oscillations in Interneuron Networks Driven by Metabotropic Glutamate Receptor Activation. *Nature* 373, 612-615.
- Wofsy, C.,** Kent, U. K., Mao, S-Y., Metzger, H., and Goldstein, B. (1995). Kinetics of Tyrosine Phosphorylation When IgE Dimers Bind to Fce Receptors on Rat Basophilic Leukemia Cells. *J. Biol. Chem.* 270, 20264-20272.
- York, D.M.,** Yang, W.T., Lee, H., Darden, T. and Pedersen, L.G. (1995). Toward the Accurate Modeling of DNA - The Importance of Long-Range Electrostatics. *J. Amer. Chem. Soc.* 117, 5001-5002.
- Zadoks, J.C.** and Van Den Bosch, F. (1994). On The Spread Of Plant Disease - A Theory On Foci. *Annual Review Of Phytopathology* 32, 503-521.