We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Strengths and Weaknesses of Selected Modeling Methods Used in Systems Biology

Pascal Kahlem et al.[*]
EMBL -European Bioinformatics Institute,
Wellcome Trust Genome Campus, Hinxton, Cambridge,
United Kingdom

1. Introduction

The development of multicellular organisms requires the coordinated accomplishment of many molecular and cellular processes, like cell division and differentiation, as well as metabolism. Regulation of those processes must be very reliable and capable of resisting fluctuations of the internal and external environments. Without such homeostatic capacity, the viability of the organism would be compromised. Cellular processes are finely controlled by a number of regulatory molecules. In particular, transcription factors are amongst the proteins that determine the transcription rate of genes, including those involved in development and morphogenesis. For this reason, the molecular mechanisms responsible for the homeostatic capacity and coordinated behavior of the transcriptional machinery have become the focus of several laboratories.

Modern techniques of molecular genetics have greatly increased the rate at which genes are recognized and their primary sequences determined. High throughput methods for the identification of gene-gene and protein-protein interactions exist and continue to be refined, raising the prospect of high-throughput determination of networks of interactions in discrete cell types. Still, classic biochemical and physiological studies are necessary to identify the targets, and to understand the functions of the encoded proteins. For such reasons, the rate at which pathways are described is much slower than the rate of genome sequencing. The large quantity of available sequences creates the challenge for molecular geneticists of linking genes and proteins into functional pathways or networks. Biologists are often interested in particular subsets of these very extensive networks obtained with

---

[*] Alessandro DiCara[2], Maxime Durot[3], John M. Hancock[4], Edda Klipp[5], Vincent Schächter[3], Eran Segal[7], Ioannis Xenarios[8], Ewan Birney[1] and Luis Mendoza[6].

[1]*EMBL -European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom*

[2] *Merck Serono, 9, chemin des Mines, Geneva, Switzerland*

[3] *UMR 8030 CNRS, Université d'Evry, CEA/Genoscope– CEA, 2Evry, France*

[4]*MRC Mammalian Genetics Unit, Harwell, Oxfordshire, United Kingdom*

[5]*Theoretical Biophysics, Institute for Biology, Humboldt University Berlin, Berlin, Germany*

[6]*Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Ciudad Universitaria, México.*

[7]*Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel*

[8]*Swiss Institute of Bioinformatic, Quartier Sorge - Batiment Genopode, Lausanne Vaud, Switzerland*

high-throughput techniques, subsets that are involved in accomplishing some specific biological objective. Such a view often does not take into account the global dynamical properties of networks, which may be finally necessary to understand the behavior of the system, for instance to correlate genotype and phenotype. This is where computational models of some particular networks will support bench biologists by providing their descriptive and predictive capacity.

With the constant development of faster and more reliable biotechnologies, the scientific community is presented with a growing collection of biological information, some qualitative, some quantitative. Formal databases include catalogs of genes (EnsEMBL (Hubbard et al. 2007)), proteins (UniProt (The UniProt Consortium 2007)), enzymes and their substrates (BRENDA (Chang et al. 2009)) and molecular reactions (Reactome (Matthews et al. 2009)), many from multiple species. Quantitative data resulting from large-scale experiments are also collected in databases; some of them are public, including gene expression (ArrayExpress (Parkinson et al. 2007)), protein interactions (IntAct (Kerrien et al. 2007)), reaction kinetics (SABIO-RK (Rojas et al. 2007)), cellular phenotypes (MitoCheck (Erfle et al. 2007)) and whole organism phenotypes (e.g. EuroPhenome (Morgan et al. 2010)) amongst others. The combination of these biological data with mathematical probabilistic methods to produce models for selected domains of biological systems becomes essential to better understand and possibly predict the behavior of systems in conditions that cannot be experimentally assayed.

However, the diversity of biological systems, the system-specific type of information and the limited availability of data, implies a requirement for the development of adapted modeling methods. Modeling methods must be tailored not only to make the best use of the information available but also to answer specific biological questions, which can span from the understanding of the function of a pathway and its evolution to the global molecular mechanisms underlying cellular events such as cell differentiation for example.

Figure 1 illustrates the tradeoff between network size and the possibility to model different aspects – from topology to dynamics – of the network. We will give here an overview of the current progress of modeling methods illustrated by chosen examples of systems.

2. Sources of data

Data acquisition has been greatly improved by the application of high-throughput screening methods in molecular biology, enabling simultaneous measurement of thousands of molecular events (data points) in a single experiment. Besides, public repositories for high-content biological data (ArrayExpress, IntAct, MitoCheck, amongst others) enable, with the increasing use of standardized annotations, data integration at the system level (Table 1).

The choice of the type of data needed to reconstruct networks depends on the type and size of system under study and is driven by the scientific rationale. Although all sources of experimental data are potentially useful for modeling a system, modeling specifically a signaling pathway will have different requirements in terms of data types to be integrated than a metabolic pathway. Additionally, the granularity of models' description can also be adjusted to incorporate the extent of experimental data available.

The TGF-beta signaling pathway has been extensively studied by the authors using both kinetic and Boolean approaches. Modeling the TGF-beta signaling pathway requires information to build the topology of the network, from recognition of the signal to its transduction to the nucleus, including the down-stream effect on the transcriptome. This information will be obtained from methods of molecular biology and biochemistry. Studying protein interactions will be necessary for predicting the topology of the model,

while assessing protein amounts and phosphorylation states, genes' expression levels, and reaction kinetics will be necessary for recording the dynamics of the model, thereby not only reducing the number of parameters to be predicted, but also defining constraints in the network. Modeling a metabolic pathway will have other data requirements, such as classical biochemistry and genetics to determine the network topology, metabolomics to inform flux parameters and, if the system allows it, phenotypes may also be integrated into the model.

Fig. 1. The size of the network and the available amount of data determine the type of approach to tackle biological questions. For very large networks, exploration is often restricted to topological aspects. Metabolic networks, even of large size can be studied with flux balance analysis, given that stoichiometry is known. For medium size networks, application of Boolean logic allows to detect different qualitative modes of dynamics. Kinetic modeling is mainly restricted to small networks that are already well characterized. Advances in data acquisition and modeling techniques will extend the applicability of different analysis methods to various biological processes in the foreseeable future.

3. Procedures of network reconstruction

Determining the network topology defines the limits of the system under study and enables the incorporation of stoichiometric information and experimental parameters (see Table 1). Over the last decades, systems' modeling has been successfully applied to formulate biological problems in a mathematically tractable form. A number of general approaches and specific models were developed to enable the understanding of biological phenomena: For example, the introduction of metabolic control analysis (Heinrich and Rapoport 1974; Kacser and Burns 1973) pinpointed the fact neglected earlier that metabolic regulation is a property of both the structure of the metabolic network and the kinetics of the individual

	Type of data	Name of the database	URL	Modeling methods		
				ODE	Boolean	Constraint-based
Knowledge Databases	Small chemical entities	ChEBI	www.ebi.ac.uk/chebi		x	
	Gene Ontology	GO	www.geneontology.org		x	
	Taxonomy	Uniprot taxonomy	www.uniprot.org/taxonomy	x		x (localization)
	Genetic diseases	OMIM	www.ncbi.nlm.nih.gov/omim	x		x
	Literature	CiteXplore	www.ebi.ac.uk/citexplore	x		x
	Genomes	Ensembl	www.ensembl.org			
	Nucleotide Sequences	EMBL-Bank	www.ebi.ac.uk/embl			x
	Proteins	UniProt	www.uniprot.org	x	x	x
	Enzymes	IntEnz	www.ebi.ac.uk/intenz	x	x	x
	Structure	MSD	www.ebi.ac.uk/msd			
	Enzymes	BRENDA	www.brenda-enzymes.info	x		x
	Pathways	Reactome	www.reactome.org	x	x	x
	Pathways	KEGG	www.genome.jp/kegg	x	x	x
	Pathways	Chilibot	www.chilibot.net	x	x	
	Pathways	iHOP	www.ihop-net.org/UniPub/iHOP		x	
	Pathways	Panther	www.pantherdb.org/panther		x	
	Pathways	Ingenuity	www.ingenuity.com		x	
	Pathways	GeneGO	www.genego.com		x	
	Pathways	BioCyc	www.biocyc.org	x		
	Genome-scale metabolic models	Thermodynamics SEED models	www.theseed.org/models	x		x
	Biological Models	BioModels	www.ebi.ac.uk/biomodels	x	x	x
	Transporters	TransportDB	www.membranetransport.org	x		x
	Thermodynamics	NIST	xpdb.nist.gov/enzyme_thermodynamics/	x		x
Quantitative Databases	Gene expression	Array Express	www.ebi.ac.uk/aerep	x		x
	Protein interaction	IntAct	www.ebi.ac.uk/intact			x
	Reaction Kinetics	SABIO-RK	sabiork.villa-bosch.de	x		x
	Cellular phenotypes	MitoCheck	www.mitocheck.org			
	Growth phenotypes	ASAP	asap.ahabs.wisc.edu/asap/home.php			x
	Metabolite concentrations	Escherichia coli Multi-omics Database	ecoli.iab.keio.ac.jp			x
	Reaction fluxes	Escherichia coli Multi-omics Database	ecoli.iab.keio.ac.jp			x
	Kinetic Parameters	KMedDB	sysbio.molgen.mpg.de/KMedDB	x		x

Table 1. Examples of sources of information (databases) related with models and methods.

enzymes, instead of only the task of a single rate-limiting enzyme. Detailed metabolic modeling has highlighted among other aspects that knowledge of *in vitro* kinetics of isolated enzymes is often not sufficient to understand metabolism (Teusink et al. 2000). The study of signaling pathways showed how cells process information and has revealed many regulatory motifs including negative or positive feedback and bistability (for overviews see e.g. (Ferrell 2002; Tyson, Chen, and Novak 2003)). Integration of signaling with gene expression, metabolism and biophysical changes demonstrated the contribution of various components of the cellular networks to stress response (e.g. (Klipp et al. 2005)). A number of problems have been tackled with modelling: Among other examples, we find

i. The question of explaining experimentally observed oscillations (e.g. in metabolism). This work resulted in various oscillator models such as the Higgins-Sel'kov oscillator (Higgins 1964; Sel'kov 1968) and more complicated models (e.g. (Hynne, Dano, and Sorensen 2001)) or application to interacting yeast cells (Wolf and Heinrich 2000; Wolf et al. 2000);

ii. Can we understand cell cycle progression from protein interactions? The approaches vary from very simple models (Goldbeter 1991) to comprehensive studies (e.g. (Chen et al. 2004));

iii. What determines robustness of bacterial chemotaxis? Barkai and Leibler proposed a mechanism for robust adaptation in simple signal transduction networks (Barkai and Leibler 1997).

Many of these examples seek to explain emergent properties whose origins are not obvious from a cursory examination of the underlying interactions. In addition, such models helped to establish an abstract language for describing biological observations and to introduce concepts such as equilibrium or steady state (early) or control, stability, robustness or signal amplification (later) into analysis of biological systems.

3.1 Types of networks

Biological networks reflect the regulatory and functional interactions between molecular components (genes, proteins, metabolites) but may also be extended to integrate information on cellular behavior and physiological impact.

The type of information usually represented in networks can be heterogeneous. For instance, the transfer of information through regulatory interactions can be distinguished from the transfer of mass during metabolic reactions.

Computational models are constrained by the amount of information available in the system of interest, which is itself limited to the fabrication of biotechnological tools enabling scientific explorations of various biological systems at the molecular scale. To date, knowledge on prokaryote metabolism is rather complete in comparison to eukaryotic organisms. Similarly, while biological data can be extracted using monocellular cultures in vitro, the transposition of knowledge to multicellular systems remains uncertain.

3.2 Modeling methods
3.2.1 ODE modeling

Among the most frequently applied techniques for modeling dynamic processes in biological networks are systems of ordinary differential equations (ODE) (Klipp 2007). These systems are used to describe the temporal changes of molecular concentrations caused by production, degradation, transport, or modification of the modeled substances. Such changes of concentration are expressed as a function of rates of reaction and appropriate

stoichiometric coefficients. Reaction rates, in turn, can be of several types, such as the mass action law (Guldberg and Waage 1879), the Michaelis-Menten rate law (Briggs and Haldane 1925; Michaelis and Menten 1913), or more complicated forms to attain some specific kinetics (Cornish-Bowden et al. 2004; Klipp et al. 2005; Koshland, Nemethy, and Filmer 1966; Liebermeister and Klipp 2006; Monod, Wyman, and Changeux 1965).

The use of kinetic ODE modeling using rate laws has been the cornerstone of our traditional biochemical terminology and thinking. Such equations have proven quite successful providing a convenient language and conveying immediate meaning since the formulation of the mass action law (Waage and Guldberg 1864), and the modeling of enzyme kinetics (Michaelis and Menten 1913). As a consequence, there is a vast amount of research and huge numbers of publications that have been devoted to the modeling of biochemical reactions using ordinary differential equations (Heinrich and Rapoport 1974; Tyson, Chen, and Novak 2003). A widely successful example of the use of the kinetic approach is the modeling of the cell-cycle control in yeast (Tyson, Csikasz-Nagy, and Novak 2002).

A severe drawback of the ODE approach is the large number of parameters involved in the system of equations. This implies that for any given biological system, there is the need of large sets of experimental data to determine the parameter values of the equations. Moreover, although a number of kinetic parameters are already available in databases, these values are not always applicable for other organisms or other experimental scenarios than for those for which they were measured. For example, kinetic constants usually are developed for metabolic reactions catalyzed by enzymes acting in a test tube, so the appropriateness for modeling in vivo reactions remains to be proven. They may also have been measured in different species from the one under consideration. Alternatively, there is the possibility of parameters estimation (Ashyraliyev, Jaeger, and Blom 2008; Sorribas and Cascante 1994), but such methodology is computationally costly and there is no guarantee that the computational result is biologically correct.

Most ODE models make use of continuous functions to describe the kinetic laws. Continuous functions, however, may not always be appropriate for describing biological processes. For example, given that molecules are discrete entities, the number of molecules as a function in time is in reality a discrete function. Hence, it is important to assess if the use of a continuous function in a given model is a reasonable approximation to reality. As a rule of thumb, if the experimental error at measuring the real value of a variable is larger than the jump in the discrete value, then it is usually harmless to replacement discrete functions by continuous ones.

The representation of chemical species as a concentration with the use of continuous variables also assumes that the system is absolutely uniform. However, in reality biochemical systems frequently exhibit a large degree of spatial heterogeneity due to processes such as compartmentalization, molecular association, or restricted diffusion. It has been mathematically demonstrated that changes in the spatial distribution of molecules have a large impact on the dynamical behavior of a biochemical system (Zimmerman and Minton 1993).

Often ODE models are deterministic, but under certain circumstances, a deterministic approach does not give an adequate representation of the biological system. At the molecular scale, individual molecules randomly collide with one another, allowing for a chemical reaction to occur only if the collision energy is strong enough. This effect is observed for small volumes; when using deterministic equations the smaller the volume the less accurate the model becomes (Ellis 2001; Erdi and Toth 1989). Therefore, it is important

to make sure that the system under study is large enough to avoid stochastic fluctuations (Fournier et al. 2007) which could be amplified and thus originate observable macroscopic effects.

3.2.2 Discrete networks

A network is a system formed by multiple nodes, each associated with a state of activation, which depends upon the state of a set of nodes. A common practice is to represent nodes with the use of continuous variables. However, usually there is only qualitative experimental data regarding the activity of most genes and/or proteins. For such reasons, a useful simplification is to suppose that genes can attain only a finite number of possible states, thus allowing their representation with discrete variables. In the simplest case, a gene might be "turned on" (or "active", or "1") or "turned off" (or "inactive", or "0") at a given time. In this case, we are dealing with binary networks, also known as Boolean networks. Boolean networks were first presented by Kauffman (Glass and Kauffman 1973; Kauffman 1969) so as to give a qualitative description of the concerted action of a group of genes during cellular differentiation. Such models were originally developed as a suitable simplification for the analysis of genetic regulation, and were originally studied exclusively from a statistical point of view (Kauffman 1993) due to the lack of biological data on experimentally validated biological networks. More recently, Boolean network models have been developed for a series of biological systems showing the suitability of this methodology to capture key aspects of cellular development and differentiation (Albert and Othmer 2003; Davidich and Bornholdt 2008; Faure et al. 2006; Gupta et al. 2007; Huang and Ingber 2000; Kervizic and Corcos 2008; Li et al. 2004; Mendoza, Thieffry, and Alvarez-Buylla 1999; Saez-Rodriguez et al. 2007; Samal and Jain 2008).

The use of Boolean models has permitted the discovery of the influence of the network topology on its dynamical behavior. Specifically, most studied networks include feedback loops or circuits. It has been shown that their presence is necessary to ensure multistationarity and homeostasis, which are particularly important properties of biological systems. The logical analysis of feedback loops decomposes any network into a well-defined set of feedback loops. It was first developed by Thomas (Thomas 1973), and formally demonstrated by others (Gouzé 1998; Plahte, Mestl, and Omholt 1995; Snoussi 1998). Negative feedback loops generate homeostasis in the form of damped or sustained oscillations. The importance of homeostasis in maintaining the internal environment of an organism is well known and dates from the work of (Cannon 1929). Conversely, positive feedback loops generate multiple alternative steady states or multistationarity. The biological interpretation of multistationarity as cellular differentiation goes back to Delbrück (Delbrück 1949), and has been developed by the group of Thomas (Thieffry et al. 1995; Thomas 1973).

Boolean networks are widely used for their computational tractability and their capability of providing qualitatively correct results; there is, however, an important issue to take into account. Boolean networks describe time as a discrete variable, hence there is the need to decide at each clock tick which nodes of the network are going to be updated. On the one hand, in the synchronous approach all nodes are updated at each time step. This methodology is the easiest to implement but also the less realistic, since it is highly unlikely that all molecules in the modeled network have the same time response. On the other hand, in the asynchronous approach only one node is updated at each time step. While this approach is closer to reality, there is usually no experimental information regarding the

correct order of response. Worse still, the group of stable states attained by the system using the synchronous and asynchronous approaches are not necessarily identical. It is therefore advisable to use both methodologies, with the aid of a modeling software (Garg, Banerjee, and De Micheli 2008; Gonzalez et al. 2006), and then use biological knowledge to decide among all possible outcomes.

3.2.3 Qualitative modeling

The use of continuous variables provides fine-granularity modeling, thus allowing for the description of a richer dynamical behavior. However, experimental data to support parameter fitting is very scarce; hence the development of quantitative models of regulatory networks is limited to a small set of experimental systems for which a large quantity of molecular data has been gathered (for an example see (Jaeger et al. 2004)). The alternative approach of modeling with the use of Boolean networks is not always possible, though, usually because of the lack of experimental information to infer the logical rules governing the response of nodes. There are, however, intermediate modeling methodologies lying between the coarse-grained binary approach and the fine-grained use of ordinary-differential equations: among them it is possible to find the use of piecewise-linear differential equations, qualitative differential equations and standardized qualitative dynamical systems.

Piecewise-linear models have been proposed for the modeling of regulatory networks. These equations, originally proposed in (Glass and Kauffman 1973) have been amply studied (Glass and Pasternack 1978; Gouzé and Sari 2002; Mestl, Plahte, and Omholt 1995; Plahte, Mestl, and Omholt 1994) from a theoretical point of view. Variables in the piecewise-linear approach represent the concentrations of proteins, while the differential equations describe the regulatory interactions among genes encoding these proteins. Each differential equation contains two terms, namely the activation part consisting of a weighted sum of products of step functions and the decay rate. The mathematical form of these equations divides the state space into multidimensional boxes, and inside the volume of each box the equations are reduced to linear ODEs, making the behavior of the system inside a given volume straightforward to analyze. Nevertheless, the global behavior of these systems of equations can be very complex, with chaotic behavior being rather common (Lewis and Glass 1995; Mestl, Bagley, and Glass 1997). Despite this drawback piecewise-linear differential equations have been used to analyze several regulatory networks of biological interest (De Jong et al. 2004; Ropers et al. 2006; Viretta and Fussenegger 2004).

In the case of the use of qualitative differential equations, the dependent variable takes a qualitative value composed of a qualitative magnitude and a direction. Here, the qualitative magnitude is a discretization of a continuous variable, while the qualitative direction is the sign of its derivative. Furthermore, each equation is actually a set of constraints that restrict the possible qualitative values of the variable. To solve the system, it is necessary to create a tree of possible sequences of transitions from the initial state. Now, this characteristic makes the methodology difficult to apply for large biological systems, since the trees describing the dynamical behavior rapidly grow out of bounds. This scalability problem has restricted the application of qualitative equations to a small number of models (Heidtke and Schulze-Kremer 1998; Trelease, Henderson, and Park 1999).

While Boolean models approximate a continuous sigmoid by a discontinuous step function, the standardized qualitative dynamical systems method goes the other way around. Within this methodology, the network is modeled as a continuous system using a set of ordinary

differential equations. These equations describe the rate of change of activation (or synthesis, or transcription) as a sigmoid function of the state of activation of the controlling input variables. The variables representing the state of activation are normalized, so that they are constrained in the range [0,1]. This feature enables a comparison of the dynamics against the results of a purely binary model. The characteristic that distinguishes this method from other approaches is that the equations are normalized and standardized, i.e. they are not network-specific. Moreover, models can be developed even in the total absence of information regarding the molecular mechanisms of a given network because no stoichiometric or rate constants are needed. Given that this qualitative method was recently developed, it has been used to model a small number of biological networks (Mendoza and Pardo 2010; Mendoza and Xenarios 2006; Sanchez-Corrales, Alvarez-Buylla, and Mendoza 2010). The standardized qualitative dynamical systems methodology has been used for the development of a software package for the automated analysis of signaling networks (http://www.enfin.org/squad).

3.2.4 Constraint-based modeling

Analyzing the dynamics of large-scale metabolic networks using kinetic modeling techniques is hampered by the size of the biological system because of the large number of parameters that need to be fitted. Constraint-based modeling bridges the gap between the mere static representation of graph-based methods and the detailed dynamic pathway analyses of kinetic modeling techniques (Feist and Palsson 2008). This framework aims at simulating at the cellular level the global dynamics of metabolism, using a limited amount of information. To that end, the method is based on the description of all reaction fluxes that are compatible with constraints deriving from basic physical assumptions, specific biological information or experimental measures. The assumption of steady state dynamics, for example, simplifies the analysis by only requiring knowledge on the reaction stoichiometries, which can be easily obtained from metabolic databases (Kanehisa et al. 2006; Karp, Paley, and Romero 2002; Matthews et al. 2009). Other information such as reaction reversibility (Joyce and Palsson 2006), flux measurements, gene expression (Shlomi et al. 2008) and metabolite concentration (Henry et al. 2006; Kummel, Panke, and Heinemann 2006) can be incorporated into the modeling if available (Durot, Bourguignon, and Schachter 2009). Compared to kinetic models, constraint-based models are therefore easily reconstructed at large scale (Feist et al. 2009) and online resources that provide automatically reconstructed models are currently being launched (Henry et al. 2010).

A set of constraints narrows the total number of possible flux states, which represent the set of possible metabolic behaviors of the system. In order to explore this set of attainable flux distributions, a number of mathematical and computational methods have been designed (Durot, Bourguignon, and Schachter 2009). Some of them attempt to describe the overall set of possibilities either by enumerating all basic independent metabolic routes, known as elementary modes or extreme pathways (Papin et al. 2004), or by randomly sampling the space of flux distributions (Schellenberger and Palsson 2009). Other methods focus on specified metabolic objectives, e.g. the production of biomass components, and look for flux distributions that optimize it (Varma and Palsson 1994).

The constraint-based modeling framework utilizes a set of equations describing metabolites, where each equation contains as many unknowns as there are fluxes for a given metabolite. Very often the number of unknowns is larger than the number of equations. This is the main drawback of the approach, since it means that the system of equations has an infinite number of solutions. This problem is usually solved by maximizing some functions, which

in turn assumes that the modeled organism has evolved optimal metabolic pathways, which is not necessarily true. Despite this limitation, the constraint-based modeling approach has been successfully applied for predicting growth phenotypes for knock-out mutants on varying environments, analyzing essentiality or dispensability, integrating genome-scale experimental data, and driving metabolic engineering designs (Durot, Bourguignon, and Schachter 2009; Feist and Palsson 2008).

3.2.5 Probabilistic graphical models

Probabilistic graphical models (Pearl 1988) represent another class of methods that have recently gained much popularity in studies of gene regulation. Such models always define a joint probability distribution over all the properties in the domain, where a property is represented by a random variable, which is either hidden (e.g., the functional module that a gene belongs to) or observed (e.g., the expression level of a gene). Random variables and the probabilistic dependencies among them define a probabilistic graphical model, which provides a statistical framework for representing complex interactions in biological domains in a compact and efficient manner.

In contrast to procedural methods, such as data clustering followed by motif finding (Tavazoie et al. 1999), probabilistic graphical models are, as their name implies, declarative, model-based approaches, where by a model we mean a simplified description of the biological process that could have generated the observed data. An important property of these models is their ability to handle uncertainty in a principled way, which is particularly useful in the biological domain, due to the stochastic nature of the biological system and due to the noise in the technology for measuring its properties. The details of the models (dependency structure and parameters) are typically learned automatically from the data, where the goal is to find a model that maximizes the probability of the model given the observed data. The main difficulty with this approach is that the learning task is challenging, as it involves several steps, some of which are computationally intractable.

Despite the computational difficulties of the probabilistic approach, it has been successfully used in several cases. One example is the reconstruction of the structure of regulatory networks (Friedman et al. 2000). The main idea is that if the expression of gene A is regulated by proteins B and C, then A's expression level is a function of the joint activity levels of B and C; this can easily be extended to networks composed of modules of co-regulated genes (Gasch et al. 2000; Segal et al. 2003). Another example is the case where probabilistic graphical models were applied to identify the cis-regulatory motifs through which transcriptional programs occur (Segal, Yelensky, and Koller 2003), with the aim of identifying modules of co-regulated genes and explain the observed expression patterns of each module via a motif profile that is common to genes in the same module.

4. Strengths and weaknesses of modeling methods

The methods chosen to model a given system depend mostly on the qualitative and quantitative features of biological data available to construct the model and on the biological question that is asked. To be accurate, an ODE model will require as much experimental information as possible since all parameters of the network are represented and missing parameters will be estimated to fit experimental values. Boolean modeling minimizes the network to the essential nodes that are subjected to decisional events and feedback loops. Although such a reductionist method does not involve kinetic parameters, it remains

capable of predicting the logical behavior of the network and identifying steady states. Although these two previous methods work with currently size-limited networks, global metabolic modeling using the constraint-based method allows the comparative analysis of thermodynamic fluxes between chosen states of a system. Probabilistic graphical models integrate different types of biological information, such as gene expression data for example, in order to reconstruct the networks and predict their behaviors.

The models studied by the authors are developed in synergy by wet and dry laboratories, which cycle between computational modeling and experimental testing. For example, the human TGF-beta signaling pathway has been modeled using ODE modeling by the group of E. Klipp (Zi and Klipp 2007) and using Boolean modeling by the group of I. Xenarios. The ODE modeling required a large amount of biochemical information, or used parameter estimation methods to predict missing values. However, since this method is based on real biological data, it allowed investigating precisely early time points of the TGF-beta signaling cascade. The Boolean method required obviously less biochemical information than the ODE, which allowed to compute accurately a large network articulated across several positive or negative feedback loops. However, qualitative models despite looking fairly simple to construct pose several issues when one wants to validate experimentally the proposed predictions. Each component (node) of a model is represented as an abstraction of a molecular entity. For example, the interferon gamma receptor complex will be represented in qualitative models as a single entity (node). However, from a biochemical point of view interferon gamma receptor is composed of three subunits (a,b,g). This raises the question whether all the subunits should be experimentally measured or only one of them. This situation was experienced by the authors who modelled the TGFb pathway: Several predictions were produced after reconstructing the Boolean network. One of them was that knocking-out one component introduced an oscillation of several components of the TGF-beta pathway. However, upon experimental validation, the nodes that were suggested to oscillate did not show any changes at the mRNA/protein expression or phosphorylation states during a time-course experiment. The model did not indicate at what molecular level the predicted oscillation would be measurable.

According to the experience acquired with the models developed by the authors as part of the ENFIN Consortium, we have assessed the relationship between modeling frameworks and the biological questions which could be answered in each case (Table 2).

Besides the biological question and the subjective choice of the modeler, the ability to access experimental data can be critical for the choice of a modeling method. This is schematically represented in the decision tree (Figure 2).

Testing entire model predictions with wet experiments is often impossible, because of simple limitations of available technologies. Experimental assays are therefore designed to target key components of models, a strategy often taken in industry to prioritize pharmaceutical targets. To our experience, the integration of several disciplines improves the coverage of computational analysis. Improving computational models can also require repetition of experiments potentially needing expensive settings, but which may not result in genuine scientific discovery. For instance, there is no point measuring some parameters that may be useful for the model if the accuracy of measurement is too poor to resolve between alternative model versions. Problems may also arise concerning the spatial resolution of experimental data. For example, how useful are whole animal metabolomics in metazoans for assessing the behavior of a particular cell type?

Biological Question	Boolean networks or qualitative models (of signaling or regulatory networks)	Probabilistic graphical models of regulatory networks	Constraint-based models (purely metabolic or complemented by a Boolean regulatory layer)	Dynamic ODE models
Reconstruct the interaction structure of the cellular process from experimental data	** using expression data, protein interaction data, perturbation experiments (siRNA, overexpression studies)	*** using dedicated learning algorithms (see III.2.d) with expression data optionally complemented w/other data types	*** using genome-annotation, physiological information and growth phenotype data	N/A : model is typically reconstructed manually (the structure of the underlying qualitative model may sometimes be reconstructed from data)
Identify the values of key parameters determining the behavior of the process, using experimental data	** using algorithms that learn Boolean rules compatible with prior knowledge	*** for parameters defining the probabilistic dependencies between steady-state expression levels	N/A, in general (no parameters to learn). *** for metabolite concentration ranges using CBMs extended with thermodynamics constraints. ** for regulated metabolic models using gene-expression data	*(**) depending on the size of the model to be identified; Different types of sensitivity analysis, such as Metabolic Control Analysis
Identify inconsistencies between model predictions and experiments, and use them to refine model	** using the algorithms mentioned above (cases where identified steady state do not correspond to prior knowledge, search of missing regulatory mechanisms)	** using the algorithms mentioned above	*** using growth-phenotype data. ** using metabolomics data and CBMs extended with thermodynamics constraints	* using simulation and comparison with experiments: (i) with data used for model fitting, (ii) with new data. Use criteria for goodness of fit
Characterize the topology of the interaction network	*** using the underlying directed interaction graph	*** using the underlying interaction graph	*** using the underlying interaction graph	*** using the underlying interaction graph
Predict steady-state behavior under well-defined environmental conditions	*** using steady-state/attractor analysis methods and feedback-loop analysis to determine stability	N/A (or use qualitative model derived from probabilistic representation)	*** using FBA or related phenotype prediction methods	*** for small networks with known quantitative parameters, or larger networks using qualitative models
Predict change in steady-state behavior after genetic or environmental perturbation	*** same as above, after modification of nodes, edges within the Boolean regulatory network	N/A	*** same as above, after inactivating reactions and/or modifying exchange fluxes	** same as above, after modification of the set of ODEs (but parameters may change with genetic perturbation)
Characterize properties of the set of attainable steady-state behaviors of the process	** Use of the synchronous and asynchronous Boolean update to identify subsets of states leading to steady states process	?	*** using methods such as elementary-mode analysis, flux-coupling analysis, probabilistic descriptions.	* using small networks: analytically, large networks by applying Monte-Carlo-type sampling of parameter space
Simulate transient dynamic behavior / reaction to external stimuli	Similar as knock-out and overexpression processes described in genetic and environmental perturbation	N/A	* using dynamic FBA growth simulations	*** by simulation, in cases supported by bifurcation analysis

Table 2. Relationships between modeling frameworks and biological questions. Each column represents a model type; each row represents a type of question. The cells contain 2 types of information: 1) the degree of relevance of the modeling framework to the question (indicated by 1 to 3 stars; N/A stands for not applicable); 2) a short explanation on how the approach enables to answer the question (which method or data are used, for instance).

Fig. 2. Decision tree: which type of data is necessary to perform a certain type of analysis?

Given the great development computational approaches in systems biology, it has become urgent to establish methods to assess the accuracy of these in specific contexts. Challenges have been organized for example in protein structure assessment (CASP (Moult et al. 2007)) or in network reconstruction (DREAM (Stolovitzky, Monroe, and Califano 2007)). In this context, one application of models can be to challenge the topology of computationally reconstructed networks *in silico*. Models can thus both be improved with new data but also be used to test the likelihood that these data make sense in the context of the already validated model's components.

5. Examples of reconstruction procedures

5.1 T-helper model

The vertebrate immune system encompasses diverse cell populations. One of these is made of CD4+ lymphocytes, or T-helper cells. While these cells have no cytotoxic or phagocytic activity, they coordinate several cellular and humoral immune responses via the release of cytokines, which influences the activity of several cell types. In vitro, T-helper cells can be further subdivided into precursor Th0 cells and effector Th1 and Th2 cells, depending on their pattern of secreted molecules. Various mathematical models have been proposed to describe the differentiation, activation and proliferation of T-helper cells. Early models aimed to describe the cellular interactions of these cells, mediated by the secretion of cytokines. More recent models have been developed to describe the molecular mechanism

that determine the differentiation process of these cells, as well as the determination of their molecular markers. There is currently a lack of quantitative data on the levels of expression of the molecules involved in the differentiation process of T-helper cells. There is, however, a vast amount of qualitative information regarding the regulatory interactions among many such molecules. As a result, it has been possible to reconstruct the basic signaling network that controls the differentiation of T-helper cells. Hence, despite the lack of quantitative data, it has been possible to model this network as a dynamical system at a qualitative level. Specifically, this network has been studied by modeling it as a discrete dynamical system (Mendoza 2006), a continuous dynamical system (Mendoza and Xenarios 2006), a Petri Net (Remy et al. 2006), and a binary decision diagram (Garg et al. 2007). Also the model has been recently updated to include the description of Th17 and Treg cell types (Mendoza and Pardo 2010). Despite the very different approaches used to model the T-helper signaling network, they all reach comparable results. Specifically, they show that the network has fixed points that correspond to the patterns of activation or expression observed in the Th0, Th1 and Th2 cell types. Furthermore, such models are capable of describing the effect of null-mutations, or over-expression of some molecules as reported by several experimental groups. The consistency among the results of several modeling approaches on the same signaling network shows that the qualitative dynamical behavior of the network is determined to a large extent by its topology. Moreover, these models show that it is possible to develop dynamical models at a qualitative level, and that such models are indeed useful to describe and predict relevant biological processes.

5.2 Genome-scale metabolic model of *Acinetobacter baylyi* ADP1

Acinetobacter baylyi ADP1, a strictly aerobic gamma-proteobacterium, is a good model organism for genetic and metabolic investigations (Metzgar et al. 2004; Young, Parke, and Ornston 2005) as well as for biotechnological applications (Abdel-El-Haleem 2003) thanks to its metabolic versatility and high competency for natural transformation. Following its sequencing and expert annotation (Barbe et al. 2004), a genome-wide collection of single-knockout mutants was generated and mutant growth phenotypes were assayed in selected environmental conditions (de Berardinis et al. 2008). In order to interpret these phenotype results and assess their consistency with the previous biochemical knowledge, a genome-scale constraint-based model of its metabolism was reconstructed (Durot et al. 2008). In a first step, an initial model was built using data from *A. baylyi*'s genome annotation, metabolic pathways databases (e.g. KEGG, BioCyc or Reactome) and physiological knowledge gathered from the literature. Such reconstruction of the overall set of biochemical reactions present in an organism has already been performed for more than 15 species (Reed et al. 2006). While facilitated by metabolism-related tools such as Pathway Tools (Karp, Paley, and Romero 2002), the task is still labor intensive and requires extensive manual curation. This initial model of *A. baylyi* included 870 reactions, 692 metabolites and 781 genes. In a second step, experimental mutant growth phenotypes were systematically compared to growth phenotypes predicted by the model. Inconsistencies between predictions and experimental results revealed potential gaps or errors in the current model: they were therefore examined to look for interpretations and possible corrections. Out of 127 inconsistencies, 60 led to a model correction. Each correction involved a modification of the gene-reaction associations, the metabolic network or the biomass requirements. Explanation of the remaining inconsistent cases would require further experimental investigations. In

several cases, hypothetical interpretations involving biological processes lying outside the model scope (e.g. regulation) could be proposed. The result of that model refinement process was an increase in growth phenotype predictive accuracy from 88% to 94%. In addition, the improved version of the model integrated a significant fraction of the additional information resulting from large-scale mutant phenotyping experiments within the metabolic knowledge on *A. baylyi* derived from its genome annotation and the literature.

5.3 Nicotinamide nucleotide transhydrogenase (Nnt) pathway

Insulin secretion in mammals is carried out by pancreatic β-cells in response to the glucose concentration in the surrounding environment. The insulin secretion response is a complex one, incorporating many well-conserved biochemical reactions coupled to a cell-type-specific insulin secretion process driven by intracellular ATP concentration. The kinetic core model of pancreatic β-cell glucose-stimulated insulin secretion (Jiang, Cox, and Hancock 2007) simulates the relationship between incoming glucose concentration and resultant cytoplasmic ATP concentration using a set of ODEs representing 44 enzymatic reactions, some of which take place in more than one compartment and simulates the concentrations of 59 metabolites. The model is divided into three compartments: cytoplasm, mitochondrial matrix and mitochondrial inter-membrane space. Information on components of the model was collected from the literature and from SABIO-RK (http://www.sabio.villa-bosch.de/SABIORK/). Initial validation of the model was based on its ability to replicate published properties of the system in vivo, such as response to changing glucose concentration and oscillation of the concentration of certain metabolites in the glycolysis pathway and elsewhere. Further characterization and optimization will require *in vivo* measurement of some of these concentrations in an appropriate cell type.

6. Discussion

We conclude along the lines of the famous quote of G.E.P. Box: "All Models Are Wrong But Some Models Are Useful" (Segal et al. 2003). Most models are often focused on particular scientific domains and integrate a limited set of information, considered being "good-enough" to answer the scientific questions of a given project. Many uncertainties still remain in the interpretation of computational models. Those uncertainties emanate from different levels of variation: technology, laboratory-to-laboratory, human-to-human specific, among others. The true biological variation or stochasticity is harder to detect. For example, when there is significant variation between individual runs of an experiment, how should the data be integrated in the model? Should individual models be created for each case, and these compared, or should data be averaged to produce a consensus model? When is a model considered to contain most of the behaviors of the real system?

The different examples represented in this review are not exhaustive by far and only reflect our limited knowledge of the interface between mechanistic modeling and "wet" experiments, based on the research performed within the ENFIN Consortium.

Ultimately providing both models and experimental data to a wider community is a way to bridge the gap and transmit our knowledge to the next generation of scientists. An effort leading in that direction is the development of databases of models such as BioModels (www.biomodels.org) (Le Novere et al. 2006), which collects molecular models by using a

standardized format. Another important development is the formulation of standards for experiments and for model formulation (such as MIRIAM (Le Novere et al. 2005)). Standards and associated controlled vocabularies will certainly contribute to unifying experimental data models that are currently scattered across several databases.

7. Acknowledgements

8. References

Abdel-El-Haleem, D. 2003. Acinetobacter: environmental and biotechnological applications. *Afr. J. Biotechnol.* 2 (4):71-4.

Albert, R., and H. G. Othmer. 2003. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *J Theor Biol* 223 (1):1-18.

Ashyraliyev, M., J. Jaeger, and J. G. Blom. 2008. Parameter estimation and determinability analysis applied to Drosophila gap gene circuits. *BMC Syst Biol* 2:83.

Barbe, V., D. Vallenet, N. Fonknechten, A. Kreimeyer, S. Oztas, L. Labarre, S. Cruveiller, C. Robert, S. Duprat, P. Wincker, L. N. Ornston, J. Weissenbach, P. Marliere, G. N. Cohen, and C. Medigue. 2004. Unique features revealed by the genome sequence of Acinetobacter sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic Acids Res* 32 (19):5766-79.

Barkai, N., and S. Leibler. 1997. Robustness in simple biochemical networks. *Nature* 387 (6636):913-7.

Briggs, G. E., and J. B. Haldane. 1925. A Note on the Kinetics of Enzyme Action. *Biochem J* 19 (2):338-9.

Cannon, W. B. 1929. Organization for physiological homeostasis. *Physiol. Rev.* 9:399-431.

Chang, A., M. Scheer, A. Grote, I. Schomburg, and D. Schomburg. 2009. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res* 37 (Database issue):D588-92.

Chen, K. C., L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson. 2004. Integrative analysis of cell cycle control in budding yeast. *Mol Biol Cell* 15 (8):3841-62.

Cornish-Bowden, A., M. L. Cardenas, J. C. Letelier, J. Soto-Andrade, and F. G. Abarzua. 2004. Understanding the parts in terms of the whole. *Biol Cell* 96 (9):713-7.

Davidich, M. I., and S. Bornholdt. 2008. Boolean network model predicts cell cycle sequence of fission yeast. *PLoS ONE* 3 (2):e1672.

de Berardinis, V., D. Vallenet, V. Castelli, M. Besnard, A. Pinet, C. Cruaud, S. Samair, C. Lechaplais, G. Gyapay, C. Richez, M. Durot, A. Kreimeyer, F. Le Fevre, V. Schachter, V. Pezo, V. Doring, C. Scarpelli, C. Medigue, G. N. Cohen, P. Marliere, M. Salanoubat, and J. Weissenbach. 2008. A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADP1. *Mol Syst Biol* 4:174.

De Jong, H., J. Geiselmann, G. Batt, C. Hernandez, and M. Page. 2004. Qualitative simulation of the initiation of sporulation in Bacillus subtilis. *Bull Math Biol* 66 (2):261-99.

Delbrück, M. 1949. *Discussion In: Unités Biologiques Douées de Continuité Génétique* Lyon: CNRS.

Durot, M., P. Y. Bourguignon, and V. Schachter. 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS microbiology reviews* 33 (1):164-90.

Durot, M., F. Le Fevre, V. de Berardinis, A. Kreimeyer, D. Vallenet, C. Combe, S. Smidtas, M. Salanoubat, J. Weissenbach, and V. Schachter. 2008. Iterative reconstruction of a global metabolic model of Acinetobacter baylyi ADP1 using high-throughput growth phenotype and gene essentiality data. *BMC systems biology* 2:85.

Ellis, R. J. 2001. Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* 26 (10):597-604.

Erdi, P., and J. Toth. 1989. *Mathematical Models of Chemical Reactions*. Manchester: Manchester University Press.

Erfle, H., B. Neumann, U. Liebel, P. Rogers, M. Held, T. Walter, J. Ellenberg, and R. Pepperkok. 2007. Reverse transfection on cell arrays for high content screening microscopy. *Nat Protoc* 2 (2):392-9.

Faure, A., A. Naldi, C. Chaouiya, and D. Thieffry. 2006. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* 22 (14):e124-31.

Feist, A. M., M. J. Herrgard, I. Thiele, J. L. Reed, and B. O. Palsson. 2009. Reconstruction of biochemical networks in microorganisms. *Nature reviews. Microbiology* 7 (2):129-43.

Feist, A. M., and B. O. Palsson. 2008. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nature Biotechnology* 26 (6):659-67.

Ferrell, J. E., Jr. 2002. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 14 (2):140-8.

Fournier, T., J. P. Gabriel, C. Mazza, J. Pasquier, J. L. Galbete, and N. Mermod. 2007. Steady-state expression of self-regulated genes. *Bioinformatics* 23 (23):3185-92.

Friedman, N., M. Linial, I. Nachman, and D. Pe'er. 2000. Using Bayesian networks to analyze expression data. *J Comput Biol* 7 (3-4):601-20.

Garg, A., D. Banerjee, and G. De Micheli. 2008. Implicit methods for probabilistic modeling of Gene Regulatory Networks. *Conf Proc IEEE Eng Med Biol Soc* 2008:4621-7.

Garg, A., I. Xenarios, L. Mendoza, and G. DeMicheli. 2007. An efficient method for dynamic analysis of gene regulatory networks and in silico gene perturbation experiments. *RECOMB 2007, Lecture Notes in Computer Science* 4453:62-76.

Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* 11 (12):4241-57.

Glass, L., and S. A. Kauffman. 1973. The logical analysis of continuous, non-linear biochemical control networks. *J Theor Biol* 39 (1):103-29.

Glass, L., and JS Pasternack. 1978. Stable oscillations in mathematical models of biological control systems. *J Math Biol* 6:207-223.

Goldbeter, A. 1991. A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase. *Proc Natl Acad Sci U S A* 88 (20):9107-11.

Gonzalez, A. G., A. Naldi, L. Sanchez, D. Thieffry, and C. Chaouiya. 2006. GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems* 84 (2):91-100.

Gouzé, J. L. 1998. Positive and negative circuits in dynamical systems. *J. Biol. Systems* 6:11-15.

Gouzé, J. L., and T. Sari. 2002. A class of piecewise linear differential equations arising in biological models. *Dyn. Syst.* 17:299-316.

Guldberg, C. M., and P. Waage. 1879. Concerning Chemical Affinity. *Erdmann's Journal für Practische Chemie* 127:69-114.

Gupta, S., S. S. Bisht, R. Kukreti, S. Jain, and S. K. Brahmachari. 2007. Boolean network analysis of a neurotransmitter signaling pathway. *J Theor Biol* 244 (3):463-9.

Heidtke, K. R., and S. Schulze-Kremer. 1998. Design and implementation of a qualitative simulation model of lambda phage infection. *Bioinformatics* 14 (1):81-91.

Heinrich, R., and T. A. Rapoport. 1974. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur J Biochem* 42 (1):89-95.

Henry, C. S., M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay, and R. L. Stevens. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology* 28 (9):977-82.

Henry, C. S., M. D. Jankowski, L. J. Broadbelt, and V. Hatzimanikatis. 2006. Genome-scale thermodynamic analysis of Escherichia coli metabolism. *Biophys J* 90 (4):1453-61.

Higgins, J. 1964. A Chemical Mechanism for Oscillation of Glycolytic Intermediates in Yeast Cells. *Proc Natl Acad Sci U S A* 51:989-94.

Huang, S., and D. E. Ingber. 2000. Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Exp Cell Res* 261 (1):91-103.

Hubbard, T. J., B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, J. Herrero, R. Holland, K. Howe, K. Howe, N. Johnson, A. Kahari, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, C. Melsopp, K. Megy, P. Meidl, B. Ouverdin, A. Parker, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, J. Severin, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. Vilella, J. Vogel, S. White, M. Wood, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, P. Flicek, A. Kasprzyk, G. Proctor, S. Searle, J. Smith, A. Ureta-Vidal, and E. Birney. 2007. Ensembl 2007. *Nucleic Acids Res* 35 (Database issue):D610-7.

Hynne, F., S. Dano, and P. G. Sorensen. 2001. Full-scale model of glycolysis in Saccharomyces cerevisiae. *Biophys Chem* 94 (1-2):121-63.

Jaeger, J., S. Surkova, M. Blagov, H. Janssens, D. Kosman, K. N. Kozlov, Manu, E. Myasnikova, C. E. Vanario-Alonso, M. Samsonova, D. H. Sharp, and J. Reinitz. 2004. Dynamic control of positional information in the early Drosophila embryo. *Nature* 430 (6997):368-71.

Jiang, N., R. D. Cox, and J. M. Hancock. 2007. A kinetic core model of the glucose-stimulated insulin secretion network of pancreatic beta cells. *Mamm Genome* 18 (6-7):508-520.

Joyce, A. R., and B. O. Palsson. 2006. The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* 7 (3):198-210.

Kacser, H., and J. A. Burns. 1973. The control of flux. *Symp Soc Exp Biol* 27:65-104.

Kanehisa, M., S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa. 2006. From genomics to chemical

genomics: new developments in KEGG. *Nucleic Acids Res* 34 (Database issue):D354-7.

Karp, P. D., S. Paley, and P. Romero. 2002. The Pathway Tools software. *Bioinformatics* 18 Suppl 1:S225-32.

Kauffman, S. A. 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 22 (3):437-67.

Repeated Author. 1993. *The origins of order: Self-organization and selection in evolution*: Oxford University Press.

Kerrien, S., Y. Alam-Faruque, B. Aranda, I. Bancarz, A. Bridge, C. Derow, E. Dimmer, M. Feuermann, A. Friedrichsen, R. Huntley, C. Kohler, J. Khadake, C. Leroy, A. Liban, C. Lieftink, L. Montecchi-Palazzi, S. Orchard, J. Risse, K. Robbe, B. Roechert, D. Thorneycroft, Y. Zhang, R. Apweiler, and H. Hermjakob. 2007. IntAct--open source resource for molecular interaction data. *Nucleic Acids Res* 35 (Database issue):D561-5.

Kervizic, G., and L. Corcos. 2008. Dynamical modeling of the cholesterol regulatory pathway with Boolean networks. *BMC Syst Biol* 2 (1):99.

Klipp, E. 2007. Modelling dynamic processes in yeast. *Yeast*.

Klipp, E., R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. 2005. *Systems Biology in Practice*. Edited by Wiley-VCH.

Koshland, D. E., Jr., G. Nemethy, and D. Filmer. 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5 (1):365-85.

Kummel, A., S. Panke, and M. Heinemann. 2006. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol Syst Biol* 2:2006 0034.

Le Novere, N., B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J. L. Snoep, and M. Hucka. 2006. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res* 34 (Database issue):D689-91.

Le Novere, N., A. Finney, M. Hucka, U. S. Bhalla, F. Campagne, J. Collado-Vides, E. J. Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J. L. Snoep, H. D. Spence, and B. L. Wanner. 2005. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* 23 (12):1509-15.

Lewis, JE, and L. Glass. 1995. Steady states, limit cycles, and chaos in models of complex biological networks. *Int J Bifurcat Chaos* 1:477-483.

Li, F., T. Long, Y. Lu, Q. Ouyang, and C. Tang. 2004. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci U S A* 101 (14):4781-6.

Liebermeister, W., and E. Klipp. 2006. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor Biol Med Model* 3:41.

Matthews, L., G. Gopinath, M. Gillespie, M. Caudy, D. Croft, B. de Bono, P. Garapati, J. Hemish, H. Hermjakob, B. Jassal, A. Kanapin, S. Lewis, S. Mahajan, B. May, E. Schmidt, I. Vastrik, G. Wu, E. Birney, L. Stein, and P. D'Eustachio. 2009. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res* 37 (Database issue):D619-22.

Mendoza, L. 2006. A network model for the control of the differentiation process in Th cells. *Biosystems* 84 (2):101-14.

Mendoza, L., and F. Pardo. 2010. A robust model to describe the differentiation of T-helper cells. *Theory in biosciences = Theorie in den Biowissenschaften* 129 (4):283-93.

Mendoza, L., D. Thieffry, and E. R. Alvarez-Buylla. 1999. Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis. *Bioinformatics* 15 (7-8):593-606.

Mendoza, L., and I. Xenarios. 2006. A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theor Biol Med Model* 3:13.

Mestl, T., RJ Bagley, and L. Glass. 1997. Common chaos in arbitrarily complex feedback networks. *Phys. Rev. Letters* 79:653-656.

Mestl, T., E. Plahte, and S. W. Omholt. 1995. A mathematical framework for describing and analysing gene regulatory networks. *J Theor Biol* 176 (2):291-300.

Metzgar, D., J. M. Bacher, V. Pezo, J. Reader, V. Doring, P. Schimmel, P. Marliere, and V. de Crecy-Lagard. 2004. Acinetobacter sp. ADP1: an ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Res* 32 (19):5780-90.

Michaelis, L., and M. Menten. 1913. Die Kinetik der Invertinwirkung. *Biochem. Z.* 49:333-369.

Monod, J., J. Wyman, and J. P. Changeux. 1965. On the Nature of Allosteric Transitions: a Plausible Model. *J Mol Biol* 12:88-118.

Morgan, H., T. Beck, A. Blake, H. Gates, N. Adams, G. Debouzy, S. Leblanc, C. Lengger, H. Maier, D. Melvin, H. Meziane, D. Richardson, S. Wells, J. White, J. Wood, M. H. de Angelis, S. D. Brown, J. M. Hancock, and A. M. Mallon. 2010. EuroPhenome: a repository for high-throughput mouse phenotyping data. *Nucleic Acids Res* 38 (Database issue):D577-85.

Moult, J., K. Fidelis, A. Kryshtafovych, B. Rost, T. Hubbard, and A. Tramontano. 2007. Critical assessment of methods of protein structure prediction-Round VII. *Proteins* 69 Suppl 8:3-9.

Papin, J. A., J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson. 2004. Comparison of network-based pathway analysis methods. *Trends Biotechnol* 22 (8):400-5.

Parkinson, H., M. Kapushesky, M. Shojatalab, N. Abeygunawardena, R. Coulson, A. Farne, E. Holloway, N. Kolesnykov, P. Lilja, M. Lukk, R. Mani, T. Rayner, A. Sharma, E. William, U. Sarkans, and A. Brazma. 2007. ArrayExpress--a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res* 35 (Database issue):D747-50.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Edited by M. Kaufmann. San Mateo, CA.

Plahte, E., T. Mestl, and S.W. Omholt. 1994. Global analysis of steady points for systems of differential equations with sigmoid interactions. *Dyn Stabil Syst* 9:275-291.

Plahte, E., T. Mestl, and S.W. Omholt. 1995. Feedback loops, stability and multistationarity in dynamical systems. *J. Biol. Systems* 3:409-413.

Reed, J. L., I. Famili, I. Thiele, and B. O. Palsson. 2006. Towards multidimensional genome annotation. *Nat Rev Genet* 7 (2):130-41.

Remy, E., P. Ruet, L. Mendoza, D. Thieffry, and C. Chaouiya. 2006. From logical regulatory graphs to standard petri nets: Dynamical roles and functionality of feedback

circuits. *Transactions on Computational Systems Biology VII, Lecture Notes in Computer Science* 4230:56-72.

Rojas, I., M. Golebiewski, R. Kania, O. Krebs, S. Mir, A. Weidemann, and U. Wittig. 2007. Storing and annotating of kinetic data. *In Silico Biol* 7 (2 Suppl):S37-44.

Ropers, D., H. de Jong, M. Page, D. Schneider, and J. Geiselmann. 2006. Qualitative simulation of the carbon starvation response in Escherichia coli. *Biosystems* 84 (2):124-52.

Saez-Rodriguez, J., L. Simeoni, J. A. Lindquist, R. Hemenway, U. Bommhardt, B. Arndt, U. U. Haus, R. Weismantel, E. D. Gilles, S. Klamt, and B. Schraven. 2007. A logical model provides insights into T cell receptor signaling. *PLoS Comput Biol* 3 (8):e163.

Samal, A., and S. Jain. 2008. The regulatory network of E. coli metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. *BMC Syst Biol* 2:21.

Sanchez-Corrales, Y. E., E. R. Alvarez-Buylla, and L. Mendoza. 2010. The Arabidopsis thaliana flower organ specification gene regulatory network determines a robust differentiation process. *Journal of theoretical biology* 264 (3):971-83.

Schellenberger, J., and B. O. Palsson. 2009. Use of randomized sampling for analysis of metabolic networks. *The Journal of biological chemistry* 284 (9):5457-61.

Segal, E., M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, and N. Friedman. 2003. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34 (2):166-76.

Segal, E., R. Yelensky, and D. Koller. 2003. Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. *Bioinformatics* 19 Suppl 1:i273-82.

Sel'kov, E. E. 1968. Self-oscillations in glycolysis. 1. A simple kinetic model. *Eur J Biochem* 4 (1):79-86.

Shlomi, T., M. N. Cabili, M. J. Herrgard, B. O. Palsson, and E. Ruppin. 2008. Network-based prediction of human tissue-specific metabolism. *Nature Biotechnology* 26 (9):1003-10.

Snoussi, E. H. 1998. Necessary conditions for multistationarity and stable periodicity. *J. Biol. Systems* 6:3-9.

Sorribas, A., and M. Cascante. 1994. Structure identifiability in metabolic pathways: parameter estimation in models based on the power-law formalism. *Biochem J* 298 (Pt 2):303-11.

Stolovitzky, G., D. Monroe, and A. Califano. 2007. Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference. *Ann N Y Acad Sci* 1115:1-22.

Tavazoie, S., J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church. 1999. Systematic determination of genetic network architecture. *Nat Genet* 22 (3):281-5.

Teusink, B., J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff, and J. L. Snoep. 2000. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem* 267 (17):5313-29.

The UniProt Consortium. 2007. The Universal Protein Resource (UniProt). *Nucleic Acids Res* 35 (Database issue):D193-7.

Thieffry, D., E. H. Snoussi, J. Richelle, and R. Thomas. 1995. Positive loops and differentiation. *J. Biol. Systems* 3:457-466.

Thomas, R. 1973. Boolean formalization of genetic control circuits. *J Theor Biol* 42 (3):563-85.

Trelease, R. B., R. A. Henderson, and J. B. Park. 1999. A qualitative process system for modeling NF-kappaB and AP-1 gene regulation in immune cell biology research. *Artif Intell Med* 17 (3):303-21.

Tyson, J. J., K. C. Chen, and B. Novak. 2003. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* 15 (2):221-31.

Tyson, J. J., A. Csikasz-Nagy, and B. Novak. 2002. The dynamics of cell cycle regulation. *Bioessays* 24 (12):1095-109.

Varma, A., and B. O. Palsson. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl Environ Microbiol* 60 (10):3724-31.

Viretta, A. U., and M. Fussenegger. 2004. Modeling the quorum sensing regulatory network of human-pathogenic Pseudomonas aeruginosa. *Biotechnol Prog* 20 (3):670-8.

Waage, P., and C. M. Guldberg. 1864. Studies Concerning Affinity. *Forhandlinger: Videnskabs – Selskabet i Christinia* 35.

Wolf, J., and R. Heinrich. 2000. Effect of cellular interaction on glycolytic oscillations in yeast: a theoretical investigation. *Biochem J* 345 Pt 2:321-34.

Wolf, J., J. Passarge, O. J. Somsen, J. L. Snoep, R. Heinrich, and H. V. Westerhoff. 2000. Transduction of intracellular and intercellular dynamics in yeast glycolytic oscillations. *Biophys J* 78 (3):1145-53.

Young, D. M., D. Parke, and L. N. Ornston. 2005. Opportunities for genetic investigation afforded by Acinetobacter baylyi, a nutritionally versatile bacterial species that is highly competent for natural transformation. *Annu Rev Microbiol* 59:519-51.

Zi, Z., and E. Klipp. 2007. Constraint-based modeling and kinetic analysis of the smad dependent tgf-Beta signaling pathway. *PLoS ONE* 2 (9):e936.

Zimmerman, S. B., and A. P. Minton. 1993. Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu Rev Biophys Biomol Struct* 22:27-65.

Systems and Computational Biology - Bioinformatics and Computational Modeling

Edited by Prof. Ning-Sun Yang

Whereas some "microarray" or "bioinformatics" scientists among us may have been criticized as doing "cataloging research", the majority of us believe that we are sincerely exploring new scientific and technological systems to benefit human health, human food and animal feed production, and environmental protections. Indeed, we are humbled by the complexity, extent and beauty of cross-talks in various biological systems; on the other hand, we are becoming more educated and are able to start addressing honestly and skillfully the various important issues concerning translational medicine, global agriculture, and the environment. The two volumes of this book present a series of high-quality research or review articles in a timely fashion to this emerging research field of our scientific community.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Pascal Kahlem, Alessandro DiCara, Maxime Durot, John M. Hancock, Edda Klipp, Vincent Schächter, Eran Segal, Ioannis Xenarios, Ewan Birney and Luis Mendoza (2011). Strengths and Weaknesses of Selected Modeling Methods Used in Systems Biology, Systems and Computational Biology - Bioinformatics and Computational Modeling, Prof. Ning-Sun Yang (Ed.), ISBN: 978-953-307-875-5, InTech, Available from: http://www.intechopen.com/books/systems-and-computational-biology-bioinformatics-and-computational-modeling/strengths-and-weaknesses-of-selected-modeling-methods-used-in-systems-biology

INTECH
open science | open minds