

Preface

Overview

Formal methods of computer science are nowadays unavoidable to model, study, and make advanced analysis of biological systems. Several formalisms are suitable to model biological systems: Petri nets, Boolean networks, reaction rules, process algebras, ordinary differential equations, timed and hybrid automata, etc. Once a biological system is encoded using one of these formalisms, some formal techniques such as model checking can be used to specify some expected properties of the system and verify whether they hold or not in the model at issue. This greatly helps in validating/refuting biological hypothesis, making predictions, and associating parameters with biological phenomena. In this book, we present and compare the main formalisms used in systems biology to model biological networks.

Organization and Features

Some crucial formal approaches used in systems biology are presented in detail, along with their advantages/drawbacks and main applications. Apart from Chapters 1 (Introduction) and 11 (Conclusion), each chapter of the book is devoted to one of the key formalisms used in the literature to model (and verify) biological systems. Each chapter includes an intuitive presentation of the targeted formalism, a brief history of the formalism and of its applications in systems biology, a formal description of the formalism and its variants, at least one realistic case study, some applications of formal techniques to validate and make deep analysis of models encoded with the formalism, and a discussion on the kind of biological systems for which the formalism is suited, along with concrete ideas on its possible evolutions.

Some chapters also include the description of a tool implementing the formalism and a sort of how-to practical guide about using the tool. The networks chosen to serve as case studies span the field of systems biology in a large way (they range from gene regulatory networks to prey-predatory networks).

Some chapters are quite technical and make use of an involved formal notation, but other chapters are more focused on the biological applications (in particular, the last chapter before the conclusion opens to some applications in precision medicine). For each chapter, the notation has been carefully chosen so that it looks the most natural and suited one to represent the formalism at issue. Please note that the authors of some chapters are the ones who first introduced the corresponding formalisms and/or tools in the literature. Also, note that all the chapters are thought to be self-contained: the reader will find in each chapter all the elements that are useful to understand and learn, without having to read other works. Some chapters contain references to other chapters to make comparisons, but there are no strong dependencies among chapters, and the reader can decide to read chapters in a different order than the one chosen in this book.

Some approaches appeared some decades ago, others are quite recent, but all of them are presented from a current and groundbreaking point of view. We describe how each formalism answers today needs in systems biology, which makes a real contribution to the scientific community.

The book is organized as follows.

Chapter 1 focuses on the necessity of using formal methods in the domain of systems biology. It introduces the main formal approaches to the modelling of biological systems and the steps to follow to improve and validate the obtained models.

Chapter 2 is devoted to Petri nets, an important tool for studying different aspects of biological systems, ranging from simple signaling pathways, metabolic networks, and genetic networks to tissues and organs. To explore such varieties of biological systems, many variants of Petri nets have been proposed. This chapter explains how these different net classes are applied to modelling and analysis of these different types of biological systems with the illustrative example of the yeast polarization model describing the pheromone-induced G-protein cycle in *Saccharomyces cerevisiae*.

Chapter 3 describes the development of Process Algebras and related analysis methods in the context of systems biology. It presents concepts that are at the basis of the application of this class of formalisms in the biological context, providing the relevant notions of biochemistry and cell biology, and discussing both qualitative and quantitative approaches. The π -calculus is chosen as representative Process Algebra in order to give

modelling examples and clarify the relationships of the process algebraic approach with the traditional modelling of cell pathways as sets of chemical reactions.

Chapter 4 describes Kappa, a site graph rewriting language. As a realistic case study, a population of hepatic stellate cells under the effect of the *tgf β* protein is modeled. Kappa offers a rule-centric approach, inspired from chemistry, where interaction rules locally modify the state of a system that is defined as a graph of components, connected or not. In this case study, the components are occurrences of hepatic stellate cells in different states and occurrences of the protein *tgf β* . The protein *tgf β* induces different behaviors of hepatic stellate cells, thereby contributing either to tissue repair or to fibrosis. Better understanding the overall behavior of the mechanisms that are involved in these processes is a key issue to identify markers and therapeutic targets likely to promote the resolution of fibrosis at the expense of its progression.

Chapter 5 introduces Pathway Logic, a formal, rule-based system and interactive viewer for developing executable models of cellular processes. It includes a curated evidence knowledge base and a diverse collection of models for evaluation by users. This chapter presents the Pathway Logic representation system and the algorithms used by the Pathway Logic Assistant. The overview discusses rewriting logic and its implementation in the Maude system, the formal basis of Pathway Logic. Other sections in the chapter present the STM8 collection of signaling response networks, provide overviews of the curation process and how rules are inferred, and illustrate the utility of Pathway Logic using the *Lps* (lipopolysaccharide) and Cell Death models.

Chapter 6 presents Boolean networks, a mathematical model that has been widely used since decades in the context of biological regulation networks qualitative modelling. They consist in collections of entities, each having two possible local states (1 – active and 0 – inactive), which interact with each other over discrete time. The simplicity of their setting together with their high abstraction level are especially convenient to focus on foundations of information transmission in genetic regulation, and on mathematical explanation and prediction of phenomenological observations. This chapter aims to present the Boolean modelling framework by developing its theoretical bases and emphasizing its usefulness for capturing biological regulation phenomena. But it goes beyond that by covering their ability to capture the information transmission and its consequences depending on the ways the entities update their local state over time.

Chapter 7 deals with Answer Set Programming (ASP), which has proven to be a strong logic programming paradigm to deal with the inherent complexity of the biological models, allowing us to quickly investigate a wide range of configurations. ASP can efficiently enumerate a large number of answer sets, as well as easily filter the results thanks to constraints based on certain properties. This chapter first motivates the merits of ASP in biological studies based on the state of the art. Then, it introduces the basic concepts about ASP and its use in systems biology. After having given an overview of the different issues that can be tackled using ASP, it then focuses on one problem that is of critical importance: model-checking with ASP, and more specifically, the identification of attractors. The merits of this study are illustrated using case studies.

Chapter 8 focuses on hybrid automata, a formalism introduced and developed with the aim of integrating discrete and continuous ingredients in a single simulation tool. This chapter introduces some key logic formalisms for systems biology; illustrates some automata-based simulation tools; discusses the role, potential, and complexity of the notion of time in automata; and presents several methodologies to integrate discrete and continuous, time-oriented, formal instruments for systems biology. Several realistic case studies are treated.

Chapter 10 discusses several network modelling methods and their applicability to precision medicine. The chapter introduces a certain number of network centrality methods (degree centrality, closeness centrality, eccentricity centrality, betweenness centrality, and eigenvector-based prestige) and two systems controllability methods (minimum dominating sets and network structural controllability). Their applicability to precision medicine on three multiple myeloma patient disease networks is demonstrated. Each network consists of protein-protein interactions built around a specific patient's mutated genes, around the targets of the drugs used in the standard of care in multiple myeloma, and around multiple myeloma-specific essential genes. For each network, it is demonstrated how the discussed network methods can be used to identify personalized, targeted drug combinations uniquely suited to that patient.

Finally, Chapter 11 underlines some key concepts of the book and opens on possible evolutions in the domain of formal methods for systems biology.

Audience

This book can be of interest to scientists at all levels, from master students to senior scientists that want to learn about formal methods for systems biology. The book is mainly addressed to bioinformaticians but can also be appreciated by biologists, medical doctors, computer scientists, and mathematicians interested in this highly interdisciplinary area.

Nice, August 2022

Elisabetta De Maria