Chapter 1

The Impact of Biology on Mathematics

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

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

The application of mathematics to biology is not new; neither is evidence of impacts on mathematics. Robert Brown, a botanist, discovered what is now called Brownian motion while watching pollen grains in water. Today, the mathematical description of such motion is central to probability theory.

Similarly, catastrophe theory is a branch of mathematics stimulated to large extent by biological theory. Inspired by Waddington's concept of an epigenetic landscape (Waddington 1957), Rene Thom generated interest in singularity theory and the bifurcations of dynamical systems (Thom 1975). Although the style of modeling used by the proponents of catastrophe theory was severely criticized, the beautiful mathematics it spawned has applications that extend far beyond those originally envisaged as part of catastrophe theory. And perhaps most importantly, the origins of the field of statistics were intimately tied up with biology.

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

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

1.1 Accomplishments of the Past

1.1.1 Statistics and Stochastic Processes

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

What were the major influences in this development? Porter (1986) introduces his history of statistics in the 19th century as follows. "This book.... is a study of the mathematical expression of what Ernst Mayr calls 'population thinking' (Porter 1986, p.Ê6; see also Mayr 1982, 1988, pp. 350-352), and "the development of statistical thinking was a truly interdisciplinary phenomenon for which mathematics had no priority of position; new ideas and approaches arose as a result of the application of techniques borrowed from one or more disciplines to the very different subject matter of another" (Porter 1986, p. 8). Porter later states "that the modern field of mathematical statistics developed out of biometry is not wholly fortuitous. The quantitative study of biological inheritance and evolution provided an outstanding context for statistical thinking, and quantitative genetics remains the best example of an area of science whose very theory is built out of the concepts of statistics. The great stimulus for modern statistics came from Galton's invention of the method of correlation, which, significantly, he first conceived not as an abstract technique of numerical analysis, but as a statistical law of heredity." (Porter 1986, p. 270). The profound problems raised by Darwin's insight have led to new fields of mathematical science. Only the surface has been scratched by these developments, and major challenges remain.

Darwin and Galton were cousins, and Darwin's ideas had a great influence on Galton (Porter 1986, p. 133 and p. 281). Likewise, problems in eugenics and plant breeding were the motivation for R. A. Fisher's statistical work (Box 1978, Fisher 1930). The analysis of variance and the theory of experimental design were developed to interpret and plan plant breeding experiments at the Experimental Station at Rothamsted, an institution that continues to be a major influence on statistical theory and practice. The benefits to mankind of these and later biometrical developments have been enormous. The "Green revolution" in agriculture would have been quite impossible without these tools. Modern medicine and public health practice depend upon carefully designed and interpreted clinical trials, and sophisticated studies of massive observational data sets.

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

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

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

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

These problems would be hopelessly intractable were it not for recent and likely advances in computational statistics. With computing power now available, we quickly can narrow our search for promising algorithms and test their effectiveness.

Other challenges involving DNA sequence data include searches of two of more pieces of data for (longest) matching subsequences. For these, new distributional results are required.

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

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

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

For the corresponding problems of statistical inference, new methods and corresponding algorithms are needed for model validation and the estimation of parameters. It can happen that models appear to fit according to currently used criteria even though they have not caught the essence of the biological phenomena of interest. A relevant question to ask is, how far off can the model be and still 'fit'? In other words, subject to fitting the data, the model should be biologically interpretable. In this area of research, collaborations between neurobiologists and statisticians have been particularly successful, as evidenced by, for example, joint work on spike train pattern recognition (Brillinger and Segundo 1979), estimation of single channel kinetic parameters (Milne et al. 1989), temporal clustering of channels (Ball and Samson 1987), estimation of open dwell time in multi-channel experiments (Yang and Swenberg, in press), and identification of kinetic states (Fredkin and Rice 1986).

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

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

1.1.2 Dynamical Systems Theory

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

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

Substantive mathematics has grown from these beginnings. Among other developments, Lanford extended Feigenbaum's arguments with numerical analysis to give a beautiful example of a rigorous "computer" proof. The study of interval maps was generalized to encompass maps of the circle. This work on circle maps has been used by Glass, Winfree and others for describing the phase responses of biological oscillators, particularly in cardiology. The work on maps of the interval has also been the starting point for the work of Carleson and Benedicks on the Henon map, a two-dimensional map that is a prototype for chaotic behavior.

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

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

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

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

1.1.3 Nonlinear Partial Differential and Functional Equations

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

Demographic methods have been applied to the study of human and nonhuman populations for centuries. These methods, which form the basis both for population projections and for understanding population consequences of life history phenomena, have had a strong impact in mathematical theory. A snapshot of the impact of demography is provided by the history of ergodic theorems. The renewal equation, a convolution integral equation that provided the first dynamical model for an age-dependent population, has roots in the work of Euler, Bortkiewicz, and Lotka (see Samuelson 1976). Sharpe and Lotka (1911) argued that most solutions to their renewal equation could be represented in a Fourier type expansion. Their argument was not accepted mathematically until Feller (1941) gave a rigorous proof for asymptotic behavior under appropriate conditions. As yet, the problem of stating conditions under which the renewal equation admits a Fourier type expansion remains partly open (see Inaba 1988).

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

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

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

Although the equations and mathematical knowledge arising from demography and epidemiology have already found applications (e.g., in evolutionary ecology, conservation biology and epidemiology), a strong need exists for new mathematics to address new pressing biologically motivated questions. For example, at the interface of social dynamics and epidemiology, new models describe "social mixing" (e.g., formation and dissolution of pairs) and its role in disease dynamics. The models are novel systems of hyperbolic partial differential equations. These models may affect practical issues of public health and broader biological issues. Since current techniques are as yet in their infancy, it is likely that new mathematics will develop from these efforts.

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

1.1.4 Classical Analysis

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

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

Vardi, Shepp, and Kaufman (1985) are responsible for a fundamental advance in positron emission tomography (PET). With emission tomography in general, a substance such as a sugar that is differentially metabolized by different tissues is tagged with an emitting molecule. In one case (PET) a positron is emitted, and in another (SPECT), a photon; with PET, each positron gives rise to two photons that move in opposite directions. In either case individual photons are counted as they hit a detector surrounding the object (for example, a human head) being imaged. The object can be modeled as a spatially inhomogeneous Poisson process, and the mathematical task is to reconstruct the intensity function from the counts. The approach of Vardi et al. was to employ an algorithm, the EM algorithm, that was developed by Harvard statisticians Dempster, Laird, and Rubin (1977); earlier basic work on EM-like algorithms was done by the mathematician Baum (1970) and others (see discussion of the paper by Vardi et al. (1985) for extensive references). A Bayesian approach to reconstruction in emission tomography utilizes Markov random fields that arise in statistical mechanics. Important contributions have been made by Geman and McClure (1985, 1987). Recently, Johnstone and Silverman (1990) have given minimax (in a statistical sense) rates of convergence for PET algorithms. The interface of emission tomography, mathematics, and statistics continues to be a particularly active area of research. It should be noted that PET permits quantitative measurements, in vivo, of local hemodynamics, metabolism, biochemistry, and pharmacokinetics (Fox et al. 1985), and that SPECT is best used for problems of perfusion rather than metabolism.

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

1.1.5 Topology and Geometry

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

1.2 Grand Challenges

Two of the most influential books in the development of biological thought should be Plato's Republic and Darwin's Origin of Species. Plato claimed that all the variation in observable horses, for example, is a mere shadow of an idealized abstract form of pure "horseness," not available to the senses. Plato's notion of idealized forms was the basis of scientific developments for two millennia. For example, Newton's concepts of absolute space and time are idealizations on the model of Plato's horseness. In biology, the Linnaean concept of species is an operational version of Plato's idea: a Linnaean species is determined by a "type specimen," deposited in a museum somewhere, and all deviations between the type specimen and real members of the species are mere irrelevant accidents. For Gauss, the variation in repeated astronomical measurements led to a theory of "error" in which variation was something to be eliminated. The influence of the Platonic theory of ideal types extended far beyond science to, for example, popular notions of national or racial "types."

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

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

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

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

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

.