Chapter 1

Introduction to Mathematical Biology

1.1 Grand Challenges in Biology

Biology is a vast field, ranging from the study of microscopic sub-cellular systems and individual molecules to the the study of ecosystems that span vast portions of the globe. Thus, to list all of the unsolved problems in biology would take several volumes at the very minimum. However, it is useful to highlight some of the "grand challenge" problems in biology today. That is, what are some of the overarching goals of a significant portion of modern biological research — problems, that if solved completely, would dramatically change the field.

1.1.1 Understanding the process of development

We begin life as a single fertilized egg cell; by the time we are an adult, each of our bodies has roughly between ten and a hundred trillion (10^{13}–10^{14}) cells, all of which are directly descended from that one egg through cell division. However, the most interesting aspect of this is not that one cell became trillions; one can see a similar process happen looking at the growth of bacterial cultures, or watching yeast grow while making bread or beer. What is most amazing is that these ten trillion cells are specialized — they have particular roles, and they are confined to particular parts of the body. In the earliest stages of development (immediately after fertilization), the egg does seem to simply just divide, and early embryos are spherical clusters of apparently identical cells. But at some point, this changes — the cells start to differentiate into skin, muscle, bone, nerves, and the other cell types of the body (of which there are more than two hundred). Additionally, the embryo begins to take on a shape, which ultimately leads to the body having a front and a back, a top and a bottom, and a left and a right, all of which are distinct; all our various tissues end up located where they need to be to form the human body.

How does this happen? What makes one cell become a nerve cell and another to become skin? What determines the top and bottom of the early embryo? Understanding this process is a key component of one of the fundamental questions that people have been asking for the entirety of history — "What makes us what we are?" Additionally, understanding the process of development would allow for the treatment and prevention of the many disorders that have their basis in improper development. On-going research has begun to answer some

of these questions — cells communicate with each other through the molecules they produce, and a cell's behavior will change in response to these different molecular cues. However, while many of the key components are known, exactly how they all work together to ensure faithful development is far from clear. Notable success has been made, however, in mathematical models that explain some of the earliest steps in the development of model organisms (such as fruit flies).

1.1.2 The biology of thought and behavior

How does the mind work? What is memory, and what determines what we remember and what we forget? What is emotion — and what underlies our feelings of happiness, sadness and love? These are among the most fundamental questions that have challenged thinkers across the ages.

We know that much of the answer must lie in the interactions of the cells of the nervous system; a human brain has roughly a hundred billion neurons, and these are connected in intricate patterns that allow them to communicate with one another. These cells are able to send electrical signals down their length, and can signal neighboring neurons by the release of particular chemicals (neurotransmitters); many drugs that affect the nervous system interfere with this communication process. What we simply don't understand is exactly how these connections are formed (both during development and in the process of learning), or how observed neural activity translates into thought and emotion. A complete understanding of how the nervous system works would have profound impacts on medicine, improving our ability to treat neurological disorders as well as the abuse of psychologically active drugs.

1.1.3 Quantitative and predictive ecological models

Human activity has had dramatic impacts on the planet, and most significantly on the other organisms that we share the globe with. Some of these have been direct, and can easily be understood; the populations of many species have been decimated by over-hunting, over-fishing, or destruction of their habitat for housing, roads and farms. However, there can also be very significant indirect effects. For example, the loss of a single species (due to direct effects), may have dramatic impact on populations that prey on (or are preyed upon by) that species. Human-influenced growth of a population — such as algal blooms spurred by fertilizer run-off or the introduction of non-native species — can additionally have huge impacts on other species. In addition to these effects, climate change is likely to have large effects on populations of numerous species across the globe.

If we were able to truly understand these ecological changes — to the extent that the effect of a given action could be predicted with reasonable accuracy — then we may be able to develop sustainable practices that balance the needs of humanity with our desire to minimally impact ecosystems. Such a predictive capability would require a detailed understanding of the relationships of different species, as well as how each species interacts with and is affected by its environment. For predictions to be quantitative (dictating the magnitude of a change, not just its direction), these relationships would need to be formulated in mathematical models. With accurate models, simulations of possible activities could be run to determine the likely effects. Such models are an active area of research, although the vast complexity

of the systems (especially when effects such as climate are factored in), make the goal of fully predictive models still far from reach.

1.1.4 Predictive personalized medicine

Modern medicine has been tremendously successful in treating many severe illnesses and diseases. However, it is also becoming more and more clear that further progress will be limited without a focus on individual differences. Many medications can have severe side effects, but these do not affect all individuals. If the occurrence of these effects could be predicted before hand, then the medication could be restricted to those for who is it the most save. Similarly, for complex illnesses like cancer, not all medications work equally well on all patients. Again, if the success of a medication could be predicted, then those patients that would not benefit from the treatment could be spared unnecessary risk or discomfort (from side effects) and expense. Currently, some of this is being done through the use of simple "markers" to classify patients into one of a few categories.

However, a more profound shift in how new therapies are developed is gaining recognition — although we are still far from implementing the new approach in most cases. This new approach includes the use of detailed computational models of many aspects of drug action: how does the drug affect the behavior of its desired target; how does the drug move through the body; how is the drug degraded; and so on. These models may then be used to develop treatments which have all the desired properties. This is a much more holistic approach than is currently used, where drugs tend to be developed first based on one key activity, and then are tested for other desirable properties. More importantly, the new approach would allow specifically for the consideration of side-effects and cross-drug interactions as part of the therapeutic design. Key to the success of this paradigm of medicine will be accurate and predictive models of disease states, of the normal functioning of the body, and the interplay of drug molecules with both. Individual variation, both genetic and metabolic, could be incorporated into models that generate personalized treatment programs for each patient.

1.1.5 Synthetic biology

The past decade has seen the rise of a completely new discipline — that of synthetic biology. This field aims to take genetic and biomolecular engineering to the next level, by truly creating novel biological systems. For example, in an experiment that that has already been performed, a new genetic circuit was inserted into bacteria, creating bacteria that fluoresce in an oscillatory pattern. Experimental techniques that facilitate this work include traditional molecular biology, but novel methods have additionally demonstrated the de novo synthesis of an active virus from non-living components, and the substitution of the fully synthetic genome of one bacterium into the cell of another. If synthetic biology is to progress to the point where novel biological systems can truly be engineered, biology will need quantitative models of how these systems work. Just as all other engineering disciplines (mechanical, electrical, and so on) are grounded in a firm mathematical basis, so does synthetic biology aim to place that same rigor into biological engineering.

1.1.6 Summary

These are just a sample of some of the major goals of biological research today — and cover a broad range of systems, from bacteria to ecosystems. However, all these "grand challenges" share an important feature — they consist of highly complex systems that simply can not be fully understood without the aid of quantitative mathematical models. Thus, the field of Mathematical Biology — also referred to as Computational, Quantitative or Systems Biology — is at the beginning of a new era of particular significance.

1.2 Mathematical Biology Past and Present

1.2.1 The Early Days

The application of mathematics to problems in biology has a long history, with examples dating to as early as the 12th century. The view of Mathematical Biology as a field can be dated to the second quarter of the 20th century, during which time several seminal works were produced. Among these were the models of Alfred Lotka and Vito Volterra who, in 1925 and 1926 respectively, independently developed a model to consider the following question — what are the dynamics of the populations of two interactions species, one a predator and the second its prey?[1] The model was very simple in its design:

Prey growth rate: Prey have a constant *per capita* growth rate, and a death rate that is proportional to the number of predators. One way to think of this is that the growth of prey is dependent on an unlimited resource (grass, for example), and that the only way prey die is to be eaten by predators.

Predator growth rate per capita: Predators have a *per capita* growth rate that is proportional to the number of prey, and a constant death rate. That is, predators reproduce depending on how much they eat (and will consume more if more prey is available), but die at a constant rate.

Since the rate of change of a variable over time is the derivative of that variable, this becomes the pair of equations:

$$\frac{dN_{prey}}{dt} = N_{prey}(A - BN_{predator}) \tag{1.1}$$

$$\frac{dN_{predator}}{dt} = N_{predator}(CN_{prey} - D) \tag{1.2}$$

Where A is the intrinsic prey growth rate, B is the rate at which an individual predator kills prey, C is the prey-dependent prey growth rate, and D is the intrinsic predator death rate. Note that both terms are the product of a *per capita* rate (in parentheses) and the current population.

[1] Lotka's work can be found in the book: Lotka, A.J., *Elements of Physical Biology*, Williams and Wilkins, (1925). Volterra's origingal work was in Italian (Volterra, V., "Variazioni e fluttuazioni del numero dindividui in specie animali conviventi", *Mem. Acad. Lincei Roma*, **2**, 31-113, (1926)), but can be found in English in: Volterra, V., "Variations and fluctuations of the number of individuals in animal species living together" in Animal Ecology, Chapman, R.N. (ed), McGrawHill, (1931).

Around the same time, Anderson McKendrick and William Kermack published a paper looking at how the number of infected people varies during an infectious disease epidemic.[2] Again, the model was quite simple — people were broken down into three groups, susceptible (S), infected (I), and recovered (R). Susceptible individuals could become infected, and infected individuals could recover, but recovered individuals would not become infected again. As the model was meant to deal with short term epidemics, there was no explicit treatment of mortality. However, it is worthwhile noting that as far as the dynamics of the model are concerned, there is no difference between a recovered individual and an individual who died from infection (neither can become reinfected). Again, the model was focused on rates of change:

Per capita infection rate: Susceptible people become infected at a rate dependent on how many infected people there are.

Per capita Recovery rate: Infected people recover at a constant rate.

This leads to the following expressions:

$$\frac{dS}{dt} = -\beta SI \qquad \frac{dI}{dt} = +\beta SI - \gamma I \qquad \frac{dR}{dt} = +\gamma I \tag{1.3}$$

There are two fundamental rates: infection is a constant (β) multiplied by the susceptible and infected populations; and recovery is a constant (γ) multiplied by the infected population. The susceptible population goes down with infection, and the recovered population increases with recovery; the infected population both grows with infection and decreases with recovery.

Both these models are very similar, containing only a few (2-3) variables with simple terms (linear in a single variable, or a product of two variables). Additionally, both describe rates of change of a variable as a function of the other variables. This type of model — a differential equation model — will be discussed more later.

In 1939, the first academic journal dedicated to the field of mathematical biology was published, "The Bulletin of Mathematical Biology". This journal published roughly twenty to thirty articles a years (in quarterly installments) on a range of topics in mathematical biology.

1.2.2 The First Burst of Growth

In the 1960's and 70's, there was a great deal of growth in the field, most readily evidenced by the emergence of several new journals:

- Journal of Theoretical Biology (1966)
- Mathematical Biosciences (1967)
- Journal of Theoretical Population Biology (1970)
- Journal of Mathematical Biology (1974)

[2] Kermack, W. O. and McKendrick, A. G. "A Contribution to the Mathematical Theory of Epidemics." *Proc. Roy. Soc. Lond. A* **115**, 700-721, 1927.

What changed to spur this large growth in interest in mathematical modeling of biological systems? While certainly there were many factors, including the growth of new fields of experimental biology (such as molecular and structural biology), one of the major reasons for this growth was the development of **computers**.

Early models had to be solved by hand, which greatly limited the complexity of the models that could be considered. While early mathematical biologists were very skilled mathematicians, certain types of problem can only be solved numerically, which is a very tedious procedure without the aid of a digital computer. The availability of computers allowed a much more detailed examination of how earlier models performed, particularly in comparison to known experimental data. Some new models were developed, but relatively few significant impacts on (experimental) biology were made. This time period also represents the dawn of *molecular* modeling — the use of computers to simulate mathematical models of biological molecules such as proteins.

Why was there not much impact at this point in time? For one thing, while computers of the day allowed much more complex systems to be studied than had previously been feasible, they were still very limited. Thus, relatively simple models were still the norm. Additionally, complex models have many *parameters* (such as A, B, C, and D in the Lotka–Volterra model, and β and γ in the McKendrick–Kermack model). In order to make meaningful comparisons to experiment, these values must be determined, typically by repeating the calculation many times with different parameters. This makes modeling even more expensive, but more importantly, it requires *quantitative* data, of which little was available at the time. Ecological modeling (one area for which a reasonable amount of quantitative data was available) made significant strides during this time.

One notable contribution from this era (or slightly before) is the Hodgkins–Huxley model for transmission of action potentials through neurons, which was published in 1952. These potentials were easily studied using squid giant axons, and the readout (electric potential) was inherently quantitative. Thus, the experimental backing for the development of an accurate model was firmly in place.

Overall, however, the field of mathematical biology throughout this time remained detached from mainstream biology; a great deal of mathematical biology was being done by mathematicians, with little involvement of biologists.

1.2.3 The Modern Era

Now, in the past decade or so, there has been a resurgence of interest in the application of mathematics to biological problems, as can again be evidenced by the rise of several new journals:

- Journal of Computational Biology (1994)
- Public Library of Science — Computational Biology (2005)
- BioMed Central — Systems Biology (2007)

It is worthwhile to note the change of naming to "Computational Biology" and "Systems Biology." In many ways, this has been a conscious decision to avoid the overly "mathematical"

associations of previous work; however, the methods applied are very similar, and many of the so-called "new" approaches to systems biology are heavily founded in the seminal work of early mathematical biologists.

Again, what has changed recently? Certainly, computers are much more powerful now than they were thirty years ago; larger and more detailed models can be considered, and parameterization can be done efficiently. Perhaps more importantly, though, there is a vastly greater amount of data available, much of it quantitative. The past several decades have seen tremendous developments in technologies to probe all sorts of biology systems, with obvious impact.

Most significantly — although whether this is an underlying cause or an effect may be debated — there has been a significant change in the culture surrounding the "new" mathematical biology. Those in the field are motivated primarily by a desire to understand the **biology**, rather than being primarily motivated by the mathematics that biological problems may inspire. As a result, the newest generation of mathematical biologists have shown a great respect for the importance of experiment, and in understanding how experimental biology is done. Reciprocally, experimental biologists have recognized the impact that the bioinformatics revolution of the 1990s (which was largely an application of computer science, rather than mathematics) had on the field, and thus are more receptive to the idea that modeling may be useful and important.

As a result, Mathematical Biology (or Computational/Systems Biology) is now a part of mainstream biological research, as is becoming more so. To quote Uri Alon (one of the leaders of the emerging field of Systems Biology), "In ten years, Systems Biology is not going to be a part of biology — all biology will be Systems Biology."[3] What does he mean by this? Quite simply, biology is becoming a quantitative science, rather than being primarily observational. What defines an "acceptable" model is changing — in the past it was a qualitative, intuitive model described best by the cartoons that are seen throughout many biology texts; we are in a transition towards requiring a quantitative models as a mathematical layer added onto those cartoons, models that allow for detailed predictions to be made.

1.2.4 Summary

Mathematical Biology has a long history, and it is important to recognize the contributions of early researchers in the field. The field has undergone several phases of growth and development, spurred partially by the growth of computers that can be harnessed in modeling, and — especially in the past decade — by quantitative experimental techniques that are producing data at an unparalleled rate. The mathematical modeling of biological systems is becoming a key component of biological research, and will become even more important in the coming years.

[3] Laufer Center Seminar, Stony Brook University, Oct. 7th, 2008.