Briefings in Bioinformatics Advance Access published October 14, 2012

doi:10.1093/bib/bbs061

Mathematical modeling of biological systems

Santo Motta and Francesco Pappalardo

Submitted: 30th April 2012; Received (in revised form): 23rd July 2012

Abstract

Mathematical and computational models are increasingly used to help interpret biomedical data produced by high-throughput genomics and proteomics projects. The application of advanced computer models enabling the simulation of complex biological processes generates hypotheses and suggests experiments. Appropriately interfaced with biomedical databases, models are necessary for rapid access to, and sharing of knowledge through data mining and knowledge discovery approaches.

Keywords: *mathematical biology; computational models; systems biology*

INTRODUCTION

Revolutions in biotechnology and information technology have produced enormous amounts of data and are accelerating the process of knowledge discovery of biological systems. These advances are changing the way biomedical research, development and applications are conducted. Clinical data complements biological data, enabling detailed descriptions of both healthy and diseased states, as well as disease progression and response to therapies. The availability of data representing various biological states, processes and their time dependencies enables the study of biological systems at various levels of organization, from molecules to organism and even up to the population level [3–5]. Multiple sources of data support a rapidly growing body of biomedical knowledge, however, our ability to analyze and interpret this data lags far behind data generation and storage capacity. Mathematical and computational models are increasingly used to help interpret biomedical data produced by high-throughput genomics and proteomics projects. The application of advanced computer models enabling the simulation of complex biological processes generates hypotheses and suggests experiments. Computational models are set to exploit the wealth of data stored on biomedical databases through text mining and knowledge discovery approaches.

Modeling is the human activity consisting of representing, manipulating and communicating real-world daily life objects. As one can easily realize, there are many ways to observe an object or, equivalently, there are many different observers for the same object. Any observer has 'different views' of the same object, i.e. 'there is no omniscient observer with special access to the truth'. Each different observer collects data and generates hypothesis that are consistent with the data. This logical process is called 'abduction'. Abduction is not infallible, though; with respect to a scientific unknown, we are all blind.

A system is a collection of interrelated objects. For example, a biological system could be a collection of different cellular compartments (e.g. cell types) specialized for a specific biological function (e.g. white and red blood cells have very different commitments). An object is some elemental unit upon which observation can be made but whose internal structure is either unknown or does not exist. The choice of the elemental unit defines the representation scale of the system. A model is a description of a system in

Corresponding author. Santo Motta, Department of Mathematics and Computer Science, University of Catania, V.le A. Doria, 6, 95125 Catania, Italy. Tel.: +39 095 7383073. Fax: +39 095 330094. Email.: motta@dmi.unict.it

Santo Motta, Laurea in Physics (University of Catania,1970) and MSc in Applied Mathematics (University of London, Queen Mary College, 1971). Associate professor of Mathematical Physics (Faculty of Pharmacy, Department of Mathematics and Computer Science, University of Catania). His present scientific interests are BioMaths, BioComputing and BioInformatics.

Francesco Pappalardo is a researcher at the University of Catania. He was a visiting research scientist at the Dana-Farber Cancer Institute in Boston and at the Molecular Immunogenetics Labs, IMGT in Montpellier. His major research area is computational biology.

terms of constitutive objects and the relationships among them, where the description itself is, in general, decodable or interpretable by humans.

Generally speaking, a system is an unknown 'black box' (S) which, under a specific external stimulus (input E) produces a response (output R) [19]. Using this general definition, one can identify three primary scientific uses of models [12]: (i) synthesis or knowledge discovery; to use the knowledge of inputs E and outputs R to infer system characteristics; (ii) analysis and prediction; to use the knowledge of the parts and their stimuli (i.e. the inputs E) to account for the observed response (i.e. the output R) and eventually, to predict response to different stimuli. (iii) Instrumentation or device; to design an 'alternative system' (i.e. hardware or software), able to reproduce the input–output relationship with the best possible adherence to the studied system.

Secondary uses of models account for conceptual frameworks to design new experiments, methods to summarize or synthesize large quantities of data, tools to discover relationships among objects.

In this article, we analyze models and modeling processes specific for the biology. We mainly focus on the use of models aiming at the points (i) and (ii) as tools for knowledge discovering in biology.

The mathematical methods used in modeling biological systems vary according to different steps of the process. We focus on the mathematical representation of the system. However, other important steps in the modeling processes are parameters fitting and model selection. We will not analyze the mathematical methods in those two important aspects as these would require separate review papers. Methods for parameters fitting refer to wide area of mathematical optimization, whereas methods for model selection mainly use statistical techniques. On top of these, sensitivity analysis and phase–space analysis of the models may be required. Interested readers may find more information in these references: [11, 15, 20].

Models for technical use are formal models, but the strategy for building them is quite different and therefore, we leave them out of the present discussion. In the following we will refer to this type of models as Black Box Models (BBM). It is worth pointing out that, as we will mention later on, alternative systems can be considered parts of a large model to account for effects whose origin can be neglected without compromising the understanding of the whole phenomena.

This article is organized as follows: in the next two sections (Models of Systems and The Modeling Process), we describe the types of models and the modeling processes in scientific investigations in a general context; then in the next section (Models in Biology: Scales and Complexity) we go more specific and talk about models in biology and medicine; few examples of models are briefly shown in the section 'Tools and Applications'; finally we draw our 'Conclusions' in the last section.

MODELS OF SYSTEMS

Not all scientific models are expressed in a precise, numerical and quantitative way. Actually, one can identify four different types of models: verbal models, conceptual or diagrammatic models, physical models and formal models.

In this article, we focus mainly on diagrammatic and formal models and we concentrate on the model building process.

Verbal models

In verbal models the system is described in words. These models, based on observations, usually evidence in a simple way the objects and relations among the objects in the system. A verbal model is a rough and sometime ambiguous qualitative representation of the knowledge of the system. These kinds of models are used in the first approach to the analysis of biological system.

Conceptual or diagrammatic models.

In conceptual or diagrammatic models the system is described by a graphical representation of the objects and the relationships describing the underlying dynamical processes. To develop these kind of models the understanding of the available data needs to be sufficient to have a detailed (even if not exhaustive) idea of the objects (or entities) and relations. A conceptual model (CM) represents 'concepts' (objects or entities) and relationships between them. In computer science, CM are also referred to as domain models. A CM is expressly independent from the design and free from implementation concerns. The aim of a CM is to convey the meaning of terms and concepts used by 'domain experts' to rationalize the problem and to find relationships among the different concepts. The CM aims to clarify the meaning of the usually ambiguous terms to minimize as much as possible problems arising from

different interpretations of terms and concepts. If a 'domain ontology' is available, then the meaning of the variety of terms used should be linked to it. Once the domain concepts have been modeled, the model becomes a stable basis for subsequent development of applications in that specific domain. Furthermore, the concepts behind the conceptual model can be mapped to physical design or implementation constructs using either manual or automated code generation approaches.

A conceptual model can be described using various notations, such as Unified Modeling Language (UML) [34, 9], Object Modeling Technique (OMT) for object modeling [32], or Information Engineering or Integration Definition for Information Modeling (IDEF1X) for Entity Relationship Modeling [6]. In UML notation, the conceptual model is often described with a class diagram in which classes represent concepts, associations represent relationships between concepts and role-type of an association represents role types taken by instances of the modeled concepts in various situations.

Physical models

In physical models the representation is done using a mock-up of a real system or object (like a scale model of an aircraft or of a ship). These type of models are mostly of interest for engineers. They are widely used when the properties of the system are almost 'scale-invariant', i.e. independent from the size of the physical model built to represent the real system to produce smaller-scale prototypes.

Formal models

Formal models represent the knowledge of the system using mathematical structures. The mathematical representation of the model depends on the knowledge of the system, on some modeling choices (for instance, the spatial scale of representation) and the aim of the modeling process. There are a large variety of mathematical/computational methods that can be used and the selection of the proper one follows rules that are often matters of experience. At a first glance, there are few questions one may ask to address the choice of the proper mathematical/computational method. Those questions are mainly related to the description of the system with respect to its different parts or components, the physical variables space and time, the type of relations between objects and the object representation *perse*. In systems biology, a system is viewed as an assembly of different parts or compartments (i.e. organs) with different functions. In this case, 'Compartment models' are widely used and each compartment may pick a different mathematical representation. Models can also represent physical variables in different ways. Besides, the model may or not consider the evolution of the system with respect to time (dynamic versus static models). Time can be treated as continuous or discrete variable (time-continuous versus time-discrete models). Likewise, spatial distribution of objects in each compartment may or not be relevant (spatially-heterogeneous versus homogeneous models). Finally, similar objects may be treated as individuals or taken in bulk (particle models versus population models). In the first case, individual objects are identified by a unique state or by a large, but finite number of states (one-state particle versus finite-state automata). Lastly, the relations between objects can be described as deterministic or stochastic rules (deterministic versus stochastic models).

According to the different modeling choices, one can get single versus multicompartments models, including transport, evolutionary differential equations versus algebraic equations or spatial partial differential equation, differential equations versus difference equations, ordinary differential equations (ODE) versus partial differential equations (PDE), kinetic methods, agent-based methods (ABM) or cellular automata (CA) versus ODE or PDE; deterministic methods (ODE or PDE, etc.) versus stochastic methods (stochastic ODE and PDE).

Statistical and artificial intelligence-based models

A statistical model is a formalization of the relationships between variables (i.e. object's measurable characteristics) in the form of mathematical equations, the only difference with the mathematical models described above is that in statistics, all variables and/or parameters of the model include a level of uncertainty. When the relationship between two objects is too complex to be easily guessed, one can resort to probabilistic measures and statistical or artificial intelligence methods to reproduce the response relationship (see e.g. refs. [21, 22, 37]). In these kind of models, the detailed analysis of the system components is usually ignored because, the objective of the model is limited to reproduce the system stimulus/response relation. Examples of this approach are the lumped models using equivalent circuits, neural network, etc.

THE MODELING PROCESS

The modeling process consists of the following steps: (i) model implementation consisting in describing by a formal language the objects/relationships identified in the system under study using a mathematical structure and/or a computer code; (ii) use the model to forecast the system behavior and (iii) evaluate the model adherence to reality by matching predictions with available data.

To find a good model is an issue. Modeling is a hard problem in itself and failure is not a rare event. The modeling procedure is a process in itself that follows a semi-formal set of rules. The methodology lean on four macro steps [31]: (i) understand the problem, i.e. clearly define the questions one asks to the model; (ii) devise a plan for solving the problem, i.e. define a series of steps to be put in practice to find an adequate model of the system under investigations. This step includes knowledge and data acquisition from field experts and literature, model structure, model hypothesis, conceptual model, choosing the appropriate mathematical formalism, solving the formal model, get the results, check model results matching to available data and so on; (iii) execute the plan, i.e. perform the steps in (ii) and (iv) check the correctness of the answer and eventually refine the model. This last point is a major test to evaluate the hypothesis formulated when setting the model.

As mentioned before, we are mostly interested in models for the analysis and predictions. For these models, the classical description of modeling process is shown in Figure 1. It is worth to mention that the schema illustrated in Figure 1 does not have the pretense to be the most general one: it is a general approach that can be used in the analysis and predictions models.

Model objectives

As we already pointed out, a proper definition of the model objectives is a fundamental step as it implies a certain level of comprehension of the problem. The reason for building a model should by clear and a proper clarification of the objective must answer to major questions: (i) what is the system to be modeled; (ii) what are the major questions to be addressed by the model; (iii) How good must the model be and to what it will be compared with? (iv) How we will analyze and use the model output? All these questions need to be clarified before we proceed in searching the current knowledge on the system we wish to model, as some informations may be more relevant than the others.

Current knowledge

A second crucial step in the modeling process is to collect the knowledge on the system under investigation. This is conducted by consulting the scientific literature, including experimental reports and/or discussing with field-experts. In the biomedical field, data sets of literature record (e.g. Pubmed) can highly facilitate the task of browsing the vast amount of information available nowadays. In this respect, methods of data mining and data extraction may be very useful [17].

Model structure

A model is a representation of a real system and has its own structure. For the sake of manageability the model structure should include, of all the acquired knowledge, only those considered to be relevant for the purpose of the study ('Realism'); the level of details of the model results should also be determined *a priori* ('Precision'); finally, a model can be general, that is applicable to other similar systems or specific to the system of interest ('Generality'). Realism, Precision and Generality are competing properties. Each of these properties trades-off against the other two. Deciding a model structure is to find a proper balance between those competing properties, which satisfy the model objectives. Decision on the model structure is crucial for defining the model hypothesis, diagrammatic model construction and mathematical formulation.

Hypotheses

The next step in the modeling process is to translate objectives and current knowledge we wish to include in the model in a list of specific working hypothesis. These are usually verbal statements, but could also be quantitative relationships. Working hypotheses are the basis of the model we are going to built and model results will depend on them. In doing the cycle refinement of model, the starting hypothesis should be critically, repeatedly analyzed.

Conceptual model

The conceptual model is a graphical representation of the relevant system knowledge and model objectives that have been identified in the hypotheses. In the conceptual model compartments, objects and

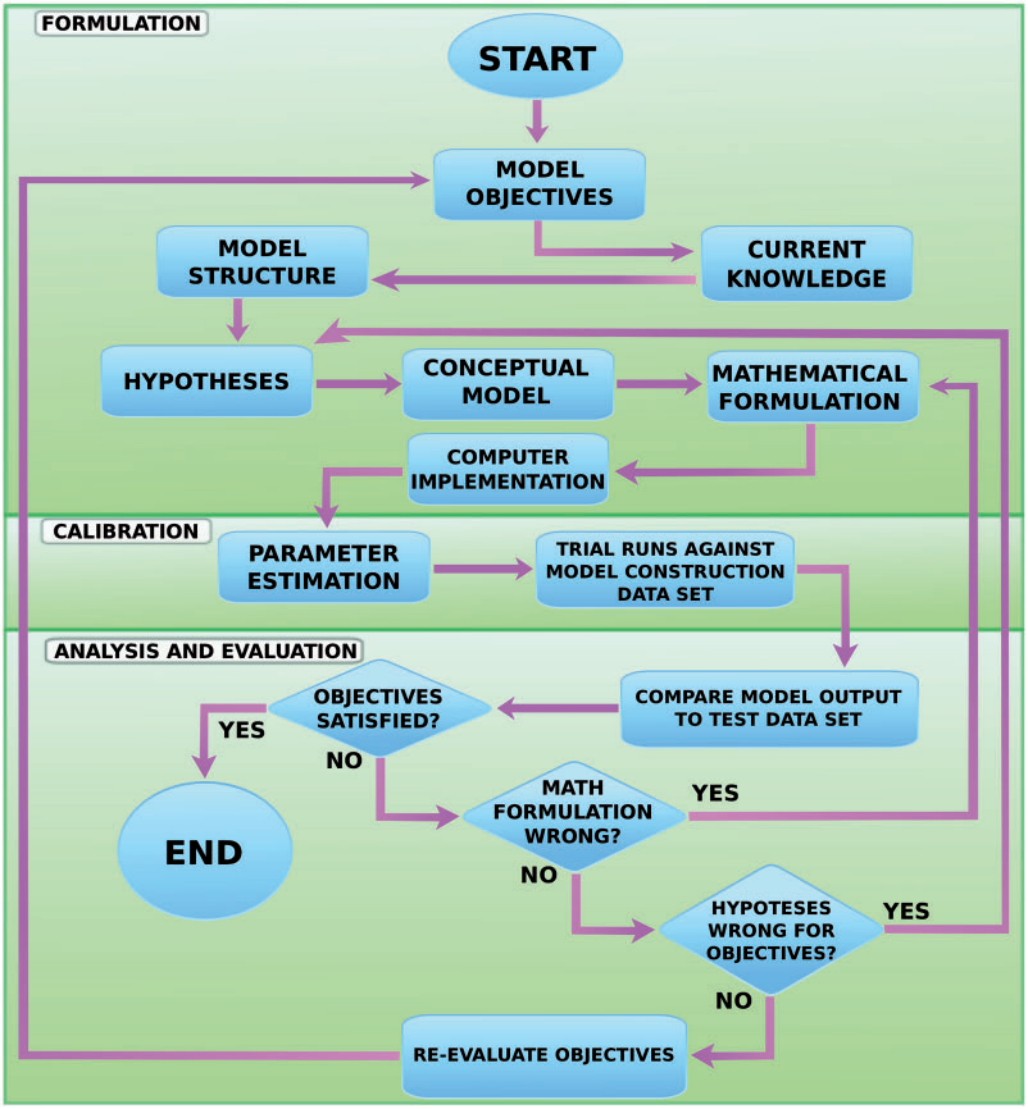

Figure 1: The description of the modeling process. The top part of the figure refers to the formulation of the model, i.e. identification of the model objectives; collection of the current knowledge about the biological system under investigation; choosing the most appropriate model structure to satisfy the model objectives; translate objectives and knowledge into model hypotheses; draw a conceptual model; identify mathematical technique and develop the formal model. Central part of the figure refers to calibration i.e. estimating and fitting model parameters. Finally, the bottom of the figure refers to analysis and evaluation, i.e. comparison of the model results against experimental data sets and analysis of discrepancies.

relations will be described in a diagram where the set of objects are fully clarified and relations bounded.

Mathematical formulation

This is usually the trickiest part of the modeling process, requiring the choice of a mathematical structure, which is appropriate for the model objectives and is able to describe in quantitative form the hypotheses. This step of the process requires a certain level of mathematical sophistication and, more importantly, it requires to define vague concepts and loose relations in strict mathematical terms. Noticeably, model objectives play an important role because, a detailed description of the biological system may turn out to be useless if not required by the model objectives.

Choosing a mathematical formulations is a mapping of the model into the mathematical domain to obtain a formal model. A good formal model must be a compromise between the competing properties of any model (Realism, Precision and Generality) and should take into account some specificity of

the mathematical domain. Accordingly, we can identify the major properties of formal models as follows: 'relevance', capturing the essential properties of the phenomenon; 'computability', transfer model hypothesis into a mathematical/computational infrastructure that can be solved to give the desired results with the required precision; 'understandability', offering a conceptual framework for thinking about the scientific domain; 'extendibility', allowing the inclusion of additional real-world objects in the same mathematical scheme. Taking into account the biological complexity, a very detailed model of a biological system may work out to be unsuitable to automatic resolution, i.e. not computable (for instance, too many equations to be solved or too many unknown parameters to be estimated); in contrast, a too simplistic model may not be able to account for the complexity of the biological system of interest.

A good formal model must be understandable by domain experts so that they could use it for their own quantitative reasoning. Finally, as biology is a fast growing science, extensibility is an important characteristic for biological models. When new objects and relations in the system are derived from laboratory experiments, it should be easy to extend the model with minor changes in the mathematical structure.

In most cases models are equipped with just one mathematical structure of those mentioned previously. However, there are also 'hybrid' models where different mathematical structures are used in combination but also models which add to a detailed description of the system of interest other types of models (e.g. BBM), which mimic the effects of other systems interlinked with the one under investigations.

We will describe some hybrid models in the section 'Tools and Applications'.

From mathematical formulation to the numerical solution

Only very simple models can be analyzed analytically (i.e. by algebraic derivation of the system properties). In most cases the model is either directly implemented as a computer code (i.e. the algorithm-like in ABM) or equations must be solved numerically. Even if there are a variety of methods for solving equations, transferring equations into computer code is a possible source of error and appropriate method to avoid errors due to numerical instabilities need to be carefully chosen. Models of biological systems can involve scores of dynamic variables and thousands of parameters, especially when spatial processes are investigated. In this respect, checking the computer results to match available data is not a trivial exercise.

Parameter estimation and tuning

Once the model formulation has been translated on a computer, a further step is necessary before a simulation can be run. Complex models usually contain many parameters whose numerical value must be determined. In biology this may not be an easy task as numerical values of parameters, if directly measurable, are often known with a large uncertainty. Methods for parameter fitting are mainly based on optimization techniques that minimize the difference between real experimental data and model output. There are a variety of techniques ranging from stochastic methods and gradient descent methods [11]. In some cases, specific experiments may be required for estimating parameters (see e.g. the model of influenza described in ref. [30]). However, a model can contain parameters that cannot be estimated *a priori*. In these cases one can have an *a posteriori* estimation. This means that the values of the parameters can be estimated by *another* model made *ad hoc* just to fit the data available [14, 29].

Model validation and cyclic refinement

Comparison of model results, or simulations are the final part of the modeling process. The general goal of a model is to reproduce data from observation or from experiments (descriptive models) or to predict the result of new observations or experiments (predictive models). Obviously, results need to be validated according to model's objectives. In some cases, a qualitative agreement between model results and experimental data is adequate, in other cases, a quantitative agreement is necessary. It is a common practice for model validation to require the model results to be validated against independent data sets. Model results that fail to fit the experimental data set notwithstanding changes in the model-free parameters, suggest a further model refinement. In this respect, in the way back and forth between model refinement and data validation one can discover interesting properties of the system of interest. The process itself leads to discovering of new knowledge.

MODELS IN BIOLOGY: SCALES AND COMPLEXITY

As already pointed out, any natural phenomena can be observed at different scales thus, in describing the phenomenon by conceptual and quantitative models one needs to choose to appropriate scale to describe the experimental data available.

However, in almost all complex natural phenomena there are aspects that cannot be even observed at just one scale of description (either temporal or spatial). To study these specific facets of reality, multiscale models that represent objects and relationships on different levels of abstraction are required.

Choosing a scale depends on which aspects of the phenomena, from 'micro' to 'macro', one is interested to analyze. In physics this is already a well defined approach that originates from different research areas, and the distinction between different scales is based upon the characteristic lengths of objects and the characteristic time of the phenomena under investigation. For instance, microphysics refers to areas of physics that study phenomena, which take place on the microscopic scale (length scales <1 mm), such as: molecular physics, atomic physics, nuclear physics and particle physics.

In the life sciences the definition of a scale is a bit more ambiguous. A basic unit available for defining a scale is the 'cell' with no regard to its physical dimension. Starting from this, one can define different scales: the 'sub-cellular' or 'intra-cellular' scale, the 'cellular', mesoscopic or 'inter-cellular scale the 'macroscopic scale' and the 'populations scale'.

Models developed at sub-cellular scale deal with the evolution of the physical and biochemical state of a single cell. This scale involves genes, proteins and signals in cell nucleus and surface, which regulate the evolution of the cell and any signaling processing operations of the cells enabling cell crosstalk.

Modeling the overall activity of a single cell is a very hard problem as many biological details of this activity are unknown.

Biologists and modelers have joined forces to develop and use mathematical and computer science techniques in modeling sub-cellular phenomena. Interested readers can find plenty of references in the scientific literature [13].

In the cellular scale, one is interested in describing the evolution of a system consisting of a large number of different interacting cells and molecules. Cell interactions are regulated by signals emitted and received by cells through complex recognition processes. Cellular scale is thus highly connected with the sub-cellular scale but, modeling at this scale, one may forget the details of single cell models and consider them as BBM. The areas of mathematical methods and tools involved in this description refer to statistical mechanics, cellular automata, lattice gas and other similar approaches.

The 'macroscopic scale' include tissues, organs (i.e. a collection of tissues joined in structural unit to serve a common function), systems (i.e. a group of organs working together to perform a certain task) and organism.

In this scale, one is interested in describing the dynamical behavior of observable quantities, in most cases, the concentrations of various entities (cells or molecules). Tissues are usually described using techniques originating from physical continuous systems, i.e. ordinary or partial differential equations or moments of kinetic equations. In describing organs, a model is required to describe both the main tissue and the sporadic tissues but also, and most importantly, the biological function. To model a system, one is required to consider a network of organs that perform a specific task. Depending on the modeling goal the model of a biological system can be arranged with different levels of details. Organs can be fully described in their components or simply as BBM performing a given task. Connections between organs (like lymphatic vessels) can be described physically (dynamical description of the fluid motion in the vessels) or simply considering the flux and the time required to move portions of fluid from different organs, i.e. though law of transport.

Finally, in the population scale, one is interested in describing the dynamics of the populations with respect to one or more characteristics. Models of epidemics or population controls are well known models at this scale. Population dynamics is extremely complex because the effects of all previously mentioned scales and the effects of the environment on a single organism can modify the overall dynamic of the populations. In this class of models, a single organism can or cannot be described in detail, according to the size of the populations one is required to describe. In both cases, changes of the major characteristics of a single organism must be taken into account. For example, to describe the response of a population to large-scale vaccinations (as required in influenza epidemics), one does not describe in detail any single organism but it may be required to

consider the age structure of the population and the effects of environments. At variance for small size populations, like modeling the effect of a new vaccine for a small trial, a sufficiently detailed description of the organism and the effect of the vaccine on each organism should be required. A variety of different techniques are available for these classes of models. In the former case, one uses both, ordinary or stochastic differential equations to describe the populations dynamics or agent-based models (simple agents representing a single organism) to study the resulting complex phenomena. In the latter case, a more detailed description of the organism is required and the population dynamics may eventually be extracted from the dynamic response of each organism.

Complexity and multiscale models

Living organisms are complex systems. Using a 'classical' definition, a complex system is a system composed of different interconnected parts that, as a whole, exhibit one or more properties which do not obviously arise from the properties of the individual parts. System complexity may be either a 'disorganized complexity' or an 'organized complexity'. In the former case complexity arises from a very large number of parts, whereas in the latter case, complexity is intrinsic to the system, eventually with a limited number of parts, and its connections rules.

In living organisms both situations occurs. A living organism is formed by a collection of different parts which are, each of them, organized complex systems. Cells, organs, systems of the human body are each of the complex systems.

As an example, the immune system is one of the very complex one where complexity arises both from a very large number of parts (organs), constituents (cells and molecules) and rules hierarchically connecting different scales of the parts.

Models including many scales of a phenomenon are now requested both for knowledge discovery and drug discovery. In life sciences not only an entire living organism, but also parts of the organism are too complex to be represented in a single, precise, multiscale model. The resulting model would certainly not be computable. Thus, one is forced to break the conceptual model in a set of models describing only part of the phenomenon (like a single organ, or a definite scale) and connect their outcomes [16]. To link models at different scales is a not an easy task. Phenomena occurring at different scales have usually different characteristic time scales and models' output should be properly fitted. Interested readers are referred to another study [3] in this issue.

TOOLS AND APPLICATIONS

Whether we investigate the growth and interactions of an entire population, the evolution of DNA sequences, the inheritance of traits, the spread of disease or the immune system response to a pathogen, biological systems are marked by change and adaptation. Even when they appear to be constant and stable, it is often the result of a balance of tendencies pushing the systems in different directions. The choice of the mathematical approach depends on the biological system one would like to model. In this section, we sketch several applications of mathematical techniques that have been effective in reproducing and in providing new insights of a particular biological problem.

Due to their incredible complexity, models that deal with an entire biological system are, to date, very few and actually incomplete. Instead there are several mathematical models that act toward single or group of components of a biological system.

Tools for bioinformatics and systems biology

Signaling network has a key role in cellular physiology and therefore, it has been widely studied in several organisms. This is due mainly because all cells interact and respond to the environment in which they live. The bad news is that such networks are very complex due to their combinatorial explosion nature. For this reason, frameworks for mapping signal-transduction networks that avoids the combinatorial explosion in some way are particularly needed. In ref. [35], a framework for mapping, visualization and automatic model creation of signal-transduction networks is presented, along with an example of its use to compile the, presently, most comprehensive map of the yeast MAP kinase network.

It is well known that the explosion of data originated from biology has made it increasingly important to provide metadata along side the core data itself. The concept of metadata derives from cart catalogs and libraries by describing the contents

and context of data files, the quality of the original data/files is greatly increased. This metadata may comprise domain-specific information as described by minimal information checklists meant to enable accurate data reuse or may be ontological in nature, specifying more precisely the kind of entities under consideration. The Minimum Information Required in the Annotation of Models Registry (http://www.ebi.ac.uk/miriam) provides unique, perennial and location-independent identifiers for data used in the biomedical domain. In [18], the authors describe the new Identifiers.org service (http://identifiers.org) that is built upon the information stored in the Registry and which provides directly resolvable identifiers, in the form of Uniform Resource Locators (URLs). In the same context, the Systems Biology Graphical Notation (SBGN) facilitates the representation and exchange of complex biological knowledge in a concise and unambiguous manner [36].

Of specific importance are the major synthetic biology platforms, for managing the data, which is used to create synthetic biological systems and to provide mechanisms to begin the process of creating standardized data, algorithms and methodologies for synthetic biology. Here we mention System Biology Workbench (http://sourceforge.net/projects/sbw/), TinkerCell (http://sourceforge.net/ projects/tinkercell), Kappa and all the tools there provided (http://kappalanguage.org/), SBGN (http://www.sbgn.org/), SBML (http://sbml.org/), Synthetic Biology Open Language (http://www.sbolstandard.org/), Clotho (http://clothocad.org/) and BEL Framework $ http://belframework.org/)

Next generation sequence analysis has become an important task both in laboratory and clinical settings. SeqAlto [25] is a new algorithm for read alignment. It is about to 10× faster than existing algorithms, while retaining high accuracy and the ability to align reads with large (up to 50 bp) indels.

Proteins execute and coordinate cellular functions by interacting with other biomolecules. Among these interactions, protein–protein (including peptide-mediated), protein–DNA and protein–RNA interactions cover a wide range of critical processes and cellular functions. Multi-VORFFIP [33] is a tool to predict protein-, peptide-, DNA- and RNA-binding sites in proteins. One of its features is the web interface to facilitate the use of the method and analysis of predictions to non-expert end-users.

Applications

Immunology

The role of mathematical modeling in immunology, one of the most complex fields in biology, were recognized early, beginning from the 1960s and the 1970s. Since then, mathematical models have been used in various domains of immunology [23]. One of the major issues in vaccine and other immunologic approaches' research is the testing of the relevant biological variables when each experiment lasts ≥1 year. One clear example is the scheduling of prolonged vaccinations. It is desirable to reduce as much as possible the number of vaccine administrations, e.g. to reduce the risk of side effects in humans. In ref. [26] the authors describe the use of a mathematical model based on ABM that faithfully reproduce *in silico* the behavior of a cancer-preventive vaccination, suggesting a possible optimized vaccine schedule [27] and highlighting certain critical issues. In particular, although vaccinations could be reduced in numbers without sacrificing efficacy, the intensity of early vaccinations was a key determinant of long-term tumor prevention needed for predictive utility in the model. Moreover, long-term studies confirmed predictions of *in silico* modeling in which an immune plateau phase, once reached, could be maintained with a reduced number of vaccinations; revealing that the accuracy of mathematical modeling of early immune responses is critical. This key example shows that an integrated *in vivo–in silico* approach could improve both mathematical and biological models of cancer immunoprevention. An example of both qualitative analysis of the asymptotic behavior and numerical simulations using nonlinear ODEs is given by the authors [2], where the mathematical modeling of the mammary carcinomaimmune system competition elicited by an external stimulus is presented. A model for keloid formation triggered by virus, their malignant effects and immune system competition have been described using a mathematical model developed by kinetic theory of active particles described in a previous study [1].

In ref. [28] the authors present a mathematical model to analyze the co-stimulatory effects of anti-CD137 monoclonal antibody (mAb) for the melanoma treatment upon synergistic adoptive transfer of activated OT-1 T cells. The reported *in vivo* experiments show that a single administration of anti-CD137 mAb plus activated OT1 T cells is sufficient to completely reject the B16-OVA, whereas

single components or not activated OT1 T cells have no success. The *in silico* experiments performed with the presented computational model show very good agreement with their *in vivo* counterpart. As many aspects of CD137 molecule biology are still not fully understood and the investigating of these aspects requires many difficult and expensive wet lab experiments, the model is a good candidate for becoming a predictive tool.

Circulatory system

Circulatory system represents a biological system made of organs that passes nutrients and other components to and from cells in the body to help fight diseases, stabilize body temperature and pH, and to maintain homeostasis. Diseases associated with this system, i.e. cardiovascular diseases, have a major impact in Western countries. Mathematical models and numerical simulations of cardiovascular system have been presented and have been demonstrated to provide help in understanding both their dynamics and possible interventions. In a previous study [10] the authors provide a general overview of mathematical representation of vascular geometries extracted from medical images, the modeling blood rheology and the complex multilayer structure of the vascular tissue, and its possible pathologies and the mechanical and chemical interaction between blood and vascular walls.

Population dynamics

In another study [8], the authors describe and analyze a periodically forced difference equation model for malaria in mosquitoes that captures the effects of seasonality and allows the mosquitoes to feed on a heterogeneous population of hosts. With the integration of the difference equation model with an individual-based stochastic simulation model for malaria in humans, they compare the effects of insecticide-treated nets (ITNs) and indoor residual spraying (IRS) in reducing malaria transmission, prevalence and incidence. They conclude showing that ITNs are more effective than IRS in reducing transmission and prevalence, proving also that the combination of both interventions is more effective than either intervention alone.

Drug efficacy

Molecular biology is the branch of biology that deals with the molecular basis of biological activity. This field share knowledge with other areas of biology and chemistry, i.e. genetics and biochemistry. Molecular biology provides the understanding of the interactions between the various systems of a cell, including the interactions between the different types of DNA, RNA and protein biosynthesis.

Understanding how drugs and diseases are associated in the molecular level, is of critical importance for better understanding of disease mechanisms and treatments. Recently in a study [38], the authors define a network-based gene closeness profile to relate drug to disease and then develop a Bayesian partition method to identify drug–gene–disease co-modules underlying the gene closeness data. Their mathematical approach and the related simulations are applied to a set consisting of 723 drugs, 275 diseases and 1442 genes. It identified new drug–disease associations and highlighted their molecular basis.

Recently in another study [7], the authors deal with drug resistance that has posed more severe and emergent threats to human health and infectious disease treatment. Due to less knowledge about the underlying mechanisms of drug resistance, wet-lab only approaches achieved limited success. With the use of interactome network of *Mycobacterium tuberculosis* and gene expression data which are treated with two kinds of antibiotics, the authors developed a mathematical workflow for giving new insights to bacterial drug resistance that can be gained by a systematic and global analysis of the bacterial regulation network.

Microbiology is the field of biology that studies microscopic organisms i.e. bacteria, viruses, fungi, prions, protists and prokaryotes. There is a huge quantity of mathematical modeling contributions to this kind of biological systems, especially in the analysis of the dynamics of pathogens. For example, in ref. [24], the authors use differential equations and computational models to characterize the *in vitro* kinetic behaviors of H5N1 avian, H1N1 seasonal and H1N1 2009 pandemic influenza virus strains. The approach provides relevant parameters for identifying and phenotyping potential pandemic strains.

CONCLUSIONS

Biological systems are complex systems and the higher levels of complexity arise from collective behavior and emerging properties at multiple levels. This requires initially the analysis of large quantities low level data either acquired by direct

measurements or by accessing a variety of sources. These data then need to be integrated into various network models or multiscale models. Models are a fundamental step in the scientific discovery. In this article, we described different types of models that have been used in biology for knowledge discovery and predictions. However, building a good model is a hard task. To help interested readers, we analyzed in detail the state-of-the-art in modeling. Furthermore, examples of recent models and applications, at different scales, are included in the final part of the article.

Key Points

- Description of models and their use.
- Analysis of the model process: form models objectives to model verification.
- Brief description of recent mathematical models in biological systems.

References

1. Bianca C. Mathematical modelling for keloid formation triggered by virus: Malignant effects and immune system competition. *Math Models Methods Appl Sci* 2011;**21**: 389–419.
2. Bianca C, Pennisi M. The triplex vaccine effects in mammary carcinoma: A nonlinear model in tune with SimTriplex. *Nonlinear Anal Real World Appl* 2012;**13**: 913–1940.
3. Alemani D., Pappalardo F., Pennisi M., Motta S., Brusic V. Combining cellular automata and lattice boltzmann method to model multiscale avascular tumor growth coupled with nutrient diffusion and immune competition. *Journal of Immunological Methods* 2012;**376**(1–2):55–68.
4. Pappalardo F., Motta S., Lollini P.-L., Mastriani E. Analysis of vaccine's schedules using models. *Cellular Immunology* 2006;**244**(2):137–140.
5. Bianca C., Pennisi M., Motta S., Ragusa M.A. Immune system network and cancer vaccine. International Conference on Numerical Analysis and Applied Mathematics: Numerical Analysis and Applied Mathematics, ICNAAM 2011, Halkidiki, Greece. AIP Conference Proceedings, Volume 1389, 945–948.
6. Chen P, Pin-Shan P. The Entity-Relationship Model - toward a unified view of data. *ACM T Database Syst* 1976; **1**:9–36.
7. Chen LC, Yeh HY, Yeh CY, *et al*. Identifying co-targets to fight drug resistance based on a random walk model. *BMC Syst Biol* 2012;**6**:5.
8. Chitnis N, Hardy D, Smith T. A periodically-forced mathematical model for the seasonal dynamics of malaria in mosquitoes. *Bull Math Biol* 2012;**74**:1098–1124.
9. Larman C. *Applying Uml and Patterns: An Introduction to Object-Oriented Analysis and Design and the Unified Process*. New Jersey, USA: Prentice Hall, 2005.
10. Formaggia L, Quarteroni A, Veneziani A. *Cardiovascular Mathematics: Modeling and Simulation of the Circulatory System*. Milano, Italy: Springer, 2009.
11. Ghosh S, Matsuoka Y, Asai Y, *et al*. Software for systems biology: from tools to integrated platforms. *Nat Rev Genet* 2011;**12**:821–32.
12. Haefner JW. *Modeling Biological Systems: Principles and Applications*. New York, USA: Springer, 2005.
13. Helm V. *Principles of Computational Cell Biology: From Protein Complexes to Cellular Networks*. Morelnbach, Germany: Wiley-VCH, 2007.
14. Ho DD, Neumann AU, Perelson AS, *et al*. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 1995;**373**:123–6.
15. Hofestadt R, Thelen S. Quantitative modeling of biochemical networks. *Stud Health Technol Inform* 2011;**162**: 3–16.
16. Hunter PJ, Crampin EJ, Nielsen PMF. Bioinformatics, multiscale modeling and the IUPS Physiome Project. *Brief Bioinform* 2008;**9**(4):333–343.
17. Younesi E, Kasam V, Hofmann-Apitius M. Direct use of information extraction from scientific text for modeling and simulation in the life sciences. In: *9th International Bielefeld Conference "Upgrading the eLibrary: enhanced information services driven by technology and economics" 2009*. Bielefeld, Germany: Bielefeld University Library. dpi:10.1108/07378830911007637.
18. Juty N, Le Novere N, Laibe C. Identifiers.org and MIRIAM Registry: Community resources to provide persistent identification. *Nucleic Acids Res* 2012;**40**: D580–6.
19. Karplus WJ. The spectrum of mathematical models. *Perspect Comput* 1983;**3**:4–13.
20. Klipp E, Herwig R, Kowald A, *et al*. *Systems Biology in Practice: Concepts, Implementation and Application*. Darmstadt, Germany: John Wiley & Sons, 2005.
21. Lin HH, Ray S, Tongchusak S, *et al*. Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunol*. 2008;**9**:8.
22. Lin HH, Zhang GL, Tongchusak S, *et al*. Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. *BMC Bioinformatics* 2008;**9(Suppl 12)**: S22.
23. Louzoun Y. The evolution of mathematical immunology. *Immunol Rev* 2007;**216**:9–20.
24. Mitchell H, Levin D, Forrest S, *et al*. Higher level of replication efficiency of 2009 (H1N1) pandemic influenza virus thanthose of seasonal and avian strains: Kinetics from epithelial cell culture and computational modeling. *J Virol* 2011;**85**:1125–35.
25. Mu JC, Jiang H, Kiani A, *et al*. Fast and accurate read alignment for resequencing. *Bioinformatics* 2012. doi: 10.1093/bioinformatics/bts450.
26. Palladini A, Nicoletti G, Pappalardo F, *et al*. In silico modeling and in vivo efficacy of cancer preventive vaccinations. *Cancer Res* 2010;**70**:7755–63.
27. Pappalardo F, Pennisi M, Castiglione F, *et al*. Vaccine protocols optimization: in silico experiences. *Biotechnol Adv* 2010; **28**:82–93.
28. Pappalardo F, Forero IM, Pennisi M, *et al*. SimB16: modeling induced immune system response against B16-melanoma. *PLoS One* 2011;**6**:e26523.

29. Perelson AS, Neumann AU, Markowitz M, *et al*. HIV-1 dynamics in vivo: virion clearance rate, infected cell lifespan, and viral generation time. *Science* 1996;**271**:1582–6.
30. Lee HY, Topham DJ, Park SY, *et al*. Simulation and prediction of the adaptive immune response to Influenza A Virus infection. *J Virol* 2009;**83**:7151–65.
31. Polya G. *How to Solve It: A New Aspect of Mathematical Method*. Princeton, USA: Princeton Science Library, 2004.
32. Rumbaugh J, Blaha M, Premerlani W, *et al*. *Object-Oriented Modeling and Design*. New Jersey, USA: Prentice Hall, 1990.
33. Segura J, Jones PF, Fernandez-Fuentes N. A holistic in silico approach to predict functional sites in protein structures. *Bioinformatics* 2012;**28**:1845–50.
34. Siau K, Halpin T. *Unified Modeling Language: Systems Analysis, Design and Development Issues*. Hershey, USA: Idea Group Publishing, 2001.
35. Tiger CF, Krause F, Cedersund G, *et al*. A framework for mapping, visualisation and automatic model creation of signal-transduction networks. *Mol Syst Biol* 2012;**8**:578.
36. van Iersel MP, Villeger AC, Czauderna T, *et al*. Software support for SBGN maps: SBGN-ML and LibSBGN. *Bioinformatics* 2012;**28**:2016–21.
37. Zhang GL, Khan AM, Srinivasan KN, *et al*. Hotspot Hunter: A computational system for large-scale screening and selection of candidate immunological hotspots in pathogen proteomes. *BMC Bioinformatics* 2008;**9**:S19.
38. Shiwen Zhao S, Li S. A co-module approach for elucidating drug-disease associations and revealing their molecular basis. *Bioinformatics* 2012;**28**:955–61.