Nothing in Biology makes Sense without the Flavour of Mathematics

H A Ranganath

H A Ranganath is at the Department of Studies in Zoology, University of Mysore, Manasagangotri, Mysore.

The elements of mathematical language are on the whole precise and unambigous. In the history of science, one can see that the development of any branch of science begins from the largely descriptive, goes on to become explanatory at the qualitative level, and finally becomes explanatory at the quantitative level. With the application of mathematical principles, physics made the transition from a qualitative to a quantitative science about 3 to 4 centuries ago; biology is making the transition now. Mathematical biology is today a fast growing, well recognized, albeit not clearly defined subject. To my mind, it is the most exciting modern application of mathematics.

The partnership between mathematics and biology began in the late 1800s, primarily through studies of human inheritance. Galton and Pearson in England developed techniques like correlation and regression to study the heritability of various traits in humans. Since then it has increasingly been realised that most biological phenomena are complex, usually non-linear, and often stochastic. Consequently, one requires a point of view rooted in mathematics, probability and statistics to fully understand and to appreciate most biological processes.

Mathematical theorists have made significant contributions to our understanding of many facets of biological systems at various levels of biological organization, and in many disciplines ranging from the structural biology to evolutionary genetics. For example, the Nobel Prize winning work of applied mathematicians like Karle and Hauptman helped construct algorithms used to reveal the structure of complex biological molecules from X-ray diffraction data. Similarly, the investigations of Cormack and Hounsfield yielded algorithms with which structures in living organisms can be determined by tomogra-

Mathematical modelling in biology has helped us towards a better understanding of phenomena as diverse as cardiac function and blood flow, muscle contraction, transmission of nerve impulses, enzyme kinetics, biological clocks, the growth and dynamics of populations, and the exquisite process of adaptive evolution.

phy. Mathematical modelling in biology has helped us towards a better understanding of phenomena as diverse as cardiac function and blood flow, muscle contraction, transmission of nerve impulses, enzyme kinetics, biological clocks, the growth and dynamics of populations, and the exquisite process of adaptive evolution.

Mathematics has had a great impact in the area of structural biology. Three disciplines namely biology, mathematics and physics, have come together to resolve the structure and functional relationship of biologically important macromolecules. Of the many methods used in structural biology, X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR) are especially important. In all such techniques, computational methods lie at the heart of their use, and new mathematical methods of analysis as well as innovative means of visualization (graphics) are continually being developed. Computer graphics can be used to visualise data and the dynamical behaviour of mathematical models. Many instruments in the biologist's arsenal (confocal scanning laser microscope, gene sequencers) gather data into a computer based graphical database. Modern computer graphics technology makes it possible to display such data pictorially and dynamically. Many physico-chemical properties of deoxyribonucleic acid (DNA), the principal hereditary material, can be explained using the mathematical concepts of differential geometry and topology (*Box* 1). The geometric concepts of tilt, roll, shear and propellar twist have been used to describe the secondary structure of DNA. These topological concepts have also helped understand the interaction of DNA with various drugs and other molecules.

The movement of cells, and their responses to changes in the environment, are very common phenomena. Mathematical modelling has helped us to learn how a cell moves, how it counts molecules in its environment, and how it uses this information. Modelling approaches, in tandem with rigorous experimental testing of theoretical predictions, have greatly enhanced our understanding of intracellular dynamics. *In vitro* experiments in

Box 1. Structure of Deoxyribonucleic acid
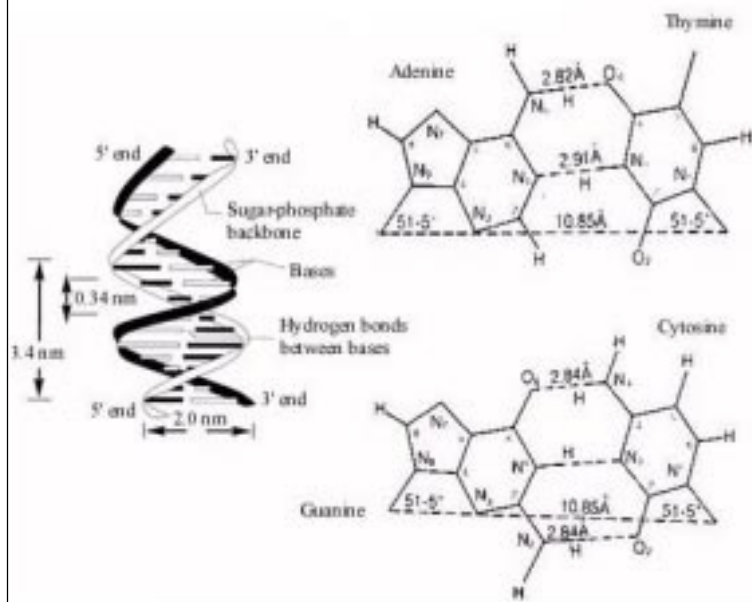
which motor molecules like myosin, dynein and kinesin, are linked to glass or plastic surfaces, along with actin and myofibrin along which these molecules move, have been key in refining the understanding of the movement of structures and organelles within cells, a topic first studied in detail theoretically. In fact, the applications of mathematics to cellular and molecular biology are so pervasive that they often go unnoticed.

The twentieth century is widely regarded to be the century of genetics, the science dealing with patterns of inheritance of characters from parents to offspring. Gregor Mendel, an Austrian monk, is considered as the father of genetics. After he was ordained a priest in 1847, he spent a few years at the University of Vienna for training in physics, mathematics and natural sciences. This training provided him with many technical and mathematical skills that were of value in performing his later experiments. The basic tenets of Mendelian principles revolve around probability, the mathematical laws of combinations. In fact, the uniqueness of Mendel's approach to the problem of inheritance at that time lay in his use of probabilistic thinking,

The applications of mathematics to cellular and molecular biology are so pervasive that they often go unnoticed.

Today we are in the era of genomics, and now possess voluminous data on the nucleotide sequence of the genomes of humans, insect (*Drosophila*) and worm (*Caenorhabditis*), as well as many bacteria and viruses.

and the quantitative analysis of data from hybridization experiments, as opposed to mere qualitative description. After the rediscovery of Mendel's laws in 1900, the field of genetics has grown and evolved in many different directions, but continues to be enriched by probabilistic mathematical models.

Today we are in the era of genomics, and now possess voluminous data on the nucleotide sequence of the genomes of humans, insect (*Drosophila*) and worm (*Caenorhabditis*), as well as many bacteria and viruses. The amount of information generated by these sequencing endeavours is staggering. For instance, the human genome has about 3 billion nucleotides and perhaps 50,000-70,000 structural genes. The nucleotide sequence in a gene (DNA) determines the amino acid sequence in a protein. The GeneBank database has information about different DNA sequences and, similarly, protein identification resources have information about different amino acid sequences in a protein. Whenever a new DNA sequence is identified, in order to identify the region in the GeneBank data, computer programmes (like FASTA, FASTN, BLAST, etc.) have to be used. If not, it will be an incredibly time-consuming process. A major challenge for the next few decades is to utilize the information generated from genomics to integrate the structure and behaviour of complex genetic regulatory systems. This job will need new mathematical and computational tools such as algorithms in the mathematical theory of directed graphs, new uses of random directed graphs theory, stochastic processes and population genetic models. The coalescing of genetics, mathematics, information science, and technology, has resulted in the emergence of a new field of study called bioinformatics. Some of the areas which come under the purview of this are biological data warehousing, data capture, data analysis, nucleic acid data banks, analysis tools for sequence data banks, structural data banks, microbial and cellular data banks, etc.

The coalescing of genetics, mathematics, information science, and technology, has resulted in the emergence of a new field of study called bioinformatics.

Ecology, today an increasingly important area of science, has been greatly transformed by the application of mathematical reasoning from a traditional and descriptive subject into a

highly developed description of causality and effect in the complex world of nature. The partnership between mathematics and ecology began in the 1920s with the Lotka–Voterra model to interpret predator-prey relationships. Prolonged interaction between mathematics and ecology has resulted in the emergence of models with predictive potential, which are of use in population biology. It has influenced the development and the utility of dynamical systems theory, differential equations and integro-differential equations. Mathematical modelling and explanation have helped ecologists to interpret and understand many important issues involved in the growth of populations such as demography, population projection and population modelling, organizing principles for collecting and analyzing data, population consequences of life history phenomena, predator-prey and

---

Box 2. "I am reluctant to intrude…"

"I am reluctant to intrude in a discussion of matters concerning which I have no expert knowledge, and I should have expected the very simple point which I wish to make to have been familiar to biologists": this is the sentence with which the mathematician G H Hardy began his brief two-page paper, published in the journal *Science* in 1908. Mendel's laws had only recently been rediscovered, and the nascent science of genetics was mired in several controversies. Punnett, a well-known geneticist and colleague of Hardy's, discussed one of these controversies with him one night after dinner. In a cross between, say, a red flowered plant and a white flowered plant, typically all progeny would be of one or the other parental type. For example, all progeny could be red flowered in which case red flower colour in that species would be considered *dominant* over white flower colour. It was typically observed that if the progeny were then crossed among themselves, the next generation would consist of dominant (red) and recessive (white) individuals in the ratio of 3:1. Several geneticists were at the time arguing that, following Mendel's laws, the frequency of occurrence of dominant traits in populations should be 75%. Yet, this was simply not so and this observation was being used by some to undermine Mendelian genetics.

Hardy, a brilliant mathematician, applied his logical mind to the problem, asked Punnett a few questions about inheritance, and then proved that in a large population, with mating being random with regard to the trait under question, and with no differences in fertility among the different genotypes, "there is not the slightest foundation for the idea that a dominant character should show the tendency to spread over a whole population, or that a recessive should tend to die out". What Hardy saw was that, if the frequencies of dominant homozygotes, heterozygotes, and recessive homozygotes were represented by P, $2Q$ and R, respectively, then the frequencies of the two alleles of that gene could easily be represented as p (frequency of dominant allele) = $P+Q$, q (frequency of recessive allele) = $R+Q$. Conceptually, the process of putting together alleles to form the diploid genotypes of the next generation, could then be thought of as pulling

alleles at random from a large pool of dominant and recessive alleles in frequencies p and q, and putting them together in pairs. The probability of pulling two dominant alleles would simply be $p \times p = p^2$, that of two recessives $q \times q = q^2$, and that of one recessive and one dominant $2 \times p \times q$, because the probability of pulling one dominant and one recessive is $p \times q$, but there are two ways one can end up getting the dominant and recessive pair (either a dominant first, and then the recessive, or the other way around). The frequencies of the three genotypes in the progeny, thus, were nothing but the expansion of $(p+q)^2$, yielding frequencies of dominant homozygotes, heterozygotes and recessive homozygotes of p^2, $2pq$ and q^2, respectively. Thereafter, these frequencies would remain the same generation after generation: this is the Hardy–Weinberg law, independently put forward by Hardy and a German physician, Wilhelm Weinberg.

Though this result is mathematically trivial, it marks the beginning of population genetics and essentially lays down the inertial state in the inheritance of traits at the population level. It tells us that there is nothing in the fundamental cycle of meiosis and fusion itself that tends to change the genetic makeup of populations over generations. In other words, Mendelian inheritance, at the populational level, is fundamentally conservative: populations maintain their genetic composition unless something else – mutation, natural selection, migration, non-random mating – is going on. The way in which these various evolutionary factors may act upon populations can be studied mathematically using the approach of Hardy, and indeed it was this approach that was used by Fisher, Wright and others to develop the entire theoretical foundations of population genetics. The Hardy–Weinberg principle, thus, provided the missing link required to weld Darwin's theory of natural selection with Mendel's laws of inheritance, leading eventually to the Neo-Darwinian Synthesis.

This story of the Hardy–Weinberg principle serves to illustrate how the application of even very simple mathematical reasoning can lead to important insights into major biological issues. Hardy may have felt that he was making a 'very simple point', and in hindsight he was, but the fact remains that at the time that point was not obvious to many leading geneticists. There are also two ironical twists to this story. Hardy was a mathematician who believed that his mathematics was completely 'useless'; he prided himself on the fact that he did pure rather than applied mathematics. Yet, to many biologists who do not know much about Hardy's contributions to mathematics, he is known largely through this very applied piece of work. The second irony is that Hardy agreed to publish in *Science*, which is today an extremely prestigious journal, largely because it was then a not very well-known American journal, and Hardy did not want to publish this fairly simple point in a journal likely to be widely read by his mathematical colleagues in England.

competitive interactions, ecological control of pests, plant-herbivore systems, etc. The realization that disease epidemics are a subset of the range of dynamic behaviour that can be exhibited by microbial populations has resulted in many refinements in the science of epidemiology, with obvious implications for public health. Population and community ecology, both fields that

have been greatly enriched by mathematical reasoning, are also integral to pressing human concerns such as sustainable harvesting of bioresources, bioresource management and conservation biology.

Evolution is the major unifying concept in biology. The mathematical contributions of Fisher, Haldane and Wright have shaped our understanding of the genetic mechanisms of evolutionary change. Very often, application of fairly elementary mathematics to biological problems can be a gold mine in terms of understanding. The Hardy–Weinberg law, with which population genetics can be said to have formally begun in 1908, is merely an application of the binomial expansion, but though mathematically trivial, it is biologically profound (*Box 2*). It has been realized that the knowledge of elementary algebra, manipulation of symbols, solution of simple equations, integration, solving of differential equations, partial differentiation, matrix algebra, etc., is required to fully appreciate the modern 'Neo-Darwinian synthesis', our present view about the mechanisms of adaptive evolution. In the words of Truman, "If you cannot stand algebra, keep out of evolutionary biology". If we believe Dobzhansky's claim that 'nothing in biology makes sense except in the light of evolution', it follows that every biologist should develop at least a working familiarity with mathematics. Comparative evolutionary biology and phylogenetics, which help to understand the interrelationships among different taxonomic groups and trends in evolutionary modification, heavily depend upon models that involve fairly elaborate mathematics. Indeed, many of those who have made major contributions to the development of evolutionary genetics have been people with a background in mathematics or physical sciences.

Stanislaw Ulam, a distinguished non-linear dynamics theorist, said at the inaugural talk at an Applied Mathematics Conference many years ago, "Ask not what mathematics can do for biology; ask, rather, what biology can do for mathematics". Biology has actually inspired and fueled the development of several areas of mathematics and statistics. For example, analysis of variance

The mathematical contributions of Fisher, Haldane and Wright have shaped our understanding of the genetic mechanisms of evolutionary change. Very often, application of fairly elementary mathematics to biological problems can be a gold mine in terms of understanding.

and covariance, correlation, regression, path analysis, etc., were all developed to tackle problems in quantitative genetics. The recent expansion of non-linear dynamics and chaos theory is in part fueled by its applications to biological problems in population and community ecology. Branching theory and the theory of the coalescent were elaborated largely to study patterns in the extinction/persistence of populations or genes over evolutionary time. Natural selection theory has now found application in computer science in the form of 'genetic algorithms'.

The examples in this article are cited not to document the accomplishments of mathematical biologists but to focus on the fact that mathematical tools are intrinsic and indispensable to most fields of biological research. In the words of Maynard Smith, the main difficulty one faces is not the mathematics itself, or the biology itself, but how to fit them together. G Odell, a renowned mathematical biologist, says 'the role of mathematical models is to tell you what you do not know'. Predictive biology is likely to be one of the major developments of the 21st century. Roger Brent of the Molecular Sciences Institute at California feels that people who understand mathematics, computation and statistics, and who are willing to apply their understanding to biological problems, are in a position to make substantial contributions to human knowledge. Biology is a demanding science – partly because living systems are so complex, and partly because biology incorporates concepts from chemistry, physics and mathematics. I believe it is fair to say that it is not only a question of using mathematics and computers in biology, it's more like a multi-way synergistic interaction where all fields and human knowledge benefit.

H A Ranganath
Drosophila Stock Centre
Department of Studies in Zoology
University of Mysore
Manasagangotri
Mysore 570 006, India.
Email:
drosrang@sancharnet.in

Suggested Reading

[1] D Brown and P Rothery, *Models in Biology: Mathematics, Statistics and Computing,* John Wiley & Sons, New York, 1993.
[2] The impact of biology on mathematics, http://www.gdb.org/Dan/mathbio/
[3] S Natarajan, *Resonance* Vol 5, No. 4, pp. 52-59, 2000.
[4] R Gadagkar, *Resonance*, Vol 5, No. 4, pp. 62-73, 2000.
[5] T Ramakrishna Rao, *Resonance*, Vol. 5, No. 4, pp. 85-90, 2000.