Systems Biology – Theory, Techniques and Applications

André X. C. N. Valente
Abhijit Sarkar · Yuan Gao
Editors

Recent Advances in Systems Biology Research

NOVA

Nova Biomedical

SYSTEMS BIOLOGY - THEORY, TECHNIQUES AND APPLICATIONS

RECENT ADVANCES IN SYSTEMS BIOLOGY RESEARCH

A. X. C. N. VALENTE
ABHIJIT SARKAR
AND
YUAN GAO
EDITORS

NOVA BIOMEDICAL
New York

Library of Congress Cataloging-in-Publication Data

Recent advances in systems biology research / editors, Andri Xavier Valente (University of Coimbra and Biocant--Biotechnology Innovation Center, Portugal), Abhijit Sarkar (The Catholic University of America, USA), and Yuan Gao (Johns Hopkins University, USA).
pages cm
Includes index.
ISBN 978-1-62948-736-6 (hardcover)
1. Biological systems. 2. Systems biology. I. Valente, Andri Xavier.
QH313.R37 2014
571.7--dc23
2013046310

Published by Nova Science Publishers, Inc. † New York

SYSTEMS BIOLOGY - THEORY, TECHNIQUES AND APPLICATIONS

RECENT ADVANCES IN SYSTEMS BIOLOGY RESEARCH

SYSTEMS BIOLOGY - THEORY, TECHNIQUES AND APPLICATIONS

Additional books in this series can be found on Nova's website under the Series tab.

Additional e-books in this series can be found on Nova's website under the e-books tab.

Preface

Systems biology has emerged as an exciting and rapidly growing area of research. This book presents 14 original chapters written by experts in the field covering topics ranging from the definition and scope of systems biology, to the experimental and computational techniques driving the data-rich biology revolution, to applications of the systems paradigm to clinical problems. The first two chapters discuss the systems approach and contrast it with the prevailing reductionist techniques. What emerges is a clear realization that the two paradigms complement each other. Neither can be dispensed off in favor of the other; however, what we learn from either approach must be combined into an integrated description of physiological processes.

The following seven chapters deal with experimental, statistical, kinetic, and graph theoretic techniques for organizing and interpreting systems data. A wide variety of computational strategies are covered along with a comprehensive but concise description of mass spectrometry methods that (combined with deep sequencing) are the source of many of the large data sets driving systems biology.

The last five chapters present applications of the systems approach to clinical problems and are a specially valuable feature of the book. These chapters provide interesting perspectives and literature reviews that should be very useful to anyone considering how to translate systems biology into clinical practice.

The book is written at a level accessible to a wide readership and presents much useful information in a concise and accessible format. In addition, each chapter contains references to important research in this field. The reader will find the material presented here to be an up-to-date introduction to this emerging discipline.

This project would not have been completed without the support of our students, postdocs, and family members. We thank them all for contributing in one way or another to our understanding of this subject. We would also like to acknowledge the assistance of the publisher in carrying out the task of preparing the individual manuscripts for publication in a book format. And, most importantly, we would like to acknowledge the time and energy spent by the contributing authors in preparing their respective chapters. Clearly, without their effort, this book would not have been possible.

A. X. C. N. Valente
Abhijit Sarkar
Yuan Gao

In: Recent Advances in Systems Biology Research
Editors: A. X. C. N. Valente, A. Sarkar and Y. Gao
ISBN: 978-1-62948-736-6

Introduction

A. X. C. N. Valente[1,2,3], Yuan Gao[5] and Abhijit Sarkar[4]
[1]Systems Biology Group, Biocant - Biotechnology Innovation Center,
Cantanhede, Portugal
[2]CNC - Center for Neuroscience and Cell Biology,
University of Coimbra, Coimbra, Portugal
[3]Institute of Fundamental Medicine and Biology,
Kazan Federal University, Kazan, Russia
[4]Physics Department and Vitreous State Laboratory,
Catholic University of America, Washington DC, US
[5]Lieber Institute for Brain Development, Johns Hopkins Medical Campus,
Baltimore, Maryland, US

Systems biology is an attempt to come to grips with the inherent complexity of biological systems. Taking the cell as an example, its physiological state is determined by the mutual interaction of a large number of proteins, polynucleotides, and small molecules. It is tempting to view the processes involved as linear cascades of chemical reactions initiated and maintained by internal and external signals. However, it has become increasing clear that a more appropriate model is that of a coupled network of protein-protein, protein-small molecule, and protein-DNA interactions. The role of chance fluctuations in determining cellular–level outcomes has also come to be recognized. In other words, to infer a cell's response to some change and to develop a predictive picture of how it responds to stimuli, studies of individual genetic processes have to be complemented by research focused on building a fuller picture of the mutual interdependence of these processes. Attention has to be equally given to how molecular-level events interact and become correlated, and how these mutual couplings respond to probabilistic variations resulting from the noisy environment inside cells. This perspective recognizes the importance, the centrality, of multi-scale and emergent phenomena in cells, and strives to incorporate them to form a more complete picture of cellular and higher-level biological systems.

Revolutionary advances in DNA and protein sequencing technologies have made possible just the sort of studies required to integrate and decipher the interdependence of individual biochemical reactions and the deterministic and chance variability in these links. The

resulting perspective has come to be called systems biology, and this book brings together the research of a number of groups at the leading edge of this rapidly growing area of biology.

We start with two chapters that consider the question "What is systems biology?" As one should expect, definitive answers to this questions are not (and perhaps cannot be) provided. However, these two chapters serve to highlight some important differences between reductionist and systems approaches. On Chapter one, Bizzarri et al. start by describing the traditional view of biological systems as systems whose behavior can be fully specified at the level of genes and their interactions. Phenotypes can then be built up from a complete understanding of genetic processes and this is akin to identifying a chain of events and the participating molecules that starts at the gene level and ends up at the phenotype level. As the authors argue, this mode of explanation does not comport with the stochastic and non-linear environment inside the cell. The focus on individual genetic events, although valuable, has to be integrated within a framework that emphasizes functional and integrative explanations. For this purpose they propose the use of the concept of morphogenetic field and a system phase space to classify the multitude of cellular dynamical configurations that give rise to observed cellular responses.

On Chapter two, Bertolaso et al. begin by arguing that the central question of systems biology is to explain how order arises in biological systems. Indeed, they want to characterize the "inter-level couplings", the regulatory processes that bind the different organizational levels in the cell, through which the ordered phenomena emerge. After reviewing existing attempts to move away from purely reductionist explanations, for example using network topological concepts to characterize interactions at the whole-cell level, they argue for the incompleteness of both bottom-up and top-down approaches, the purely reductionist and the fully holistic approach. Instead, the authors present the merits of an intermediate approach, which they call the mesoscopic level, which places explanatory weight on how different organizational levels "communicate" with each other and lead them to cohere together into an integrated system. They develop their proposal in a variety of interesting ways and provide some guidance on how to select, for any given problem, the appropriate mesoscopic level that provides maximum explanatory leverage.

The next seven chapters discuss different modeling methodologies and experimental approaches that are useful for generating and analyzing large-scale data sets of the sort that are driving the systems approach. We start with the contribution of Valente et al. who present on Chapter three a novel methodology for analyzing genome-wide association study (GWAS) data. These datasets are generated by sequencing single nucleotide polymorphisms (SNPs) in both a control group and a group of subjects afflicted with a particular illness. The goal is to identify SNPs that differentially modulate the incidence of the condition. A key problem in these studies is how to convert existing biological insight about the condition into a more effective statistical search for the genetic risk factors. The authors present a so called hypothesis-rich approach, based on the construction of Rational Classes, sets of candidate risk factors that share an underlying common rationale. Of note, the authors point out that many insights and hypotheses that can straightforwardly be incorporated via Rational Classes, would be of no or difficult use under a classical Bayesian approach. The power of the method is illustrated by an application to a Parkinson's GWAS.

On Chapter four, Rietman et al. present a thermodynamic framework for analyzing the dynamics of molecular interaction networks consisting of mutating and replicating units. They show how a description of these dynamics can be formulated in terms of a stochastic

kinetic model and its time evolution mapped by using concepts from irreversible thermodynamics. This wide-ranging chapter illustrates how the basic concepts of statistical mechanics and phase transition theory can be used to generate predictive models of the time course of interaction networks inside cells. An interesting example of the transition from a normal to a cancerous cell is presented as a case study of how the formalism can be used to develop a systemic perspective on the transition to cancer.

Chapter five by Tolga Can discusses an algorithmic way to indentify functional subunits in protein-protein interaction networks. This chapter begins to address the more general question of how to partition complex systems – systems consisting of many strongly interacting parts where the dynamics of the individual parts are well-characterized but cannot be aggregated in a simple way to furnish a predictive dynamical model for the system as whole – and partition it into subcomponents whose interactions are known and do provide a basis for constructing useful models at the system-wide level. After a brief but useful introduction to web-accessible interaction datasets, the author discusses general graph-theoretical ways to identify sub-graphs or modules defined according to some criteria. This is followed by a review of various algorithms developed specifically for studying protein-protein interaction datasets. Finally, the chapter turns to a discussion of model-based approaches to module identification. The key concept here is to utilize known signaling or interaction pathways to guide or bias the search for functional subunits in non-annotated networks. This chapter is a useful, non-technical review of the methods developed so far and some of the remaining challenges in the field.

Chapter six by Cheng et al. reviews advanced protein mass spectrometry (MS) methods, specifically quantified liquid chromatography (LC) MS/MS techniques, and how they are used to generate protein-protein interaction pathways. These techniques together with advances in so-called deep sequencing technologies for DNA have greatly advanced the availability of datasets that, when combined with appropriate analytical methods, are well-suited to provide a systemic perspective on cellular processes. After reviewing the steps involved in preparing samples for MS, the authors discuss how these techniques are used to identify the protein elements in a pathway. This is followed by a discussion of strategies for quantifying changes in protein abundance from MS data. Next, the authors provide a thorough discussion of how LC-MS/MS can be used to discover which proteins have been subject to post-translational modifications. The chapter finishes by describing how protein-protein interactions can be identified by LC-MS/MS.

The theory of reliability has been developed as a way to formalize and predict the effect of a component failure on the functioning of a network or system to which it belongs. Most applications have been confined to engineering analyses of the robustness of networks; however, this framework seems appropriate to describe the response of cells or even of the whole organism to dysfunctional elements within it. This is especially true if one considers that the design mechanism in biology – Natural Selection – would be particularly sensitive to reliability features of an imperfectly self-replicating system. This topic is taken up in Chapter seven by Koltover, who reviews how reliability concepts and formalisms can be used in the systems biology context. After discussing the basic theoretical terms, these concepts are applied to a study of enzymes to discover general principles of enzymatic design that may reflect reliability considerations. Next, an interesting application at the whole organismal level is presented. This is followed by a discussion of how differential reliability

characteristics may show themselves in the birth-death processes of heterogeneous populations. The chapter ends with a discussion on reliability testing in biology.

Chapter eight by Moskon et al. presents a comprehensive review of various computational approaches that have been used to model the evolutions of molecular species in cells as well as interaction network dynamics. This wide-ranging chapter usefully summarizes a number of graph-theoretic techniques as well as concepts and strategies from deterministic and stochastic kinetic models. This is followed by a discussion of various statistical parameter estimation methods that may be used to develop estimates for rate and other constants that appear in these equations. Taking up a theme from the previous chapter on reliability and robustness, the authors present a number of criteria that may be used to quantify and evaluate these characteristics of biological systems. The chapter ends with a discussion of how techniques from dynamical systems theory, especially concepts associated with the stability of dynamical systems, can be applied in the systems biology context.

Chapter 9 by Davies and Husmeier discusses how Bayesian statistics can be used to generate a model for transcriptional regulation using infinite-parameter Gaussian Process (GP) priors. After defining GP priors, the authors present calculations describing how the posterior probability may be computed for models generated from stochastic kinetic differential equations using GPs as priors. This technical section is followed by a discussion of identifiability and noise assumptions. The authors present simulation testing of their model by considering the case of transcriptional dynamics under Michaelis-Menten kinetics. A brief section on extending the model to non-linear regulation rounds out the chapter.

On chapter 10, Blair et al. present a review of various strategies that may be used to model cancer at the systems-level. While focusing on one area – cancer biology – the techniques and approaches they present are equally applicable to the study of other biological processes. After describing the context of the chapter, the author review differential-equation-based kinetic models and well as graphical representations of systems data. This is followed by a review of databases available to provide inputs to these modeling approaches. This chapter is marked by the absence of technical discussions and should be useful to a wide range of readers.

We end with a group of four chapters presenting applications of the systems approach to clinical problems. Chapter 11 by Louridas and Louridas describes how a systems approach can provide a new perspective in cardiology. After describing the context of the discussion, the authors review how large scale genomic and proteomic – primarily transcriptomics and metabolomics – data sets have led to the possibility of individualized biomarkers for heart disease and related conditions. These results they have termed personalized cardiology. They end with a discussion of how clinical care can also be facilitated by the systems-level data they describe earlier.

Chapter 12 by Petrasek and Petrasek presents some thoughts on how a systems-level perspective may be relevant to the understanding of diabetes. After presenting their views on the reductionist versus systems paradigms, the authors provide an eclectic argument for their preferred approach. This conceptually wide ranging article describes the role of uncertainty in biology and how it should be handled. Examples of the DCCT and ACCORD trials are presented as case studies.

Chapter 13 by Hong et al. presents another perspective on cardiac systems biology. The authors begin by describing the various genome-wide association studies (GWAS) related to heart disease. They discuss –omics data, and then summarize relevant protein structural

biology information. The rest of the chapter focuses on modeling and integration of these data streams and reviews studies in this direction.

Chapter 14 by Saidi and Kenari, the last chapter of the book, reviews systems and other approaches to the study of transplant immune tolerance. After defining tolerance in the clinical context, the authors describe the results of clinical tolerance studies. This is followed by a review of clinically-relevant biomarkers and system-type studies that have been undertaken to analyze immune tolerance.

Contents

In: Recent Advances in Systems Biology Research
Editors: A. X. C. N. Valente, A. Sarkar and Y. Gao
ISBN: 978-1-62948-736-6

Chapter 1

The Conceptual Foundations of Systems Biology: An Introduction

Mariano Bizzarri[1], Alessandra Cucina[2] and Alessandro Palombo[3]
[1]Department of Experimental Medicine, "Sapienza" University of Rome, Roma, Italy;
[2]Department of Surgery "P. Valdoni", "Sapienza" University of Rome, Roma, Italy.
[3]Department of Clinical and Molecular Medicine, "Sapienza"
University of Rome, Roma, Italy

Abstract

Systems Biology is more than just a summary of different sciences, given that Systems Biology deals with 'systems', and it is concerned with the complex, emergent properties that arise from the relationship between molecules, cells and tissues. Functional properties are not yet in the 'molecules', instead they 'emerge' from a self-organized process, which geometrically shape the living structure into a system, characterized by hierarchical levels. Interaction among them leads to both top- and downward causation. Thus, biology is now facing a crisis — or is it an opportunity — reminiscent of the state of biology in pre-double-helix time. The time is gone for biology to "concern itself with the great 'non-reductionist' 19th century biological problems that molecular biology left untouched: e.g., the nature of the complex organization of the living matter. Precisely, Systems Biology affords that task.

Introduction: What We Do Mean by Systems Biology?

Despite the widespread use of the term 'Systems Biology' - a word coined by Mesarovic in 1968[1] - a shared and unified consensus about what Systems Biology is is still lacking. At a glance Systems Biology may be considered a kind on 'unusual' combination of biology with a great many other sciences, from physics to ecology, mathematics, medicine, chemistry

and even sociology and linguistic. According to this 'minimalist' approach, Systems Biology has been viewed as the 'successor' of Molecular Biology[2], mainly characterized by the use of large-scale molecular interactions ('omic' framework) focused on building complex signalling networks by applying mathematical modelling, and thus showing how cells make decisions based on the information flow through those networks[3]. Eventually, this approach still relies on the molecular level thought as the privileged level of explanation and, despite its 'practical' usefulness, it has to be considered nothing more than a mere 'epicycle' of the existing, reductionistic theories[4].

Yet, several reports[5,6] have clearly showed that Systems Biology is more than just a sum up of different sciences, given that Systems Biology deals with 'systems', and it is concerned with the complex, emergent properties that arise from the relationship between molecules, cells and tissues. Functional properties are not yet in the 'molecules', instead they 'emerge' from a self-organized process, which shape geometrically the living structure into a system, characterized by hierarchical levels. Interaction among them lead to both top- and down-ward causation. Since the behaviour of such, complex systems is mainly ruled out by non-linear dynamics, it follows henceforth that the classical concept of linear causality, is meaningless in Biology[7]. Thus, *ad-hoc* approaches are clearly insufficient to cope with complexity, and classical physics as well as Molecular Biology are unfit to address such issues, given that "the existing philosophy of biology fails to address the rather profound issue of what distinguishes the living from the non living, except to say that something lives because its ancestors lived"[8]. Therefore, Systems Biology should have a unique and robust philosophical foundation and a different methodological approach in order to take up the challenge of understanding living organisms as wholes.

We need to build a reliable approach, allowing us to copy with the intrinsic 'disorder' of living processes, e.g., with the chaotic fluctuations underlying biological function[9]. What we are used to call 'biological noise', may be not really noise, as it actually reflects normal biological fluctuations. A great part of "relevant biological information is embedded in this part of the data: the small variations of small signals, which collectively could contain more information than the large variations of a handful of genes"[10]. Therefore, this is the challenge: how to include chaotic and non-linear, unpredictable, processes into our comprehension of Biology. Systems Biology addresses such issue, and should be defined as the study of phenomena in terms of *how the objects are related*, rather than what they are composed of. Systems Biology is "a science no more confined to simplified and idealised phenomenon: but a science facing itself with the complexity of the real world"[11]. We posit hence that a new approach, a shift from reductionism to an "integrated" perspective is by now required.

However, more than just a pronouncement of a new approach is required. What is needed is to provide a conceptual framework able to integrate some entrenched aspects, as complexity, hierarchical structured levels of observation, geometrical relationships, non-linear dynamics, network modelling, influence of biophysical constraints, operating on different scales, rather than solely focusing on building numerical mathematical or computer models[12]. Those aspects must be collectively considered in order to find organizing principles that exactly outline the evolution of systems in space and time[13].

The Newtonian Paradigm and the Emergence on Non-Linear Dynamics

Science is generally viewed as strongly grounded in the Newtonian paradigm of mechanics. According to that paradigm (to which Descartes proved a robust philosophical underpinning), every phenomenon we observed can be analytically reduced to its components. No matter the number of constituents are, you only need to take your analysis one step further, and look at their interactions. By continuing this subdivision long enough, you will end up with the smallest possible parts (atoms) or 'elementary particles', that are seen as 'separate' pieces of the same substance (matter). A fundamental property that distinguishes particles is their position in the space and its change along time. Any development or evolution is conceived as geometrical rearrangement due to the movement, and the movement is entirely governed by deterministic laws of cause and effect. By knowing the initial positions and the velocities of each system's particles, together with the forces acting on the system, you will be able (at least in principle) to predict the evolution of the system with complete certainty and accuracy, as claimed by the well known Laplace's aphorism[1]. The trajectories of the system are not only determined towards the future, but also towards the past. Therefore, you can theoretically 'reverse' the process to its past states: within that frame there is no room for the time's arrow[14]. The mathematical description of such kind of evolution can be depicted by a phase space diagram, in which changes are represented as phase transitions. A phase transition is described by a point along the line where the control parameter(s) runs. A phase transition is critical, i.e., it allows the system to acquire a different configuration, when some known (or eventually, hidden) observables diverge, thus leading to the appearance of a coherent structure. In Newtonian physics, the dynamics of the system within the phase space diagram is entirely described by linear mathematic equations providing a full knowledge of the object. Randomness ("noise") in such a framework is thought to be due to unknown ('hidden') variables, from whose interactions unpredictability comes. Therefore, according to that classical mechanistic view, variability in physical processes is due to a lack of *epistemic* knowledge: a physical phenomenon is "per se, a geodesic. This is a unique, optimal and 'critical' path, completely determined by the Hamiltonian, and may be computed as an optimum of a Lagrangian functional"[15]. As a consequence, such process possesses a unique basin of attraction. However, in the real world such 'beautiful' paradigm, rarely works, given that its validity is limited to systems behaving according to a linear dynamics. Indeed, most of the systems modelled by traditional science are linear. This means basically that effects are proportional to their causes and the system is thought to adopt only one configuration. Those systems are symmetrical. In some instances, the symmetry can be broken when one or more system's parameters fluctuate over a well-defined threshold value, reaching thereafter a bifurcation point where it experiences a symmetry breaking[16]. The system loses it homogeneity, and

[1] "We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes". P.-S. Laplace, A Philosophical Essay, New York, 1902, p. 4.

can acquire an efficient notion of distance in space. The symmetry breaking discloses different solutions for the same parameters values (hysteresis and bistability), therefore opening the system evolution towards novelty and variability. The system drives along different trajectories, thus converging into one or more 'attractors'. An attractor is a stable solution to the set of mathematical equations that describe a dynamical system, representing the state of equilibrium to which the system will tend to move. Attractors are distributed along a complex landscape, in which stable (valleys), as well as metastable or unstable (hills) states are depicted [17]. By this way new configurations and self-organized structures arise. In non-linear systems randomness is an intrinsic property. A quite similar situation happen in Quantum Mechanics, where variables cannot always be associated to single points (quantum entanglement) and unpredictability is intrinsic, given that – according to Heinsenberg's principle and Schrödinger equation - the parameter value is expressed as a *probability*. In Quantum mechanics, some quanta interact and form one system in which they are 'entangled': this means that they are *non-separable* by measurements. By analogy, a complex, living system is characterized by non-separable, 'entangled' processes, behaving according to non-local rules, and henceforth requiring for a deeply understanding, a global, 'systemic', view.

Science is generally used to deal with (ideal) non-linear, isolated systems. Such models have played an important role in enabling us to conceptualise the rules governing their physico-chemical behaviour. Those systems are accurately described by equilibrium Thermodynamics, given that they behave like 'isolated' systems: they are insensitive to their environment. This 'insensitivity' implies that the state of equilibrium is stable (small perturbations are quickly damped and the system is driven back to the equilibrium), and external influences (like that exerted by gravitational or electromagnetic field) are small compared to thermal energy, and cannot trigger any significant effect. The state of equilibrium is characterized by maximum entropy, and there is no cooperativity within the system, except near a phase transition. Let us to consider a gas whose molecules have a masse m enclosed in a volume of length l, in a gravitational field of strength g: the maximum energy of interaction is therefore mgl. That energy will likely produce a density variation proportional to mgl/kT. As $mgl/kT<<1$, such effects of external fields on equilibrium systems can be ignored: they might be observed only when they are larger than the thermal fluctuations. However, for both non-equilibrium and equilibrium systems, there exist situations in which small external cues could cause large effects. Usually, for equilibrium systems, this happens when cooperative effects underwent near a symmetry breaking (bifurcation point), leading thereafter the system towards a phase-transition and, subsequently, allowing it to acquire a 'new' macroscopic order[18]. In non-equilibrium systems such phenomenon are more frequently encountered, given the richness of instabilities (due to non-linearity), and the consequent transition to very complex dissipative structures. By this way, sensitivity to the external field allows systems to become correlated whit the environmental cues. Keeping in mind the number of transitions that a non-equilibrium system can experience, such sensitivity becomes a basic and relevant feature, unlike in equilibrium systems in which phase transitions are isolated occurrences. By this way the system becomes complex. This is a remarkable and central aspect that permits to link the 'birth' of complexity to the emergence of different stable and unforeseen solutions. This is the background that has allowed Longo and Montévil to claim living, dissipative systems as characterized by extended critical transitions and symmetry changes: "The unifying theoretical framework in biology