

# Computational Biology

**Michelle LaBrunda**

*Health First Medical Group, USA*

**Mary Jane Miller**

*University of Guam, Guam*

**Andrew La Brunda**

*University of Guam, Guam*

## INTRODUCTION

There are many challenges that science researchers face today. Among these are tasks of data acquisition and management, data transformation and modeling, and combining different data sources to focus on a single problem. Researchers are also challenged to make use of evolving machine learning methods, and to develop new modeling strategies (Gentleman, 2004). Computational biology involves high performance computing that addresses the needs of advanced scientific study. It is often employed in solving complex problems that could not otherwise be effectively attempted. The use of computational biology has become imperative in many areas of scientific research and has evolved into a discipline itself, within biological and computer sciences (Kingsbury, 96).

Modern machine learning techniques have proven to be extremely valuable in analyzing data in computational biology problems (Scholkopf, Koji, & Vert, 1995) and the field itself has been revolutionized by advances in both hardware and software (Bader, 2004). Biology today is in an era that is yielding many remarkable discoveries with the promise of many more to come. Leading the way to successfully making new discoveries, modeling scientific outcomes, contributing solutions to complex problems and revealing links between solutions is the exponential growth in the size of information-packed databases. Critical to the current and future success of numerous research projects is the ability to analyze and use all available data (Scholkopf et al., 1995), making computational biology an increasingly important tool for researchers.

## BACKGROUND

Computational biology is a relatively new branch of modern biology. It employs aspects of biology, computer science, and mathematics to solve problems that are unworkable with traditional biological techniques. It is only in the last 30 years that the tools needed by computational biologist have been available. With the advent and advancement of complex data processing systems, computational biology has emerged.

The term “computational biology” is often confused with “bioinformatics” because of the similarity of tools which are employed to solve problems in their respective areas. Both use advanced computational processing and mathematical modeling to explain phenomena and predict outcomes. They both involve the use of techniques from computer science, statistics, and applied mathematics to model living systems and solve biological problems. The primary difference between computational biology and bioinformatics is that the former tends to focus on the testing of hypotheses and new discoveries within the realm of biology and the latter tends to focus on the development of mathematical techniques and algorithms that can be applied to the simulation of biological systems (Ouzounis, 2012).

The field of computational biology offers a nontraditional approach to investigating complex biological problems. Typical topics within the field include gene finding, genome assembly, protein structure prediction and alignment, and the modeling of biological systems and processes over long periods of time. Examples of processes over time suitable for such models would include evolutionary trends, gene expression through

DOI: 10.4018/978-1-4666-5888-2.ch043

multiple generations, and perhaps long-term biological consequences of climate change (Ouzounis, 2012).

Computational biology includes many traditional areas such as systems biology, molecular biology, biochemistry, biophysics, statistics, and computer science as well as recently developed disciplines including bioinformatics and computational genomics. Algorithms which are able to closely model biological behavior help to validate the medical understanding of the observed processes and can be used to model scenarios that might not be able to be physically reproduced.

The ultimate goal of computational biology would be to create a high level software based organism that is comprised of a collection of biological subsystems which would include circulatory, digestive, endocrine, integumentary, lymphatic, muscular, nervous, reproductive, respiratory, skeletal, reproductive, and urinary systems. Within each of these systems, software based cells and their biological, mechanical, and chemical behaviors would be programmed to interact with the environment and subsystems with which the cell functions.

If such software could be created it would be possible to develop and administer experimental medication to a software organism without endangering the host. Creating a system this complex is currently an overwhelming task. Instead the discipline of computational biology focuses on smaller pieces with specific goals in a highly controlled environments. At this level of simplicity it's possible to create simulations that can be verified and compared to physical experimental results.

In this article we will examine a few interesting branches in more detail, including computational genomics, systems biology, protein structure prediction, and evolutionary biology, all of which model microscopic structures.

## COMPUTATIONAL GENOMICS

To simplify the complexity of a biology simulation a distinction is made between internal and external stimuli. Computational genomics focuses on the processes of internal stimuli, specifically processing DNA and RNA (Koonin, 2001).

An organism's heredity is stored as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). Each of the storage methods contains a linear chain of finite elements called bases. In the case of DNA the chain is composed of four bases, adenine, cytosine, guanine, and thymine. RNA is composed of four bases, adenine, cytosine, guanine, and uracil. This information can be represented in a computer model through a series of computer data structures.

Certain regions of DNA, called genes, are used by the body to create proteins. These proteins are used to construct and maintain the organism. Genes can be thought of as blueprint instructions for how to make each unique individual. There are long stretches of DNA between the genes. The function of these lengths of DNA is not well understood. The sum of all the genetic information about an organism is called a genome.

For the purpose of biological simulation, computational genomics views DNA as one very large input parameter used to chemically drive cellular behavior. The challenge is that without a precise understanding of each DNA and RNA component, creating a behavior model for an entire organism isn't possible. There are two major challenges in creating a fully defined genomic structure. The first is genetic sequencing, (determining the order of the bases that make a strand of genetic material) and the second is localizing the genes within the genome. Sequence comparison is probably the most useful computational tool for molecular biologists. Repositories containing hundreds of genomes have been established and are available for public access. A biologist now has the ability to compare a unique sequence of DNA with the already known genetic sequences from this massive repository. Prior to the application of computers to the genomic problems, scientist had to manually attempt to align sequences using ill-suited tools such word processors (Koonin, 2001).

One of the most important modern molecular genetic advances is the sequencing of specific genes and computational biology is becoming of increasing importance in such studies. In the future, medications may be tailor made to the needs of each individual based on their specific genetic makeup. One of the hurdles that must be crossed to reach this point is a cost

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/computational-biology/112356](http://www.igi-global.com/chapter/computational-biology/112356)

## Related Content

---

### Resource Allocation in OFDMA as an Optimization Problem

Álvaro Pachón, Ubaldo García-Palomares and Juan Manuel Madrid (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 5730-5740).

[www.irma-international.org/chapter/resource-allocation-in-ofdma-as-an-optimization-problem/113028](http://www.irma-international.org/chapter/resource-allocation-in-ofdma-as-an-optimization-problem/113028)

### Detecting Communities in Dynamic Social Networks using Modularity Ensembles SOM

Raju Enugala, Lakshmi Rajamani, Sravanthi Kurapati, Mohammad Ali Kadampur and Y. Rama Devi (2018). *International Journal of Rough Sets and Data Analysis* (pp. 34-43).

[www.irma-international.org/article/detecting-communities-in-dynamic-social-networks-using-modularity-ensembles-som/190889](http://www.irma-international.org/article/detecting-communities-in-dynamic-social-networks-using-modularity-ensembles-som/190889)

### Marketing Vulnerabilities in an Age of Online Commerce

Robert S. Owen (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 5403-5411).

[www.irma-international.org/chapter/marketing-vulnerabilities-in-an-age-of-online-commerce/112989](http://www.irma-international.org/chapter/marketing-vulnerabilities-in-an-age-of-online-commerce/112989)

### Big Data Summarization Using Novel Clustering Algorithm and Semantic Feature Approach

Shilpa G. Kolte and Jagdish W. Bakal (2017). *International Journal of Rough Sets and Data Analysis* (pp. 108-117).

[www.irma-international.org/article/big-data-summarization-using-novel-clustering-algorithm-and-semantic-feature-approach/182295](http://www.irma-international.org/article/big-data-summarization-using-novel-clustering-algorithm-and-semantic-feature-approach/182295)

### Fuzzy Decision Support System for Coronary Artery Disease Diagnosis Based on Rough Set Theory

Noor Akhmad Setiawan (2014). *International Journal of Rough Sets and Data Analysis* (pp. 65-80).

[www.irma-international.org/article/fuzzy-decision-support-system-for-coronary-artery-disease-diagnosis-based-on-rough-set-theory/111313](http://www.irma-international.org/article/fuzzy-decision-support-system-for-coronary-artery-disease-diagnosis-based-on-rough-set-theory/111313)