

## MATHEMATICAL AND THEORETICAL BIOLOGY: AN INTERDISCIPLINARY SCIENTIFIC RESEARCH FIELD THAT LINKS BIOLOGY, MEDICINE AND BIOTECHNOLOGY

**Abalaka, M. E. and Danboyi, D.**

Department of Microbiology, Federal University of Technology,  
Minna Niger state, Nigeria.

### ABSTRACT

This review underscores the importance and link between mathematics biology, medicine and biotechnology. The application of molecular set theory, population dynamics, and mathematical biophysics was also highlighted. The use of Pearson's statistical correlation mathematics to solve the mystery of gene expression and drug activity in human cancer was also discussed. To realize the possibilities of effective synergy between biology and mathematics will require both avoiding potential problems and seizing potential opportunities. These potential problems were identified as: education, intellectual property, and national security. The study of mathematics and biology will not be complete without the mention of bioinformatics. Bioinformatics is the application of computer science and information technology to the field of biology and medicine. Various forms of bioinformatics was discussed such as computational evolutionary biology and structural bioinformatics approaches.

**Key words:** Theoretical biology, Scientific Research, Medicine, Biotechnology, Mathematical biology

### INTRODUCTION

Mathematical and theoretical biology is an interdisciplinary scientific research field with a range of applications in biology, medicine and biotechnology. The field may be referred to as mathematical biology or biomathematics to stress the mathematical side, or as theoretical biology to stress the biological side.. It includes at least four major subfields: *biological mathematical modeling*, *relational biology/complex systems biology (CSB)*, *bioinformatics* and *computational biomodeling/biocomputing*. [1,2]. Mathematical biology aims at the mathematical representation, treatment and modeling of biological processes, using a variety of applied mathematical techniques and tools. It has both theoretical and practical applications in biological, biomedical and biotechnology research. For example, in cell biology, protein interactions are often represented as "cartoon" models, which, although easy to visualize, do not accurately describe the systems studied. In order to do this, precise mathematical models are required. By describing the systems in a quantitative manner, their behavior can

be better simulated, and hence properties can be predicted that might not be evident to the experimenter.

Such mathematical areas as calculus, probability theory, statistics, linear algebra, abstract algebra, graph theory, combinatorics, algebraic geometry, topology, dynamical systems, differential equations and coding theory are now being applied in biology.[3].

#### **Computer models and automata theory**

A monograph on this topic summarizes an extensive amount of published research in this area up to 1986 [4-6], including subsections in the following areas: computer modeling in biology and medicine, arterial system models, neuron models, biochemical and oscillation networks, quantum automata, quantum computers in molecular biology and genetics[7], cancer modeling[8], neural nets, genetic networks, abstract categories in relational biology[9], metabolic-replication systems, category theory[10], applications in biology and medicine,[25], automata theory, cellular automata, tessellation models [11,

12], and complete self-reproduction, chaotic systems in organisms, relational biology and organismic theories [1, 13]. This published report also includes 390 references to peer-reviewed articles by a large number of authors. [14,5,15] .

### **Molecular set theory**

Molecular set theory was introduced by Anthony Bartholomew, and its applications were developed in mathematical biology and especially in Mathematical Medicine. [16]. Molecular set theory (MST) is a mathematical formulation of the wide-sense chemical kinetics of biomolecular reactions in terms of sets of molecules and their chemical transformations represented by set-theoretical mappings between molecular sets. In a more general sense, MST is the theory of molecular categories defined as categories of molecular sets and their chemical transformations represented as set-theoretical mappings of molecular sets. The theory has also contributed to biostatistics and the formulation of clinical biochemistry problems in mathematical formulations of pathological, biochemical changes of interest to Physiology, Clinical Biochemistry and Medicine [16].

### **Population dynamics**

Population dynamics has traditionally been the dominant field of mathematical biology. Work in this area dates back to the 19th century. The Lotka–Volterra predator-prey equations are a famous example. In the past 30 years, population dynamics has been complemented by evolutionary game theory, developed first by John Maynard Smith. Under these dynamics, evolutionary biology concepts may take a deterministic mathematical form. Population dynamics overlap with another active area of research in mathematical biology: mathematical epidemiology, the study of infectious disease affecting populations. Various models of the spread of infections have been proposed and analyzed, and provide important results that may be applied to health policy decisions.

### **Mathematical methods**

A model of a biological system is converted into a system of equations, although the word 'model' is often used synonymously with the system of

corresponding equations. The solution of the equations, by either analytical or numerical means, describes how the biological system behaves either over time or at equilibrium. There are many different types of equations and the type of behavior that can occur is dependent on both the model and the equations used. The model often makes assumptions about the system. The equations may also make assumptions about the nature of what may occur.

### **Mathematical biophysics**

The earlier stages of mathematical biology were dominated by mathematical biophysics, described as the application of mathematics in biophysics, often involving specific physical/mathematical models of biosystems and their components or compartments.

### **The cell cycle**

The eukaryotic cell cycle is very complex and is one of the most studied topics, since its misregulation leads to cancers. It is possibly a good example of a mathematical model as it deals with simple calculus but gives valid results. Two research groups [17], have produced several models of the cell cycle simulating several organisms. They have recently produced a generic eukaryotic cell cycle model which can represent a particular eukaryote depending on the values of the parameters, demonstrating that the idiosyncrasies of the individual cell cycles are due to different protein concentrations and affinities, while the underlying mechanisms are conserved. By means of a system of ordinary differential equations these models show the change in time (dynamical system) of the protein inside a single typical cell; this type of model is called a deterministic process (whereas a model describing a statistical distribution of protein concentrations in a population of cells is called a stochastic process). To obtain these equations an iterative series of steps must be done: first the several models and observations are combined to form a consensus diagram and the appropriate kinetic laws are chosen to write the differential equations, such as rate kinetics for stoichiometric reactions, Michaelis-Menten kinetics for enzyme substrate reactions and Goldbeter–Koshland kinetics for ultrasensitive

transcription factors, afterwards the parameters of the equations (rate constants, enzyme efficiency coefficients and Michealis constants) must be fitted to match observations; when they cannot be fitted the kinetic equation is revised and when that is not possible the wiring diagram is modified. The parameters are fitted and validated using observations of both wild type and mutants, such as protein half-life and cell size. In order to fit the parameters the differential equations need to be studied. This can be done either by simulation or by analysis. In a simulation, given a starting vector (list of the values of the variables), the progression of the system is calculated by solving the equations at each time-frame in small increments.

In analysis, the proprieties of the equations are used to investigate the behavior of the system depending of the values of the parameters and variables. A system of differential equations can be represented as a vector field, where each vector described the change (in concentration of two or more protein) determining where and how fast the trajectory (simulation) is heading. Vector fields can have several special points: a stable point, called a sink, that attracts in all directions (forcing the concentrations to be at a certain value), an unstable point, either a source or a saddle point which repels (forcing the concentrations to change away from a certain value), and a limit cycle, a closed trajectory towards which several trajectories spiral towards (making the concentrations oscillate).

A better representation which can handle the large number of variables and parameters is called a bifurcation diagram (Bifurcation theory): the presence of these special steady-state points at certain values of a parameter (e.g. mass) is represented by a point and once the parameter passes a certain value, a qualitative change occurs, called a bifurcation, in which the nature of the space changes, with profound consequences for the protein concentrations: the cell cycle has phases (partially corresponding to G1 and G2) in which mass, via a stable point, controls cycling levels, and phases (S and M phases) in which the concentrations change independently, but once

the phase has changed at a bifurcation event (Cell cycle checkpoint), the system cannot go back to the previous levels since at the current mass the vector field is profoundly different and the mass cannot be reversed back through the bifurcation event, making a checkpoint irreversible. In particular the S and M checkpoints are regulated by means of special bifurcations called a Hopf bifurcation and an infinite period bifurcation.

## **MATHEMATICS AND BIOLOGY; THE PAST, PRESENT AND FUTURE.**

### **The Past**

The interactions between mathematics and biology at present follow from their interactions over the last half millennium. The discovery of the New World by Europeans approximately 500 years ago—and of its many biological species not described in religious Scriptures—gave impetus to major conceptual progress in biology.

From the time of the ancient Greek physician Galen (131–201 C.E.) until William Harvey studied medicine in Padua (1600–1602, while Galileo was active there), it was believed that there were two kinds of blood, arterial blood and venous blood. Both kinds of blood were believed to ebb and flow under the motive power of the liver, just as the tides of the earth ebbed and flowed under the motive power of the moon. Harvey became physician to the king of England. He used his position of privilege to dissect deer from the king's deer park as well as executed criminals. Harvey observed that the veins in the human arm have one-way valves that permit blood to flow from the periphery toward the heart but not in the reverse direction. Hence the theory that the blood ebbs and flows in both veins and arteries could not be correct. Harvey also observed that the heart was a contractile muscle with one-way valves between the chambers on each side. He measured the volume of the left ventricle of dead human hearts and found that it held about two ounces (about 60 ml), varying from 1.5 to three ounces in different individuals. He estimated that at least one-eighth and perhaps as much as one-quarter of the blood in the left ventricle was expelled with each stroke of the heart. He

measured that the heart beat 60–100 times per minute. Therefore, the volume of blood expelled from the left ventricle per hour was about  $60 \text{ ml} \times 1/8 \times 60 \text{ beats/minute} \times 60 \text{ minutes/hour}$ , or 27 liters/hour. However, the average human has only 5.5 liters of blood (a quantity that could be estimated by draining a cadaver). Therefore, the blood must be like a stage army that marches off one side of the stage, returns behind the scenes, and reenters from the other side of the stage, again and again. The large volume of blood pumped per hour could not possibly be accounted for by the then-prevalent theory that the blood originated from the consumption of food. Harvey inferred that there must be some small vessels that conveyed the blood from the outgoing arteries to the returning veins, but he was not able to see those small vessels. His theoretical prediction, based on his meticulous anatomical observations and his mathematical calculations, was spectacularly confirmed more than half a century later when Marcello Malpighi saw the capillaries under a microscope[18]. Harvey's discovery illustrates the enormous power of simple, off-the-shelf mathematics combined with careful observation and clear reasoning. It set a high standard for all later uses of mathematics in biology.

### **The Present**

#### **Gene expression and drug activity in human cancer.**

Suppose a person has a cancer. Could information about the activities of the genes in the cells of the person's cancer guide the use of cancer-treatment drugs so that more effective drugs are used and less effective drugs are avoided? To suggest answers to this question, Scherf et al. (2000) [19] ingeniously applied off-the-shelf mathematics, specifically, correlation—invented nearly a century earlier by Karl Pearson (Pearson and Lee, 1903) [20] in a study of human inheritance—and clustering algorithms, which apparently had multiple sources of invention, including psychometrics [21]. They applied these simple tools to extract useful information from, and to combine for the first time, enormous databases on molecular pharmacology and gene expression (<http://discover.nci.nih.gov/arraytools/>). They used

two kinds of information from the drug discovery program of the National Cancer Institute. The first kind of information described gene expression in 1,375 genes of each of 60 human cancer cell lines. A target matrix  $T$  had, as the numerical entry in row  $g$  and column  $c$ , the relative abundance of the mRNA transcript of gene  $g$  in cell line  $c$ . The drug activity matrix  $A$  summarized the pharmacology of 1,400 drugs acting on each of the same 60 human cancer cell lines, including 118 drugs with “known mechanism of action.” The number in row  $d$  and column  $c$  of the drug activity matrix  $A$  was the activity of drug  $d$  in suppressing the growth of cell line  $c$ , or, equivalently, the sensitivity of cell line  $c$  to drug  $d$ . The target matrix  $T$  for gene expression contained 82,500 numbers, while the drug activity matrix  $A$  had 84,000 numbers.

If a person's cancer cells have high expression for a particular gene, and the correlation of that gene with drug activity is highly positive, then that gene may serve as a marker for tumor cells likely to be inhibited effectively by that drug. If the correlation with drug activity is negative, then the marker gene may indicate when use of that drug is contraindicated.

While important scientific questions about this approach remains open, its usefulness in generating hypotheses to be tested by further experiments is obvious. It is a very insightful way of organizing and extracting meaning from many individual observations.

### **The Future**

To realize the possibilities of effective synergy between biology and mathematics will require both avoiding potential problems and seizing potential opportunities.

#### **Potential problems.**

The productive interaction of biology and mathematics will face problems that concern education, intellectual property, and national security. Educating the next generation of scientists will require early emphasis on quantitative skills in primary and secondary schools and more opportunities for training in both biology and

mathematics at undergraduate, graduate, and postdoctoral levels [22].

Intellectual property rights may both stimulate and obstruct the potential synergy of biology and mathematics. Science is a potlatch culture. The bigger one's gift to the common pool of knowledge and techniques, the higher one's status, just as in the potlatch culture of the Native Americans of the northwest coast of North America. In the case of research in mathematics and biology, intellectual property rights to algorithms and databases need to balance the concerns of inventors, developers, and future researchers [23].

The future of a scientific field is probably less predictable than the future in general. Doubtless, though, there will be exciting opportunities for the collaboration of mathematics and biology. Mathematics can help biologists grasp problems that are otherwise too big (the biosphere) or too small (molecular structure); too slow (macroevolution) or too fast (photosynthesis); too remote in time (early extinctions) or too remote in space (life at extremes on the earth and in space); too complex (the human brain) or too dangerous or unethical (epidemiology of infectious agents).

## BIOINFORMATICS

Bioinformatics is the application of computer science and information technology to the field of biology and medicine. Bioinformatics deals with algorithms, databases and information systems, web technologies, artificial intelligence and soft computing, information and computation theory, software engineering, data mining, image processing, modeling and simulation, signal processing, discrete mathematics, control and system theory, circuit theory, and statistics, for generating new knowledge of biology and medicine, and improving & discovering new models of computation (e.g. DNA computing, neural computing, evolutionary computing, immuno-computing, swarm-computing, cellular-computing). Java, XML, Perl, C, C++, Python, R, SQL and Mat Lab are the programming languages popularly used in this field. A bioinformatician needs to have a basic

and general sense of the ideas and approaches of science and engineering [24].

Bioinformatics was applied in the creation and maintenance of a database to store biological information at the beginning of the "genomic revolution", such as nucleotide and amino acid sequences. Development of this type of database involved not only design issues but the development of complex interfaces whereby researchers could both access existing data as well as submit new or revised data.

In order to study how normal cellular activities are altered in different disease states, the biological data must be combined to form a comprehensive picture of these activities. Therefore, the field of bioinformatics has evolved such that the most pressing task now involves the analysis and interpretation of various types of data, including nucleotide and amino acid sequences, protein domains, and protein structures. The actual process of analyzing and interpreting data is referred to as computational biology. Important sub-disciplines within bioinformatics and computational biology include:

- the development and implementation of tools that enable efficient access to, and use and management of, various types of information.
- the development of new algorithms (mathematical formulas) and statistics with which to assess relationships among members of large data sets, such as methods to locate a gene within a sequence, predict protein structure and/or function, and cluster protein sequences into families of related sequences.

### Computational evolutionary biology

Evolutionary biology is the study of the origin and descent of species, as well as their change over time. Informatics has assisted evolutionary biologists in several key ways; it has enabled researchers to:

- trace the evolution of a large number of organisms by measuring changes in their DNA, rather than through physical taxonomy or physiological observations alone,
- more recently, compare entire genomes, which permits the study of more complex evolutionary

events, such as gene duplication, horizontal gene transfer, and the prediction of factors important in bacterial speciation,

- build complex computational models of populations to predict the outcome of the system over time
- track and share information on an increasingly large number of species and organisms

### **Structural Bioinformatics Approaches**

#### **Prediction of protein structure**

Protein structure prediction is another important application of bioinformatics. The amino acid sequence of a protein, the so-called primary structure, can be easily determined from the sequence on the gene that codes for it. In the vast majority of cases, this primary structure uniquely determines a structure in its native environment. (Of course, there are exceptions, such as the bovine spongiform encephalopathy – a.k.a. Mad Cow Disease – prion.) Knowledge of this structure is vital in understanding the function of the protein. For lack of better terms, structural information is usually classified as one of *secondary*, *tertiary* and *quaternary* structure.

One of the key ideas in bioinformatics is the notion of homology. In the genomic branch of bioinformatics, homology is used to predict the function of a gene: if the sequence of gene *A*, whose function is known, is homologous to the sequence of gene *B*, whose function is unknown, one could infer that *B* may share *A*'s function. In the structural branch of bioinformatics, homology is used to determine which parts of a protein are important in structure formation and interaction with other proteins. In a technique called homology modeling, this information is used to predict the structure of a protein once the structure of a homologous protein is known. This currently remains the only way to predict protein structures reliably.

### **CONCLUSION**

This suffices to say that the relationship between mathematics and biology is seamless. Rather than making biology and medicine difficult, mathematics makes it more realistic. Mathematics has played

exceptionally important roles throughout the history of biology. However there is a need to identify a succinct list of achievements that represent the power of mathematics in biology. Mathematical biology has changed a great deal in the last decade, mainly for the better. Whereas twenty years ago, it was peripheral to many sub disciplines of biology, its success is marked today by ever increasing integration into areas of application. Indeed of a truth, mathematical and theoretical biology is an interdisciplinary research field that links biology, medicine, and biotechnology.

### **REFERENCES**

1. Baianu, I. C. 1987, Computer Models and Automata Theory in Biology and Medicine., in M. Witten (ed.), *Mathematical Models in Medicine*, vol. 7., Ch.11 Pergamon Press, New York, 1513-1577. <http://cogprints.org/3687/>
2. Preziosi L. (2003). Cancer Modeling and Simulation. Chapman Hall/CRC Press, [ISBN 1-58488- 361-8](https://doi.org/10.1080/15475270310001628888). <http://library.bjcancer.org/ebook/109.pdf>
3. Robeva, R. et al (Fall 2010). "Mathematical Biology Modules Based on Modern Molecular Biology and Modern Discrete Mathematics". *CBE Life Sciences Education* (The American Society for Cell Biology) **9** (3): 227–240. doi:10.1187/cbe.10-03-0019. PMC 2931670.
4. Witten M. Ed. (1986). "Computer Models and Automata Theory in Biology and Medicine" In: *Mathematical Modeling: Mathematical Models in Medicine*, volume 7:1513-1577, Pergamon Press: New York.
5. Lin, H.C.( 2004). "Computer Simulations and the Question of Computability of Biological Systems": 1-15,doi=10.1.1.108.5072. <https://tspace.library.utoronto.ca/bitstream/1807/2951/2/compauto.pdf>
6. "Computer Models and Automata Theory in Biology and Medicine" (1986).(*Abstract*) [http://biblioteca.universia.net/html\\_bura/ficha/params/title/computer-models-and-automata-theory-in-biology-and-medicine/id/3920559.html](http://biblioteca.universia.net/html_bura/ficha/params/title/computer-models-and-automata-theory-in-biology-and-medicine/id/3920559.html)
7. "Natural Transformations Models in Molecular Biology"(1983). In: *SIAM and Society of Mathematical Biology, National Meeting*, Bethesda,MD:1-12. <http://citeseerx.ist.psu.edu/showciting;jsessionid=BD12D600C39F9979633DB877CA74212B?cid=642862>
8. "Quantum Interactomics and Cancer Mechanisms" (2004): *Research Report communicated to the Institute of Genomic Biology, University of Illinois at Urbana*;1-16 [https://tspace.library.utoronto.ca/retrieve/4969/QuantumInteractomicsInCancer\\_Sept13k4\\_E\\_cuteprt.pdf](https://tspace.library.utoronto.ca/retrieve/4969/QuantumInteractomicsInCancer_Sept13k4_E_cuteprt.pdf)

9. Kainen, P.C.( 2005)."Category Theory and Living Systems", In: Charles Ehresmann's Centennial Conference Proceedings: University of Amiens, France, October 7-9th, 2005, A. Ehresmann, Organizer and Editor. 1-5. <http://vbm-ehr.pagesperso-orange.fr/ChEh/articles/Kainen.pdf>
10. "bibliography for category theory/algebraic topology applications in physics". PlanetPhysics. Retrieved 2011-10-21.
11. Preston K. and Duff M. J. B. *Modern Cellular Auom*[http://books.google.co.uk/books?id=10\\_0qu\\_UC&dq=cellular+automata+and+tessalation&pg=PP1&ots=ciXYCF3AYm&source=citation&sig=CtaUDhisM7MalS7rZfXvp689y-8&hl=en&sa=X&oi=book\\_result&resnum=12&ct=result](http://books.google.co.uk/books?id=10_0qu_UC&dq=cellular+automata+and+tessalation&pg=PP1&ots=ciXYCF3AYm&source=citation&sig=CtaUDhisM7MalS7rZfXvp689y-8&hl=en&sa=X&oi=book_result&resnum=12&ct=result)
12. Wells D.(1991). *The Penguin Dictionary of Curious and Interesting Geometry*. London, Penguin, pp 60-61.
13. "Computer models and automata theory in biology and medicine | KLI Theory Lab". Theorylab.org. 2009-05-26. Retrieved 2011-10-21.
14. Baianu, I. C.; Brown, R.; Georgescu, G.; Glazebrook, J. F. (2006). "Complex Non-linear Biodynamics in Categories, Higher Dimensional Algebra and Łukasiewicz–Moisil Topos: Transformations of Neuronal, Genetic and Neoplastic Networks". *Axiomathes* 16: 65. doi:10.1007/s10516-005-3973-8.
15. "bibliography for mathematical biophysics and mathematical medicine". PlanetPhysics. 2009-01-24. Retrieved 2010-03-17.
16. PlanetPhysics (2009)"molecular set category". Retrieved 2011-10-21.
17. Zhang T, Brazhnik P, and Tyson J.J.(2009). Computational Analysis of Dynamical Responses to the Intrinsic Pathway of Programmed Cell Death. *Biophysical Journal* 97(2)415-434.
18. Adelman, H. (1966) *Marcello Malpighi and the Evolution of Embryology* 5 vol., Cornell University Press, Ithaca, N.Y.
19. Scherf U., Ross D.T., Waltham M., Smith L.H., Lee J.K. et al. (2000) A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 24: 236–244.
20. Pearson K. and Lee A. (1903) On the laws of inheritance in man. *Biometrika* 2: 357–462.
21. Johnson S.C (1967) Hierarchical clustering schemes. *Psychometrika* 2: 241–254.
22. [CUBE] Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century, Board on Life Sciences, Division on Earth and Life Studies, National Research Council of the National Academies (2003) *BIO 2010: Transforming undergraduate education for future research biologists*. Washington (D.C.): National Academies Press. 191 p.
23. Rai A.K. and Eisenberg R.S. (2003) Bayh-Dole reform and the progress of biomedicine. *Am Sci* 91: 52–59. PMID 20810955.
24. Appel R.D and Freytmans E.(2011). *Bioinformatics*. World Scientific Publishing Co.
25. "bibliography for mathematical biophysics and mathematical medicine". PlanetPhysics. 2009-01-24. Retrieved 2010-03-17.