

# System biology: Mathematical modeling of biological systems

Jose Arturo Molina Mora

**Abstract**— Biological systems have evolved optimally, which makes them highly complex. Advances in experimental techniques provide an opportunity for the development of mathematical models of biochemical networks, including signal transduction pathways and metabolic networks. Systems biology is an area of active research in which various modeling techniques have been proposed to analyze and identify a wide range of biological networks and that is achieved with the use of algorithms to aid in the task of integrating expert knowledge and get a synthetic view of the existing biological knowledge in a network and achieve the prediction of the dynamic behavior of a system. A review of the generalities of systems biology and mathematical modeling in the context of biological systems is presented.

**Index Terms**— Systems biology, mathematical modelong, biological systems.

## I. INTRODUCTION: BIOLOGICAL SYSTEMS

Progress in the study of cellular and molecular biology has led to the identification of complex systems involved in the physiology of cells, tissues and organs. In this context, a biological system can be considered a collection of different specialized individual compartments (cell level, tissue, etc) for a specific biological function [1]. Biological systems store large amounts of information. For example, at the cellular level and only through the coordinated activity of enzymes and other proteins involved in a series of biochemical reactions, biological information can be used and archived [2]. To understand the information processing and generation of a behavior within cells, the study of complex regulatory networks is often feasible with mathematical models [3].

Thus advances in experimental techniques provide an opportunity for the development of mathematical models of biochemical networks, including signal transduction pathways and metabolic networks [4]. Signal transduction is the process by which a cell converts an external signal or stimulus into an appropriate cellular response [5]. In contrast, in metabolic networks it is given a chemical transformation of molecules or substances, affecting its structure, amount and function within a system [6]. In either case, the events triggered the start and individually are not sufficient to explain how cell fate is controlled but requires an integrated analysis system [3].

*Manuscript received Aug, 2016.*

*Jose Arturo Molina Mora, School of Mathematic, University of Costa Rica San Jose, Costa Rica. Phone: +506 8885 9445.*

As biology information processing of living cells integrate complex paths, rather than a single path, mathematical models become indispensable tools [7]. Due to the inherent complexity of biological systems, mathematical models are a central tool for understanding and predicting the behavior of integration of these systems [8].

## II. MATHEMATICAL MODELING

Modeling as activity represents, manipulates and communicates real-world objects of daily life [9]. The main reasons for the use of models can be summarized as: organizing information that is out of a coherent system, calculate the components and interactions in a complex system, to simulate, predict and optimize procedures, experiments and therapies, prove or disprove hypotheses and/or improved define hypotheses, and finally understand the essential characteristics of a system [4].

The modeling process consists of the following steps [1]: (i) the implementation of the model, which is the description of a formal language of objects/relationships identified in the system under study, with the use of a mathematical structure; (ii) use the model to predict system behavior and (iii) evaluate the model to reality adherence according to these predictions.

Models of dynamic biological systems refer to attempts aimed to analyze and simulate biological systems and processes with the use of techniques of mathematics, computing and information technology, they must be validated to determine if the model does not contradict the knowledge of a biological system [5].

Given the potential magnitude and complexity of the experiments and the resulting data sets, it takes a modeling and simulation techniques for optimally designing experiments and interpreting results. Modeling should suggest a range of hypotheses that can potentially explain results of an experiment and select optimally the following experiment, in order to reduce the number of possible alternative hypothesis and that can be used to predict the effect of molecular perturbations [10]. In this regard, the *in silico* modeling can be used so that the activity of biological systems can be manipulated under controlled conditions [11].

However, some potential drawbacks of these considerations are, first, that a system based on the scheme correct reaction to the system model may fail to fit the experimental data, either because there is an absence of any unknown regulatory signal in a structure network or inappropriate use kinetic representations to write the model.

Second, the models use complex mathematical representations have the flexibility to adapt to a wider range of dynamic behaviors that models created using simple mathematical representations. One consequence of this is that, depending on the mathematical representation used for the model, can lead to "false positives", which fits with the experimental measurements but incorrectly represents the mechanism of the process [8]. This complexity makes the management, analysis and visualization of knowledge available is a challenge in research [12].

### III. SYSTEM BIOLOGY

Systems biology is an area of active research in which different modeling techniques have been proposed to analyze and identify a wide range of biological networks such as gene regulatory networks, metabolic pathways and cascades of signal transduction [13]. The aim of systems biology is to design models using computer algorithms to assist in the task of integrating the expert knowledge, largely obtained data and published results to create new hypotheses [14]. Thus, part of the motivation of network modeling and biological systems is to get a synthetic view of the existing biological knowledge in a network and achieve the prediction of the dynamic behavior of a system [13].

The traditional study of the dynamics of cellular processes, particularly metabolism and signaling, identification and characterization of each of the molecular components usually have the disadvantage that does not analyze the properties that emerge from interactions between different components [1]. That is why systems biology comes to solve this deficiency, in which the whole system and the properties that emerge as a whole is studied. Thus, the approach of mathematical modeling and computational systems biology can be thought of two philosophical categories [4]:

□ Bottom-up, which builds models with the literature and from experiments designed to provide information about specific aspects of the model.

□ Top-down approach, trying to make use of the data collected through a high-performance measurement technology to infer the structures and relationships in the system, which is also known as "reverse engineering".

Both approaches require integration of signals at different time scales, generating different outputs depending on the amplitude and duration that characterize the inputs, and the presence of feedback loops that act as switches to process information through system.

Many of these models seek a quantitative description allowing conducting biological "virtual" or *in silico* experiments, to be more effective and faster than *in vitro* or *in vivo*. However, the results of *in silico* experiments depend directly on the model that simulates [11].

The choice of a mathematical formulation is a mapping model in the mathematical domain. A good model must have a compromise between the competing properties of any model (realism, accuracy and generality) and should take into account some of the mathematical specifications domain [1].

There is an inherent difficulty in the design of living systems, arising from the large number of cellular components and the enormous complexity of its network (evolutionarily optimized). The simulation of a large number of interacting molecules requires the evaluation of the probability and magnitude of all possible interactions between non-identical components and that, unfortunately, the kinetic laws available in the literature do not always consider all effectors simultaneously, and much information exists in a qualitative or semi manner [15].

### IV. SYSTEM'S KNOWLEDGE DEFINES THE MATHEMATICAL MODEL

Mathematical models of biological problems require quantitative information on the reaction rates and molecular concentrations. For most processes, these parameters are not directly accessible *in vivo*, making use of the experimental approach so made that models based on mechanisms predict system behavior and dynamics for each fixed set of parameters [4].

These models usually are based on ordinary differential equations and can be considered as the pillars of systems biology, ultimately providing quantitative and scientific explanations of the behavior of biological systems in health and biology. However, in many cases the number of evaluable parameters and, therefore, the maximum size of the models, have been very limited due to the large amount of experimental data necessary for parameter estimation. Some proposals to combat the high dimensionality of the problem of optimization of parameters on a large scale include: (i) infer the state trajectories of the components that have not been measured and then use these profiles as additional data; (ii) to include known biological characteristics (semi quantitative information) of the behavior of the system in parameter estimation algorithm and (iii) make use of the modular topology with biochemical networks [1], [8].

In addition, models based on quantitative mechanisms could allow researchers to predict the overall behavior of the specified time system to track its dynamics, which can meet a key question in systems biology: what signaling components and rate constants are most critical to the system output behavior? [4].

However, as previously mentioned, calculate parameters or constant rates and concentrations of the initial components must be experimentally measured or inferred by specifying the model, and it is not always easy; also their approach does not guarantee that reflects the real biological system, since measurements are made with *in vitro* models and varying conditions [16], [17]. For example, in the study of biological networks with differential equations the number of equations needed to describe the system can become huge, and analytically difficult and expensive to solve [7].

Due to the incomplete knowledge of the kinetic parameters it is that many qualitative or semi-quantitative methods have been explored in systems biology [6]. A variety of modeling methods can be applied to the understanding of signaling networks and relationships between signals and phenotypes [18], [19]. Some of them include Boolean models, Bayesian models, models with fuzzy logic and hybrid models with neural networks to counter models differential equations, which are computationally expensive and sensitive to inaccurate measurement of the parameters [10]. However, models based on differential equations are highly specified and dependent on extensive prior knowledge about the components and their interactions, and have the advantage that capture temporary and spatial dynamics at the level of individual reactions [18].

In the case of Boolean networks, they can simplify the data structure through continuous conversion of data discretely by placing data in one of two states "On" or "Off", but this simplification ignores valuable information [20]. Furthermore, it has reported a possible resolution inadequate to describe the system to see the data in an exclusionary way these two categories [10].

Meanwhile, fuzzy logic represents a compromise between discrete and continuous models where the observations are classified in a range that goes from 0 to 1 and can maintain their partial membership in more than one category of the respective variable. One of the observations to these models is that the guidelines for building are open and that means it is a highly subjective process [18], [20].

In addition, one of the limitations due to non-linear scalability of the method of computational modeling with fuzzy logic is the cost in generating rules multiple entries (exponential order) and also that only one possible upregulation and negative regulator allows, which means few biologically significant ideas and not so easy to check experimentally [10].

Dynamic Bayesian networks are capable of handling data in a non-discrete and have been widely used for reverse engineering of biological networks and model uncertainty signaling networks. Bayesian networks are based on probability distributions, in contrast to the membership functions in fuzzy logic [18], [21].

Other approaches to modeling of systems biology have explored stochastic methods, which are based on representing the behavior of individual molecules and analyze the variability in overall system performance. In others systems, it has been incorporated cellular automata as a versatile, simple and scalable tool that maps the space of interest as a regular lattice in one dimension and that is optimal for the requirements of the model [5].

## V. CONCLUSION

Advances in experimental techniques provide an opportunity for the development of mechanistic mathematical models of biological systems, including signal transduction, gene expression and metabolic pathways. Mathematical models can be used to understand biological systems, which are characterized by being highly complex, allowing the integration of different levels of response and create a network representing the dynamic behavior of genes, proteins or metabolites. However, despite the amount and quality of experimental data increase rapidly, quantitative measurements of many cellular components in time and space are still scarce for models and parameters, which has staged different types of alternative mathematical models, such as those that manage the vagueness and uncertainty, and that may explain the biology of the system and make predictions that can be verified experimentally.

## REFERENCES

- [1] S. Motta and F. Pappalardo, "Mathematical modeling of biological systems," *Brief. Bioinform.*, vol. 14, no. 4, pp. 411–22, Jul. 2012.
- [2] S. Zhang, R. Wang, X. Zhang, and L. Chen, "Fuzzy System Methods in Modeling Gene Expression and Analyzing Protein Networks," *Fuzzy Syst. Methods Model. Gene Expr.*, vol. 9, pp. 165–189, 2009.
- [3] S. Huang and D. Ingber, "Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks.," *Exp. Cell Res.*, vol. 261, no. 1, pp. 91–103, Nov. 2000.
- [4] N. a W. van Riel, "Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments.," *Brief. Bioinform.*, vol. 7, no. 4, pp. 364–74, Dec. 2006.
- [5] D. Gilbert, H. Fuss, X. Gu, R. Orton, S. Robinson, V. Vysheirsky, M. J. Kurth, C. S. Downes, and W. Dubitzky, "Computational methodologies for modelling, analysis and simulation of signalling networks.," *Brief. Bioinform.*, vol. 7, no. 4, pp. 339–53, Dec. 2006.
- [6] A. Sackmann, M. Heiner, and I. Koch, "Application of Petri net based analysis techniques to signal transduction pathways.," *BMC Bioinformatics*, vol. 7, p. 482, Jan. 2006.
- [7] S. Bornholdt, "Systems biology. Less is more in modeling large genetic networks.," *Science (80-. )*, vol. 310, no. 5747, pp. 449–51, Oct. 2005.
- [8] R. Alves, E. Vilaprinyo, B. Hernández-Bermejo, and A. Sorribas, "Mathematical formalisms based on approximated kinetic representations for modeling genetic and metabolic pathways," *Biotechnol. Genet. Eng. Rev.*, vol. 25, no. 1, pp. 1–40, Jan. 2008.
- [9] R. D. King, S. M. Garrett, and G. M. Coghill, "On the use of qualitative reasoning to simulate and identify metabolic

- pathways.,” *Bioinformatics*, vol. 21, no. 9, pp. 2017–26, May 2005.
- [10] B. Sokhansanj, J. P. Fitch, J. N. Quong, and A. a Quong, “Linear fuzzy gene network models obtained from microarray data by exhaustive search.,” *BMC Bioinformatics*, vol. 5, p. 108, Aug. 2004.
- [11] A. K. Chakraborty, M. L. Dustin, and A. S. Shaw, “In silico models for cellular and molecular immunology: successes, promises and challenges.,” *Nat. Immunol.*, vol. 4, no. 10, pp. 933–6, Oct. 2003.
- [12] C.-F. Tiger, F. Krause, G. Cedersund, R. Palmér, E. Klipp, S. Hohmann, H. Kitano, and M. Krantz, “A framework for mapping, visualisation and automatic model creation of signal-transduction networks.,” *Mol. Syst. Biol.*, vol. 8, no. 578, p. 578, Jan. 2012.
- [13] S.-J. Wu, C.-T. Wu, and J.-Y. Chang, “Adaptive neural-based fuzzy modeling for biological systems.,” *Math. Biosci.*, vol. 242, no. 2, pp. 153–60, Apr. 2013.
- [14] W. J. Bosl, “Systems biology by the rules: hybrid intelligent systems for pathway modeling and discovery.,” *BMC Syst. Biol.*, vol. 1, p. 13, Jan. 2007.
- [15] B. Lee, J. Yen, L. Yang, and J. Liao, “Incorporating qualitative knowledge in enzyme kinetic models using fuzzy logic,” *Biotechnol. Bioeng.*, vol. 62, no. 6, pp. 722–729, 1999.
- [16] D. Faller, U. Klingmuller, and J. Timmer, “Simulation Methods for Optimal Experimental Design in Systems Biology,” *Simulation*, vol. 79, no. 12, pp. 717–725, Dec. 2003.
- [17] Z. Zi, K.-H. Cho, M.-H. Sung, X. Xia, J. Zheng, and Z. Sun, “In silico identification of the key components and steps in IFN-gamma induced JAK-STAT signaling pathway.,” *FEBS Lett.*, vol. 579, no. 5, pp. 1101–8, Feb. 2005.
- [18] B. B. Aldridge, J. Saez-Rodriguez, J. L. Muhlich, P. K. Sorger, and D. a Lauffenburger, “Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling.,” *PLoS Comput. Biol.*, vol. 5, no. 4, p. e1000340, Apr. 2009.
- [19] M. U. Khan, J. P. Choi, H. Shin, and M. Kim, “Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare.,” *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2008, pp. 5148–51, Jan. 2008.
- [20] G. N. Brock, W. D. Beavis, and L. S. Kubatko, “Fuzzy logic and related methods as a screening tool for detecting gene regulatory networks,” *Inf. Fusion*, vol. 10, no. 3, pp. 250–259, Jul. 2009.
- [21] N. Friedman, “Inferring cellular networks using probabilistic graphical models.,” *Science (80-. )*, vol. 303, no. 5659, pp. 799–805, Feb. 2004.

**Jose Arturo Molina Mora** is microbiologist and mathematician of the University of Costa Rica. He has a Master degree in Bioinformatics and currently is working as professor and researcher in the same university. Main research work focus on mathematical models of medical/biological issues, including study of metabolic routes with dynamical modeling and soft computing.