



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

In Silico Models: From Simple Networks to Complex Diseases

Debmalya Barh^{*}, Vijender Chaitankar[†], Eugenia Ch Yiannakopoulou^{**}, Emmanuel O Salawu^{††}, Sudhir Chowbina^{***}, Preetam Ghosh[†] and Vasco Azevedo^{†††}

^{*}Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, India, [†]Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia, USA, ^{**}Department of Basic Medical Lessons Faculty of Health and Caring Professions, Technological Educational Institute of Athens, Athens, Greece, ^{††}Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan, PhD Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan, and Institute of Information Science, Academia Sinica, Taipei, Taiwan, ^{***}Advanced Biomedical Computing Center, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, Maryland, USA, ^{†††}Laboratorio de Genetica Celular eMolecular, Departamento de Biologia Geral, Instituto de Ciencias Biologicas, Universidade Federal de Minas Gerais Belo Horizonte, Minas Gerais, Brazil

Chapter Outline

Summary	385	Specific Examples of Infectious Diseases in the <i>In Silico</i> Model	396
What You Can Expect to Know	385	Challenges in <i>In Silico</i> Modeling of Infectious Diseases	397
History and Methods	386	Neuronal Diseases	397
Bioinformatics in Animal Biotechnology	386	Pathophysiology of Neuronal Diseases as the Source of Parameters for <i>In Silico</i> Modeling of Neuronal Diseases	397
Bioinformatics and Systems Biology	387	Parameters for <i>In Silico</i> Modeling of Neuronal Diseases	398
Common Computational Methods in Systems Biology	387	Neuronal Disease <i>In Silico</i> Model Proper (a Typical Approach/Scenario)	398
Experimental Methods in Systems Biology	388	Specific Examples of Neuronal Disease <i>In Silico</i> Models	398
Protein–Protein Interactions (PPIs)	388	Possible Limitations of <i>In Silico</i> Modeling of Neuronal Diseases	399
Transcriptional Control Networks	388	Conclusion	400
Signal Transduction Networks	388	Ethical Issues	400
Mathematical Modeling Techniques	389	Translational Significance	400
The Concept of Modeling	390	World Wide Web Resources	401
<i>In Silico</i> Models of Cells	391	References	401
Advantages and Disadvantages of <i>In Silico</i> Modeling	392	Further Reading	402
Applications of <i>In Silico</i> Disease Modeling in Practice	392	Glossary	403
<i>In Silico</i> Models of Cancer	392	Abbreviations	403
Types of Cancer Models	393	Long Answer Questions	403
<i>In Silico</i> Models and Inflammatory Response Syndrome in Trauma and Infection	394	Short Answer Questions	403
Infectious Diseases	394	Answers to Short Answer Questions	403
Triad of Infectious Diseases as the Source of Parameters for <i>In Silico</i> Modeling of Infectious Diseases	394		
Parameters for <i>In Silico</i> Modeling of Infectious Diseases	395		
Infectious Diseases <i>In Silico</i> Model Proper, a Typical Approach/Scenario	395		

SUMMARY

In this chapter, we consider *in silico* modeling of diseases starting from some simple to some complex (and mathematical) concepts. Examples and applications of *in silico* modeling for some important categories of diseases (such as for cancers, infectious diseases, and neuronal diseases) are also given.

WHAT YOU CAN EXPECT TO KNOW

Recent advances in bioinformatics and systems biology have enabled modeling and simulation of sub-cellular and cellular processes, and disease using primarily methods from dynamical systems theory. In this approach, all interactions among all components in a system are described

mathematically and computed models are established. These *in silico* models encode and test hypotheses about mechanisms underlying the function of cells, the pathogenesis and pathophysiology of disease, and contribute to identification of new drug targets and drug design. The development of *in silico* models is facilitated by rapidly advancing experimental and analytical tools that generate information-rich, high-throughput biological data. Bioinformatics provides tools for pattern recognition, machine learning, statistical modeling, and data extraction from databases that contribute to *in silico* modeling. Dynamical systems theory is the natural language for investigating complex biological systems that demonstrate nonlinear spatio-temporal behavior. Most *in silico* models aim to complement (and not replace) experimental research. Experimental data are needed for parameterization, calibration, and validation of *in silico* models. Typical examples in biology are models for molecular networks, where the behavior of cells is expressed in terms of quantitative changes in the levels of transcripts and gene products, as well as models of cell cycle. In medicine, *in silico* models of cancer, immunological disease, lung disease, and infectious diseases complement conventional research with *in vitro* models, animal models, and clinical trials. This chapter presents basic concepts of bioinformatics, systems biology, their applications in *in silico* modeling, and also reviews applications in biology and disease.

HISTORY AND METHODS

BIOINFORMATICS IN ANIMAL BIOTECHNOLOGY

Biotechnology will be the most promising life science frontier for the next decade. Together with informatics, biotechnology is leading revolutionary changes in our society and economy. This genomic revolution is global, and is creating new prospects in all biological sciences, including medicine, human health, disease, and nutrition, agronomy, and animal biotechnology.

Animal biotechnology is a source of innovation in production and processing, profoundly impacting the animal husbandry sector, which seeks to improve animal product quality, health, and well-being. Biotechnological research products, such as vaccines, diagnostics, *in vitro* fertilization, transgenic animals, stem cells, and a number of other therapeutic recombinant products, are now commercially available. In view of the immense potential of biotechnology in the livestock and poultry sectors, interest in animal biotechnology has increased over the years.

The fundamental requirement for modern biotechnology projects is the ability to gather, store, classify, analyze, and distribute biological information derived from genomics projects. Bioinformatics deals with methods for

storing, retrieving, and analyzing biological data and protein sequences, structures, functions, pathways, and networks, and recently, *in silico* disease modeling and simulation using systems biology. Bioinformatics encompasses both conceptual and practical tools for the propagation, generation, processing, and understanding of scientific ideas and biological information.

Genomics is the scientific study of structure, function, and interrelationships of both individual genes and the genome. Lately, genomics research has played an important role in uncovering the building blocks of biology and complete genome mapping of various living organisms. This has enabled researchers to decipher fundamental cellular functions at the DNA level, such as gene regulation or protein–protein interactions, and thus to discover molecular signatures (clusters of genes, proteins, metabolites, etc.), which are characteristic of a biological process or of a specific phenotype. Bioinformatics methods and databases can be developed to provide solutions to challenges of handling massive amounts of data.

The history of animal biotechnology with bioinformatics is to make a strong research community that will build the resources and support veterinary and agricultural research. There are some technologies that were used dating back to 5,000 B.C. Many of these techniques are still being used today. For example, hybridizing animals by crossing specific strains of animals to create greater genetic varieties is still in practice. The offspring of some of these crosses are selectively bred afterward to produce the most desirable traits in those specific animals.

There has been significant interest in the complete analysis of the genome sequence of farm animals such as chickens, pigs, cattle, sheep, fish, and rabbits. The genomes of farm animals have been altered to search for preferred phenotypic traits, and then selected for better-quality animals to continue into the next generation. Access to these sequences has given rise to genome array chips and a number of web-based mapping tools and bioinformatics tools required to make sense of the data. In addition, organization of gigabytes of sequence data requires efficient bioinformatics databases. [Fadiel et al. \(2005\)](#) provides a nice overview of resources related to farm animal bioinformatics and genome projects.

With farm animals consuming large amounts of genetically modified crops, such as modified corn and soybean crops, it is good to question the effect this will have on their meat. Some of the benefits of this technology are that what once took many years of trial and error is now completed in just months. The meats that are being produced are coming from animals that are better nourished by the use of biotechnology. Biotechnology and conventional approaches are benefiting both poultry and livestock producers. This will give a more wholesome affordable product that will meet growing population demands.

Moreover, bioinformatics methods devoted to investigating the genomes of farm animals can bring eventual economic benefits, such as ensuring food safety and better food quality in the case of beef. Recent advances in high-throughput DNA sequencing techniques, microarray technology, and proteomics have led to effective research in bovine muscle physiology to improve beef quality, either by breeding or rearing factors. Bioinformatics is a key tool to analyze the huge datasets obtained from these techniques. The computational analysis of global gene expression profiling at the mRNA or protein level has shown that previously unsuspected genes may be associated either with muscle development or growth, and may lead to the development of new molecular indicators of tenderness. Gene expression profiling has been used to document changes in gene expression; for example, following infection by pathological organisms, during the metabolic changes imposed by lactation in dairy cows, in cloned bovine embryos, and in various other models.

Bioinformatics enrichment tools are playing an important role in facilitating the functional analysis of large gene lists from various high-throughput biological studies. Huang *et al.* discusses 68 bioinformatics enrichment tools, which helps us understand their algorithms and the details of a particular tool. However, in biology genes do not act independently, but in a highly coordinated and interdependent manner. In order to understand the biological meaning, one needs to map these genes into Gene-Ontology (GO) categories or metabolic and regulatory pathways. Different bioinformatics approaches and tools are employed for this task, starting from GO-ranking methods, pathway mappings, and biological network analysis (Werner, 2008). The awareness of these resources and methods is essential to make the best choices for particular research interests.

Knowledge of bioinformatics tools will facilitate their wide application in the field of animal biotechnology. Bioinformatics is the computational data management discipline that helps us gather, analyze, and represent this information in order to educate ourselves, understand biological processes in healthy and diseased states, and to facilitate discovery of better animal products. Continued efforts are required to develop cost-effective and efficient computational platforms that can retrieve, integrate, and interpret the knowledge behind the genome sequences. The application of bioinformatics tools for biotechnology research will have significant implications in the life sciences and for the betterment of human lives. Bioinformatics is being adopted worldwide by academic groups, companies, and national and international research groups, and it should be thought of as an important pillar of current and future biotechnology, without which rapid progress in the field would not be possible. Systems approaches in combination with genomics, proteomics, metabolomics, and kinomics data have tremendous potential for providing insights into various biological mechanisms, including the most important human diseases.

BIOINFORMATICS AND SYSTEMS BIOLOGY

We are witnessing the birth of a new era in biology. The ability to uncover the genetic code of living organisms has dramatically changed the biological and biomedical sciences approach towards research. These new approaches have also brought in newer challenges.

One such challenge is that recent and novel technologies produce biological datasets of ever-increasing size, including genomic sequences, RNA and protein abundances, their interactions with each other, and the identity and abundance of other biological molecules. The storage and compilation of such quantities of biological data is a challenge: the human genome, for example, contains 3 billion chemical units of DNA, whereas a protozoan genome has 670 billion units of DNA. Data management and interpretation requires development of newly sophisticated computational methods based on research in biology, medicine, pharmacology, and agricultural studies and using methods from computer science and mathematics – in other words, the multi-disciplinary subject of bioinformatics.

Bioinformatics enables researchers to store large datasets in a standard computer database format and provides tools and algorithms scientists use to extract integrated information from the databases and use it to create hypotheses and models. Bioinformatics is a growth area because almost every experiment now involves multiple sources of data, requiring the ability to handle those data and to draw out inferences and knowledge. After 15 years of rapid evolution, the subject is now quite ubiquitous.

Another challenge lies in deciphering the complex interactions in biological systems, known as systems biology. Systems biology can be described as a biology-based interdisciplinary field of study that focuses on complex interactions of biological systems. Those in the field claim that it represents a shift in perspective towards holism instead of reductionism.

Systems biology has great potential to facilitate development of drugs to treat specific diseases. The drugs currently on the market can target only those proteins that are known to cause disease. However, with the human genome now completely mapped, we can target the interaction of genes and proteins at a systems biology level. This will enable the pharmaceutical industry to design drugs that will only target those genes that are diseased, improving healthcare in the United States. Like two organs in one body, systems analysis and bioinformatics are separate but interdependent.

COMMON COMPUTATIONAL METHODS IN SYSTEMS BIOLOGY

Computational methods take an interdisciplinary approach, involving mathematicians, chemists, biologists, biochemists, and biomedical engineers. The robustness of datasets

related to gene interaction and co-operation at a systems level requires multifaceted approaches to create a hypothesis that can be tested. Two approaches are used to understand the network interactions in systems biology, namely Experimental and Theoretical and Modeling techniques (Choi, 2007). In the following sections is a detailed overview of the different computational or bioinformatics methods in modern systems biology.

EXPERIMENTAL METHODS IN SYSTEMS BIOLOGY

Experimental methods utilize real situations to test the hypothesis of mined data sets. As such, living organisms are used whereby various aspects of genome-wide measurements and interactions are monitored. Specific examples on this point include:

PROTEIN–PROTEIN INTERACTIONS (PPIs)

Protein–protein interaction predictions are methods used to predict the outcome of pairs or groups of protein interactions. These predictions are done *in vivo*, and various methods can be used to carry out the predictions. Interaction prediction is important as it helps researchers make inferences of the outcomes of PPI. PPI can be studied by phylogenetic profiling, identifying structural patterns and homologous pairs, intracellular localization, and post-translational modifications, among others. A survey of available tools and web servers for analysis of protein–protein interactions is provided by Tuncbag et al., 2009.

TRANSCRIPTIONAL CONTROL NETWORKS

Within biological systems, several activities involving the basic units of a gene take place. Such processes as DNA replication, and RNA translation and transcription into proteins must be controlled; otherwise, the systems could yield numerous destructive or useless gene products. Transcriptional control networks, also called gene regulatory networks, are segments within the DNA that govern the rate and product of each gene.

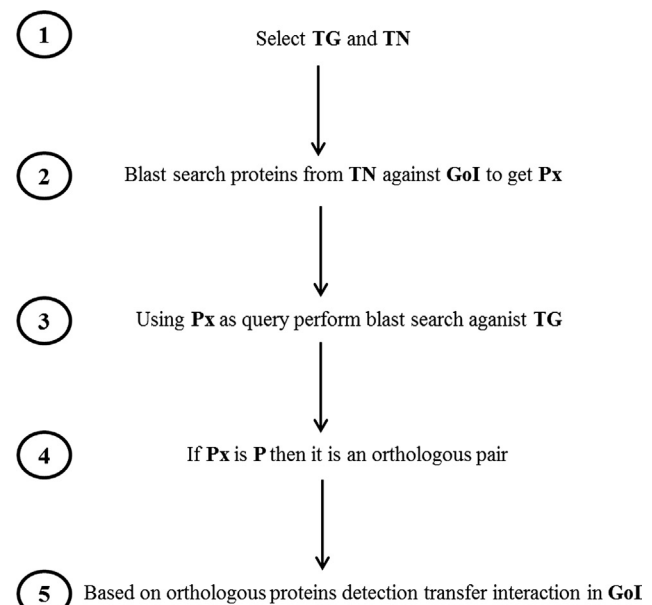
Bioinformatics have devised methods to look for destroyed, dormant, or unresponsive control networks. The discovery of such networks helps in corrective therapy, hence the ability to control some diseases resulting from such control network breakdowns. There has also been rapid progress in the development of computational methods for the genome-wide “reverse engineering” of such networks. ARACNE is an algorithm to identify direct transcriptional interactions in mammalian cellular networks, and promises to enhance the ability to use microarray data to elucidate cellular processes and to identify molecular targets of pharmacological drugs in mammalian cellular networks.

In addition to methods like ARACNE, systems biology approaches are needed that incorporate heterogeneous data sources, such as genome sequence and protein–DNA interaction data. The development of such computational modeling techniques to include diverse types of molecular biological information clearly supports the gene regulatory network inference process and enables the modeling of the dynamics of gene regulatory systems. One such technique is the template-based method to construct networks. An overview of the method is shown in Flow Chart 21.1.

The template-based transcriptional control network reconstruction method exploits the principle that orthologous proteins regulate orthologous target genes. Given a genome of interest (GoI), the first step is to select the template genome (TG) and known regulatory interactions (i.e. template network, TN) in this genome. In step 2, for every protein (P) in TN, a blast search is performed against GoI to obtain the best hit sequences (Px). In step three these Px are then used as a query to perform a blast search against TG. If the best hit using Px as a query happens to be P, then both P and Px are selected as orthologous proteins in step four. If orthologs were detected for an interacting P and target gene then the interaction is transferred in GoI in the final step. Note that this automated way of detecting orthologs can infer false positives.

SIGNAL TRANSDUCTION NETWORKS

Signal transduction is how cells communicate with each other. Signal transduction pathways involve interactions between proteins, micro- and macro-molecules, and DNA. A breakdown in signal transduction pathways could lead



FLOW CHART 21.1 Template-based method for regulatory network reconstruction.

to detrimental consequences within the system due to lack of integrated communication. Correction of broken signal transduction pathways is a therapeutic approach researched for use in many areas of medicine.

High-throughput and multiplex techniques for quantifying signaling and cellular responses are becoming increasingly available and affordable. A high-throughput quantitative multiplex kinase assay, mass spectrometry-based proteomics, and single-cell proteomics are a few of the experimental methods used to elucidate signal transduction mechanisms of cells. These large-scale experiments are generating large data sets on protein abundance and signaling activity. Data-driven modeling approaches such as clustering, principal components analysis, and partial least squares need to be developed to derive biological hypotheses. The potential of data-driven models to study large-scale data sets quantitatively and comprehensively will make sure that these methods will emerge as standard tools for understanding signal-transduction networks.

MATHEMATICAL MODELING TECHNIQUES

The systems biology and mathematical biology fields focus on modeling biological systems. Computational systems biology aims to develop computational models of biological systems. Specifically, it focuses on developing and using efficient algorithms, data structures, visualization tools, and communication tools. A mathematical model can provide new insights into a biological model of interest and help in generating testable predictions.

Modeling or simulation can be viewed as a way of creating an artificial biological system *in vitro* whose properties can be changed or made dynamic. By externally controlling the model, new datasets can be created and implemented at a systems level to create novel insights into treating gene-related problems. In modeling and simulation, sets of differential equations and logic clauses are used to create a dynamic systems environment that can be tested.

Mathematical models of biochemical networks (signal transduction cascades, metabolic pathways, gene regulatory networks) are a central component of modern systems biology. The development of formal methods adopted from theoretical computing science is essential for the modeling and simulation of these complex networks. The computational methods that are being employed in mathematical biology and bioinformatics are the following: (a) directed graphs, (b) Bayesian networks, (c) Boolean networks and their generalizations, (d) ordinary and partial differential equations, (e) qualitative differential equations, (f) stochastic equations, and (g) rule-based formalisms. Below are a few specific examples of the applications of these methods.

Mathematical models can be used to investigate the effects of drugs under a given set of perturbations based on specific tumor properties. This integration can help in the

development of tools that aid in diagnosis and prognosis, and thus improve treatment outcome in patients with cancer. For example, breast cancer, being a well-studied disease over the last decade, serves as a model disease. One can thus apply the principles of molecular biology and pathology in designing new predictive mathematical frameworks that can unravel the dynamic nature of the disease. Genetic mutations of BRCA1, BRCA2, TP53, and PTEN significantly affect disease prognosis and increase the likelihood of adverse reactions to certain therapies. These mutations enable normal cells to become self-sufficient in survival in a stepwise process. Enderling et al. (2006) modeled this mutation and expansion process by assuming that mutations in two tumor-suppressor genes are sufficient to give rise to a cancer. They modified Enderling's earlier model, which was based on an established partial differential equation model of solid tumor growth and invasion. The stepwise mutations from a normal breast stem cell to a tumor cell have been described using a model consisting of four differential equations.

Recently, Woolf et al. (2005) applied a novel graphical modeling methodology known as Bayesian network analysis to model discovery and model selection for signaling events that direct mouse embryonic stem cells (an important preliminary step in hypothesis testing) in protein signaling networks. The model predicts bidirectional dependence between the two molecules ERK and FAK. It is interesting to appreciate that the apparent complexity of these dynamic ERK–FAK interactions is quite likely responsible for the difficulty in determining clear “upstream” versus “downstream” influence relationships by means of standard molecular cell biology methods. Bayesian networks determine the relative probability of statistical dependence models of arbitrary complexity for a given set of data. This method offers further clues to apply Bayesian approaches to cancer biology problems.

Cell cycle is a process in which cells proliferate while collectively performing a series of coordinated actions. Cell-cycle models also have an impact on drug discovery. Chassagnole et al. (2006) used a mathematical model to simulate and unravel the effect of multi-target kinase inhibitors of cyclin-dependent kinases (CDKs). They quantitatively predict the cytotoxicity of a set of kinase inhibitors based on the *in vitro* IC₅₀ measurement values. Finally, they assess the pharmaceutical value of these inhibitors as anti-cancer therapeutics.

In cancer, avascular tumor growth is characterized by localized, benign tumor growth where the nearby tissues consume most of the nutrients. Mathematical modeling of avascular tumor growth is important to understanding the advanced stages of cancer. Kiran et al. (2009) have developed a spatial–temporal mathematical model classified as a different zone model (DZM) for avascular tumor growth based on diffusion of nutrients and their consumption, and

it includes key mechanisms in the tumor. The diffusion and nutrient consumption are represented using partial differential equations. This model predicts that onset of necrosis occurs when the concentrations of vital nutrients are below critical values, and also the overall tumor growth based on the size effects of the proliferation zone, quiescent zone, and necrotic zone.

The mathematical approaches towards modeling the three natural scales of interest (subcellular, cellular, and tissue) are discussed above. Developing models that can predict the effects across biological scales is a challenge. The long-term goal is to build a “virtual human made up of mathematical models with connections at the different biological scales (from genes to tissue to organ).”

THE CONCEPT OF MODELING

A model is an optimal mix of hypotheses, evidence, and abstraction to explain a phenomenon. Hypothesis is a tentative explanation for an observation, phenomenon, or scientific problem that can be tested by further investigation. Evidence describes information (i.e. experimental data) that helps in forming a conclusion or judgment. Abstraction is an act of filtering out the required information to focus on a specific property only. For example, archiving books based on the year of publication, irrespective of the author name, would be an example of abstraction. In this process, some detail is lost and some gained. Predictions are made through modeling that can be tested by experiment. A model may be simple (e.g. the logistic equation describing how a population of bacteria grows) or complicated. Models may be mathematical or statistical.

Mathematical models make predictions, whereas statistical models enable us to draw statistical inferences about the probable properties of a system. In other words, models can be deductive or inductive. If the prediction is necessarily true given that the model is also true, then the model is a deductive model. On the other hand, if the prediction is statistically inferred from observations, then the model is inductive. Deductive models contain a mathematical description; for example, the reaction–diffusion equation that makes predictions about reality. If these predictions do not agree with experiment, then the validity of the entire model may be questioned. Mathematical models are commonly applied in physical sciences. On the other hand, inductive models are mostly applied in the biological sciences. In biology, models are used to describe, simulate, analyze, and predict the behavior of biological systems. Modeling in biology provides a framework that enables description and understanding of biological systems through building equations that express biological knowledge. Modeling enables the simulation of the behavior of a biological system by performing *in silico* experiments (i.e. numerically solving the equations or rules that describe the model). The results of these

in silico experiments become the input for further analysis; for example, identification of key parameters or mechanisms, interpretation of data, or comparison of the ability of different mechanisms to generate observed data.

In particular, systems biology employs an integrative approach to characterizing biological systems in which interactions among all components in a system are described mathematically to establish a computable model. These *in silico* models complement traditional *in vivo* animal models and can be applied to quantitatively study the behavior of a system of interacting components. The term “*in silico*” is poorly defined, with several researchers claiming their role in its origination (Ekins et al., 2007). Sieburg (1990) and (Danchin et al. 1991) were two of the earliest published works that used this term. Specifically, *in silico* models gained much interest in the early stages by various imaging studies (Chakroborty et al., 2003). As an example, microarray analysis that enabled measurement of genome-scale expression levels of genes provided a method to investigate regulatory networks. Years of regulatory network studies (that included microarray-based investigations) led to the development of some well-characterized regulatory networks such as *E. coli* and yeast regulatory networks. These networks are available in the GeneNetWeaver (GNW) tool. GNW is an open-source tool for *in silico* benchmark generation and performance profiling of network inference methods. Thus, the advent of high-throughput experimental tools has allowed for the simultaneous measurement of thousands of biomolecules, opening the way for *in silico* model construction of increasingly large and diverse biological systems. Integrating heterogeneous dynamic data into quantitative predictive models holds great promise for significantly increasing our ability to understand and rationally intervene in disease-perturbed biological systems. This promise – particularly with regards to personalized medicine and medical intervention – has motivated the development of new methods for systems analysis of human biology and disease. Such approaches offer the possibility of gaining new insights into the behavior of biological systems, of providing new frameworks for organizing and storing data and performing statistical analyses, suggesting new hypotheses and new experiments, and even of offering a “virtual laboratory” to supplement *in vivo* and *in vitro* work.

However, *in silico* modeling in the life sciences is far from straightforward, and suffers from a number of potential pitfalls. Thus, mathematically sophisticated but biologically useless models often arise because of a lack of biological input, leading to models that are biologically unrealistic, or that address a question of little biological importance. On the other hand, models may be biologically realistic but mathematically intractable. This problem usually arises because biologists unfamiliar with the limitations of mathematical analysis want to include every known biological effect in the model. Even if it were possible to produce

such models, they would be of little use since their behavior would be as complex to investigate as the experimental situation. These problems can be avoided by formulating clear explicit biological goals before attempting to construct a model. This will ensure that the resulting model is biologically sound, can be experimentally verified, and will generate biological insight or new biological hypotheses. The aim of a model should not simply be to reproduce biological data. Indeed, often the most useful models are those that exhibit discrepancies from experiment. Such deviations will typically stimulate new experiments or hypotheses. An iterative approach has been proposed, starting with a biological problem, developing a mathematical model, and then feeding back into the biology. Once established, this collaborative loop can be traversed many times, leading to ever increasing understanding.

The ultimate goal of *in silico* modeling in biology is the detailed understanding of the function of molecular networks as they appear in metabolism, gene regulation, or signal transduction. This is achieved by using a level of mathematical abstraction that needs a minimum of biological information to capture all physiologically relevant features of a cellular network. For example, ideally, for *in silico* modeling of a molecular network, knowledge of the network structure, all reaction rates, concentrations, and spatial distributions of molecules at any time point is needed. Unfortunately, such information is unavailable even for the best-studied systems. *In silico* simulations thus always have to use a level of mathematical abstraction, which is dictated by the extent of our biological knowledge, by molecular details of the network, and by the specific questions that are addressed. Understanding the complexity of the disease and its biological significance in health can be achieved by integrating data from the different functional genomics experiments with medical, physiological, and environmental factor information, and computing mathematically. The advantage of mathematical modeling of disease lies in the fact that such models not only shed light on how a complex process works, which could be very difficult for inferring an understanding of each component of this process, but also predict what may follow as time evolves or as the characteristics of particular system components are modified. Mathematical models have generally been utilized in association with an increased understanding of what models can offer in terms of prediction and insight.

The two distinct roles of models are prediction and understanding the accuracy, transparency, and flexibility of model properties. Prediction of the models should be accurate, including all the complexities and population-level heterogeneity that have an additional use as a statistical tool. It also provides the understanding of how the disease spreads in the real world and how the complexity affects the dynamics. Model understanding aids in developing sophisticated predictive models, along with gathering more

relevant epidemiological data. A model should be as simple as possible and should have balance in accuracy, transparency, and flexibility; in other words, a model should be well suited for its purpose. The model should be helpful in understanding the behavior of the disease and able to simplify the other disease condition.

IN SILICO MODELS OF CELLS

Several projects are proceeding along these lines, such as E-CELL (Tomita, 2001) and simulations of biochemical pathways. Whole cell modeling integrates information from metabolic pathways, gene regulation, and gene expression. Three elements are needed for constructing of a good cell model: precise knowledge of the phenomenon, an accurate mathematical representation, and a good simulation tool.

A cell represents a dynamic environment of interaction among nucleic acids, proteins, carbohydrates, ions, pH, temperature, pressure, and electrical signals. Many cells with similar functionality form tissue. In addition, each type of tissue uses a subset of this cellular inventory to accomplish a particular function. For example, in neurons, electro-chemical phenomena take precedence over cell division, whereas, cell division is a fundamental function of skin, lymphocytes, and bone marrow cells. Thus, an ideal virtual cell not only represents all the information, but also exhibits the potential to differentiate into neuronal or epithelial cells. The first step in creating a whole cell model is to divide the entire network into pathways, and pathways into individual reactions. Any two reactions belong to a pathway if they share a common intermediate. *In silico* modeling consists not only of decomposing events into manageable units, but also of assembling these units into a unified framework. In other words, mathematical modeling is the art of converting biology into numbers.

For whole cell modeling, a checklist of biological phenomena that call for mathematical representation is needed. Biological phenomena taken into account for *in silico* modeling of whole cells are the following:

1. DNA replication and repair
2. Translation
3. Transcription and regulation of transcription
4. Energy metabolism
5. Cell division
6. Chromatin modeling
7. Signaling pathways
8. Membrane transport (ion channels, pump, nutrients)
9. Intracellular molecular trafficking
10. Cell membrane dynamics
11. Metabolic pathways

The whole cell metabolism includes enzymatic and non-enzymatic processes. Enzymatic processes cover most of the metabolic events, while non-enzymatic processes include gene expression and regulation, signal transduction, and diffusion.

In silico modeling of whole cells not only requires precise qualitative and quantitative data, but also an appropriate mathematical representation of each event. For metabolic modeling, the data input consists of kinetics of individual reactions and also effects of cofactors, pH, and ions on the model. The key step in modeling is to choose an appropriate assumption. For example, a metabolic pathway may be a mix of forward and reverse reactions. Furthermore, inhibitors that are part of the pathway may influence some reactions. At every step, enzymatic equations are needed that best describe the process. *In silico* models are built because they are easy to understand, controllable, and can store and analyze large amounts of information. A well-built model has diagnostic and predictive abilities. A cell by itself is a complete biochemical reactor that contains all the information one needs to understand life. Whole cell modeling enables investigation of the cell cycle, physiology, spatial organization, and cell–cell communication. Sequential actions in whole cell modeling are the following:

1. Catalog all the substances that make up a cell.
2. Make a list of all the reactions, enzymes, and effectors.
3. Map the entire cellular pathway: gene regulation, expression, metabolism, etc.
4. Build a stoichiometric matrix of all the reactions vs. substances (for qualitative modeling).
5. Add rate constants, concentration of substances, and strength of inhibition.
6. Assume appropriate mathematical representations for individual reactions.
7. Simulate reactions with suitable simulation software.
8. Diagnose the system with system analysis software.
9. Perturb the system and correlate its behavior to an underlying genetic and/or biochemical factor.
10. Predict phenomenon using a hypothesis generator.

Advantages and Disadvantages of *In Silico* Modeling

In silico modeling of disease combines the advantages of both *in vivo* and *in vitro* experimentation. Unlike *in vitro* experiments, which exist in isolation, *in silico* models provide the ability to include a virtually unlimited array of parameters, which render the results more applicable to the organism as a whole. *In silico* modeling allows us to examine the workings of biological processes such as homeostasis, reproduction, evolution, etc. For example, one can explore the processes of Darwinian evolution through *in silico* modeling, which are not practical to study in real time.

In silico modeling of disease is quite challenging. Attempting to incorporate every single known interaction rapidly leads to an unmanageable model. Furthermore, parameter determination in such models can be a frightening experience. Estimates come from diverse experiments,

which may be elegantly designed and well executed but can still give rise to widely differing values for parameters. Data can come from both *in vivo* and *in vitro* experiments, and results that hold in one medium may not always hold in the other. Furthermore, despite the many similarities between mammalian systems, significant differences do exist, and so results obtained from experiments using animal and human tissue may not always be consistent. Also there are many considerations that cannot be applied. For example, one cannot investigate the role of stochastic fluctuations by removing them from the system, or one cannot directly explore the process that gave rise to current organisms.

APPLICATIONS OF *IN SILICO* DISEASE MODELING IN PRACTICE

In silico modeling has been applied in cancer, systemic inflammatory response syndrome, immune diseases, neuronal diseases, and infectious diseases (among others). *In silico* models of disease can contribute to a better understanding of the pathophysiology of the disease, suggest new treatment strategies, and provide insight into the design of experimental and clinical trials for the investigation of new treatment modalities.

In Silico Models of Cancer

In silico modeling of cancer has become an interesting alternative approach to traditional cancer research. *In silico* models of cancer are expected to predict the complexity of cancer at multiple temporal and spatial resolutions, with the aim of supplementing diagnosis and treatment by helping plan more focused and effective therapy via surgical resection, standard and targeted chemotherapy, and novel treatments. *In silico* models of cancer include: (a) statistical models of cancer, such as molecular signatures of perturbed genes and molecular pathways, and statistically-inferred reaction networks; (b) models that represent biochemical, metabolic, and signaling reaction networks important in oncogenesis, including constraint-based and dynamic approaches for the reconstruction of such networks; and (c) models of the tumor microenvironment and tissue-level interactions (Edelman et al., 2010).

Statistical models of cancer can be broadly divided into those that employ unbiased statistical inference, and those that also incorporate a priori constraints of specific biological interactions from data. Statistical models of cancer biology at the genetic, chromosomal, transcriptomic, and pathway levels provide insight about molecular etiology and consequences of malignant transformation despite incomplete knowledge of underlying biological interactions. These models are able to identify molecular signatures that can inform diagnosis and treatment selection, for example with molecular targeted therapies such as Imatinib (Gleevec) (Edelman et al., 2010).

However, in order to characterize specific biomolecular mechanisms that drive oncogenesis, genetic and transcriptional activity must be considered in the context of cellular networks that ultimately drive cellular behavior. In microbial cells, network inference tools have been developed and applied for the modeling of diverse biochemical, signaling, and gene expression networks. However, due to the much larger size of the human genome compared to microbes, and the substantially increased complexity of eukaryotic genetic regulation, inference of transcriptional regulatory networks in cancer presents increased practical and theoretical challenges.

Biochemical reaction networks are constructed to represent explicitly the mechanistic relationships between genes, proteins, and the chemical inter-conversion of metabolites within a biological system. In these models, network links are based on pre-established biomolecular interactions rather than statistical associations; significant experimental characterization is thus needed to reconstruct biochemical reaction networks in human cells. These biochemical reaction networks require, at a minimum, knowledge of the stoichiometry of the participating reactions. Additional information such as thermodynamics, enzyme capacity constraints, time-series concentration profiles, and kinetic rate constants can be incorporated to compose more detailed dynamic models (Edelman et al., 2010).

Microenvironment-tissue level models of cancer apply an “engineering” approach that views tumor lesions as complex micro-structured materials, where three-dimensional tissue architecture (“morphology”) and dynamics are coupled in complex ways to cell phenotype, which in turn is influenced by factors in the microenvironment. Computational approaches of *in silico* cancer research include continuum models, discrete models, and hybrid models.

Types of Cancer Models

In **continuum models**, extracellular parameters can be represented as continuously distributed variables to mathematically model cell–cell or cell–environment interactions in the context of cancers and the tumor microenvironment. Systems of partial differential equations have been used to simulate the magnitude of interaction between these factors. Continuum models are suitable for describing the individual cell migration, change of cancer cell density, diffusion of chemo-attractants, heat transfer in hyperthermia treatment for skin cancer, cell adhesion, and the molecular network of a cancer cell as an entire entity. However, these types of *in silico* models have limited ability for investigating single-cell behavior and cell–cell interaction.

On the other hand, “**discrete**” models (i.e. cellular automata models) represent cancer cells as discrete entities of defined location and scale, interacting with one another and external factors in discrete time intervals according to

predefined rules. Agent-based models expand the cellular automata paradigm to include entities of divergent functionalities interacting together in a single spatial representation, including different cell types, genetic elements, and environmental factors. Agent-based models have been used for modeling three-dimensional tumor cell patterning, immune system surveillance, angiogenesis, and the kinetics of cell motility.

Hybrid models have been created which incorporate both continuum and agent-based variables in a modular approach. Hybrid models are ideal for examining direct interactions between individual cells and between the cells and their microenvironment, but they also allow us to analyze the emergent properties of complex multi-cellular systems (such as cancer). Hybrid models are often multi-scale by definition, integrating processes on different temporal and spatial scales, such as gene expression, intracellular pathways, intercellular signaling, and cell growth or migration. There are two general classes of hybrid models, those that are defined upon a lattice and those that are off-lattice.

The classification of hybrid models on these two classes depends on the number of cells these models can handle and the included details of each individual cell structure, i.e. models dealing with large cell populations but with simplified cell geometry, and those that model small colonies of fully deformable cells.

For example, a hybrid model investigated the invasion of healthy tissue by a solid tumor. The model focused on four key parameters implicated in the invasion process; tumor cells, host tissue (extracellular matrix), matrix-degradative enzymes, and oxygen. The model is actually hybrid, wherein the tumor cells were considered to be discrete (in terms of concentrations), and the remaining variables were in the continuous domain in terms of concentrations. This hybrid model can make predictions on the effects of individual-based cell interactions (both between individuals and the matrix) on tumor shape. The model of Zhang et al. (2007) incorporated a continuous model of a receptor signaling pathway, an intracellular transcriptional regulatory network, cell-cycle kinetics, and three-dimensional cell migration in an integrated, agent-based simulation of solid brain tumor development. The interactions between cellular and microenvironment states have also been considered in a multi-scale model that predicts tumor morphology and phenotypic evolution in response to such extracellular pressures.

The biological context in which cancers develop is taken into consideration in *in silico* models of the tumor microenvironment. Such complex tumor microenvironments may integrate multiple factors including extracellular biomolecules, vasculature, and the immune system. However, rarely have these methods been integrated with a large cell–cell communication network in a complex tumor microenvironment. Recently, an interesting effort of *in silico* modeling

was described in which the investigators integrated all the intercellular signaling pathways known to date for human glioma and generated a dynamic cell–cell communication network associated with the glioma microenvironment. Then they applied evolutionary population dynamics and the Hill functions to interrogate this intercellular signaling network and execute an *in silico* tumor microenvironment development. The observed results revealed a profound influence of the micro-environmental factors on tumor initiation and growth, and suggested new options for glioma treatment by targeting cells or soluble mediators in the tumor microenvironment (Wu et al., 2012).

***In Silico* Models and Inflammatory Response Syndrome in Trauma and Infection**

Trauma and infection can cause acute inflammatory responses, the degree of which may have several pathological manifestations like systemic inflammatory response syndrome (SIRS), sepsis, and multiple organ failure (MOF). However, an appropriate management of these states requires further investigation. Translating the results of basic science research to effective therapeutic regimes has been a longstanding issue due in part to the failure to account for the complex nonlinear nature of the inflammatory process wherein SIRS/MOF represent a disordered state. Hence, the *in silico* modeling approach can be a promising research direction in this area. Indeed, *in silico* modeling of inflammation has been applied in an effort to bridge the gap between basic science and clinical trials. Specifically, both agent-based modeling and equation-based modeling have been utilized (Vodovotz et al., 2008). Equation-based modeling encompasses primarily ordinary differential equations (ODE) and partial differential equations (PDE). Initial modeling studies were focused on the pathophysiology of the acute inflammatory response to stress, and these studies suggested common underlying processes generated in response to infection, injury, and shock. Later, mathematical models included the recovery phase of injury and gave insight into the link between the initial inflammatory response and subsequent healing process. The first mathematical models of wound healing dates back to the 1980s and early 1990s. These models and others developed in the 1990s investigated epidermal healing, repair of the dermal extracellular matrix, wound contraction, and wound angiogenesis. Most of these models were deterministic and formulated using differential equations. In addition, recent models have been formulated using differential equations to analyze different strategies for improved healing, including wound VACs, commercially engineered skin substitutes, and hyperbaric oxygen. In addition, agent-based models have been used in wound healing research. For example, Mi et al. (2007) developed an agent-based model to analyze different treatment strategies with wound debridement

and topical administration of growth factors. Their model produced the expected results of healing when analyzing for different treatment strategies including debridement, release of PDGF, reduction in tumor necrosis factor- α , and increase of TGF- β 1. The investigators suggested that a drug company should use a mathematical model to test a new drug before going through the expensive process of basic science testing, toxicology and clinical trials. Indeed, clinical trial design can be improved by prior *in silico* modeling. For example, *in silico* modeling has led to the knowledge that patients who suffered from the immune-suppressed phenotype of late-stage multiple organ failure, and were susceptible to usually trivial nosocomial infections, demonstrated sustained elevated markers of tissue damage and inflammation through two weeks of simulated time. However, anti-cytokine drug trials with treatment protocols of only one dose or one day had not incorporated this knowledge into their design, with subsequent failure of candidate treatments.

Infectious Diseases

By now the reader is expected to be familiar with the meaning and the basics of *in silico* modeling. In this section we discuss the application of *in silico* modeling in the understanding of infectious diseases and in the proposition/development of better treatments for infectious diseases. In fact, the applications of *in silico* modeling can help far beyond just the understanding of the dynamics (and sometimes, statistics) of infectious diseases, and far beyond the proposition/development of better treatments for infectious diseases. The modeling can be helpful even in the understanding of better prevention of infectious diseases.

The level of pathogen within the host defines the process of infection; such pathogen levels are determined by the growth rate of the pathogen and its interaction with the host's immune response system. Initially, no pathogen is present, but just a low-level, nonspecific immunity within the host. On infection, the parasite grows abundantly over time with the potential to transmit the infection to other susceptible individuals.

Triad of Infectious Diseases as the Source of Parameters for In Silico Modeling of Infectious Diseases

To comprehensively understand *in silico* modeling in the domains of infectious diseases, one should first understand the “triad of infectious diseases,” and the characteristics of “infectious agent,” “host,” and “environment” on which the models are always based. In fact, modeling of infectious diseases is just impossible without this triad; after all, the model would be built on some parameters (also called variables in more general language), and those parameters

always have their origin from the so-called “triad of infectious diseases.” At this point, a good question would be: What is a “triad of infectious diseases?”

“Triad of infectious diseases” means the interactions between (1) agent, which is the disease causing organism (the pathogen); (2) host, which is the infected organism, or in the case of pre-infection, the organism to be infected is the host (thus in this case the host is the animal the agent infects); and (3) environment, which is a kind of link between the agent and the host, and is essentially an umbrella word for the entirety of the possible media through which the agent reaches the host.

Now that we have an idea on what *in silico* modeling of infectious diseases are generally based on, we will outline a better understanding of the parameters that are considered in most *in silico* disease models. To discuss the parameters in an orderly manner, we just categorize them under each of the three components of the “triad of infectious diseases,” and summarize them in the next sub-section. It must be emphasized at this point that (1) even though all the possible parameters for *in silico* modeling of infectious diseases can be successfully categorized under the characteristics of one of any of the three components of the “triad of infectious diseases” (agent, host, and environment), (2) the parameters discussed in the next sub-section are by no means the entirety of all the possible parameters that can be included in *in silico* modeling of infectious diseases. In fact, several parameters exist, and this section cannot possibly enumerate them all. That is why we have discussed the parameters using a categorical approach.

Parameters for In Silico Modeling of Infectious Diseases

Parameters Derived from Characteristics of the Agent

Some of the parameters for *in silico* modeling of infectious diseases are essentially a measure of infectivity (ability to enter the host), pathogenicity (ability to cause divergence from homeostasis/disease), virulence (degree of divergence from homeostasis caused/ability to cause death), antigenicity (ability to bind to mediators of the host’s adaptive immune system), and immunogenicity (ability to trigger adaptive immune response) of the concerned infectious agent. The exact measure (and thus the units) used can vary markedly depending on the intentions for which the *in silico* infectious diseases model is built, as well as the assumptions on which the *in silico* disease model is based. From the knowledge of the agent’s characteristics, one should know that unlike parameters related to the other characteristics of the agent, the parameters related to infectivity find their most important use only in the modeling of the pre-infection stage in infectious disease modeling.

Finally, some of the agent-related parameters of great importance in *in silico* modeling of infectious diseases are concentration of the agent’s antigen–host antibody complex, case fatality rate, strain of the agent, other genetic information of the agent, etc.

Parameters Derived from Characteristics of the Host

The parameters originating from characteristics of the host can also be diversified and based on the intentions for which the *in silico* infectious diseases model are built and the assumptions on which the *in silico* disease model are based; however, the parameters could then be grouped and explained under the host’s genotype (the allele at the host’s specified genetic locus), immunity/health status (biological defenses to avoid infection), nutritional status (feeding habits/food intake characteristics), gender (often categorized as male or female), age, and behavior (the host’s behaviors that affect its resistance to homeostasis disruptors).

Typical examples of host-related parameters are the alleles at some specifically targeted genetic loci; the total white blood cell counts; differential white blood cell counts, and/or much more sophisticated counts of specific blood cell types; blood levels of some specific cytokines, hormones, and/or neurotransmitters; daily calories, protein, and/or fat intake; daily amount of energy expended and/or duration of exercise; etc.

Parameters Derived from Characteristics of Environment

At first parameters originating from the environment might seem irrelevant to the *in silico* modeling of infectious diseases, but they are relevant. Even after the pre-infection stage, the environment still modulates the host–agent intersections. For example, the ability (and thus the related parameters) of the agent to multiply and/or harm the host are continually influenced by the host’s environmental conditions, and in a similar way the host’s defense against the adverse effects of the agents are modulated by the host’s environmental conditions. But somehow, not so many of these parameters have been included in *in silico* infectious disease models in the recent past. A few examples of these parameters are the host’s ambient temperature, the host’s ambient atmospheric humidity, altitude, the host’s light–dark cycle, etc.

Infectious Diseases In Silico Model Proper, a Typical Approach/Scenario

Now that we know the parameters for *in silico* infectious disease modeling, the next reasonable question would be “What form does a typical *in silico* infectious disease model take?” So, this sub-section attempts to answer this very important question.

Let us view the *in silico* model as a system of well-integrated functional equations or formulae. Such

well-integrated functional equations can be viewed or approximated as a single, albeit more complex, functional equation/formula. It is hence possible to vary any (or a combination) of the variables contained in this equation by running numerical simulations on a computer depending on the kind of prediction one wants to make. Such *in silico* models can hence investigate several (maybe close to infinite) possible data points within reasonable limits that one sets depending on the nature of the variables considered.

So the equations behind a typical infectious disease *in silico* model could take the form (Equation 21.1):

$$H = \beta \begin{matrix} \text{(link function)} \\ \text{(link function)} \end{matrix} f(A) g(E) + \varepsilon \dots \quad (21.1)$$

where H is the output from a smaller equation that is based on host parameters; β is a constant; f and g are link functions which may be the same as or different from each other and other link functions in this system of equations; A is the output from a smaller equation that is based on agent parameters; g is a link function which may be the same or different from other link functions in this system of equations; E is the output from a smaller equation that is based on environment parameters; and ε is a random error parameter.

Readers should know that we use the term “link function” to refer to any of the various possible forms of mathematical operations or functions. This means that based on the complexity of the model, a particular “link function” might be as simple as a mere addition or as complex as several combinations of operators with high degree polynomials.

H in Equation 21.1 could have resulted from a smaller model/function of the form (Equation 21.2):

$$H = \beta_h \begin{matrix} \text{(link function)} \\ \text{(link function)} \\ \text{(link function)} \end{matrix} f_{h1}(h_1) f_{h2}(h_2) \dots f_{hx}(h_x) + \varepsilon \dots \quad (21.2)$$

where β_h is a constant; f_{h1} , f_{h2} , ... f_{hx} are link functions that may be the same or different (individually) from (every) other link function in this system of equations; h_1 , h_2 , ... h_x are a set of the host's parameters (e.g. age, gender, white blood cell count, cytokine level, etc); and ε is a random error parameter.

A in Equation 21.1 could have resulted from a smaller model/function of the form (Equation 21.3):

$$A = \beta_a \begin{matrix} \text{(link function)} \\ \text{(link function)} \\ \text{(link function)} \end{matrix} f_{a1}(a_1) f_{a2}(a_2) \dots f_{ax}(a_x) + \varepsilon \dots \quad (21.3)$$

where β_a is a constant; f_{a1} , f_{a2} , ... f_{ax} are link functions that may be same or different (individually) from (every) other link function in this system of equations; a_1 , a_2 , ... a_x are a set of the agent's parameters (e.g. case fatality rate, agent's genotype, etc); and ε is a random error parameter.

In a similar way, E in Equation 21.1 could have resulted from a smaller model/function of the form (Equation 21.4):

$$E = \beta_e \begin{matrix} \text{(link function)} \\ \text{(link function)} \\ \text{(link function)} \end{matrix} f_{e1}(e_1) f_{e2}(e_2) \dots f_{ex}(e_x) + \varepsilon \dots \quad (21.4)$$

where β_e is a constant; f_{e1} , f_{e2} , ... f_{ex} are link functions which may be the same or different (individually) from (every) other link function in this system of equations; e_1 , e_2 , ... e_x are a set of environmental parameters (e.g. host's ambient temperature, host's ambient atmospheric humidity, etc.); and ε is a random error parameter.

Specific Examples of Infectious Diseases in the In Silico Model

Muñoz-Elías et al. (2005) documented (through their paper “Replication Dynamics of Mycobacterium Tuberculosis in Chronically Infected Mice”) a successful *in silico* modeling of infectious diseases (specifically, tuberculosis). In their *in silico* modeling of tuberculosis in mice, the researchers investigated both the static and dynamic host–pathogen/agent equilibrium (i.e. mice–mycobacterium tuberculosis static and dynamic equilibrium). The rationale behind their study was that a better understanding of host–pathogen/agent interactions would make possible the development of better anti-microbial drugs for the treatment of tuberculosis (as well as provide similar understanding for the cases of other chronic infectious diseases). They modeled different types of host–pathogen/agent equilibriums (ranging from completely static equilibrium, all the way through semi-dynamic, down to completely dynamic scenarios) by varying the rate of multiplication/growth and the rate of death of the pathogen/agent (Mycobacterium tuberculosis) during the infection's chronic phase. Through their *in silico* study (which was also verified experimentally), they documented a number of remarkable findings. For example, they established that “viable bacterial counts and total bacterial counts in the lungs of chronically infected mice do not diverge over time,” and they explained that “rapid degradation of dead bacteria is unlikely to account for the stability of total counts in the lungs over time because treatment of mice with isoniazid for 8 weeks led to a marked reduction in viable counts without reducing the total count.

Readers who are interested in further details on the generation of this *in silico* model for the dynamics of Mycobacterium tuberculosis infection, as well as the complete details of the parameters/variables considered, and the comprehensive findings of the study, should refer to the article of Ernesto et al. published in infection and immunity.

Another one of the many notable works in the domain of infectious disease *in silico* modeling is the study by Navratil et al. (2011). Using protein–protein interaction

data that the authors obtained from available literature and public databases, they (after first curating and validating the data) computationally (*in silico*) re-examined the virus–human protein interactome. Interestingly, the authors were able to show that the onset and pathogenesis of some disease conditions (especially chronic disease conditions) often believed to be of genetic, lifestyle, or environmental origin, are, in fact, modulated by infectious agents.

Model of Bacterial and Viral Dynamics

Models have been constructed to simulate bacterial dynamics, such as growth under various nutritional and chemical conditions, chemotactic response, and interaction with host immunity. Clinically important models of bacterial dynamics relating to peritoneal dialysis, pulmonary infections, and particularly of antibiotic treatment and bacterial resistance, have also been developed.

Baccam et al. (2006) utilized a series of mathematical models of increasing complexity that incorporated target cell limitation and the innate interferon response. The models were applied to examine influenza A virus kinetics in the upper respiratory tracts of experimentally infected adults. They showed the models to be applicable for improving the understanding of influenza A virus infection, and estimated that during an upper respiratory tract infection, the influenza virus initially spreads rapidly with one cell, infecting (on average) about 20 others (Daun and Clermont, 2007).

Model parameter and spread of disease: Model parameters are one of the main challenges in mathematical modeling since all models do not have a physiological meaning. Sensitivity analysis and bifurcation analysis give us the opportunity to understand how model outcome and model parameters are correlated, how the sensitivity of the system is with respect to certain parameters, and the uncertainty in the model outcome yielded by the uncertainties in the parameter values. Uncertainty and sensitivity analysis was used to evaluate the input parameters play in the basic productive rate (R_0) of severe acute respiratory syndrome (SARS) and tuberculosis. Control of the outbreak depends on identifying the disease parameters that are likely to lead to a reduction in R .

Challenges in *In Silico* Modeling of Infectious Diseases

Difficulty in finding the most appropriate set of parameters for *in silico* modeling of infectious diseases is often a challenge. It is hoped this challenge will subside with the advancement in infectonomics and high-throughput technology. However, another important challenge lies in the understanding (and the provision of reasonable interpretations for) the results from all the complex interactions of parameters considered.

Neuronal Diseases

In this sub-section we focus on the application of *in silico* modeling to improve knowledge of neuronal diseases, and thus improve the applications of neurological knowledge for solving neuronal health problems. It is not an overstatement to say that one of the many aspects of life sciences where *in silico* disease modeling would have the biggest applications is in the better understanding of the pathophysiology of nervous system (neuronal) diseases. This is basically because of the inherent delicate nature of the nervous system and the usual extra need to be sure of how to proceed prior to attempting to treat neuronal disease conditions. By this we mean that the need to first model neuronal disease conditions *in silico* prior to deciding on or suggesting (for example) a treatment plan is, in fact, rising. This is not unexpected; after all, it is better to be sure of what would work (say, through *in silico* modeling) than to try what would not work.

Pathophysiology of Neuronal Diseases as the Source of Parameters for In Silico Modeling of Neuronal Diseases

Obtaining appropriate parameters for the *in silico* modeling of a nervous system (neuronal) disease is rooted in a good understanding of the pathophysiology of such neuronal disease. Since comprehensive details of pathophysiology of neuronal diseases is beyond the scope of this book, we only present the basic idea that would allow the reader to understand how *in silico* modeling of a nervous system (neuronal) disease can be done.

To give a generalized explanation and still concisely present the basic ideas underlying the pathophysiology of neuronal diseases, we proceed by systematically categorizing the mediators of neuronal disease pathophysiology: (1) nerve cell characteristics, (2) signaling chemicals and body electrolytes, (3) host/organism factors, and (4) environmental factors. Readers need to see all these categories as being highly integrated pathophysiologically rather than as separate entities, and also that we have only grouped them this way to make simpler the explanation of how the parameters for *in silico* modeling of neuronal diseases are generated.

When something goes wrong with (or there is a marked deviation from equilibrium in) a component of any of the four categories above, the other components (within and/or outside the same category) try hard to make adjustments so as to annul/compensate for the undesired change. For example, if the secretion of a chemical signal suddenly becomes abnormally low, the target cells for the chemical signal may develop mechanisms to use the signaling chemical more efficiently, and the degradation rate of the signaling chemical may be reduced considerably. Through these, the potentially detrimental effects of reduced secretion of

the chemical signal are annulled via compensation from the other components. This is just a simple example; much more complex regulatory and homeostatic mechanisms exist in the neuronal system. Despite the robustness of those mechanisms, things still get out of hand sometimes, and disease conditions result. The exploration of what happens in (and to) each and all of the components of this giant system of disease conditions is called the pathophysiology of neuronal disease, and it is this pathophysiology that “provides” parameters for the *in silico* modeling of neuronal diseases.

Parameters for In Silico Modeling of Neuronal Diseases

Parameters Derived from Characteristics of Nerve Cells

Some of the important parameters (that are of nerve cell origin) for a typical *in silico* modeling of a neuronal disease (say, Alzheimer’s disease) are the population (or relative population) of specific neuronal cells (such as glial cells: microglia, astrocytes, etc.), motion of specific neuronal cells (e.g. microglia), amyloid production, aggregation and removal of amyloid, morphology of specific neuronal cells, status of neuronal cell receptors, generation/regeneration/degeneration rate of neuronal cells, status of ion neuronal cell channels, etc. Based on their relevance to the pathophysiology of the neuronal disease being studied, many of these parameters are often considered in the *in silico* modeling of the neuronal disease. More importantly, their spatiotemporal dynamics are often seriously considered.

Parameters Derived from Characteristics of Signaling Chemicals and Body Electrolytes

The importance of signaling chemicals and electrolytes in the nervous system makes parameters related to them very important. The secretion, uptake, degradation, and diffusion rates of various neurotransmitters and cytokines are often important parameters in the *in silico* modeling of neurodiseases. Other important parameters are the concentration gradients of the various neurotransmitters and cytokines, the availability and concentration of second messengers, and the electrolyte status/balance of the cells/systems. The spatiotemporal dynamics of all of these are also often seriously considered.

Parameters Derived from Host/Organism Factors

The parameters under host/organism factors can be highly varied depending on the intentions and the assumptions governing the *in silico* disease modeling. Nonetheless, one could basically group and list the parameters collectively under genotype (based on the allele at a specified genetic locus), nutritional status (feeding habits/food intake characteristics; e.g. daily calories, protein intake, etc.), gender (male or female), age, and behavior (host’s behaviors/lifestyle that influences homeostasis and/or responses to stimuli).

Parameters Derived from Environmental Factors

A few examples of these parameters are ambient temperature, altitude, light–dark cycle, social network, type of influences from people in the network, etc.

Neuronal Disease In Silico Model Proper (a Typical Approach/Scenario)

Just like other *in silico* models, a neuronal disease *in silico* model is also based on what could be viewed as a single giant functional equation, which is composed of highly integrated simpler functional equations.

So the equations behind a typical neuronal disease *in silico* model could take the form (Equation 21.5):

$$N = \beta (\text{link function}) f(C) (\text{link function}) g(S) (\text{link function}) j(H) (\text{link function}) k(E) + \varepsilon \dots \quad (21.5)$$

where N could be a parameter that is a direct measure of the disease manifestation; β is a constant; f , g , j , and k are link functions which may be the same or different from other link functions in this system of equations; C , S , H , and E are the outputs from smaller equations that are based on parameters from neuronal cell characteristics, signaling molecule and electrolyte parameters, host parameters, and environment parameters, respectively; and ε is a random error parameter.

The reader should know that each of N , C , S , H and E could have resulted from smaller equations that could take forms similar to those (Equations 21.2 to 21.4) described under *in silico* modeling of infectious diseases (previous sub-section).

Specific Examples of Neuronal Disease In Silico Models

Edelstein-Keshet and Spiros (2002) used *in silico* modeling to study the mechanism and/formation of Alzheimer’s disease. The target of their *in silico* modeling was to explore and demystify how various parts implicated in the etiology and pathophysiology of Alzheimer’s disease work together as a whole. Employing the strength of *in silico* modeling, the researchers were able to transcend the difficulty of identifying detailed disease progression scenarios, and they were able to test a wide variety of hypothetical mechanisms at various levels of detail. Readers interested in the complete details of the assumptions that govern *in silico* modeling of Alzheimer’s disease, the various other aspects of the model, and more detailed accounts of the findings should look at the article by Edelstein-Keshet and Spiros.

Several other interesting studies have applied *in silico* modeling techniques to investigate various neuronal diseases. A few examples include the work of Altmann and Boyton (2004), who investigated multiple sclerosis (a very common disease resulting from demyelination in the central

nervous system) using *in silico* modeling techniques; Lewis et al. (2010), who used *in silico* modeling to study the metabolic interactions between multiple cell types in Alzheimer's disease; and Raichura et al. (2006), who applied *in silico* modeling techniques to dynamically model alpha-synuclein processing in normal and Parkinson's disease states.

Alzheimer's Disease Model

A more specific example of a molecular level *in silico* Alzheimer's disease model can be found in Ghosh et al. (2010). Among the amyloid proteins, Amyloid- β ($A\beta$) peptides ($A\beta_{42}$ and $A\beta_{40}$) are known to form aggregates that deposit as senile plaques in the brains of Alzheimer's disease patients. The process of $A\beta$ -aggregation is strongly nucleation-dependent, and is inferred by the occurrence of a "lag-phase" prior to fibril growth that shows a sigmoidal pattern. Ghosh et al. (2010) dissected the growth curve into three biophysically distinct sections to simplify modeling and to allow the data to be experimentally verifiable. Stage I is where the pre-nucleation events occur whose mechanism is largely unknown. The pre-nucleation stage is extremely important in dictating the overall aggregation process where critical events such as conformation change and concomitant aggregation take place, and it is also the most experimentally challenging to decipher. In addition to mechanistic reasons, this stage is also physiologically important as low-molecular-weight (LMW) species are implicated in AD pathology. The rate-limiting step of nucleation is followed by growth. The overall growth kinetics and structure and shape of the fibrils are mainly determined by the structure of the nucleating species. An important intermediate along the aggregation pathway, called "protofibrils," have been isolated and characterized that have propensities to both elongate (by monomer addition) as well as to laterally associate (protofibril–protofibril association) to grow into mature fibrils (Stage III in the growth curve).

Simulation of the Fibril Growth Process in $A\beta_{42}$ Aggregation Ghosh et al. (2010) generated an ODE-based molecular simulation (using mass–kinetics methodology) of this fibril growth process to estimate the rate constants involved in the entire pathway. The dynamics involved in the protofibril elongation stage of the aggregation (Stage III of the process) were estimated and validated by *in vitro* biophysical analysis.

Preliminary Identification of Nucleation Mass Ghosh et al. (2010) next used the rate constants identified from Stage III to create a complete aggregation pathway simulation (combining Stages I, II, and III) to approximately identify the nucleation mass involved in $A\beta$ -aggregation.

In order to model the $A\beta$ -system, one needs to estimate the rate constants involved in the complete pathway and the nucleation mass itself. It is difficult to iterate through different values for each of these variables to

get close to the experimental plots (fibril growth curves measured via fluorescence measurements with time) due to the large solution space; also, finding the nucleation phase cannot be done independently without estimating the rate constants alongside. However, having separately estimated the post-nucleation stage rate constants (as mentioned above) reduces the overall parameter estimation complexity.

The complete pathway simulation was used to study the lag times associated with the aggregation pathway, and hence predict possible estimates of the nucleation mass. The following strategy was used: estimate the pre-nucleation rate constants that give the *maximum lag times* for each possible estimate of the nucleation mass. This led to four distinctly different regimes of possible nucleation masses corresponding to four different pairs of rate constants for the pre-nucleation phase (Regime 1, where $n = 7, 8, 9, 10, 11$; Regime 2, where $n = 12, 13, 14$; Regime 3, where $n = 15, 16, 17$; and Regime 4, where $n = 18, 19, 20, 21$). However, it was experimentally observed that the semi-log plot of the lag times against initial concentration of $A\beta$ is linear, and this characteristic was used to figure out what values of nucleation mass are most feasible for the $A\beta_{42}$ -aggregation pathway. The simulated plots show a more stable relationship between the lag times and the initial concentrations, and the best predictions for the nucleation mass were reported to be in the range 10–16.

Such molecular pathway level studies are extremely useful in understanding the pathogenesis of AD in general, and can motivate drug development exercises in the future. For example, characterization of the nucleation mass is important as it has been observed that various fatty acid interfaces can arrest the fibril growth process (by stopping the reactions beyond the pre-nucleation stage). Such in depth modeling of the aggregation pathway can suggest what concentrations of fatty acid interfaces should be used (under a given $A\beta$ concentration in the brain) to arrest the fibril formation process leading to direct drug dosage and interval prediction for AD patients.

Possible Limitations of In Silico Modeling of Neuronal Diseases

Despite the fact that we have mentioned several possible parameters for *in silico* modeling of neuro-diseases, it is noteworthy that finding a set of the most reasonable parameters for the modeling is in fact a big challenge. On the other hand, understanding (and thus finding reasonable biological interpretations for) the results from the complex interaction of all parameters considered is also a big challenge. In addition, a number of assumptions that models are sometimes based on still have controversial issues. Accurately modeling spatio-temporal dynamics of neurons and neurotransmitters (and other chemicals/ligands) also constitutes a huge challenge.

CONCLUSION

Understanding the complex systems involved in a disease will make it possible to develop smarter therapeutic strategies. Treatments for existing tumors will use multiple drugs to target the pathways or perturbed networks that show an altered state of activity. In addition, models can effectively form the basis for translational research and personalized medicine.

Biological function arises as the result of processes interacting across a range of spatiotemporal scales. The ultimate goal of the applications of bioinformatics in systems biology is to aid in the development of individualized therapy protocols to minimize patient suffering while maximizing treatment effectiveness. It is now being increasingly recognized that multi-scale mathematical and computational tools are necessary if we are going to be able to fully understand these complex interactions (e.g. in cancer and heart diseases).

With the bioinformatics tools, computational theories, and mathematical models introduced in this article, readers should be able to dive into the exhilarating area of formal computational systems biology. Investigating these models and confirming their findings by experimental and clinical observations is a way to bring together molecular reductionism with quantitative holistic approaches to create an integrated mathematical view of disease progression. We hope to have shown that there are many interesting challenges yet to be solved, and that a structured and principled approach is essential for tackling them.

Systems biology is an emerging field that aims to understand biological systems at the systems level with a high degree of mathematical and statistical modeling. *In silico* modeling of infectious diseases is a rich and growing field focused on modeling the spread and containment of infections with model designs being flexible and enabling adaptation to new data types.

ETHICAL ISSUES

The advantages of avoiding animal testing have often been seen as one of the advantages offered by *in silico* modeling; the biggest advantage is that there are no ethical issues in performing *in silico* experiments as they don't require any animals or live cells. Furthermore, as the entire modeling and analysis are based on computational approaches, we can obtain the results of such analysis even within an hour. This saves huge amounts of time and reduces costs, two major factors associated with *in vitro* studies. However, a key issue that needs to be considered is whether *in silico* testing will ever be as accurate as *in vitro* or *in vivo* testing, or whether *in silico* results will always require non-simulated experimental confirmation.

TRANSLATIONAL SIGNIFICANCE

Tracqui et al. (1995) successfully developed a glioma model to show how chemo-resistant tumor sub-populations cause treatment failure. Similarly, a computational model of tumor

invasion by Frieboes et al. (2006) is able to demonstrate that the growth of a tumor depends on the microenvironmental nutrient status, the pressure of the tissue, and the applied chemotherapeutic drugs. The 3D spatio-temporal simulation model of a tumor by Dionysiou et al. (2004) was able to repopulate, expand, and shrink tumor cells, thus providing a computational approach for assessment of radiotherapy outcomes. The glioblastoma model of Kirby et al. (2007) is able to predict survival outcome post-radiotherapy. Wu et al. (2012) has also developed an *in silico* glioma microenvironment that demonstrates that targeting the microenvironmental components could be a potential anti-tumor therapeutic approach.

The *in silico* model-based systems biology approach to skin sensitization (TNF-alpha production in the epidermis) and risk of skin allergy assessment has been successfully carried out; it can replace well known *in vitro* assays, such as the mouse local lymph node assay (LLNA) used for the same purpose (by Maxwell and Mackay (2008) at the Unilever Safety and Environmental Assurance Centre). Similarly, Davies et al. (2011) effectively demonstrated an *in silico* skin permeation assay based on time course data for application in skin sensitization risk assessment. Kovatchev et al. (2012) showed how the *in silico* model of alcohol dependence can provide virtual clues for classifying the physiology and behavior of patients so that personalized therapy can be developed.

Pharmacokinetics and pharmacodynamics are used to study absorption, distribution, metabolism, and excretion (ADME) of administered drugs. *In silico* models have tremendous efficacy in early estimation of various ADME properties. Quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) models have been commonly used for several decades to predict ADME properties of a drug at early phases of development. There are several *in silico* models applied in ADME analysis, and readers are encouraged to read the review by van de Waterbeemd and Gifford (2003). GastroPlus™, developed at Simulations Plus (www.simulations-plus.com), is highly advanced, physiologically based rapid pharmacokinetic (PBPK) simulation software that can generate results within 5 seconds, thus saving huge amounts of time and cost in clinical studies. The software is an essential tool to formulation scientists for *in vitro* dose disintegration and dissolution studies. Towards next-generation treatment of spinal cord injuries, Novartis (www.novartis.com) is working to model the human spinal cord and its surrounding tissues *in silico* to check the feasibility of monoclonal antibody-based drug administration and their pharmacokinetics and pharmacodynamics study results.

The *in silico* “drug re-purposing” approach by Bisson et al. (2007) demonstrated how phenothiazine derivative antipsychotic drugs such as acetophenazine can cause endocrine side effects. Recently Aguda et al. (2011) reported a computational model for sarcoidosis dynamics that is useful for pre-clinical therapeutic studies for assessment of dose optimization of targeted drugs used to treat sarcoidosis.

Towards designing personalized therapy of larynx injury leading to acute vocal fold damage, Li et al. (2008) developed agent-based computational models.

In a further advancement, Entelos® (www.entelos.com) has developed “virtual patients,” *in silico* mechanistic models of type-2 diabetes, rheumatoid arthritis, hypertension, and atherosclerosis for identification of biomarkers, drug targets, development of therapeutics, and clinical trial design, and patient stratification. Entelos’ virtual Idd9 mouse (NOD mouse) can replace diabetes resistance type-1 diabetes live mice for various *in vivo* experiments.

Apart from diseases, systems level modeling of basic biological phenomena and their applications in disease have also been reported. An *in silico* model to mimic the *in vitro* rolling, activation, and adhesion of individual leukocytes has been developed by Tang et al. (2007). Developing virtual mitochondria, Cree et al. (2008) have demonstrated how the *in silico* modeling can assist in predicting the severity of mitochondrial diseases. Sobie and Wehrens, 2010 have developed *in silico* cardiac models and demonstrated how calcium mediated arrhythmias can develop. Interested readers in computational cardiology can consult with review of Trayanova, 2012.

WORLD WIDE WEB RESOURCES

Virus Pathogen Database and Analysis Resource (ViPR)

ViPR (<http://www.viprbrc.org/brc/home.do?decorator=vipr>) is one of the five Bioinformatics Resource Centers (BRCs) funded by the National Institute of Allergy and Infectious Diseases (NIAID). This website provides a publicly available database and a number of computational analysis tools to search and analyze data for virus pathogens. Some of the tools available at ViPR are the following:

1. GATU (Genome Annotation Transfer Utility), a tool to transfer annotations from a previously annotated reference to a new, closely related target genome.
2. PCR Premier Design, a tool for designing PCR primers.
3. A sequence format conversion tool.
4. A tool to identify short peptides in proteins.
5. A meta-driven comparative analysis tool.

As there are many different kinds of tools available the tools on the website are organized by the virus family.

The Rat Genome Database (RGD)

The Rat Genome Database (<http://rgd.mcw.edu/wg/home>) is funded by the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH). The goal of this project is to consolidate research work from various institutes to generate and maintain a rat genomic database (and make it available to the scientific community). The website provides a variety of tools to analyze data.

Influenza Research Database (IRD)

IRD (<http://www.fludb.org/brc/home.do?decorator=influenza>) is one of the five Bioinformatics

Resource Centers (BRCs) funded by the National Institute of Allergy and Infectious Diseases (NIAID). This website provides a publicly available database and a number of computational analysis tools to search and analyze data for influenza virus. This website provides many of the same tools that are provided at ViBR. There are numerous other tools such as Models of Infectious Disease Agent Study (MIDAS), which is an *in silico* model for assessing infectious disease dynamics. MIDAS assists in preparing, detecting, and responding to infectious disease threats.

The Wellcome Trust Sanger Institute

The Sanger Institute (<http://www.sanger.ac.uk/>) investigates genomes in the study of diseases that have an impact on global health. The Sanger Institute has made a significant contribution to genomic research and developing a new understanding of genomes and their role in biology. The website provides sequence genomes for various bacterial, viral, and model organisms such as zebrafish, mouse, gorilla, etc. A number of open source software tools for visualizing and analyzing data sets are available at the Sanger Institute website.

REFERENCES

- Aguda, B. D., Marsh, C. B., Thacker, M., & Crouser, E. D. (2011). An *in silico* modeling approach to understanding the dynamics of sarcoidosis. *PLoS One*, 6, (5)<http://dx.doi.org/10.1371/journal.pone.0019544>. e19544.
- Altmann, D., & Boyton, R. (2005). Models of multiple sclerosis. Auto-immune diseases., *Drug Discovery Today. Disease Models*, 11, 405–410.
- Baccam, P., Beauchemin, C., Macken, C. A., Hayden, F. G., & Perelson, A. S. (2006). Kinetics of influenza A virus infection in humans. *Journal of Virology*, 80, 7590–7599.
- Bisson, W. H., Cheltsov, A. V., Bruey-Sedano, N., Lin, B., Chen, J., Goldberger, N., May, L. T., Christopoulos, A., Dalton, J. T., Sexton, P. M., Zhang, X. K., & Abagyan, R. (2007). Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. *Proceedings of the National Academy of Sciences of the United States of America*, 104(29), 11927–11932.
- Chassignole, C., Jackson, R. C., et al. (2006). Using a mammalian cell cycle simulation to interpret differential kinase inhibition in anti-tumour pharmaceutical development. *Biosystems*, 83(2-3), 91–97.
- Chakraborty, Arup K., Dustin, Michael L., & Shaw, Andrey S. (2003). *In silico* models for cellular and molecular immunology: successes, promises and challenges. *Nature immunology*, 4, 10, 933–936.
- Choi, S. (2007). *Introduction to Systems Biology*. Humana Press Inc. SBN 9781597455312.
- Cree, L. M., Samuels, D. C., de Sousa Lopes, S. C., Rajasimha, H. K., Wonnapijit, P., Mann, J. R., Dahl, H. H., & Chinnery, P. F. (2008). A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nature genetics*, 40(2), 249–254.
- Danchin, A., et al. (1991). From data banks to data bases. *Research in microbiology*, 142.7, 913–916.
- Daun, S., & Clermont, G. (2007). *In Silico* modeling in infectious disease. *Drug Discov. Today Dis. Models*, 4(3), 117–122.

- Davies, M., Pendlington, R. U., Page, L., Roper, C. S., Sanders, D. J., Bourner, C., Pease, C. K., & MacKay, C. (2011; Feb). Determining epidermal disposition kinetics for use in an integrated nonanimal approach to skin sensitization risk assessment. *Toxicological Sciences*, 119(2), 308–318. <http://dx.doi.org/10.1093/toxsci/kfq326>.
- Dionysiou, D. D., Stamatakos, G. S., Uzunoglu, N. K., Nikita, K. S., & Marioli, A. (2004; Sep 7). A four-dimensional simulation model of tumour response to radiotherapy *in vivo*: parametric validation considering radiosensitivity, genetic profile and fractionation. *Journal of Theoretical Biology*, 230(1), 1–20.
- Edelman, L. B., Eddy, J. A., & Price, N. D. (2010). *In silico* models of cancer. *Wiley Interdiscip. Rev. Syst. Biol. Mol.* 4, 438–459.
- Enderling, H., Anderson, A. R. A., et al. (2006). Mathematical modeling of radiotherapy strategies for early breast cancer. *Journal of Theoretical Biology*, 241(1), 158–171.
- Edelstein-Keshet, L., & Spiros, A. (2002). Exploring the Formation of Alzheimer's Disease Senile Plaques *in Silico*. *Journal of Theoretical Biology*, 216, 301–326.
- Ekins, S., Mestres, J., & Testa, B. (2009). *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British Journal of Pharmacology*, 152.1, 9–20.
- Fadiel, A., Anidi, I., et al. (2005). Farm animal genomics and informatics: an update. *Nucleic Acids Research*, 33(19), 6308–6318.
- Frieboes, H. B., Zheng, X., Sun, C. H., Tromberg, B., Gatenby, R., & Cristini, V. (2006 Feb 1). An integrated computational/experimental model of tumor invasion. *Cancer Research*, 66(3), 1597–1604.
- Ghosh, P., Kumar, A., Datta, B., & Rangachari, V. (2010). Dynamics of protofibril elongation and association involved in Abeta42 peptide aggregation in Alzheimer's disease. *BMC Bioinformatics*, 6 (11 Suppl), S24.
- Huang da, W., Sherman, B. T., et al. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1), 1–13.
- Kiran, K. L., Jayachandran, D., et al. (2009). Mathematical modeling of avascular tumour growth based on diffusion of nutrients and its validation. *Canadian Journal of Chemical Engineering*, 87(5), 732–740.
- Kovatchev, B., Breton, M., & Johnson, B. (2012). *In silico* models of alcohol dependence and treatment. *Front Psychiatry*, 3, 4.
- Lewis, N. E., Schramm, G., Bordbar, A., Schellenberger, J., Andersen, M. P., Cheng, J. K., Patel, N., Yee, A., Lewis, R. A., Eils, R., & König, R. A., PalssonBØ. (2010). Large-scale *in silico* modeling of metabolic interactions between cell types in the human brain. *Nature Biotechnology*, 28, 1279–1285.
- Li, N. Y., Verdolini, K., Clermont, G., Mi, Q., Rubinstein, E. N., Hebda, P. A., & Vodovotz, Y. (2008; Jul 30). A Patient-Specific *in silico* Model of Inflammation and Healing Tested in Acute Vocal Fold Injury. *PLoS One*, 3(7), e2789.
- Maxwell, G., & Mackay, C. (2008 Nov). Application of a systems biology approach to skin allergy risk assessment. *Alternatives to Laboratory Animals*, 36(5), 521–556.
- Mi, Q., Rivière, B., Clermont, G., Steed, D. L., & Vodovotz, Y. (2007). Agent-based model of inflammation and wound healing: insights into diabetic foot ulcer pathology and the role of transforming growth factor- β 1. *Wound Repair Regeneration*, 15, 671–682.
- Muñoz-Eliás, E. J., Timm, J., Botha, T., Chan, W. T., Gomez, J. E., & McKinney, J. D. (2005). Replication dynamics of mycobacterium tuberculosis in chronically infected mice. *Infection and Immunity*, 73(1), 546–551.
- Navratil, V., de Chassey, B., Combe, C. R., & Lotteau, V. (2011). When the human viral infectome and disease networks collide: towards a systems biology platform for the aetiology of human diseases. *BMC Systems Biology*, 5, 13.
- Raichura, A., Valia, S., & Gorinb, F. (2006). Dynamic modeling of alpha-synuclein aggregation for the sporadic and genetic forms of Parkinson's disease. *Neuroscience*, 142(3), 859–870.
- Sieburg, Hans B. (1990). Physiological studies *in silico*. *Studies in the Sciences of Complexity*, 12.2, 321–342.
- Sobie, A. E., & Wehrens, Z. H. T. (2010). Computational and experimental models of Ca²⁺-dependent arrhythmias. *Drug Discov Today Dis Models*, 6, 5.
- Tang, J., Ley, K. F., & Hunt, C. A. (2007; Feb 19). Dynamics of *in silico* leukocyte rolling, activation, and adhesion. *BMC Systematic Biology*, 1, 14.
- Tomita, M. (2001). Whole-cell simulation: a grand challenge of the 21st century. *Trends of Biotechnology*, 19(6), 205–210.
- Tracqui, P., Cruywagen, G. C., Woodward, D. E., Bartoo, G. T., Murray, J. D., & Alvord, E. C., Jr (1995 Jan). A mathematical model of glioma growth: the effect of chemotherapy on spatio-temporal growth. *Cell Prolif*, 28(1), 17–31.
- Trayanova, N. A. (2012). Computational cardiology: The heart of the matter. *ISRN Cardiol*, 2012, 269680.
- Tuncbag, N., Kar, G., Keskin, O., Gursoy, A., & Nussinov, R. (2009 May). A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Brief Bioinform*, 10(3), 217–232.
- van de Waterbeemd, H., & Gifford, E. (2003 Mar). ADMET *in silico* modelling: towards prediction paradise? *Nat Rev Drug Discov*, 2(3), 192–204.
- Vodovotz, Y., Csete, M., Bartels, J., Chang, S., & An, G. (2008). Translational systems biology of inflammation. *PLoS Computational Biology*, 4, e1000014.
- Werner, T. (2008). Bioinformatics applications for pathway analysis of microarray data. *Current Opinion in Biotechnology*, 19(1), 50–54.
- Woolf, P. J., Prudhomme, W., et al. (2005). Bayesian analysis of signaling networks governing embryonic stem cell fate decisions. *Bioinformatics*, 21(6), 741–753.
- Wu, Y., Lu, Y., Chen, W., Fu, J., & Fan, R. (2012 Feb). *In silico* experimentation of glioma microenvironment development and anti-tumor therapy. *PLoS Computing Biology*, 8(2), e1002355.
- Zhang, Y., Athale, C. A., & Deisboeck, T. S. (2007). Development of a three-dimensional multiscale agent-based tumor model: simulating gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer. *Journal of Theoretical Biology*, 244, 96–107.

FURTHER READING

- Bock, Gregory, & Goode, Jamie (Eds.), (2002). *'In Silico' Simulation of Biological Processes* (Vol. 247). Wiley.
- Cronin, Mark TD., & Madden, Judith C. (2010). *Silico Toxicology: Principles and Applications* (Vol. 7). Royal Society of Chemistry.
- Deisboeck, Thomas S., & Stamatakos, Georgio S. (2010). *Multiscale Cancer Modeling* (Vol. 34). CRC Press LLC.
- Flower, Darren R., Timmis, Jon, & Timmis, Jonathan (Eds.), (2007). *In Silico Immunology*. Springer.
- Sharpe, Jason, John Lumsden, Charles, & Woolridge, Nicholas (2008). *In Silico: 3D Animation and Simulation of Cell Biology with Maya and MEL*. Morgan Kaufmann.

GLOSSARY

Algorithm Any well-defined computational procedure that takes some values, or set of values, as input, and produces some value, or set of values, as output.

Bioinformatics Bioinformatics is the application of statistics and computer science to the field of molecular biology.

Biotechnology The exploitation of biological processes for industrial and other purposes.

Data Structures A way to store and organize data on a computer in order to facilitate access and modifications.

Genome The complete set of genetic material of an organism.

Genomics The branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes.

Gene Ontology A major bioinformatics initiative to unify the representation of gene and gene product attributes across all species.

Informatics The science of processing data for storage and retrieval; information science.

In Silico *In silico* is an expression used to mean “performed on a computer or via computer simulation.”

In Vivo In microbiology *in vivo* is often used to refer to experimentation done in live isolated cells rather than in a whole organism.

In Vitro *In vitro* studies in experimental biology are those that are conducted using components of an organism that have been isolated from their usual biological surroundings in order to permit a more detailed or more convenient analysis than can be done with whole organisms.

Kinomics Kinomics is the study of kinase signaling within cellular or tissue lysates.

Oncogenesis The progression of cytological, genetic, and cellular changes that culminate in a malignant tumor.

Pathophysiology The disordered physiological processes associated with disease or injury.

Proteomics The branch of genetics that studies the full set of proteins encoded by a genome.

Sequencing The process of determining the precise order of nucleotides within a DNA molecule.

Systems Biology An inter-disciplinary field of study that focuses on complex interactions within biological systems by using a more holistic perspective.

mRNA Messenger Ribonucleic Acid

ODE Ordinary Differential Equations

P Protein

PCR Polymerase Chain Reaction

PDE Partial Differential Equations

PDGF Platelet-Derived Growth Factor

PPI Protein–Protein Interaction

RGD Rat Genome Database

RNA Ribonucleic Acid

SARS Severe Acute Respiratory Syndrome.

SIRS Systematic Inflammatory Response System

TG Template Genome

TN Template Network

ViPR Virus Pathogen Database and Analysis Resource

LONG ANSWER QUESTIONS

1. Explain the role of bioinformatics in animal biotechnology.
2. Explain the common computational methods in systems biology.
3. Explain the concept of *in silico* modeling.
4. Discuss the advantages, disadvantages, and ethical issues of *in silico* modeling.
5. What are the different application areas of *in silico* modeling? Discuss in detail how *in silico* modeling is applied in one application area.

SHORT ANSWER QUESTIONS

1. Describe the template-based methods to reconstruct Transcriptional Regulatory Networks.
2. What is the goal of *in silico* modeling?
3. What are the challenges in *in silico* modeling of infectious diseases?
4. What are the three types of cancer models discussed in the chapter?
5. Discuss the parameters considered for *in silico* modeling of infectious diseases.

ABBREVIATIONS

AD Alzheimer Disease

ARACNE Algorithm for the Reconstruction of Accurate Cellular Networks

BRC Bioinformatics Resource Center

CDKs Cyclin-Dependent Kinases

DNA Deoxyribonucleic acid

DZM Different Zone Model

GNW GeneNetWeaver

GO Gene Ontology

GoI Genome of Interest

IRD Influenza Research Database

LMW Low Molecular Weight

MOF Multi Organ Failure

NIAID National Institute of Allergy and Infectious Diseases

MIDAS Models of Infectious Disease Agent Study

ANSWERS TO SHORT ANSWER QUESTIONS

1. The template-based transcriptional control network reconstruction method exploits the principle that orthologous proteins regulate orthologous target genes. In this approach, regulatory interactions are transferred from a genome (such as a genome of a model organism or well studied organism) to the new genome.
2. The ultimate goal of *in silico* modeling in biology is the detailed understanding of the function of molecular networks as they appear in metabolism, gene regulation, or signal transduction.

3. There are two major challenges in modeling infectious diseases:
 - a. Difficulty in finding the most appropriate set of parameters for the *in silico* modeling of infectious diseases is often a challenge.
 - b. Understanding the results from all the complex interactions of parameters considered.
4. There are three types of cancer models. **Continuum models:** In these models extracellular parameters can be represented as continuously distributed variables to mathematically model cell–cell or cell–environment interactions in the context of cancers and the tumor microenvironment. **Discrete models:** These models represent cancer cells as discrete entities of defined location and scale, interacting with one another and external factors in discrete time intervals according to predefined rules. **Hybrid models:** These models incorporate both continuum and discrete variables in a modular approach.
5. There are three types of parameters considered for *in silico* modeling of infectious diseases:
 - a. **Parameters derived from characteristics of agent:** Examples: concentration of the agent’s antigen–host antibody complex; case fatality rate; strain of the agent; other genetic information of the agent; etc.
 - b. **Parameters derived from characteristics of host:** Examples: the total white blood cell counts; differential white blood cell counts, and/or much more sophisticated counts of specific blood cell types; blood levels of some specific cytokines, hormones, and/or neurotransmitters; daily calories, protein, and/or fat intake; daily amount of energy expended and/or duration of exercise; etc.
 - c. **Parameters derived from characteristics of environment:** Examples: host’s ambient temperature; host’s ambient atmospheric humidity; altitude; host’s light–dark cycle; etc.