

Pathway modeling and simulation analysis

Gitanjali Tandon¹, Sunita Yadav¹ and Sukhdeep Kaur²

¹Centre for Agricultural Bioinformatics (CABin), ICAR—Indian Agricultural Statistics Research Institute (IASRI), New Delhi, India, ²Department of Computational Biology and Bioinformatics, Sam Higginbottom University of Agriculture, Technology and Sciences (SHUATS), Prayagraj, India

24.1 Introduction

Systems biology is an emerging field of life sciences that has evinced interest among mathematicians, chemists, physicists, and many researchers who are extensively involved in computer modeling. Although it is an emerging field, one can easily find a lot of literature related to system biology as it has been attracted by many fields of science. After studying it in detail, one can easily conclude that it is fully dedicated to the study of systems. Thus it can be stated that “System biology is a discipline that is involved in the study of systems related to life sciences.” It involves the extensive usage of computational technology so that clear and reliable interpretations can be made with the high-throughput data. Earlier, the systems biology studies were based on their static behaviors, but now these studies have been emphasized their dynamic behaviors (Alon, 2019; Kitano, 2002).

For a deeper understanding of biological systems, the following properties of the system are to be considered:

1. Systems structure: gene—gene interaction/protein—protein interactions, intercellular mechanism, and biochemical pathways.
2. Systems dynamics: dynamic, sensitivity, and metabolic analysis for studying systems behavior.
3. Control methods: modulation behavior that regulates cell states under controlled conditions.
4. Design method: construction of system with desired properties and carry out simulations to study the effects.

By studying the system biology, it helps in:

- a better understanding of the biochemical structure and the network architecture of the system,
- quantitative as well as the qualitative dynamic behavior of the system,
- predicting the control points of the system, and
- knowledge of the design of methodologies for the system.

Being so insightful helps in the understanding of the interactions and functionality of the complex systems (Ross & Arkin, 2009). Biological units are considered as complex and holistic structures in systems biology, whose behavior cannot be diminished to a linear sum of the functions (Palsson, 2015). To understand the biological complexity in a better way, computer modeling and simulation could aid in better understanding of the experimental data set as well as the ability to generate and test the hypothesis for them (Macklin, Ruggero, & Covert, 2014). Given the enormous complexity and special properties of these systems require a thorough understanding of the specific modeling demands to elucidate what a good systems biology model should look like. For accurate modeling and simulation of the metabolic pathways, a prerequisite is pathway landscaping. Despite so much availability of the annotated genomes, large information about the biological pathways is still missing. So, to fill these knowledge gaps, system biology will play an important role as pathway study will help in mining the actual functioning of the system as a whole (Raman, Rajagopalan, & Chandra, 2006).

24.2 Computational modeling of a pathway

The modeling of biological systems must include some important aspects. How these aspects are considered is dependent on the context. Several biological phenomena correspond better to discrete and qualitative information, while in

other cases, biological components are characterized by continuous quantities. Thus an ideal model should be able to treat both continuous and discrete variables as quantitative and qualitative information.

24.2.1 Type of modeling

In general, models of biological systems can be distinguished mathematically as well as computationally. A computational model is a formal model in which primary semantics are operational. In other words, the model prescribes a sequence of steps or instructions that can be executed by an abstract machine implemented on a real computer. A mathematical model is a formal model whose primary semantics are denotation; that is, the model uses equations to describe a relationship between quantities and how they change over time (Fisher & Henzinger, 2007). However, this criterion is not exactly that is followed. Simulation of the mathematical models is based on the algorithm selection and the computational power is used to solve the model. Knowledge about the computational models can be obtained by running and analyzing them. Mathematical models can be used as a reference for gaining information and can be simulated as well as analyzed.

Mathematical and computational models' classification is broadly based on similar opposite characteristics; both of them may be quantitative or qualitative, continuous or discrete, deterministic or stochastic. They are further classified as ordinary differential equations (ODEs), partial differential equations (PDEs), or stochastic differential equations (Fig. 24.1) (Bardini, Politano, Benso, & Di Carlo, 2017). Using the above approaches, biological computational models can be formulated. These models could be elicited from Petri nets, Boolean nets, Interactive state machines, rule-based systems process algebra, or on state graphics. If we consider the spatial representation of the system, these models can be based on compartments, agents, or grids (Machado et al., 2011).

24.2.2 Approaches of modeling

Systems biology models should handle several types of illustrations to design the whole system or its subparts in a complicated hierarchical structure. They should also be able to process different types of information that are represented in different forms. A specific class of models is required, commonly referred to as hybrid and multilevel models. Multilevel models represent the system with at least two different levels. Interactions take place between these levels (Uhrmacher, Degenring, & Zeigler, 2005). Multilevel models represent the “bottom-up” and “top-down” relationships. Bottom-up relationships shape the fact that the system is somewhat limited by the behavior of its parts. Concurrently, top-down relationships model how the performance of each party is influenced through the conduct of the whole system.

While examining multilevel models, it becomes very important to design explicit representations that distinguish between the concept of level and the concept of scale (Gil-Quijano, Louail, & Hutzler, 2010). More particularly, the idea of scale denotes a quantifiable dimension for the study of the phenomenon being considered. These dimensions

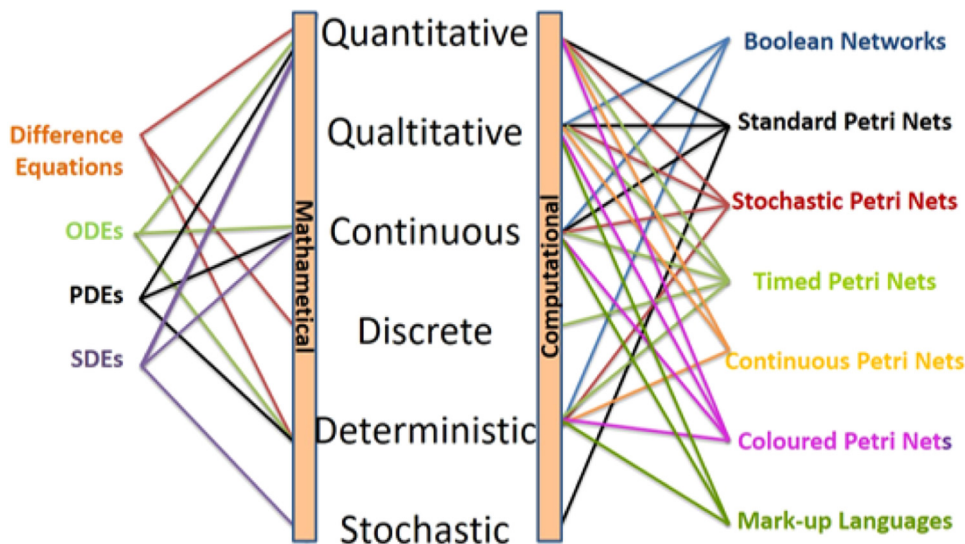


FIGURE 24.1 Types of modeling with their main features.

could be temporal, quantitative, and spatial. The spatial dimension signifies the magnitude of the entities that are participating in the phenomenon, while the temporal dimension refers to the time allied with the behavior of all the entities as well as their relations. Instead, the measurable dimension states the number of entities that are occupied in the phenomenon. A level generally resembles all the biological entities whose characteristic evolutionary time and size have an identical or similar order of magnitude. Such as, a system can be depicted at atomic, molecular, cellular, organic, and population stage. Methods according to their granularity are depicted in Fig. 24.2 (National Research Council, 2006).

The concept of multilevel models could be merged with the concept of hybrid models. As stated by, Stéphanou, a hybrid as in most common definition “The model represents to the interaction or the coupling within two or additional models which is not based upon the alike formalism” (Stéphanou & Volpert, 2016). When creating a hybrid and multi-level model, the formalities to be described must be selected, in addition to the selection of interesting organizational levels in the various components of the general structure of the model. In that way, it can be useful for a quick review of the most common forms used for modeling of biological systems such that their strengths, as well as limitations, could be considered while choosing hybrid blends for several organizational stages that are to be constructed.

Hybrid models are the models that consist of objects that have altered mathematical representations. These different mathematical models allow the modeler with the flexibility to blend modeling paradigms to define different parts of the complex systems. Let us consider a hybrid model with a signal transduction pathway that can be defined by a certain set of differential equations. This pathway can be connected to the graphical model of the genetic regulatory network that it affects. The main benefit of the hybrid based model is that model modules can be developed from abstract high-level descriptions to detailed low-level descriptions, as these components are well characterized and very well understood (National Research Council, 2006).

McAdams and Shapiro (1995) provide an example of the use of hybrid models, which indicates that genetic networks comprising a large number of genes (generally more than 10) are tough to analyze. The hybrid modeling approach assimilates conventional biochemical kinetic modeling with the circuit simulation framework. Diagrammatic representation of the circuit represents the connectivity in the signal pathways with the biochemical components. The operons act as an active integrated logical component, introducing signal time delays that are vital to the in vivo behavior of the lambda phage, which is the key feature of the lambda genetic circuit.

24.3 General diagram and language used in pathway modeling

24.3.1 Systems Biology Graphical Notation

During studying the pathway modeling, graphical representation helps in a better understanding of the system. In system biology, the pathways are represented under Systems Biology Graphical Notation (SBGN). SBGN is considered as

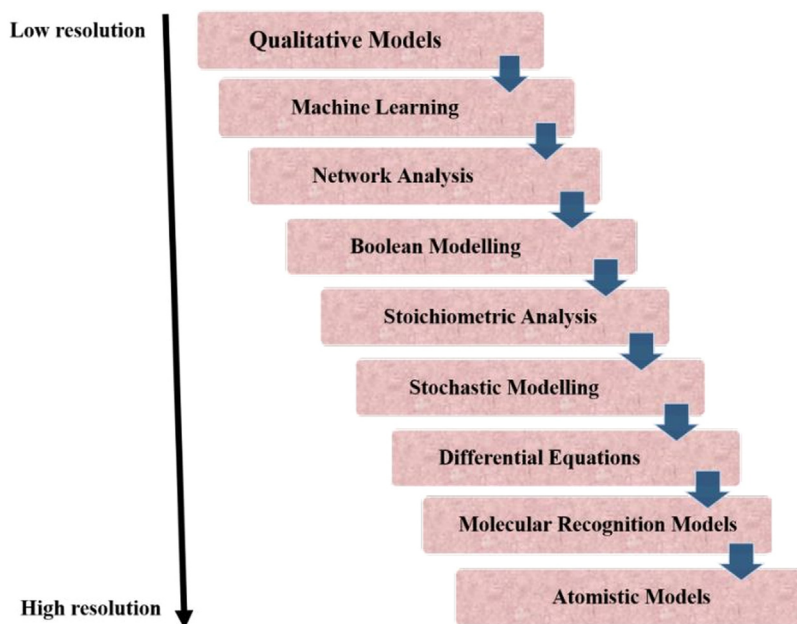


FIGURE 24.2 Modeling techniques used in systems biology. Several methods have been depicted according to their granularity (or resolution).

a standard graphical representation that is designed to promote exchange, efficient storage, and reprocess of information through signaling pathways, metabolic networks, and gene regulatory networks between communities of biochemists, biologists, and theoretical. The system was developed over several years by a community of biochemists, modelers, and computer scientists (Le Novère et al., 2009). SBGN comprises three orthologous languages for the representation of various views for biological systems:

- process descriptions (PDs)
- entity relationships
- activity flows.

Using these three notations various networks of biochemical interactions can be studied.

1. PD language:

PD language depicts the temporal courses for the network of biological interactions. It shows the complete set of molecular interactions that are taking place between the biological entities. The same entity appearing with multiple lines in the same diagram represents it as an important entity (Moodie, Le Novère, Demir, Mi, & Villegier, 2011).

2. Entity relationship language:

This language allows in studying all the relationships or the participation for a given entity (Le Novère, Moodie, Sorokin, Schreiber, & Mi, 2010).

3. Activity flow language:

Activity flow language indicated the flow of information among various biological entities (Mi, Schreiber, Le Novère, Moodie, & Sorokin, 2009).

24.3.2 Systems Biology Markup Language

Earlier the major drawback of various pathway modeling and simulation tools was the inability to exchange the model between different platforms. To address this problem, a Software Platforms for Systems Biology forum under ERATO Kitano Systems Biology Project was formed. During the first meeting of this forum in April 2000, it was decided to develop a simple Extensible Markup Language (XML)-based language for the representation and the exchange of the system biology models. This XML language came to know as System Biology Markup Language (SBML). As a result, during the 2nd Workshop on Software Platforms for Systems Biology in August 2000, a draft of SBML was presented. After various revisions and discussions, version 1 of SBML level 1 was produced in March 2001.

SBML level 2 was introduced by the University of Hertfordshire, UK, in July 2002 at the 5th Workshop on Software Platforms for Systems Biology. Thereafter, in September 2006 version 2 of SBML level 2 was issued. Till these versions, there were three main editors for SBML language were Michael Hucka, Andrew Finney, and Nicolas Le Novère. For the version 3 edition, two more editors were added. After various versions of SBML level 2, SBML level 3 version 1 was released in 2010. The latest version 5 for SBML level 2 is available that has significant changes over SBML level 2 version 4. This XML language obtained wide acceptance as it is free and due to its portability. Currently it is considered as a standard language for the representation of the computational models in system biology (Hucka et al., 2004).

Some important tags that are used in SBML are listed below:

- **Body:**
Defines the detailed information of the model.
- **Function:**
List of defined functions.
- **Compartments:**
It represents the bounded volume where species/entities are located.
- **Reactant species:**
These represent the entities, such as ions or molecules, that take part in the reactions.
- **Reactions:**
A reaction represents the transformation, a biological process, and in general, a chemical reaction. It depicts the change of species and, in SBML, they represent the list of reactant species and their products.
- **Parameters:**
The parameter element in SBML is used to associate a name with a floating point value so that the name can be used in formulas instead of value.

- Unit definitions:
Some basic units are included in SBML by default but it also provides the provision of defining new and redefining the units (Hucka et al., 2003).

Steps of modeling in SBML:

- to create a biochemical pathway network,
- setting ODEs,
- assigning values for each object of the model, and
- using simulators for equation solving (Chaouiya et al., 2013).

24.4 Pathway simulations analysis

Biological pathways represent the series of molecular interactions inside a cell under different conditions. So, the study of these pathways acts as a crucial step for a better understanding of the biological processes. To achieve this various experimental approaches have been collaborated with computational methods to reconstruct these pathways (De Jong, 2002). For a better understanding of the biological processes, the dynamic behavior of biological systems at the pool (a pool can be defined as a set of molecules that cannot be distinguished from each other) level of the molecular species needs to be identified. A network is formed by the interconnection of the various pools joined together via biochemical reactions. So, studying the dynamic behavior of these pools would be of great advantage for a better understanding of the biological systems.

The foremost step for modeling and simulating metabolic pathways is to construct an accurate landscape description of the pathways. Mathematical models are developed by collecting the model components and their interaction information. In the first step, the topological structure of the network and their reaction stoichiometries are designed (Visser & Heijnen, 2003). After the reaction network and its stoichiometry are defined, the detailed mathematical model can be constructed. Generally, ODEs are utilized for this construction purpose. For all this, detailed information about the kinetics of all the reactions in the pathway is required. If no such information is available, then appropriate assumptions are done and models are designed accordingly (Mendes & Kell, 1998).

Apart from this, various simulation techniques are used in system biology, such as stochastic methods, Boolean models, Petri nets, cellular automata, agent-based systems, PDEs, and hybrid approaches. Among all these methods, ODEs are the most common and have been widely used in computational systems biology.

In system biology, computers have been extensively used for a better understanding of biological systems. The major benefit of using computational simulations for biochemical networks is that it aids in a detailed exploration of how the system behaves in the models very easily (Bock & Goode, 2003; McCulloch & Huber, 2002). Another advantage of computational simulations is that it helps in studying the behavioral changes in a very short time and various “what-if” related questions can be answered easily with better understanding. For obtained effective results, it is a must that the models of the systems are constructed correctly. This will help in generating the hypothesis that can be very similar to the real life systems. Suppose, if the optimum results are not obtained then one can go back and reconstruct the model or simulate the models again. Thus we can say that simulations are the most essential part of the scientific process.

24.4.1 Ordinary differential equations

ODEs, particularly systems, have been used in many fields, such as molecular biology, medicine, and bioengineering. Mathematical and theoretical biology is an interdisciplinary field of scientific research with several applications in molecular biology. Therefore finding and interpreting the solutions to these differential equations is a central part of applied mathematics, and a thorough understanding of differential equations is essential for any applied mathematician. For this, precise mathematical models are required. By describing systems quantitatively, their behavior can be better simulated, and thus properties can be predicted that may not be obvious to the experimenter. To have a quantitative understanding of a particular molecular biological process, it is necessary to make a mathematical description of the process we are thinking of. If we want to solve a problem in molecular biology, we must first formulate the problem as a mathematical expression in the form of variables, functions, equations. This expression is known as the mathematical model of the given problem, one has to analyze it to understand it a little. The above rules generally apply to mathematics. They are called mathematical models. One of these important models is ODEs. Describe the relationships between variables and their derivatives. Such models appear everywhere (Van Kampen, 1992).

In 1768 Euler designed the first numerical solution of ODEs. For increasing the computational accuracy, various other numerical solutions of ODEs have been implemented. One of the most popular methods is the fourth-order Runge–Kutta algorithm that can handle stiff differential equations. In ODEs, the numerical calculations are done studying the discrete time steps with step size h . So for evaluating the system using ODEs at time t according to Runge–Kutta algorithm is done by studying at t , $t + 1/2h$, and $t + h$. In the case of Runge–Kutta algorithm, the error scale reaches up to the fourth power of h . To diminish this effect, the step size is abridged by a factor of 10, thus leading to the diminishing of error by 10^4 . If the step size is smaller, it leads to high accuracy, and also the computational time is also decreased. ODEs are used for systems in which components are distributed homogeneously and system biology also assumes that cell components are homogeneously distributed. However, in reality, the molecules are synthesized at a specific location and are transported to some other locations via active or passive diffusion mechanisms of the cytosol. Moreover, this transport depends on various factors, such as high viscosity and various intracellular structures, which leads to nonhomogeneous distribution. If nonhomogeneous conditions are to be considered then, PDEs can be useful. But the model construction, as well as simulations of such models, would be significantly difficult.

The ODE to be used for studying the whole system can be described by the following equation:

$$\frac{dxi}{dt} = \sum f(x) - \sum g(x) \quad (24.1)$$

In this equation, $\frac{dxi}{dt}$ represents the net pool content, while $f(x)$ represents the positive steps that lead to addition to the pool and $g(x)$ represents the negative steps that lead to negative steps that decrease the pool content. It depicts the state of the system in a particular duration of time and also the change of pool sizes (concentrations of molecular species) with time. This equation also helps in identifying the steady state if $\frac{dxi}{dt} = 0$.

24.4.1.1 Integration of ODEs

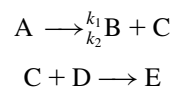
The systems based on ODEs for molecular networks are very complex to solve analytically, so numerical integrations are used to carry out simulations on the computer. Many algorithms do this but almost all are constructed from discretizing the differential equations and move forward with a small step size. The simplest among these is the Euler step method (Jones, Plank, & Sleeman, 2009). Euler's method is a first order method in which the local error (error per step) is proportional to the square of the size of the step and the global error (error at a given time) is proportional to the size of the no (Ascher & Petzold, 1998). The simplest variant of the small stepping can be represented by the Euler method.

24.4.2 Stochastic simulation

When the entities are described by the model variables are present in large numbers such that they can be considered as continuous variables, then differential equations are used. Natural systems are heterogeneous by default: depending on the situation, this can be a boon or a hurdle that must be overcome. Traditionally, while building mathematical models of these systems, heterogeneity has generally been overlooked, in spite of its crucial role. In recent years, however, stochastic calculation methods have obtained commonplace in science. This one can adequately account for heterogeneity; in fact, they are based on the principle that systems intrinsically contain at least one source of heterogeneity (i.e., intrinsic heterogeneity).

Stochastic simulation can be referred to as a type of simulation in which variables can be changed randomly or stochastically based on individual probabilities. Changes based on the random variables are generated and inserted into the models of the system. Outputs based on the changes in the model are noted down and the process is repeated with a new set of random values. There is the repetition of all these steps up to such an extent so that a sufficient amount of data is obtained. After analyzing the distribution of the output, ranges of the values corresponding to the changes in the variables are estimated.

Let's consider the following reactions:



Suppose the initial condition is referred to as S and the molecular species be A , B , C , D , and E . After the occurrence of the first reaction S will change to S^* and the molecular will change to $A - 1$, $B + 1$, $C + 1$, D , and D and E . The probability of the change in the reaction with time t will be given by the following equation:

$$P(S^*, t + dt | S, t) = a_{\mu} dt + o(dt) \quad (24.2)$$

In the above equation, a_{μ} is referred to as a product of mesoscopic rate constant with the current molecular species number while $o(dt)$ can be stated as other terms that can be neglected. This equation is known as the master equation and this state will soon become intractable as each state need a separate variable.

The stochastic algorithms were first applied in the 1970s by Gillespie for simulating chemical reactions. Gillespie stochastic algorithm is based on master equation and instead of calculating probabilities of all the trajectories, it generates the single trajectories for all the reactions. Gillespie used two stochastic algorithms named the direct method and the first reaction method. In the direct method, two random numbers in each iteration are generated and the computational time for calculations is directly proportional to the number of all the possible reactions. In the case of the first reaction method, it produces the putative waiting time period for every possible reaction. This algorithm requires n random numbers for each iteration along the computation time for calculations is proportional to a total number of possible reactions. This first reaction method was further improved by Gibson and Bruck and this improvised version is known as the next reaction method. This updated method utilizes only a single number per iteration while the computation time is directly proportional to the logarithm of the total number of reactions (Gibson & Bruck, 2000). These changes help in dealing with the calculations of larger systems.

24.5 Platforms used for modeling and simulations

To visualize and process the large dataset, several computational tools have been designed for carrying out modeling and simulations. These are based on different algorithms and have been designed in various programming languages. The tools available for pathway analysis can be categorized into general purpose and special programs. General purpose tools, such as MATLAB, Mathematica, and R, are although very powerful packages but they have a limitation of requiring great efforts to get started. As a result, this limitation specialized tools have been developed that have overcome the shortcomings of the general purpose tools. Some of the important tools used for modeling and simulations are discussed below.

24.5.1 Pathway designing tools

24.5.1.1 CellDesigner

CellDesigner is a biological image editor tool that is used for drawing biochemical and gene regulatory networks (<http://www.celldesigner.org/>). The networks designed by this tool can be associated with various simulations and other analysis tools that are available via Systems Biology Workbench (SBW) (Funahashi et al., 2008). Many important databases, such as KEGG (Kanehisa & Goto, 2000), BioModels (Le Novere et al., 2006), and PubMed (Canese & Weis, 2013), are been associated with this software. Various biological entities are represented with different shapes as given in Fig. 24.3. Biological pathways can be either constructed or the downloaded pathways can be visualized in Cell Designer. One of the well-known TCA (Tricarboxylic acid cycle) is shown in Fig. 24.4. In this cycle, the reactants and the products can be easily visualized and their connectivity can be studied. Fig. 24.5 represents the simulation plot of species, such as adenine, hypoxanthine, inosine, adenosine, AMP, IMP, XMP, and GMP, over 500-s simulation time in a pathway related to Leishmaniasis (Bora & Jha, 2020). Variation in the concentration of different species with respect to time can be seen and can be used to interpret the behavior of the system at different conditions.

24.5.1.2 Cytoscape

Cytoscape is an offline podium for constructing gene regulatory networks, which is used for visualizing and constructing molecular communication networks, thus relating these connections with the profiles of gene expression (<http://www.cytoscape.org/>). The networks in this tool are depicted in form of the nodes and the edges. The nodes represent the biological entities, such as genes and proteins, whereas the edges represent the connectivity of these biological entities. This tool works best when used in combination with the huge databases corresponding to protein–DNA, protein–protein, gene expression data, and also the genetic connections (Shannon et al., 2003). There are several options for adding for constructing pathways. Fig. 24.6 shows the network generated by expression data. Various interaction networks obtained from interaction databases can also be imported via several plugin options available in Cytoscape.

24.5.2 Pathway Tools

The Pathway Tools is a software package used for pathway study of the genomes (<http://bioinformatics.ai.sri.com/ptools>). It assimilates the data obtained from genomic resources information with the functional annotation of the

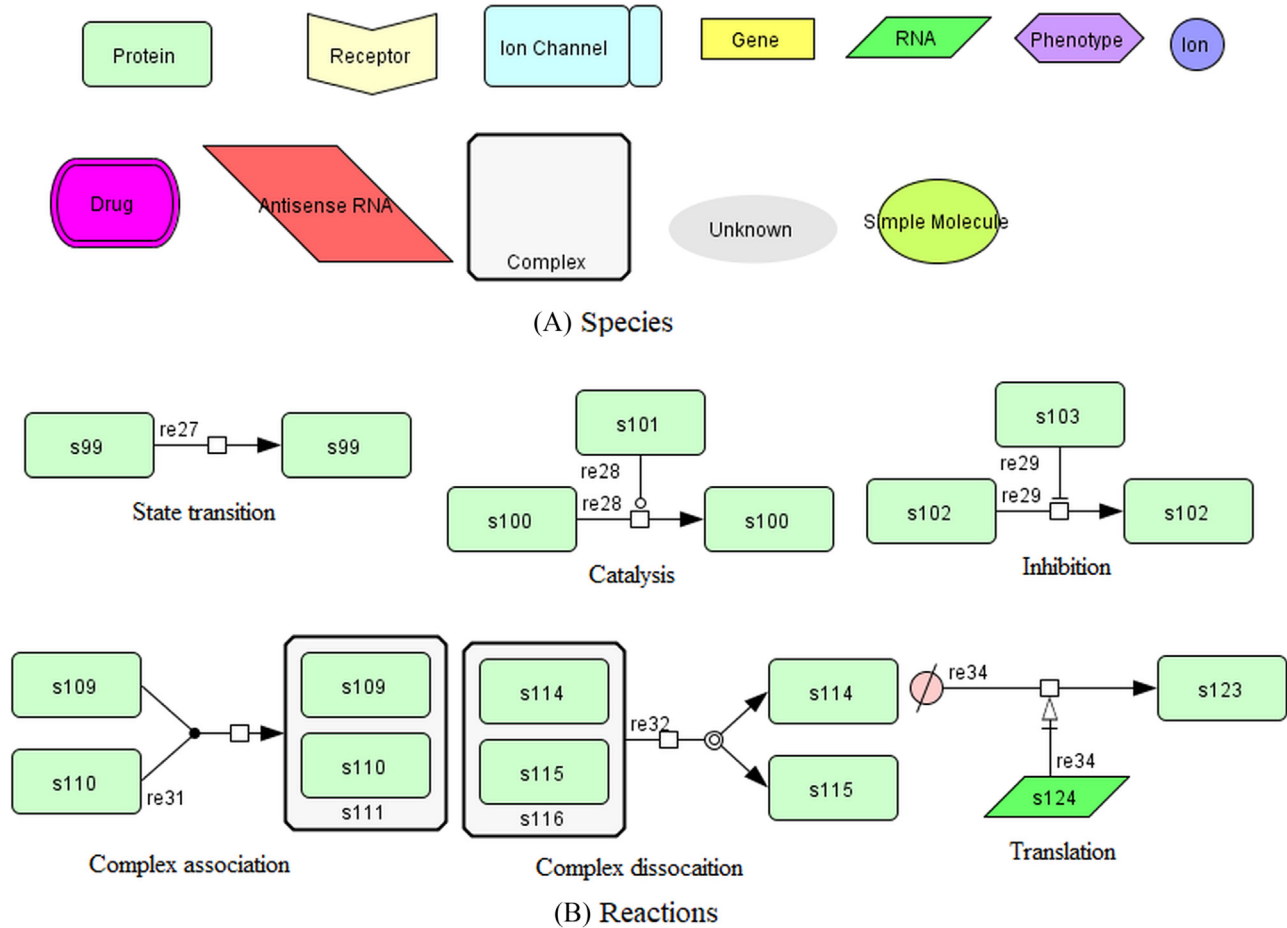


FIGURE 24.3 CellDesigner, some systems biology graphical notations used for the representation of (A) species and (B) reaction.

genome. These annotations include descriptions of signaling and metabolic pathways. This tool also helps in prediction, querying, interactive editing, and finally visualization of the metabolic pathways. The data types that are included in Pathway Tools are reactions, enzymes, and metabolites (Karp, Paley, & Romero, 2002).

24.5.3 Simulation tools

24.5.3.1 GEPASI

General Pathway Simulator (GEPASI) is a software suite that is used for the construction of biochemical systems (available on <http://www.gepasi.org/>). The kinetics of the biochemical reactions for the system is simulated by using GEPASI. It also provides various numbers of options that aid in the fitting of the data to the best model, optimizing functions of the model, linear stability analysis, and performing metabolic control analysis (Mendes, 1993).

24.5.3.2 Systems Biology Workbench

Systems Biology Workbench is a dynamical and extensible, componential message-transient structure for simplifying the message among various sources, which help in research related to systems biology. The main features included in SBW are simply a stochastic simulator, and SBML-to-Simulink, MATLAB ODE file decoder, and the plotting component for the time and series dataset (Bergmann & Sauro, 2006).

24.5.3.3 COPASI

Complex Pathway Simulator (COPASI) is a package used for carrying out simulation as well as the study of biological networks (<http://www.copasi.org>). This software includes a model constructor, sensitivity analysis, a variety of

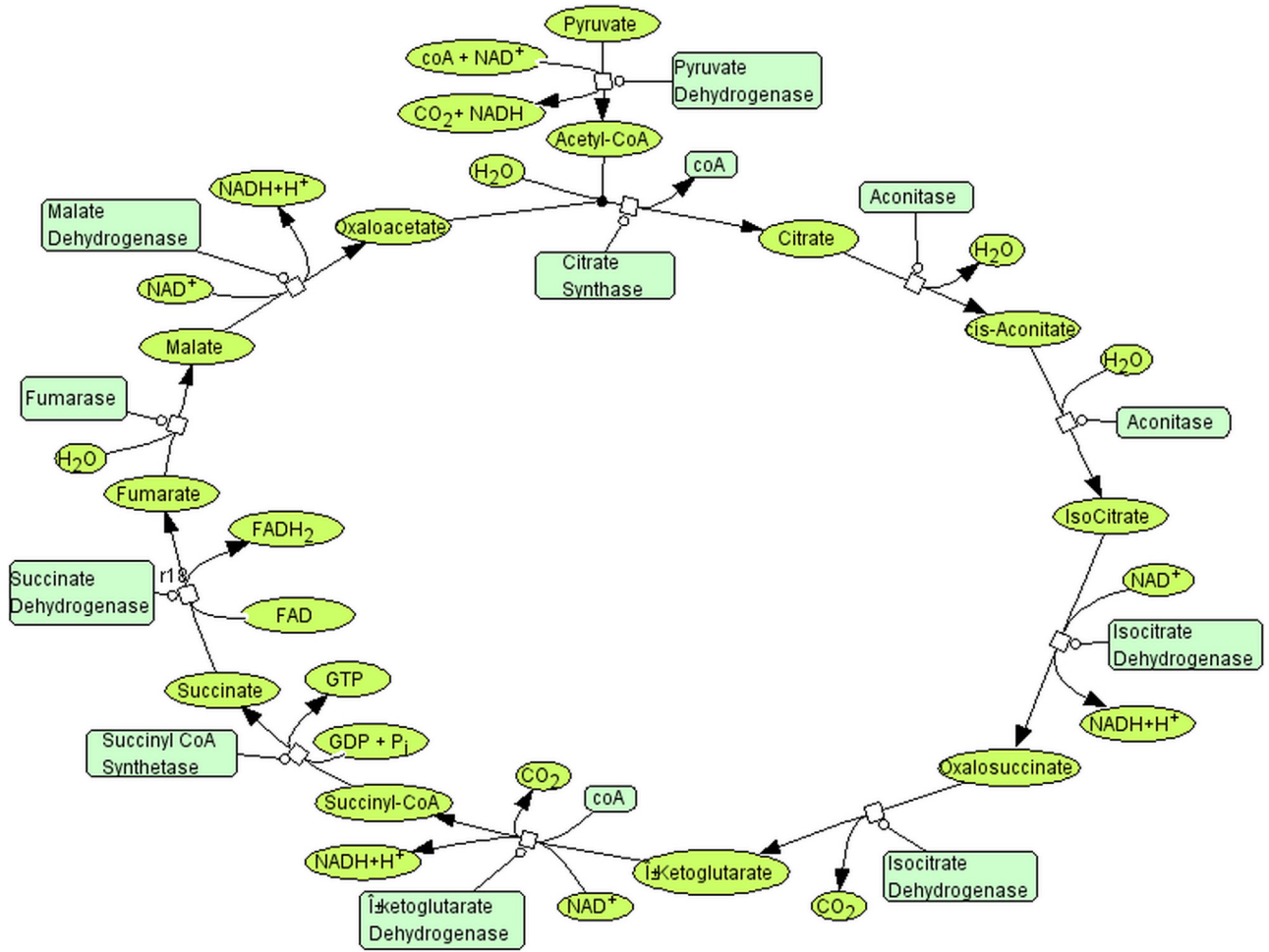


FIGURE 24.4 TCA (Tricarboxylic acid cycle) pathway representation in CellDesigner showing different species and their relationship using notations.

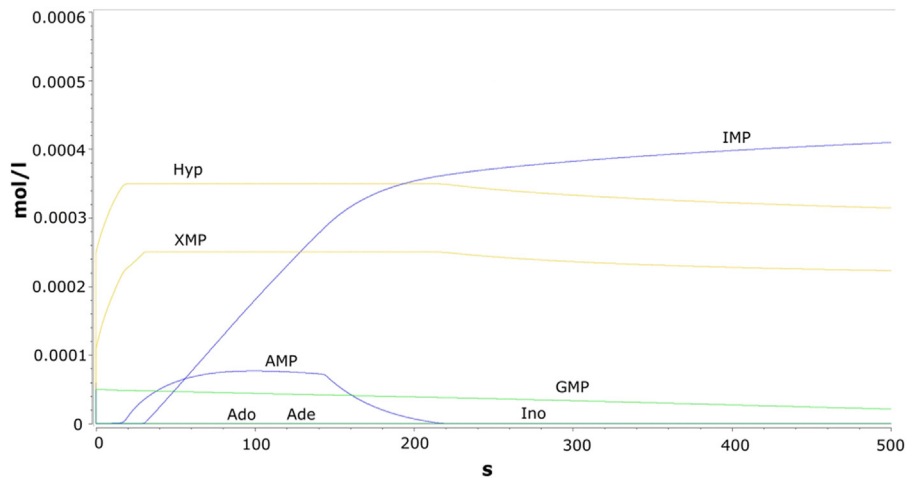


FIGURE 24.5 Simulation result of the pathway, concentration of metabolites during simulation time for 500 s: adenine, hypoxanthine, inosine, adenosine, AMP (Adenosine monophosphate), IMP (Inosine monophosphate), XMP (Xanthosine monophosphate), and GMP (Guanosine monophosphate).

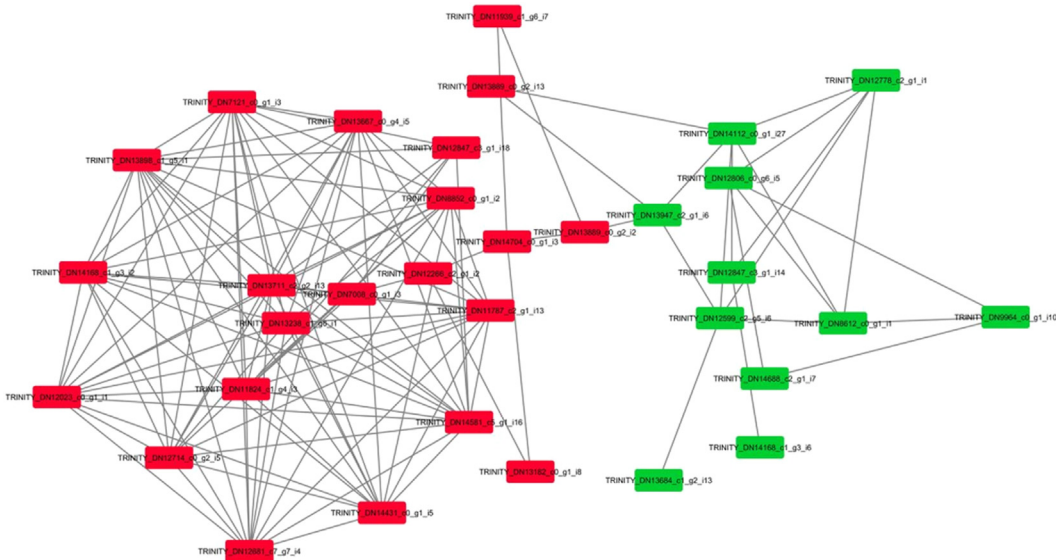


FIGURE 24.6 Expression correlation network generated by Cytoscape.

simulation techniques (such as stochastic and deterministic), optimization routines, and visualization techniques. This software allows the import as well as the export of SBML. It also facilitates the provision of metabolic control analysis and elementary mode analysis (Hoops et al., 2006) (Table 24.1).

24.6 Applications of pathway modeling and simulations

Construction of in silico models for metabolic pathways aids in studying the systems via simulations under varied experimental conditions. This detailed system information provided by the models is of great utility to bioprocesses industries, for the improvement of annotation of the genome and most importantly for the pharmaceutical industries. Moreover, missing links among the metabolic pathways can also be deduced. By iterative modeling, gene insertion or deletion can be done and the corresponding effect can be studied. Some of the major implementations of pathway modeling are stated below:

24.6.1 Metabolic engineering

In quest of enhanced production of the desired metabolite, metabolic engineering acts as an important role player, which helps in the alteration of the metabolic pathways. Basically, it allows studying the phenotype changes with the change in the genotypes. Hal Alper and coworkers have reported a whole genome stoichiometric flux balance analysis to decipher putative genes in *Escherichia coli* that affect network attributes and thus cell phenotype, particularly for reengineering production of glycopene. To achieve this, computational studies have been extensively used, which has helped in studying FBA upon maximizing the cell growth (Alper, Jin, Moxley, & Stephanopoulos, 2005). Apart from transferring few genes, a cluster of genes has also been transferred for producing the desired product with metabolic engineering. For instance, cluster of genes for the production of antibiotic Erythromycin has been transferred from *Streptomyces erythreus* to *Streptomyces lividans* (Vara, Lewandowska-Skarbek, Wang, Donadio, & Hutchinson, 1989). Thus pathway modeling has been very important for metabolic engineering.

24.6.2 Drug designing

System biology has also revolutionized the study of drug discovery as it helps in studying the pathways of the cell, thus identifying the important drug targets. Apart from identifying the drug targets, it also assists in the study of the interaction of drug molecule, drug target, and the system as a whole. Some of the study related to this has been published by Gupta, Singh, Shukla, and Misra (2013) and Gupta et al. (2013) in which they have studied the thyroid disorder

TABLE 24.1 Some important pathway databases.

| S. no. | Database name | Description | Web link | References |
|--------|--|--|---|--------------------------|
| 1. | BioCyc | Pathway and genomic information of thousands of organisms | https://biocyc.org/ | Karp et al. (2019) |
| 2. | MetaCyc | Largest database with metabolic pathway along with enzyme information | https://metacyc.org/ | Krieger et al. (2004) |
| 3. | EcoCyc | Pathway and genomic information for <i>Escherichia coli</i> K-12 MG1655 | https://ecocyc.org/ | Keseler et al. (2005) |
| 4. | KEGG (Kyoto Encyclopedia of Genes and Genomes) | A database for the understanding of high-level function along with utilities of the biological systems | https://www.genome.jp/kegg/ | Kanehisa and Goto (2000) |
| 5. | BRENDA | Most extensive enzyme repository with every relevant information for all the enzymes | https://www.brenda-enzymes.org/ | Scheer et al. (2010) |
| 6. | BIGG | Knowledgebase of biochemically, genetically, and genomically structured genome-scale metabolic network reconstructions. Consists of 70 genome-scale metabolic networks | http://bigg.ucsd.edu/ | King et al. (2016) |
| 7. | BIND | Biomolecular interaction database of many organisms | http://bind.ca/ | Bader et al. (2001) |
| 8. | PANTHER | Signaling pathway information | http://pantherdb.org/ | Mi et al. (2005) |

pathways, which will be helpful in identifying the drug targets. With the help of system study, the possible side effects of the drugs and the toxicity of the drug can also be studied (Apic, Ignjatovic, Boyer, & Russell, 2005). The mechanisms that lead to pathogen survival can also be revealed. Eissenthal and Cornish-Bowden have identified drug targets in the *Trypanosoma brucei* via (African trypanosome). Pathway modeling study is thus able to reveal that it needs only glycolysis for its energy for its survival (Eissenthal & Cornish-Bowden, 1998). In another study, drug targets for the disease tuberculosis that is centered on the flux balance analysis study for the mycolic acid pathway in *Mycobacterium tuberculosis* have been identified by Raman and coworkers. This has been achieved by in silico gene deletion followed by a sequence scrutiny study (Raman, Rajagopalan, & Chandra, 2005). So, pathway analysis study has facilitated in all directional study for drug designing.

24.6.3 Study of phenomics

The phenomics discipline deals with the systemic study of phenotypes. It involves the measurement of phenotypes of a particular trait. Phenomics is a basic study that is extensively used in agricultural research, pharmaceutical research, and functional genomics. With the help of pathway modeling, modulation of phenotypes can be studied under varied conditions (O'Leary et al., 2013).

24.6.4 Flux balance analysis

Pathway modeling and simulation analysis also allows to perform the flux balance analysis study. Generally, linear programming method is used. It allows studying the changes in the flux upon change in the metabolite concentrations at various steps of the pathway. It provides the approximate idea for effect of changes in the experimental studies (Ranganathan, Suthers, & Maranas, 2010).

24.7 Challenges

Pathway modeling and simulation have facilitated in better understanding of the biochemical systems. It has led to the combination of theoretical, experimental, and computational aspects. No doubt it has proven to be the best solution for many biological questions, but certain challenges are faced in the path of system study.

24.7.1 Knowledge gaps between computationalists and experimentalists

With reductionism, many achievements have been obtained in the field of biology and various folds have been uncovered. This has been mainly possible because of the application of computational techniques for a better understanding of biology. Many success stories are available proving that a combination of computational technology has proved to be a boon. [Lahav et al. \(2004\)](#) have studied the dynamic behavior of p53 and how the p53–Mdm2 complex generates a feedback loop that creates a “digital” clock to release timely quanta of p53 till the damage is repaired or until the cell dies ([Lahav et al., 2004](#)). Another study by [Sha et al. \(2003\)](#) has shown that hysteresis acts as a driving force in cell-cycle transitions by studying the biochemical oscillations in *Xenopus laevis*. Things seem to be easy but actually, the ground reality is that there is a wide gap of knowledge existing between computationalists and experimentalists. Both computationalists and experimentalists have to face difficulties at their end. Being a computationalist it is very difficult to understand the complexity of the biological system and the difficulties that experimentalists have to face during experiments.

24.7.2 Theory development

When developing a new theory, many challenges are to face. For instance, we can never understand cellular networks as a whole. Looking at cellular networks in this way leads to the general observation that these cellular networks are very complex. Complexity often arises because there is no way to view the entire network at once. In engineering, especially electrical engineering, large systems are modularly divided into separate functional subsystems. The subsystems are capable of carrying out a relative and well-defined process. By understanding the functions of these subsystems, a hierarchy of modules can be constructed, which will aid in rationalizing the complex systems. Now the important question arises that what are the functional modules with respect to biological networks? So for studying the large biological networks, it becomes mandatory to identify the functional modules ([Hartwell, Hopfield, Leibler, & Murray, 1999](#); [Newman & Girvan, 2004](#); [Tyson, Chen, & Novak, 2003](#); [Wolf & Arkin, 2003](#)). After studying the detailed aspects of these modules, next combining of these modules in the right manner becomes a major challenge and major theoretical aspects are to be developed for achieving this.

In the present scenario, the modularization is based on topological modularity, whereas the network is to be functionally modularized. So, modularization and then a combination of these modules are the toughest challenges that are faced by computationalists and biologists.

24.7.3 Miscellaneous computational challenges

Matching a man-made system with a natural system is very difficult. The biological systems respond to so many internal as well as external factors. Every factor plays an important role in the changes in the biological systems. Many tools that are developed for the study of biochemical systems try to incorporate all these factors with the support of various versions of languages as well as of SBML. Considering all the factors is not feasible in the man-made systems and if any of the factors becomes missing, it can lead to the generation of wrong results ([Cuellar et al., 2003](#); [Hucka et al., 2003](#)).

24.8 Conclusion

System biology is seeking the attention of the scientific community and this can be evidenced by the increase of no of publications. As per the present scenario, most of the biological components are known but the problem lies in the understanding of how these components communicate with each other in a biological system. For this attempts are being conducted to understand the pathways operating in the systems. Pathway modeling and simulation studies have revolutionized the study of the biochemical system on living organisms. With this now it has become easy to understand the working of all

the pathways as a whole. The modulating behavior of the system based on various factors can also be checked. This computational technology has played a vital role as all this study can be done in a short time period. As every new technology has some pros and cons, likewise this study is also faced with some pros and cons. Upon improvement of various algorithms and techniques, this study can be streamlined and can further lead to better results.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Alon, U. (2019). *An introduction to systems biology: Design principles of biological circuits*. CRC Press.
- Alper, H., Jin, Y. S., Moxley, J. F., & Stephanopoulos, G. (2005). Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metabolic Engineering*, 7(3), 155–164.
- Apic, G., Ignjatovic, T., Boyer, S., & Russell, R. B. (2005). Illuminating drug discovery with biological pathways. *FEBS Letters*, 579(8), 1872–1877.
- Ascher, U. M., & Petzold, L. R. (1998). *Computer methods for ordinary differential equations and differential-algebraic equations* (Vol. 61). Siam.
- Bader, G. D., Donaldson, I., Wolting, C., Ouellette, B. F., Pawson, T., & Hogue, C. W. (2001). BIND—The biomolecular interaction network database. *Nucleic Acids Research*, 29(1), 242–245.
- Bardini, R., Politano, G., Benso, A., & Di Carlo, S. (2017). Multi-level and hybrid modelling approaches for systems biology. *Computational and Structural Biotechnology Journal*, 15, 396–402.
- Bergmann, F. T., & Sauro, H. M. (2006). SBW—A modular framework for systems biology. In *Proceedings of the 2006 winter simulation conference* (pp. 1637–1645). IEEE.
- Bock, G. R., & Goode, J. A. (Eds.), (2003). *'In silico' simulation of biological processes* (Vol 247). John Wiley & Sons.
- Bora, N., & Jha, A. N. (2020). In silico metabolic pathway analysis identifying target against leishmaniasis—A kinetic modeling approach. *Frontiers in Genetics*, 11, 179.
- Canese, K., & Weis, S. (2013). *PubMed: The bibliographic database. The NCBI handbook* (2nd ed.). National Center for Biotechnology Information (United States).
- Chaouiya, C., Bérenguier, D., Keating, S. M., Naldi, A., Van Iersel, M. P., Rodriguez, N., & Wicks, B. (2013). SBML qualitative models: A model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools. *BMC Systems Biology*, 7(1), 1–15.
- Cuellar, A. A., Lloyd, C. M., Nielsen, P. F., Bullivant, D. P., Nickerson, D. P., & Hunter, P. J. (2003). An overview of CellML 1.1, a biological model description language. *Simulation*, 79(12), 740–747.
- De Jong, H. (2002). Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1), 67–103.
- Eisenenthal, R., & Cornish-Bowden, A. (1998). Prospects for antiparasitic drugs the case of *Trypanosoma brucei*, the causative agent of African sleeping sickness. *Journal of Biological Chemistry*, 273(10), 5500–5505.
- Fisher, J., & Henzinger, T. A. (2007). Executable cell biology. *Nature Biotechnology*, 25(11), 1239–1249.
- Funahashi, A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N., & Kitano, H. (2008). CellDesigner 3.5: A versatile modeling tool for biochemical networks. *Proceedings of the IEEE*, 96(8), 1254–1265.
- Gibson, M. A., & Bruck, J. (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry. A*, 104(9), 1876–1889.
- Gil-Quijano, J., Louail, T., & Hutzler, G. (2010). From biological to urban cells: Lessons from three multilevel agent-based models. In *International conference on principles and practice of multi-agent systems* (pp. 620–635). Berlin, Heidelberg: Springer.
- Gupta, M. K., Singh, D. B., Shukla, R., & Misra, K. (2013). A comprehensive metabolic modeling of thyroid pathway in relation to thyroid pathophysiology and therapeutics. *OmicS*, 17(11), 584–593.
- Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, 402(6761), C47–C52.
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., & Kummer, U. (2006). COPASI—A complex pathway simulator. *Bioinformatics (Oxford, England)*, 22(24), 3067–3074.
- Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., & Cuellar, A. A. (2003). The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics (Oxford, England)*, 19(4), 524–531.
- Hucka, M., Finney, A. B. B. J., Bornstein, B. J., Keating, S. M., Shapiro, B. E., Matthews, J., & Kitano, H. (2004). Evolving a lingua franca and associated software infrastructure for computational systems biology: The Systems Biology Markup Language (SBML) project. *Systems Biology*, 1(1), 41–53.
- Jones, D. S., Plank, M., & Sleeman, B. D. (2009). *Differential equations and mathematical biology*. CRC Press.
- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27–30.
- Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., & Subhraveti, P. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, 20(4), 1085–1093.
- Karp, P. D., Paley, S., & Romero, P. (2002). The pathway tools software. *Bioinformatics (Oxford, England)*, 18(suppl_1), S225–S232.

- Keseler, I. M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I. T., & Karp, P. D. (2005). EcoCyc: A comprehensive database resource for *Escherichia coli*. *Nucleic Acids Research*, *33*(suppl_1), D334–D337.
- King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., & Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, *44*(D1), D515–D522.
- Kitano, H. (2002). Systems biology: A brief overview. *Science (New York, N.Y.)*, *295*(5560), 1662–1664.
- Krieger, C. J., Zhang, P., Mueller, L. A., Wang, A., Paley, S., Arnaud, M., & Karp, P. D. (2004). MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Research*, *32*(suppl_1), D438–D442.
- Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A. J., Elowitz, M. B., & Alon, U. (2004). Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nature Genetics*, *36*(2), 147–150.
- Le Novère, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., & Hucka, M. (2006). BioModels Database: A free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, *34*(suppl_1), D689–D691.
- Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., & Bergman, F. T. (2009). The systems biology graphical notation. *Nature Biotechnology*, *27*(8), 735–741.
- Le Novère, N., Moodie, S., Sorokin, A., Schreiber, F., & Mi, H. (2010). Systems biology graphical notation: Entity relationship language level 1. *Nature Precedings*, 1.
- Machado, D., Costa, R. S., Rocha, M., Ferreira, E. C., Tidor, B., & Rocha, I. (2011). Modeling formalisms in systems biology. *AMB Express*, *1*(1), 1–14.
- Macklin, D. N., Ruggero, N. A., & Covert, M. W. (2014). The future of whole-cell modeling. *Current Opinion in Biotechnology*, *28*, 111–115.
- McAdams, H. H., & Shapiro, L. (1995). Circuit simulation of genetic networks. *Science (New York, N.Y.)*, *269*(5224), 650–656.
- McCulloch, A. D., & Huber, G. (2002). *Integrative biological modelling in silico*. *Novartis foundation symposium* (1999, pp. 4–19). Chichester, New York: John Wiley.
- Mendes, P. (1993). GEPASI: A software package for modelling the dynamics, steady states and control of biochemical and other systems. *Bioinformatics (Oxford, England)*, *9*(5), 563–571.
- Mendes, P., & Kell, D. B. (1998). Numerical optimisation and simulation for rational metabolic engineering. *BioThermoKinetics in the Postgenomic Era*, 345–349.
- Mi, H., Schreiber, F., Le Novère, N., Moodie, S., & Sorokin, A. (2009). Systems biology graphical notation: Activity flow language level 1. *Nature Precedings*, 1–1.
- Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., & Thomas, P. D. (2005). The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Research*, *33*(suppl_1), D284–D288.
- Moodie, S., Le Novère, N., Demir, E., Mi, H., & Villeger, A. (2011). Systems biology graphical notation: Process description language level 1. *Nature Precedings*, 1–1.
- National Research Council. (2006). *Catalyzing inquiry at the interface of computing and biology*. National Academies Press.
- Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113.
- O’Leary, M. A., Bloch, J. I., Flynn, J. J., Gaudin, T. J., Giallombardo, A., Giannini, N. P., & Ni, X. (2013). The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science (New York, N.Y.)*, *339*(6120), 662–667.
- Palsson, B. (2015). *Systems biology*. Cambridge University Press.
- Raman, K., Rajagopalan, P., & Chandra, N. (2005). Flux balance analysis of mycolic acid pathway: Targets for anti-tubercular drugs. *PLoS Computational Biology*, *1*(5), e46.
- Raman, K., Rajagopalan, P., & Chandra, N. (2006). Principles and practices of pathway modelling. *Current Bioinformatics*, *1*(2), 147–160.
- Ranganathan, S., Suthers, P. F., & Maranas, C. D. (2010). OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Computational Biology*, *6*(4), e1000744.
- Ross, J., & Arkin, A. P. (2009). Complex systems: From chemistry to systems biology. *Proceedings of the National Academy of Sciences of Sciences of the United States of America*, *106*(16), 6433–6434.
- Scheer, M., Grote, A., Chang, A., Schomburg, I., Mumenthaler, C., Rother, M., & Schomburg, D. (2010). BRENDA, the enzyme information system in 2011. *Nucleic Acids Research*, *39*(suppl_1), D670–D676.
- Sha, W., Moore, J., Chen, K., Lassaletta, A. D., Yi, C. S., Tyson, J. J., & Sible, J. C. (2003). Hysteresis drives cell-cycle transitions in *Xenopus laevis* egg extracts. *Proceedings of the National Academy of Sciences of Sciences of the United States of America*, *100*(3), 975–980.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504.
- Stéphanou, A., & Volpert, V. (2016). Hybrid modelling in biology: A classification review. *Mathematical Modelling of Natural Phenomena*, *449*(1), 37–48.
- Tyson, J. J., Chen, K. C., & Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: Dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, *15*(2), 221–231.
- Uhrmacher, A. M., Degenring, D., & Zeigler, B. (2005). *Discrete event multi-level models for systems biology*. *Transactions on computational systems biology I* (pp. 66–89). Berlin, Heidelberg: Springer.
- Van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry* (Vol 1). Elsevier.

- Vara, J., Lewandowska-Skarbek, M., Wang, Y. G., Donadio, S., & Hutchinson, C. R. (1989). Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *Journal of Bacteriology*, 171(11), 5872–5881.
- Visser, D., & Heijnen, J. J. (2003). Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metabolic Engineering*, 5(3), 164–176.
- Wolf, D. M., & Arkin, A. P. (2003). Motifs, modules and games in bacteria. *Current Opinion in Microbiology*, 6(2), 125–134.