A Survey on Methods for Modeling and Analyzing Integrated Biological Networks

Nuno Tenazinha and Susana Vinga

Abstract—Understanding how cellular systems build up integrated responses to their dynamically changing environment is one of the open questions in Systems Biology. Despite their intertwinement, signaling networks, gene regulation and metabolism have been frequently modeled independently in the context of well-defined subsystems. For this purpose, several mathematical formalisms have been developed according to the features of each particular network under study. Nonetheless, a deeper understanding of cellular behavior requires the integration of these various systems into a model capable of capturing how they operate as an *ensemble*. With the recent advances in the "omics" technologies, more data is becoming available and, thus, recent efforts have been driven toward this integrated modeling approach. We herein review and discuss methodological frameworks currently available for modeling and analyzing integrated biological networks, in particular metabolic, gene regulatory and signaling networks. These include network-based methods and Chemical Organization Theory, Flux-Balance Analysis and its extensions, logical discrete modeling, Petri Nets, traditional kinetic modeling, Hybrid Systems and stochastic models. Comparisons are also established regarding data requirements, scalability with network size and computational burden. The methods are illustrated with successful case studies in large-scale genome models and in particular subsystems of various organisms.

Index Terms—Systems biology, survey, modeling methodologies, integrated biological networks.

1 Introduction

As Denis Noble elegantly states, biological systems function as a full orchestra with its different elements playing *ensemble* the score of life [1]. Understanding how cellular systems build up these integrated responses when subjected to dynamically changing surroundings is one of the open questions in Systems Biology. The apparent linear flow from DNA, to mRNA and to proteins and their functions, expressed in the "central dogma of molecular biology" [2], [3], unveiled an unexpectedly intricate web of relationships.

1.1 Biological Networks: Working Ensemble

Nowadays, biological systems can be generally described as networks of interacting elements, hierarchically organized and tightly regulated [4], [5]. The intertwined behavior of signaling, regulatory and metabolic layers allows cells, or higher-level biological organizations, to develop response phenotypes when faced with external stimuli (Fig. 1).

Signaling networks allow the cell to monitor environmental cues such as nutrient availability, cell density or temperature. Series of biochemical reactions such as phosphorelay or kinase cascades are triggered allowing the information flow to the intracellular milieu. Some of these signals will eventually lead to the activation of transcription factors that recognize specific regulatory consensus in the DNA sequences and activate or repress target genes. In their turn, changes in enzymatic gene expression modulate the enzyme concentrations available to catalyze metabolic reactions. The operation of metabolism is ultimately responsible for supplying the cell with free energy, redox power and the precursor molecules required for growth or cell maintenance. The groups of genes, regulatory proteins and their interactions are often referred to as regulatory networks, whereas the complete set of metabolites and the enzyme-driven reactions constitute the metabolic networks. Therefore, when subjected to external stimuli, the cell's metabolic processes can be readily readjusted by changes in metabolite concentrations and fluxes (allosteric regulation within the metabolic network) or, in the long-term, by changes in gene expression, affecting enzyme activities (hierarchical regulation between gene regulatory networks and metabolism) [6]. The orchestrated behavior of these different layers of cellular organization is essential in order to maintain the intracellular physiological equilibrium (homeostasis).

1.2 Mathematical Modeling in Biology

The qualitative understanding of cellular responses has been achieved from extensive research in molecular biology and, more recently, from the endeavor of Integrative Systems Biology [5], [7]. In the last decade, advances in the "omics" high-throughput technologies led to an unprecedented increase in the amount and quality of available biological information. It also fostered the development of bioinformatics methods to interpret these data as well as the emergence of mathematical modeling as an essential tool to understand cell behavior. Mathematical models have proven invaluable in several fields of science and engineering [8]. In biology, they provide structured abstractions that enable studying the evolution, organization and design principles of cellular systems, ultimately providing insights into the control of living organisms.

• *The authors are with the Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento (INESC-ID), R Alves Redol 9, 1000-029 Lisboa, Portugal.*
E-mail: {ntenazinha, svinga}@kdbio.inesc-id.pt.

Manuscript received 5 May 2010; revised 3 Aug. 2010; accepted 19 Aug. 2010; published online 24 Nov. 2010.
For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBB-2010-05-0113.
Digital Object Identifier no. 10.1109/TCBB.2010.117.

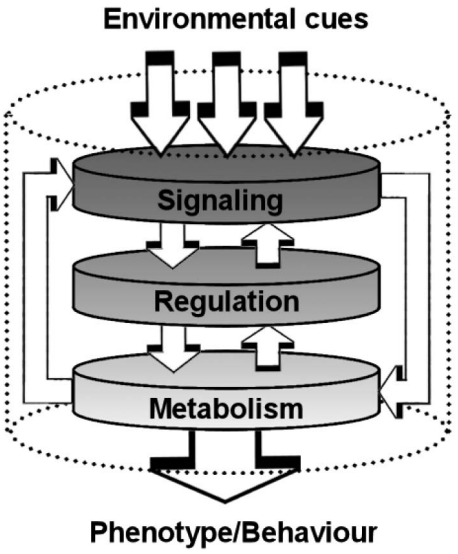

Fig. 1. Integrated operation of different cellular layers to produce several phenotypes when the biological system perceives signals from the surrounding environment.

Deterministic ordinary differential equations (ODEs[1]) are one of the widely used tools for modeling dynamic systems and their application also extends to Systems Biology. If the hierarchy of biochemical processes is known, the biological network can be generally represented by a stoichiometric matrix $\boldsymbol{S}$ where every element S_{ij}, $i = 1, \ldots, n, j = 1, \ldots, m$, corresponds to the stoichiometric coefficient of species or metabolite i in process j. Under this framework, a dynamic model for the system can be built such that:

$$\dot{\boldsymbol{X}} = \boldsymbol{S} \cdot \boldsymbol{v}(\boldsymbol{X}, \boldsymbol{u}, \boldsymbol{\theta}_{kin}, \boldsymbol{\theta}_{env}), \quad (1)$$

where $\dot{\boldsymbol{X}} = (dX_1/dt, \ldots, dX_i/dt, \ldots, dX_n/dt)$ is the vector formed by the time derivatives of the state variables, $X_i, i = 1, \ldots, n$, which represent concentrations of molecular components. In order to have the full description of the system, the vector of fluxes $v = (v_1, \ldots, v_j, \ldots, v_m), j = 1, \ldots, m$ must be defined. This vector depends on the actual state of the system as defined by the state vector $\boldsymbol{X}$, on possible input (external) functions $\boldsymbol{u}(t)$, on a given internal kinetic parameterization $\boldsymbol{\theta}_{kin}$ and also on physico-chemical parameters concerning the cellular environment, $\boldsymbol{\theta}_{env}$, such as pressure, temperature or pH. Hence, defining $\boldsymbol{v}$ requires the knowledge of the intervening processes so that the mathematical structure of the kinetic rates v_j can be appropriately chosen and the corresponding parameters fixed.

In spite of being a powerful modeling tool, the use of ODE-based modeling in Integrative Systems Biology has been frequently hindered. The reasons behind this are diverse, the most relevant being the high-dimensionality of biological systems, as illustrated by a recent joint effort that allowed the reconstruction of a consensus metabolic network for the yeast *Saccharomyces cerevisiae* [9]. This system comprised 2,153 components, including 1,168 metabolites, 832 genes, 888 proteins, 96 catalytic protein complexes and 1,857 involved reactions. Building a fully mechanistic model, based on ODEs, is currently unfeasible for a system of this size. Not only we lack the complete knowledge on the totality of processes of the system and their details, but also a kinetic model would require a critical amount of experimental parameters to define every single reaction rate involved. This information is simply not available at the moment. Most of the times, we can only have approximated values, educated guesses or estimations for the systems' parameters, and this is a major obstacle in ODE models. Even if, ideally, *in vitro* data were available for all the system's enzymes, it would not mean that the resulting model would describe the *in vivo* kinetics [10]. Despite the awareness of this problem in the Systems Biology community, its solution is not straightforward, and is highly dependent on experimental advances and new technologies.

Technical limitations also play an important role. As systems grow in size, so does the computational burden associated to their simulation and analysis. The different time-scales inherent to biochemical processes can also pose significant challenges to integrated models. Typically, signaling and metabolic reactions occur in the order of seconds, whereas regulatory responses or cellular growth can go from minutes to hours. Such slow-fast dynamics when described through ODEs lead to "stiff systems" that can exhibit severe numerical instability problems when one tries to integrate or solve them, while also being very sensitive to modeling errors [11], [12].

For the aforementioned reasons, most computational-driven studies have focused in specific subsystems of well-known model organisms (such as the central carbon metabolism [13], [14]), where plenty of quantitative experimental data already exists. To deal with the incompleteness of the data, different modeling approaches were developed according to the systems under study.

1.3 Different Systems, Different Tools

Given that biochemical networks have distinct properties, modelers resorted to use different abstraction levels in model building. These translate into approaches that range from more qualitative to fully quantitative and mechanicistic descriptions, and with different information requirements.

Regulatory networks have typically been described using discrete logical networks [15], [16], [17], nonlinear ODEs [18], [19], [20], piecewise-linear differential equations [21], [22] or stochastic approaches [23], [24], [25]. For a review on these methods refer to [26], [27]. Recently, a quasi-stoichiometric method was also suggested using a matrix formalism to describe small systems and evaluate their functional states [28].

The frameworks for modeling metabolic systems were developed independently based either in stoichiometric analysis or in ODEs, using models ranging from detailed kinetics to simple linearizations and approximate formalisms. Metabolic Control Theory has been used to analyze steady-state sensitivity to changes in the system [29], whereas Flux-Balance Analysis (FBA) and Elementary Flux

1. **Abbreviations:** ODEs, ordinary differential equations; EM, Elementary Flux Modes; EP, Extreme Pathways; OT, Chemical Organization Theory; FBA, Flux Balance Analysis; rFBA, regulatory FBA; srFBA, steady-state regulatory FBA; iFBA, integrated FBA; idFBA, integrated dynamic FBA; MRS, metabolic-regulatory steady-state; MILP, mixed integer linear programming.

Modes/Extreme Pathway Analysis have been used to characterize possible steady-state distributions [30], [31]. Continuous models based in enzyme kinetics were built for several well-characterized systems such as the *lac* operon [32] or central carbon metabolism [13], [33].

For comprehensive reviews on these frameworks and their potential refer to [34] and [35]. Reconstruction of genome-scale models of metabolism using some of these methods is also reviewed elsewhere [36], [37].

1.4 The Aim of This Survey

A recent review reinforces the necessity to integrate information from gene expression and metabolic flux data, how it can be rationalized and what potential insights can be unveiled from the integration of both [38]. Since it becomes clear that a vast repertoire of methodologies is already available for studying different biochemical networks, the main question is to choose a framework for integrating these systems—and, more importantly, what answers can be provided by this framework that could not be inferred by simply modeling the separate subsystems.

The current paper reviews several methodologies that have been proposed for modeling and analysis of integrated biochemical networks. Focus is given to their basic principles and how they accommodate the system's integration but also, whenever available, to representative case studies where they were successfully applied and demonstrated. The methods are exposed according to the underlying level of abstraction. Structural and stoichiometric approaches assume either static or pseudo steady-state conditions, respectively. The first allow the analysis of topological features useful for network reconstruction (network-based analysis) while the latter use constrained linear relations among fluxes together with regulatory rules to predict cellular responses (extensions of Flux Balance Analysis). Kinetic modeling methods allow to study the time evolution of the system but require explicit flux specifications, assuming discrete states (logical modeling and Standard Petri Nets), continuous states (ODE-based modeling) or methods that couple both (hybrid modeling). Finally, and even if their use in integrated models is still limited, stochastic methods are also discussed given their expected future impact. The various methods are further compared regarding data requirements, scalability with network size, computational burden, and the biological problems each approach is more adequate to address.

2 Structural and Stoichiometric Modeling

2.1 Network-Based Analysis

2.1.1 (Hyper)Graph-Based Analysis

Using annotated genome sequences, biochemical data and physiological aspects, it is conceptually possible to reconstruct an integrated genome-scale model for a biological system. The simplest mathematical abstraction for achieving this is the graph representation of the network of interacting species whose establishment is the necessary first step before any other type of analysis can be performed. For an introductory survey on graph theory and its application in Biology please refer to [39]. These methods have been thoroughly used for studying interesting properties of several biological networks such as connectivity, motifs, modularity, or robustness (for a review refer to [40], [41]). An hypergraph $\boldsymbol{H} = (V, E)$ is a generalization of the graph model, where V is a set of vertices or nodes and E are hyperedges, i.e., a generalization of the edge concept, allowing an arbitrary number of nodes to be connected. In what regards network integration, the concept of hypergraph can be a more realistic representation of the diversity of interactions among biological networks, and one that is not yet fully explored in the field. Klamt et al. [42] present a recent review on hypergraph theory applied to biological network modeling.

In metabolic networks, the edges and the nodes of the hypergraph represent, respectively, reactions and the metabolites involved in these reactions. To integrate regulatory networks into these models, we further need to identify links between metabolites and transcription factors controlling enzyme gene expression. A method for learning these relationships was proposed [43] that gradually adds edges between metabolites and transcription factors validating the relation through the explanatory power of the increasingly intertwined network against a set of experimental perturbation data. Additionally, the authors explicitly added regulatory motifs such as feedback regulation, to test if the augmented joint network was capable of explaining the data. A joint model for the central carbon metabolism of *Escherichia coli* was built in which, in spite of the simplicity and level of abstraction of this type of model, perturbation data was significantly better explained than in the separate subsystems.

2.1.2 Elementary Flux Modes/Extreme Pathway Analysis

Elementary Flux Modes (EM) and Extreme Pathway (EP) analysis allow the study of the space of flux distributions of a network. They have been successfully used to study systemic features of metabolic networks such as pathway length and network redundancy [44], [45], [46]. Both approaches use convex analysis to study the null space of the stoichiometric matrix $\boldsymbol{S}$. This corresponds to computing the steady-state solutions of (1):

$$\dot{\boldsymbol{X}} = 0 \Leftrightarrow \boldsymbol{S} \cdot \boldsymbol{v}_0 = 0. \tag{2}$$

In (2), $\boldsymbol{v}_0$ becomes the flux-vector evaluated at steady-state. EMs are a subset of the nullspace basis vectors of $\boldsymbol{S}$, that verifies thermodynamical feasibility ($v_0 \geq 0$ for irreversible reactions). They are also *minimal* in the sense that no other subset of reactions exists that verifies (2). Defined in this way, EMs represent basic, nondecomposable, subnetworks of the system that can operate at steady-state, and are the edges of a convex polyhedral cone in the space of flux distributions. EPs are conceptually closely related to EMs, but they rely on a network reconfiguration, where reversible reactions are decoupled into the corresponding forward and backward reactions. Every possible steady-state flux distributions allowed for a given network can thus be represented by a linear combination of EMs/EPs.

Covert and Palsson [47] use EP analysis to study how gene regulation affects the space of available metabolic flux

distributions in a network. For a system comprised of the core metabolic processes (e.g., glycolysis, TCA cycle, and fermentation pathways), they show that adding the corresponding regulatory phenomena results in a 67.5-97.5 percent pruning of the EPs of the nonregulated system. This clearly highlights the importance of integrating regulation in models of metabolism. The space of physiological behaviors in the cell is more restricted than the ones allowed in nonregulated models.

Despite their good results on metabolic network analysis, to our knowledge, there are no other EM/EP-based models explicitly applied to integrated biological networks. Part of the problem relies in the known combinatorial complexity associated to EMs/EPs' computation, resulting in a poor scaling of the algorithms for medium and large-scale networks (refer to [48], [49] for reviews on some advances in the field and applications).

2.2 Constraints-Based Modeling

Biological systems operate at several time-scales, e.g., metabolism involves fast reactions when compared to regulatory events. As a consequence, an acceptable simplification can be made stating that, in the long-term, metabolite concentrations are approximately constant. This is known as the (pseudo) steady-state assumption, underlying stoichiometric models [50], and is expressed in (2). Equation (2) also represents one of the most basic constraints in biological systems, the mass conservation within the network.

Flux-Balance Analysis (FBA) is a well-grounded framework built upon these concepts, assuming that cellular systems operate subjected to several governing constraints [49]. These can be established based on thermodynamic assumptions to describe reaction irreversibility ($v_j \geq 0$) or limiting flux capacities, which can be imposed to enzymes or transporters ($v_{\min} \leq v_j \leq v_{\max}$). These sets of constraints restrict the solution space of all possible flux distributions in the network to a bounded convex space (see Fig. 2) where potential phenotypes are located. While EMs/EPs allow the enumeration of all the possible distributions in this space, FBA uses linear programming methods to find particular solutions according to some optimality criteria. Target optimal functions are expressed as a linear equation Z [51]:

$$Z = \boldsymbol{w}^T \cdot \boldsymbol{v}, \tag{3}$$

where $\boldsymbol{w}$ stands for the weight vector applied to the flux vector $\boldsymbol{v}$. For metabolic networks these target functions are typically maximum energy production, growth optimization or a combination of flux-modes with efficient biomass production [52]. In the case of signaling networks or gene regulatory networks, establishing these optimizing functions is, however, not so straightforward. Consequently, FBA is not *per se* amenable to study biochemical networks in an integrated fashion because the stoichiometric matrix does not usually contain information on the regulatory relationships, but only on the transformations between the network constituents. In the next subsection, we present methodologies that extend the FBA framework in order to deal with regulated metabolic systems. Fig. 2 summarizes the basic principles underlying these extensions.

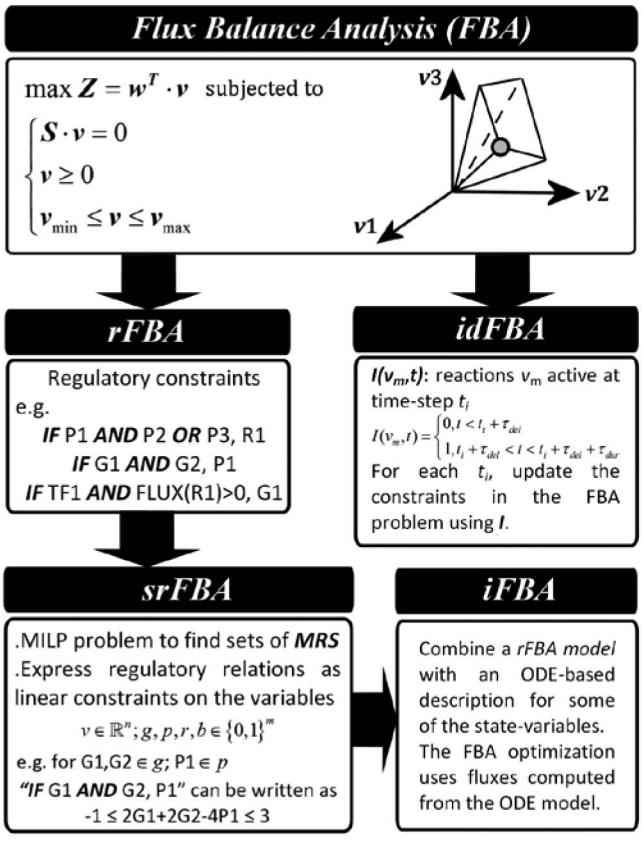

Fig. 2. Flux Balance Analysis and derived methods. FBA constraints transform the flux distribution search space into a convex bounded space where the optimum of the target function Z can be found using linear optimization routines; *rFBA* constrains even more this space by adding regulatory constraints in the form of Boolean logical rules; using this description and logic operators one can express reaction dependencies on proteins, protein dependencies on gene expression and gene expression dependencies on other genes, proteins or metabolites; in the example, reaction R1 occurs only if its catalyzing enzymes are present, the complex P1-P2 or the alternative protein P3; protein P1 is present only if gene G1 and gene G2 are expressed and translated and gene G1 is expressed only if the protein TF1 is present and the flux of reaction R1 is positive (i.e., the flux predicate FLUX(R1) > 0 is active); *srFBA* uses Mixed-Integer Linear Programming (*MILP*) to search for consistent metabolic and regulatory steady-states *(MRS*) in the flux distribution; *iFBA* uses available information to build ODE-based models that are embedded in the *rFBA* framework by computing the ODE-model derived fluxes and using these values in the FBA optimization; *idFBA* circumvents the different time-scales problem from biological networks by introducing the Boolean matrix I that describes which reactions are active/inactive at each time-step of a time-course simulation; matrix I is used to change the structure of the stoichiometric matrix by shutting down some of the reactions; parameters τ_{del} and τ_{dur} deal with response delays and the duration of each reaction, respectively, and are used to update matrix I during the time-course simulation algorithm.

2.2.1 Regulated Flux Balance Analysis (rFBA)

Incorporation of regulatory phenomena into FBA models was addressed by embedding Boolean logics as additional constraints [53]. This reduces the convex solution space, increasing the model accuracy for predicting phenotypes as shown in a follow-up work where the solution space reduction is analyzed with extreme-pathway analysis [47]. Boolean logic is used to describe the transcriptional regulatory structure, through operators such as AND, OR, and NOT, used to express regulatory dependencies between genes (G), proteins/enzymes (P), transcription

factors (TF) and metabolic fluxes' (R) predicates (example in Fig. 2). Furthermore, dynamic cell growth profiles can be simulated from a given time-point by iteratively computing regulatory states consistent with the metabolic flux-distribution of the previous time-step (for details see [53], [54]). The updated metabolic state can then lead to a new regulatory state defined by the Boolean rules.

This framework, named *rFBA*, was used to model a reconstructed joint regulatory and metabolic network in *E. coli* [55]. The model comprised 149 genes whose expression included 16 transcription factors and 73 enzymes involved in 113 reactions. The expression of 43 of the enzymes was under transcriptional regulation. Including the regulatory effects into metabolism improved the accuracy in a mutant study, correctly predicting $\sim 91\%$ of the 116 mutant phenotypes (nine more when compared with the FBA stand-alone model).

The model also yielded more meaningful simulations of several *in silico* experiments such as glucose/acetate growth or the glucose-lactose diauxie. The scalability of the *rFBA* approach was illustrated in the reconstruction of the first genome-scale integrated network in *E. coli* [56]. This model included 1,010 genes from which 104 coded for regulatory proteins involved in the control of 479 of the remaining 906 genes in the metabolic network. Later on, the first eukaryotic genome-scale model of regulated metabolism in *S. cerevisiae* was also published [57]. It included 805 genes (from which 750 were metabolic genes), 775 regulatory interactions and 55 nutrient-regulated transcription factors. In both cases, the discrepancies found between the model predictions and the actual gene expression data allowed the formulation of hypotheses concerning novel regulatory mechanisms not obvious from the available data at the time.

2.2.2 Steady-State Regulated Flux Balance Analysis (srFBA)

A potential criticism for the *rFBA* approach is that a single metabolic steady-state is chosen at every time point, leaving a myriad of potential trajectories in the solution space unexplored. To avoid this, an extension of the formalism was suggested, based on previous work [47], to include the concept of MRS [58]. MRS represents a pair of consistent metabolic and regulatory states, satisfying both the metabolic and regulatory constraints. Under the name of *srFBA*, the authors devised the problem as a MILP strategy to identify MRS solutions. The Boolean constraints of the regulatory network together with the mapping of the genes into reactions are translated to sets of linear equations and constraints. Auxiliary variables must be defined, that express the dependencies between the fluxes, $v \in \mathbf{R}^n$, and the gene state, $g \in \{0,1\}^k$. These are the protein state $p \in \{0,1\}^j$, the reaction state $r \in \{0,1\}^m$ and the flux predicate state $b \in \{0,1\}^l$ (for a more detailed description of the method please refer to [58]). MILP optimization is then applied to maximize the physiologically relevant objective function, e.g., cell growth, on the rephrased constraints v, g, p, r, and b. Alternative expression states of a given gene g_i are explored by solving two MILP problems for both values of g_i. If a feasible MRS solution is found for one of the problems then the gene is considered either expressed or nonexpressed. If a feasible solution exists for both cases, the gene is considered undetermined.

The authors applied this modeling approach to rebuild the genome-scale *E. coli* model published in [56] and used it to study the variability of expression and metabolic activity states within and between different growth media. Quantifying the fraction of reactions whose fluxes were determined by the *srFBA* integrated model but not by the FBA stand-alone counterpart helped to assess the roles of metabolism and regulatory phenomenon in their control. Their results suggested that transcriptional regulation can have an indirect impact in metabolic fluxes not subjected to hierarchical regulation, by acting on neighboring processes, as previously suggested elsewhere [59]. It is shown in this work that, in general, studying the space of MRS solutions allows the modeler to explore a wider range of questions, not available through *rFBA* modeling.

2.2.3 Integrated Flux Balance Analysis (iFBA)

iFBA is an extension of *rFBA* [60] that takes advantage of detailed kinetic information from ODE-based models, when available. Kinetic models outperform FBA whenever genes show multistability or when certain regulatory proteins have dependencies on internal metabolite concentrations that cannot be computed with the latter (given the pseudo steady-state assumption). After reconstruction of a system partially described in the *rFBA* framework and with some of the state variables defined as in (1), the authors propose an algorithmic approach for simulating the systems' dynamic behavior. Initially, time is discretized and a time interval t_N must be chosen as a compromise between the steady-state assumption of *rFBA* and the necessary accuracy to avoid numerical error accumulation when computing the dynamic time-profiles. After setting the initial conditions for the system, the algorithm follows an iterative process that can be generally described in three steps: 1) in the current time-step t_i, the ODE model is numerically integrated and the regulatory constraints computed based on the conditions of the previous time-step, t_{i-1}. ODE fluxes are calculated at the end of this time-step; 2) the FBA optimization at t_i is formulated using the flux vector at t_{i-1} together with the previously computed ODE fluxes; 3) the objective function value and external metabolite concentrations are updated and the algorithm proceeds to time-step t_{i+1}. Note that eventually, there will be fluxes described by both FBA and the ODE-model. To deal with this, the authors suggest the notion of *ODE matching constraints* by setting the boundaries of the FBA fluxes equal to the corresponding rate values as computed from the ODE model.

The method was evaluated in an extended version of the *E. coli* model in ref. [55] combined with a kinetic model for PTS-mediated catabolite repression [61]. In some cases, the *iFBA* model yielded different predictions regarding gene lethality in comparison to the original *rFBA* model and the authors attributed these differences to the effect of internal metabolite concentrations accounted in *iFBA*. Their results also suggested that *iFBA* can even show advantages over the kinetic model because, it allowed the study of systemic global effects of the catabolite repression dynamics, not evident from the latter—some phenotype predictions were captured by *iFBA* but not in the kinetic

model. The authors also suggest that *iFBA* can still be improved to account for multiple flux distributions using the MRS concept and also by studying different objective functions for optimization. This framework is particularly interesting for gradually incrementing the accuracy of genome-scale FBA-based models, by embedding dynamic properties of parts of the system, as these become available.

2.2.4 Integrated Dynamic Flux Balance Analysis (idFBA)

The approaches hitherto presented have been mostly dedicated to transcriptional regulation of metabolic networks. To address the problems of modeling signaling networks, *idFBA* was presented as a new FBA formulation for analyzing integrated biochemical networks [62]. To deal with the problem of different time-scales, pseudo steady-state is assumed for "fast reactions" whereas "slow reactions" are incorporated into the stoichiometric matrix with a time-delay. This requires a discretization of the time domain into small time-steps t_N, just as in *iFBA*. An algorithmic procedure for phenotype simulation is proposed and can be generally described as follows:

1. given a stoichiometric matrix $\boldsymbol{S}$, as in (2), initialize an auxiliary binary incidence matrix $\boldsymbol{I}$ with dimensions $m \times t_N$, m being the number of processes in $\boldsymbol{S}$. Matrix $\boldsymbol{I}$ describes the participation (denoted by 1) or absence (denoted by 0) of every reaction in the system at a defined time-step. It does so by updating the constraints imposed to the flux boundaries of the FBA problem and shutting down some of the reactions;
2. at t_i, solve the FBA optimization problem with the updated constraints from t_{i-1};
3. given the optimized flux vector, compute the value of the phenotype variable at the end of t_i;
4. update the incidence matrix $\boldsymbol{I}$ based on the optimized flux vector and accounting for the time-delays (τ_{del}) and reaction duration parameters (τ_{dur}): $I = 0$ for $t < t_i + \tau_{del}$; $I = 1$ for $t_i + \tau_{del} < t < t_i + \tau_{del} + \tau_{dur}$;
5. finally update the time-step and iterate from step 2.

One of the problems of this strategy is obviously the definition of matrix $\boldsymbol{I}$, since it gives a qualitative description of the systems dynamics by expliciting which reactions are active in the network during the time course. Nevertheless, the authors argued that from experimental data and some assumptions on the systems behavior, $\boldsymbol{I}$ can serve as a first guess to the systems' dynamics. They compare *idFBA* derived models against mechanistic kinetic models of a toy example and of an integrated network from the high-osmolarity glycerol pathway in *S. cerevisiae* [63]. Using biomass production as a phenotype variable, they achieve similar dynamic time-courses with much fewer parameters, supporting the use of time-delays and reaction duration parameters to give satisfactory predictions of the dynamic behaviors of systems with different time-scales. Furthermore, they argue that the discrepancies found could be diminished by using a finer grained time-interval for the simulation.

2.3 Chemical Organization Theory

Chemical Organization Theory (OT) is a framework to represent integrated networks built upon concepts of constructive dynamical systems [64]. Let the 2-tuple $(\mathcal{M}, \mathcal{R})$ represent a reaction system where, $\mathcal{M}$ is the set of all biochemical species and $\mathcal{R}$ is the set of reactions involving the species in $\mathcal{M}$, such that $\mathcal{R} \subseteq P_{\mathrm{M}}(\mathcal{M}) \times P_{\mathrm{M}}(\mathcal{M})$. Then $P_{\mathrm{M}}(\mathcal{M})$ will be the power set of all species in $\mathcal{M}$. The reaction system can also be expressed using the stoichiometric matrix $\boldsymbol{S}$, as in (1). Regulatory effects are embedded in OT by mapping regulatory Boolean rules into chemical reaction within the set $\mathcal{R}$: activations are described as nonconsuming species required for a reaction to proceed (e.g., $\mathrm{R_A}$ being the activator, the reaction rule would be $\mathrm{A} + \mathrm{R_A} \rightarrow \mathrm{B} + \mathrm{R_A}$; in the absence of $\mathrm{R_A}$ the reaction has zero flux); inhibitions impose the definition of pseudo-species whose "nonexistence" drives the reaction (e.g., $\mathrm{A} + \mathrm{R_I^*} \rightarrow \mathrm{B} + \mathrm{R_I^*}$; where $\mathrm{R_I^*}$ represents the absence of inhibitor $\mathrm{R_I}$).

The concept of *organization* is suggested to study both static and dynamic aspects of the system [65]. Formally, an organization is a set $\mathcal{S} \subseteq \mathcal{M}$ that fulfills the conditions of closure and self-maintenance. The set $\mathcal{S}$ is closed if it contains all the biochemical species that can be produced by reactions in $\mathcal{R}$, involving species from $\mathcal{S}$. Formally, $\mathcal{S}$ is closed if, for all reactions $(\mathrm{A} \rightarrow \mathrm{B}) \in \mathcal{R}$ with $\mathrm{A} \in P_{\mathrm{M}}(\mathcal{S})$, then $\mathrm{B} \in P_{\mathrm{M}}(\mathcal{S})$. It is also said to be self-maintained if there is a guarantee that any species consumed within the set, can be recovered from other species within $\mathcal{S}$. This is to say that $\mathcal{S} \subseteq \mathcal{M}$ is self-maintained if there exists a flux vector $\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbf{R}^n_{\geq 0}$ verifying the following conditions: 1) for every reaction $(\mathrm{A} \rightarrow \mathrm{B}) \in \mathcal{R}$ with $\mathrm{A} \in P_{\mathrm{M}}(\mathcal{S})$, its flux is such that $\mathbf{v}_{\mathrm{A}\rightarrow\mathrm{B}} \geq 0$; 2) for reactions $(\mathrm{A} \rightarrow \mathrm{B}) \in \mathcal{S}$ where $\mathrm{A} \notin P_{\mathrm{M}}(\mathcal{S})$, then $\boldsymbol{v}_{\mathrm{A}\rightarrow\mathrm{B}} = 0$; 3) for every species $i \in \mathcal{S}$ its concentration change is nonnegative, $(\boldsymbol{S}.\boldsymbol{v})_{\mathrm{i}} \geq 0$. Thus, organizations represent sets of species in the network that can be maintained over long periods of time in the system, both at steady-state and at growth states. Therefore, the concepts of EM/EP are naturally embodied in this framework. As shown by Kaleta et al. [66] both approaches can complement each other and OT can be used to analyze and categorize EM solutions.

A partially ordered set of organizations represents the structural organization of a system and movements along this set can give insights into its dynamics, through a mapping of trajectories in the state-space (refer to [65] for further details and extensive theory presentation). In fact, for a system described by ODEs, Theorem 1 in [65] shows that all the fixed-points of those ODEs are instances from organizations of the system, thus relating network topology to potential dynamic behaviors. It remains an open problem whether all the attractors of the system can be mapped to organizations.

OT has been used in photochemical models of the Martian atmosphere [67], models of HIV infection [68] and of the central carbon metabolism of *E. coli* [69]. In the context of network integration, Kaleta et al. [70] successfully applied this framework to mimic the stoichiometric *E. coli* model by Covert et al. [55] and found that the 16 different wildtype growth scenarios under study were totally covered by 16 single organizations. Also, in knockout experiments, OT generally yielded results similar to the *rFBA* model, but with higher accuracy in some cases [70].

A major limitation in OT modeling is the computation of organizations which, can grow exponentially with network

size and number of reactions (e.g., in [71] a network with 31 species and 103 reactions shows 1,088,640 organizations). This presents a computational challenge for genome-scale models of regulated metabolism. Nevertheless, efforts have been taken to circumvent this drawback with heuristic approaches being proposed as an alternative to exhaustive deterministic algorithms [71]. These methods compute a significant subset of the total organizations in the system and can be efficiently used in networks up to 1,000 metabolites.

3 Kinetic Modeling

3.1 Discrete Models

3.1.1 Logical-Based Models

Equation (1) is the general continuous ODE-based description for any biochemical network relating how the state variables of the system evolve with time, depending on the mass balance of the processes involved. It can, however, be formulated in a more simplistic manner by establishing the dependencies between the fluxes v_j and the state-variables X_i by means of logical relationships, in order to have a qualitative description of the systems' behavior [71]. For an integrated biochemical network the system is described by a list of statements given by the discretized set $\{\boldsymbol{v}, \boldsymbol{g}\}$ composed of metabolite fluxes, $\boldsymbol{v}$, and genes, $\boldsymbol{g}$. Logical activation functions are used to relate the values of $\boldsymbol{v}$ and $\boldsymbol{g}$ with discretized metabolite concentrations, $\boldsymbol{m}$. These can be generally built from a combination of mathematical functions such as the Heaviside or the signum function, enabling the state computation for every element in the network. In this way gene regulators depend on metabolite concentrations, and the metabolic fluxes are functions of the activation state of genes.

In this context, a discrete mathematical model was proposed for the regulated carbon metabolism of *E. coli* under mixed substrates from three different carbon sources (glucose, glycerol, and acetate) [72]. The model included 19 metabolites and regulators to which were attributed discretized concentration levels, 28 genes in their active or inactive form and 20 enzymes with discretized flux values. The dynamics of the network were studied by assigning random values to the variables and then computing the state of the system according to the activation functions iteratively until a stable-state (or cycle) was reached. These described potential qualitative phenotypes that were later compared to experimental data. An attractor was always found for analysis under eight different mixed substrate configurations. However, this was only possible if the hierarchical preference of carbon source seen in *E. coli* was input into the model. Solely based on qualitative relationships, the authors give general insights on the system's behavior. This type of models can also be easily and automatically analyzed using methods such as reachability analysis and model checking.

A major shortcoming for this approach is that it constitutes a time-consuming process since all the logical activation functions for updating the network must be explicited, rendering the method unfeasible for large systems.

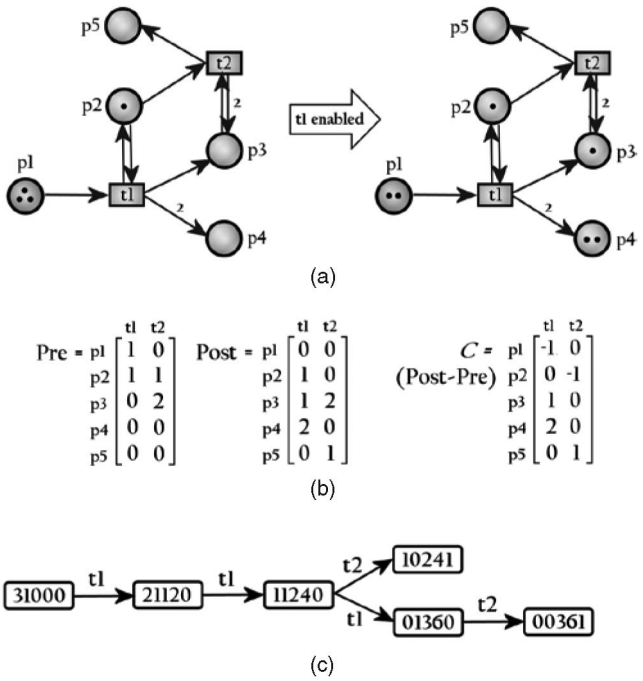

Fig. 3. Petri Net representation. (a) Places are represented by the circles whereas, transitions are represented by rectangles; dots represent the tokens assigned to each place; for the initial marking shown, transition t1 is enabled and fires, transferring tokens to the output places according to the weights in the incoming and outgoing arcs. (b) The incidence matrix C of the Petri Net. It can be computed by subtracting matrices Post and Pre. (c) The marking graph of the available transitions for the given initial marking; numbers in each node represent the numbers of tokens associated to each place; when several transitions are enabled, the marking graph exhibits different branching paths corresponding to alternative dynamics. This example mimics a biochemical reaction catalyzed by an enzyme that is inhibited by one of the reaction products when its concentration reaches a threshold value.

3.1.2 Petri Net Modeling

Standard Petri Nets are a discrete mathematical formalism initially developed for the modeling of distributed systems that have now turned into a useful tool for biochemical model building in Systems Biology (for reviews see [73] or [74] and for the basics on Petri Net Theory refer to [75]). They have been used for modeling metabolic networks [76], [77], signaling pathways [78], [79] and gene regulatory networks [80], [81].

Formally Petri Nets are described by a directed bipartite graph $G = (P, T, F, M_0)$, where nodes represent places p $\in P$ (the actual state of the system) and transitions t $\in T$ (discrete events changing the state of the system). Relationships between places and transitions are mediated by arcs (directed and weighted edges) described in the multiset F. The Petri Net can also be represented as an incidence matrix $\boldsymbol{C}$, equivalent to the stoichiometric matrix of metabolic networks (Fig. 3b). Species concentrations are expressed in this formalism as the discrete number of tokens, at a given instant, assigned to a place. The distribution of tokens along the places in the graph corresponds to the state of the system, and is known as the Petri Net marking. M_0 is the initial marking, i.e., the vector of the initial token distribution. A transition is said to fire when its input places have the minimum number of tokens specified in the corresponding arc weights (enabled transition), resulting in the

token transfer to the output place. Fig. 3a exemplifies a graphical representation of a Petri Net model where two different behaviors can develop. For a given initial marking M_0 one can study the system's dynamics and identify stable token distributions corresponding to attractors in the model. The possible firing sequences from an initial marking can be represented by its marking graph (Fig. 3c).

Petri Nets allow the modeling of regulatory effects by translating logical rules into the same qualitative framework. A systematic approach for translating Boolean rules into Standard Petri Nets is described in [82], [83]. Activations and inhibitions can be included by adding places accounting for the complementary state of the regulators.

A case study of the tryptophan biosynthetic pathway in *E. coli* was used to illustrate this method [84]. By including the tryptophan allosteric inhibition on the first enzyme in tryptophan and the transcriptional inhibition of the operon through its repressor complex TrpR, a sound qualitative description of the dynamical behavior of the metabolic system was obtained depending on the low, moderate, or high concentrations of extracellular tryptophan.

Nevertheless, in this formulation, one can only obtain qualitative analyses of the asymptotic behaviors of the system with time-implicit in the firing of the transitions. To overcome this limitation, semiquantitative analyses can be sought with several extensions of the standard framework. Coloured Petri Nets [85], Timed Petri Nets [86], and Stochastic Petri Nets [80] have been used for modeling particular biological systems but their application on integrated systems is yet to be demonstrated (for further details and applications refer to [74]). Hybrid Functional Petri Nets [87], on the contrary, have already proven as a valuable approach for this purpose (see Section 3.3).

Another advantage of Petri Net modeling is the series of methods useful to study both static and dynamic properties with meaningful biochemical interpretations. Conservation relations can be obtained from P-invariant analysis while T-invariants describe system's behaviors, e.g., at steady-state [75]. *Minimal* T-invariants are nonnegative solutions y of the linear equation $\boldsymbol{C}.y = 0$, where $\boldsymbol{C}$ is the Petri Net incidence matrix and where no subset of y is also a T-invariant. They are closely related to the concept of EM, the main difference being that T-invariants only apply to irreversible reactions. Structural properties can be assessed with this method and recent progress has been done for network modularization based on T-invariant analysis [88].

Dynamic properties can also be studied using reachability analysis to identify possible trajectories leading to a desired state from a defined initial marking while model-checking methods are being proposed for the formal verification of Petri Net models [89], [90].

3.2 Continuous Models

Continuous modeling has been used thoroughly for many subnetworks of biological systems. It provides the possibility to simulate the systems behavior in a quantitative way through time, and also to apply well established mathematical results from dynamical systems theory that allow a series of analytical methods. Sensitivity analysis can be used to study the robustness of the system upon changes in concentrations/fluxes and bifurcation analysis helps in assessing changes in dynamic properties associated to parameter intervals [91].

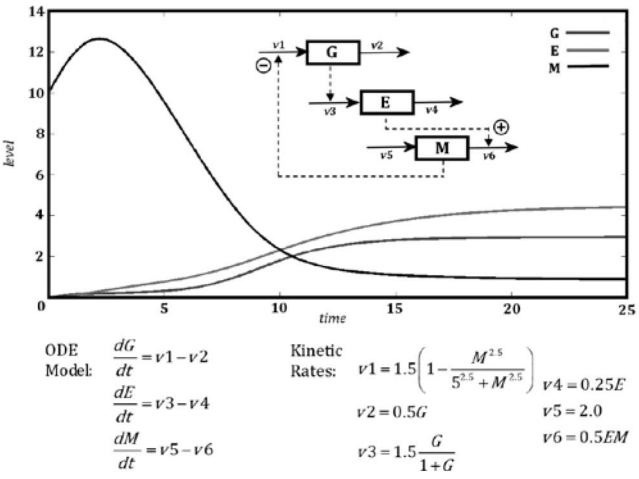

Fig. 4. ODE model for an artificial gene regulated metabolic network. Metabolite M indirectly controls the expression of gene G, coding for an enzyme E responsible for the degradation of M. If G is initially not expressed and M has accumulated in the cell, the system will evolve toward the degradation of this metabolite, with an increase in the expression levels of G. Rate v_1 is described with a Hill equation, v_3 with a Michaelis-Menten process and the remaining rates with mass-action kinetics. The simulation was performed in MATLAB with the parameterization shown above, using arbitrary units.

If formulated as a system of ODEs as in (1), the mathematical rate equation to describe the flux v must be chosen appropriately in order to describe the mechanistic aspects of the system. The inherent nonlinearities of enzyme-driven reactions have been commonly captured by mass action or Michaelis-Menten like kinetic rates, accounting for the presence of inhibitors or effectors that can compete with the substrate for the catalytic site (see Section 5.1 in [92] for a comprehensive presentation of biochemical kinetic laws). Cooperativity or switch-like kinetics associated to regulatory events have been mostly described in a phenomenological way by the Hill equation [93].

As an alternative, approximate formalisms have been developed to describe biochemical processes when their underlying mechanisms are unknown. From these, the most well known are the power-law formalism within the Biochemical Systems Theory [94], [95] and the lin-log/log(linear) approximation [96], [97]. Recently, the Saturable and Cooperative formalism was proposed as an alternative to the previous [98]. They all provide a way to perform systems level analyses, sometimes with analytical solutions, with rates that generally have fewer parameters that can be inferred from experimental time-series using optimization algorithms [99].

Fig. 4 illustrates a simple ODE-based model for a three-variable artificial network that mimics a small integrated-biological system along with its simulation. Different types of the aforementioned kinetic rates are also illustrated (see caption of Fig. 4).

Probably one of the first examples of continuous modeling applied to an integrated biochemical network is the regulation of the *lac* operon involved in lactose uptake and metabolism. The apparently simple regulation involves a repressor protein that blocks the operon transcription, unless

it is inactivated when binding a substrate of lactose metabolism. This first model goes back to 1972 [100] and is based solely on five ODEs with nonlinear feedback regulation. In spite of its simplicity, it reproduces the well-known bistability of the system. Other authors have since then proposed several models for this system for studying different aspects: the diauxic growth of *E. coli* in glucose and lactose, using a detailed model of the *lac* operon regulation (with 13 state variables) and including phenomena such as catabolite repression or inducer exclusion [32]; the influence of these phenomena in the bistability behavior, considering the cooperative effect between the operators which was lacking in the previous model [101]. More recently, other evolutionary studies emerged using *in silico* simulations in bacterial populations [102], [103].

Building continuous models requires significant amounts of knowledge on the system under study and this is a strict limitation for large networks. However, efforts have focused in medium-sized subsystems with successful results. Klipp et al. [104] presented a modular model containing the signaling processes (phosphorelay and the MAP kinase cascade), transcriptional regulation, carbon metabolism, and glycerol production involved in the yeast response to osmotic stress. Overall the model contained 32 ODEs and 70 parameters with three additional ODEs and two algebraic relationships accounting for changes in volume, osmotic pressure, and turgor pressure. Process rates were described using either linear or Michaelis-Menten kinetics. Through *in silico* simulation the authors suggested some new hypotheses regarding the regulatory design of the HOG pathway, which were partially supported by experimental data. Also in yeast, the metabolic shift from glucose to galactose due to the GAL regulatory switch and the phenomenon of catabolite repression was also modeled [105]. The three integrated modules consisted of glycolysis, galactose metabolism through the Leloir pathway, and the GAL genetic network with its transcriptional regulators. The model reproduced the carbon source preference of yeast due to glucose repression of the transcriptional regulators that control the expression of the Leloir genes. It also predicted that when galactose is used, carbon flux is shifted toward the production of glycerol and glycogen, instead of ethanol and carbon dioxide as in glucose.

However, useful as illustrated by the latter examples, resorting to this modeling approach is still unfeasible for large-scale models because it is still very difficult, if not impossible, to gather the necessary parameters for the complete system. Using approximate formalisms can provide an interesting solution for this problem, modeling the processes as black boxes with parameters inferred from experimental data. Still, the systems analyses become increasingly difficult due to the computational burden of simulating a large number of nonlinear ODEs, arising mainly from stiffness problems.

3.3 Hybrid Modeling

Hybrid Systems are commonly used to model dynamic systems showing continuous and discrete dynamic behaviors, i.e., when part of the system's states assume only finite values whereas, the remaining can assume values in $\mathbf{R}^n$ (Lygeros et al. [106] present the main concepts and properties of Hybrid Systems). They have been successfully applied in engineering for mechanical systems, electrical circuits, or chemical process control. Certain aspects of biochemical processes have a clear translation into a Hybrid Systems framework, e.g., to deal with different time-scales, to approximate sigmoidal nonlinearities with piecewise step functions, to account for regulatory effects that result in altered dynamic modes or even to deal with stochasticity when the assumptions for continuous modeling do not hold [107].

Hybrid Systems have been proposed as a conceptual framework for modeling integrated biochemical systems in [108]. Following ideas from Rosen and Casti [109], [110], the authors propose a formalism based on hybrid automata that is able to describe intracellular dynamics, coupled with their regulatory counterparts and discrete events promoting regime switches. The formalism is illustrated with a model of the xanthophyll cycle reaction, a mechanism that higher plants use as protection against excess light, by converting it into heat. De Jong et al. [21] present a model for the initiation of sporulation in *Bacillus subtilis* including the phosphorelay system and the underlying genetic regulatory network. This model is based on piecewise-linear differential equations, where regulation is described through regulatory functions that can assume Boolean values embedded in the rate. The approach is able to provide a qualitative simulation of the sporulation behavior without requiring precise numerical values to parameterize the system.

Hybrid rectangular multiaffine systems are proposed elsewhere to build a model of the stringent response (nutrient deprivation) in bacteria [111]. Instead of considering only Boolean regulatory relationships, product-type nonlinearities between the regulators are allowed. The model comprises nine state-variables plus three conservation laws, with two modes where a transition occurs depending on the concentration value of guanosine pentaphosphate, a mediator of the response to amino acid deprivation. Performing both analytical and numerical analyses of the system, the authors study how a nutritional perturbation reflects in increased levels of guanosine pentaphosphate to drive the stringent response. By using reachability analysis they further explored the possible stable steady-states from a given departing initial state. The convenience of this method to study Hybrid Systems models is also illustrated in [112] where the authors compare hybrid approximations to a fully continuous model of the *lac* operon with the original version. Nonlinear rate laws are approximated from the original model with piecewise-affine linear abstractions. From their analysis they shed new insights on the importance of the basal transcription rate for the typical bistability of the *lac* operon as well as its induction capability when lactose is available.

Another example of hybrid modeling is the extension of Petri Nets known as Hybrid Functional Petri Nets [113]. The underlying idea is to include continuous places and transitions where the firing process is assigned with a rate $\boldsymbol{v}$ that describes its speed. Discrete transitions can be set to fire with an associated time-delay. This is suited for modeling different time-scales of cellular processes and

allows the generation of time-courses describing the systems dynamics. They have been used in several integrated models as illustrated in [114] for the regulation of the urea cycle in the liver. Other examples include the control of commitment and sporulation phenomena in *Physarum polycephalum* [115], the regulation of early human haematopoiesis for studying the effect of interleukin-6 cytokine [116] or, more recently, a Hybrid Functional Petri Net model for the genetic toggle associated to the SOS signaling pathway in *E. coli*, within Biochemical Systems Theory [117], [118].

Despite its potential, this sort of approach is not yet fully explored. Hybrid approaches require less initial input and are not so computationally demanding as fully kinetic models which renders them attractive for modeling medium-scale systems. Furthermore, they come with a series of methods such as reachability analysis or temporal logics applied to model checking that enable a semiquantitative study of biochemical systems.

3.4 Stochastic Modeling

Stochastic events arise naturally in biological systems. When the number of molecules on a contained volume is large, the law of mass action (Section 5.1 in [92]) allows us to use an averaged concentration as an accurate measure of this amount, and stochastic effects can be disregarded. However, in regulatory networks the number of molecules is often low, and the effects of stochasticity can have a significant impact—intrinsic noise [119]. Furthermore, random fluctuations in the cell's environment can also contribute to change dynamic behaviors—extrinsic noise [120].

The importance of this sort of approach has been illustrated by several applications in regulatory networks (e.g., [121], [122], [123], [124]). For this reason, modelers should also be concerned of whether or not to include these effects when integrating regulation into metabolic networks. In spite of this awareness, there is not, to our knowledge, any published attempt to include stochastic effects into integrated metabolic and regulatory systems. Recently, Prasad and Venkatesh [125] studied the hierarchy of glucose repression in the well-known GAL system of yeast, using stochastic analysis. Given the small system's size they were able use Gillespie's Stochastic Simulation Algorithm (SSA) (refer to [126] for the basics of this method) in reasonable time. Unfortunately, this is not the case for higher-scale networks. Current stochastic methods have a very poor scaling which strictly limits their application and, together with the scarcity of experimental data, this is probably why they have not been used to model integrated systems. Much effort has been put into finding exact or approximate stochastic simulation methods (e.g., the tau-leaping method) to circumvent these computational limitations [127], [128], [129], [130]. More recently, the moment-closure approximations [131] and the finite state projection method [132], [133] were suggested and show good prospects for improving the computation speed of stochastic simulation, motivating their future application for integrated modeling.

4 Conclusion and Future Perspectives

The available knowledge on how cells operate has been mostly derived from accumulated experimental data on the several constituents of the cellular machinery. Signaling cascades, gene regulatory networks and metabolism have been thoroughly described in model organisms such as *E. coli* or the yeast *S. cerevisiae*. However, the sheer complexity of biological systems often hinders how these elements interact at several levels and how their integrated functioning results in complex cell behaviors.

Making sense of this intertwinement becomes harder as systems grow in size, to the whole-genome level. At this point, mathematical models can be a valuable tool to understand biological phenomena because they provide useful abstractions to represent them. Furthermore, they can offer means to study biological principles and their evolution or to predict phenotypes, along with the formulation of new hypotheses that can be very helpful in experimental design.

In this review, we aimed at compiling various methodologies used to approach the problem of modeling integrated biological networks in recent years. Table 1 summarizes these methods together with examples where they were successfully illustrated.

The choice of a particular formalism must be guided by several factors that range from the type and quality of the available data, to the ultimate goal of the modeling procedure and the computational demand for running the model and analyzing it. These factors should all be weighed in order to decide the level of abstraction of the model.

4.1 Different Data, Complementary Methods

A major concern in integrated modeling is the diversity of biological data, e.g., qualitative phenotypic observations, microarray data, metabolic flux analysis, or dynamic time profiles of metabolites.

An ideal modeling tool must be able to handle this diversity, yielding models that reproduce biological reality and predict potential behaviors.

Developing a tool of this sort is a time-consuming process, and no "solves-it-all" solution has yet been found. In spite of the advances in experimental Systems Biology, the current knowledge on biological systems is limited to some specific organisms and quantitative data on cellular processes still lacks. Therefore, to build an integrated model one must rely on several tools, whose outcomes, altogether, represent a space of different-level descriptions of the system.

A key issue of all the presented methods relates with the effect of a priori information on the model which includes previous knowledge about the system, either structural or data/parameter related. In the former case, prior to any other type of analysis, it is essential to establish the network structure (identify its nodes and edges) to represent the system. Incomplete knowledge in this case can sometimes be overcome with structural methods (e.g., hypergraph-based, EM/EP, OT) that search the space of unknown potential regulatory interactions between and within the cell layers. Given a set of alternative network structures, model selection strategies can be applied in order to fill in the knowledge gaps in unknown biological systems. Furthermore, coupled

TABLE 1
Examples of Different Modeling Approaches for Integrated Biochemical Networks

Method	Application	Refs.
	STRUCTURAL AND STOICHIOMETRIC MODELING	
Network-based		
Hypergraph-based	Reconstruction of the central carbon metabolism of *E. coli;*	[43]
EM/EP Analysis	Reduction of the flux distribution space when the core metabolic processes are subjected to gene regulation	[47]
OT-based	Mimics the *E. coli* rFBA model in [55];	[70]
Constraints-based		
rFBA	Regulated metabolic model of *E. coli* metabolism including glucose/acetate growth and glucose-lactose diauxie;	[55]
	Integrated genome-scale metabolic model of *E. coli*;	[56]
	Integrated genome-scale metabolic model of *S. cerevisiae*;	[57]
srFBA	Mimics the whole-genome *E. coli* model in [47];	[58]
iFBA	Extension of the model in [55] including the phosphotransferase catabolite repression system;	[60]
idFBA	Model of the HOG pathway in *S. cerevisiae*;	[62]
	KINETIC MODELING	
Discrete		
Logical-based	Regulated carbon metabolism of *E. coli* subjected to three different carbon sources: glucose, glycerol and acetate;	[72]
Petri Nets	Tryptophan biosynthetic pathway in *E. coli*;	[84]
Continuous		
ODE-based	Regulation of the *lac* operon involved in lactose uptake and metabolism;	[32, 100-103]
	Yeast response to osmotic shock (including phosphorelay, MAP kinase cascade, transcriptional regulation, carbon metabolism and glycerol production);	[104]
	Glucose-Galactose metabolism and catabolite repression in *S. cerevisiae*;	[105]
Hybrid modeling		
Hybrid systems	Initiation of sporulation in *B. subtilis*;	[21]
	Nutrient deprivation in bacteria;	[111]
	Regulation of the *lac* operon in lactose uptake and metabolism in *E. coli*;	[112]
	Xanthophyll cycle reaction in plants;	[108]
Hybrid Functional Petri Nets	Regulation of the urea cycle in the liver;	[114]
	Regulation of early human haematopoiesis;	[116]
	Genetic toggle switch associated to the SOS signaling pathway in *E. coli*;	[117-118]

to efficient learning algorithms, these methods can prove essential for integrated genome-scale model reconstruction, but only for organisms which have already large amounts of biochemical data publicly available.

Another important problem is the overall impact that data incompleteness and quality (e.g., sampling frequency, biological error magnitude, and missing parameters) can have on the procedure. This is a well known open problem with no straightforward answer. Generally speaking, unknown information usually leads to a large parameter solution space with many possible sets, whereas noise and poor data can deform the error surface (which represents model's deviation to reality as a function of its parameters) in such a way that the interception with the real system might be severely compromised. This inevitably hampers the generalization of the model to new experimental conditions and its applicability to predict beyond known training sets.

The problems of a priori information are easily illustrated in FBA-derived methods, such as *rFBA,* where we deal with an optimization procedure within a polyhedral cone (defined by the metabolic and regulatory constraints). Wrong or uncertain information will trim this set in a way that can potentially leave aside the desired flux distribution solution. In the same way, partial regulatory constraints might lead us to a larger set where potential solutions can lay. For this reason, it can be desirable to use pathway-based methods to find alternative potential regulatory structures that can help to strengthen the results from *rFBA* and shrink a potentially large solution space.

Accuracy and quantitative predictions come with the cost of additional information requirements. Kinetic models are considered to be nearest to a good description of biological reality. Parametric models (e.g., ODE-based) are most of the times defined based on kinetic parameters experimentally determined. But, whenever this information is insufficient, these parameters can also be inferred from time-series data in a reverse modeling approach. This usually involves the nonlinear optimization of a large number of parameters for which several methods have been applied (refer to [99], [134] for comprehensive reviews). This is an extremely challenging problem, highly dependent on the quality of the data, due to the complexity of the error surfaces involved and the existence of local extrema. In this context, parameter identifiability has emerged as a key topic to understand model properties [135], [136], [137]. A

fundamental issue in model identification is whether its structure and the available experimental data allow for the determination of a unique solution for each of the model parameters. It is known that some models are not a priori or structurally identifiable, which means that, even with noise-free data, it is impossible to obtain a unique solution. In this case, data incompleteness/quality may affect the specific solution space but several sets of parameters will always be obtained that equally describe the original data. However, practical identifiability also considers data dependency, and it usually relies on statistical methods [138] to assess solution existence and uniqueness. This multiplicity of solutions might hinder parameter estimation algorithms and it is known to have an impact on the generalization capability of the obtained models.

Since poor data results in poor kinetic models, it can be preferrable to have qualitative descriptions of parts of the model (using logical statements) and semiquantitative outcomes than to rely on models that may suffer from overfitting problems. Heuristic approaches can also be sought to deal with missing data on parts of the system, e.g., FBA-derived methods can be coupled to kinetic models, and, using stoichiometric computations alternative flux distributions can be proposed that are coherent with the ODE description and the available data.

Given the current state-of-the-art, we believe that hybrid modeling shows good prospects, giving the best compromise between information requirements and expected model outcomes. Either in the form of the Hybrid Systems framework or through Hybrid Functional Petri Nets, these modeling approaches can be used to study biological systems and the interplay of their different operatory modes in a semiquantitative way. The Hybrid Functional Petri Net framework can also be very advantageous, in comparison to Hybrid Systems, because it additionally presents a quite intuitive visualization system.

The problem of identification in Hybrid Systems is, however, intricate, since it involves the identification of both the operatory modes of the system and their respective parameterizations. It constitutes a topic of current intensive research [139], [140], [141], whose outcomes will greatly benefit the field of Systems Biology.

Since no unifying modeling framework is expected to be developed in the short-term, the Systems Biology modeler must resort to use the available methodologies in a complementary way to build integrated models. Fig. 5 gives a qualitative comparison of the different methodologies reviewed regarding some of the aforementioned criteria that should be considered when choosing the modeling framework for a given biological system.

Petri Nets were not explicitly represented in this figure, since they can be included in several subgroups such as Discrete/Logical, Stochastic, or Hybrid. It is worth remembering that, as models grow in computational demand, their scalability becomes a crucial issue.

It is worth noting that in this review we focused on methods which have been used to model case studies where there is a clear interaction between metabolism and gene regulation, and sometimes signaling pathways. Other methods exist, such as the ones based on concepts of cybernetic modeling [142], [143], or the previously discussed stochastic approaches which remain to be fully explored.

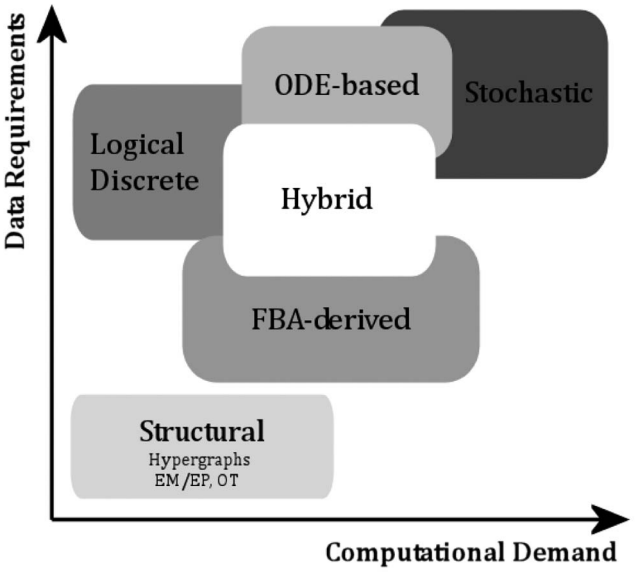

Fig. 5. Qualitative comparison of the different modeling approaches. Comparison relies on data requirements and computational demand. The level of abstraction of each method, the predictive power of the resulting models, and its scalability with system's size are criteria that depend on the chosen variables. Petri Nets are not explicitly represented, since they can be included within Logical/Discrete, Hybrid or Stochastic models.

A complete understanding of how cells operate is still far from being fully accomplished. Integrating information from signaling, gene regulation, and metabolic networks is easily recognized as an important step in this direction and models contemplating the interplay between these processes are essential. Besides helping to rationalize the complexity of cellular systems, integrated models are also expected to become a useful tool for researchers in the wet lab to guide the experimental design of new and informative experiments.

ACKNOWLEDGMENTS

The authors would like to acknowledge financial support from the PhD Program in Computational Biology from Instituto Gulbenkian de Ciencia (sponsored by FCG, Siemens SA and FCT) (SFRH/BD/33209/2007) and by projects DynaMo (PTDC/EEA-ACR/69530/2006, S.Vinga, PI) and PneumoSys (PTDC/SAU-MII/100964/2008, A.R.Neves, PI), from the Portuguese Science Foundation (FCT). This work was partially supported by FCT (INESC-ID multiannual funding) through the PIDDAC Program funds. The authors also thank Dr. Ana Rute Neves (ITQB/UNL, Portugal) and the referees' comments for insightful suggestions that greatly improved the comprehensiveness of the manuscript.

REFERENCES

[1] D. Noble, *The Music of Life: Biology Beyond the Genome*. Oxford Univ. Press, 2006.
[2] F. Crick, "Central Dogma of Molecular Biology," *Nature*, vol. 227, no. 5258, pp. 561-563, Aug. 1970.

[3] F.H. Crick, "On Protein Synthesis," *Proc. Symp. Soc. Experimental Biology,* vol. 12, pp. 138-163, 1958.

[4] A.L. Barabasi and Z.N. Oltvai, "Network Biology: Understanding the Cell's Functional Organization," *Nature Rev. Genetics,* vol. 5, no. 2, pp. 101-113, Feb. 2004.

[5] H. Kitano, "Systems Biology: A Brief Overview," *Science,* vol. 295, no. 5560, pp. 1662-1624, Mar. 2002.

[6] B.H. ter Kuile and H.V. Westerhoff, "Transcriptome Meets Metabolome: Hierarchical and Metabolic Regulation of the Glycolytic Pathway," *Federation of European Biochemical Soc. Letters,* vol. 500, no. 3, pp. 169-171, July 2001.

[7] U. Sauer, M. Heinemann, and N. Zamboni, "Genetics. Getting Closer to the Whole Picture," *Science,* vol. 316, no. 5824, pp. 550-551, Apr. 2007.

[8] A. Quarteroni, "Mathematical Models in Science and Engineering," *Notices of the AMS,* vol. 56, no. 1, pp. 10-19, 2009.

[9] M.J. Herrgard, N. Swainston, P. Dobson, W.B. Dunn, K.Y. Arga, M. Arvas, N. Bluthgen, S. Borger, R. Costenoble, M. Heinemann, M. Hucka, N. Le Novere, P. Li, W. Liebermeister, M.L. Mo, A.P. Oliveira, D. Petranovic, S. Pettifer, E. Simeonidis, K. Smallbone, I. Spasic, D. Weichart, R. Brent, D.S. Broomhead, H.V. Westerhoff, B. Kirdar, M. Penttila, E. Klipp, B.O. Palsson, U. Sauer, S.G. Oliver, P. Mendes, J. Nielsen, and D.B. Kell, "A Consensus Yeast Metabolic Network Reconstruction Obtained from a Community Approach to Systems Biology," *Nature Biotechnology,* vol. 26, no. 10, pp. 1155-1160, Oct. 2008.

[10] B. Teusink, 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, "Can Yeast Glycolysis Be Understood in Terms of in Vitro Kinetics of the Constituent Enzymes? Testing Biochemistry," *European J. Biochemistry,* vol. 267, no. 17, pp. 5313-5329, Sept. 2000.

[11] S. Rinaldi and M. Scheffer, "Geometric Analysis of Ecological Models with Slow and Fast Processes," *Ecosystems,* vol. 3, no. 6, pp. 507-521, Nov./Dec. 2000.

[12] A. Kumar, P.D. Christofides, and P. Daoutidis, "Singular Perturbation Modeling of Nonlinear Processes with Nonexplicit Time-Scale Multiplicity," *Chemical Eng. Science,* vol. 53, no. 8, pp. 1491-1504, Apr. 1998.

[13] M. Rizzi, M. Baltes, U. Theobald, and M. Reuss, "In Vivo Analysis of Metabolic Dynamics in Saccharomyces Cerevisiae.2. Mathematical Model," *Biotechnology and Bioeng.,* vol. 55, no. 4, pp. 592-608, Aug. 1997.

[14] U. Theobald, W. Mailinger, M. Baltes, M. Rizzi, and M. Reuss, "In Vivo Analysis of Metabolic Dynamics in Saccharomyces Cerevisiae.1. Experimental Observations," *Biotechnology and Bioeng.,* vol. 55, no. 2, pp. 305-316, July 1997.

[15] R. Thomas, "Boolean Formalization of Genetic Control Circuits," *J. Theoretical Biology,* vol. 42, no. 3, pp. 563-585, Dec. 1973.

[16] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, "The Yeast Cell-Cycle Network is Robustly Designed," *Proc. Nat'l Academy of Sciences USA,* vol. 101, no. 14, pp. 4781-4786, Apr. 2004.

[17] A. Faure, A. Naldi, C. Chaouiya, and D. Thieffry, "Dynamical Analysis of a Generic Boolean Model for the Control of the Mammalian Cell Cycle," *Bioinformatics,* vol. 22, no. 14, pp. e124-e131, July 2006.

[18] B.C. Goodwin, "Oscillatory Behavior in Enzymatic Control Processes," *Advances in Enzyme Regulation,* vol. 3, pp. 425-438, 1965.

[19] S. Li, P. Brazhnik, B. Sobral, and J.J. Tyson, "A Quantitative Study of the Division Cycle of Caulobacter Crescentus Stalked Cells," *Plos Computational Biology,* vol. 4, no. 1, article no. e9, Jan. 2008.

[20] K.C. Chen, L. Calzone, A. Csikasz-Nagy, F.R. Cross, B. Novak, and J.J. Tyson, "Integrative Analysis of Cell Cycle Control in Budding Yeast," *Molecular Biology of the Cell,* vol. 15, no. 8, pp. 3841-3862, Aug. 2004.

[21] H. De Jong, J. Geiselmann, G. Batt, C. Hernandez, and M. Page, "Qualitative Simulation of the Initiation of Sporulation in Bacillus Subtilis," *Bull. Math. Biology,* vol. 66, no. 2, pp. 261-299, Mar. 2004.

[22] D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann, "Qualitative Simulation of the Carbon Starvation Response in Escherichia Coli," *Biosystems,* vol. 84, no. 2, pp. 124-152, May 2006.

[23] H.H. McAdams and A. Arkin, "Stochastic Mechanisms in Gene Expression," *Proc. Nat'l Academy of Sciences USA,* vol. 94, no. 3, pp. 814-819, Feb. 1997.

[24] J. Paulsson, "Models of Stochastic Gene Expression," *Physics of Life Rev.,* vol. 2, no. 2, pp. 157-175, June 2005.

[25] E. Cinquemani, A. Milias-Argeitis, S. Summers, and J. Lygeros, "Stochastic Dynamics of Genetic Networks: Modelling and Parameter Identification," *Bioinformatics,* vol. 24, no. 23, pp. 2748-2754, Dec. 2008.

[26] H. De Jong, "Modeling and Simulation of Genetic Regulatory Systems: A Literature Review," *J. Computational Biology,* vol. 9, no. 1, pp. 67-103, 2002.

[27] G. Karlebach and R. Shamir, "Modelling and Analysis of Gene Regulatory Networks," *Nature Rev. Molecular Cell Biology,* vol. 9, no. 10, pp. 770-780, Oct. 2008.

[28] E.P. Gianchandani, J.A. Papin, N.D. Price, A.R. Joyce, and B.O. Palsson, "Matrix Formalism to Describe Functional States of Transcriptional Regulatory Systems," *Plos Computational Biology,* vol. 2, no. 8, pp. 902-917, Aug. 2006.

[29] R. Heinrich and S. Schuster, *The Regulation of Cellular Systems.* Chapman & Hall, 1996.

[30] J.S. Edwards, R.U. Ibarra, and B.O. Palsson, "In Silico Predictions of Escherichia Coli Metabolic Capabilities Are Consistent with Experimental Data," *Nature Biotechnology,* vol. 19, no. 2, pp. 125-130, Feb. 2001.

[31] C.H. Schilling, J.S. Edwards, D. Letscher, and B.O. Palsson, "Combining Pathway Analysis with Flux Balance Analysis for the Comprehensive Study of Metabolic Systems," *Biotechnology and Bioeng.,* vol. 71, no. 4, pp. 286-306, 2000.

[32] P. Wong, S. Gladney, and J.D. Keasling, "Mathematical Model of the Lac Operon: Inducer Exclusion, Catabolite Repression, and Diauxic Growth on Glucose and Lactose," *Biotechnology Progress,* vol. 13, no. 2, pp. 132-143, Mar./Apr. 1997.

[33] E.O. Voit, J. Almeida, S. Marino, R. Lall, G. Goel, A.R. Neves, and H. Santos, "Regulation of Glycolysis in Lactococcus Lactis: An Unfinished Systems Biological Case Study," *Systems Biology,* vol. 153, no. 4, pp. 286-298, July 2006.

[34] H.V. Westerhoff, A. Kolodkin, R. Conradie, S.J. Wilkinson, F.J. Bruggeman, K. Krab, J.H. van Schuppen, H. Hardin, B.M. Bakker, M.J. Mone, K.N. Rybakova, M. Eijken, H.J. van Leeuwen, and J.L. Snoep, "Systems Biology towards Life in Silico: Mathematics of the Control of Living Cells," *J Math. Biology,* vol. 58, nos. 1/2, pp. 7-34, Jan. 2009.

[35] R. Alves, E. Vilaprinyo, B. Hernandez-Bermejo, and A. Sorribas, "Mathematical Formalisms Based on Approximated Kinetic Representations for Modeling Genetic and Metabolic Pathways," *Biotechnology and Genetic Eng. Rev.,* vol. 25, pp. 1-40, 2008.

[36] M. Durot, P.Y. Bourguignon, and V. Schachter, "Genome-Scale Models of Bacterial Metabolism: Reconstruction and Applications," *FEMS Microbiology Rev,* vol. 33, no. 1, pp. 164-190, Jan. 2009.

[37] M. Terzer, N.D. Maynard, M.W. Covert, and J. Stelling, "Genome-Scale Metabolic Networks," *WIREs Systems Biology and Medicine,* vol. 1, no. 3, pp. 285-297, 2009.

[38] M.A. Ovacik and I.P. Androulakis, "On the Potential for Integrating Gene Expression and Metabolic Flux Data," *Current Bioinformatics,* vol. 3, no. 3, pp. 142-148, Sept. 2008.

[39] O. Mason and M. Verwoerd, "Graph Theory and Networks in Biology," *IET Systems Biology,* vol. 1, no. 2, pp. 89-119, Mar. 2007.

[40] T. Aittokallio and B. Schwikowski, "Graph-Based Methods for Analysing Networks in Cell Biology," *Brief Bioinformatics,* vol. 7, no. 3, pp. 243-255, Sept. 2006.

[41] R. Albert, "Network Inference, Analysis, and Modeling in Systems Biology," *Plant Cell,* vol. 19, no. 11, pp. 3327-3338, Nov. 2007.

[42] S. Klamt, U.U. Haus, and F. Theis, "Hypergraphs and Cellular Networks," *Plos Computational Biology,* vol. 5, no. 5, article no. e1000385, May 2009.

[43] C.H. Yeang and M. Vingron, "A Joint Model of Regulatory and Metabolic Networks," *BMC Bioinformatics,* vol. 7, article no. 332, July 2006.

[44] S. Schuster, D.A. Fell, and T. Dandekar, "A General Definition of Metabolic Pathways Useful for Systematic Organization and Analysis of Complex Metabolic Networks," *Nature Biotechnology,* vol. 18, no. 3, pp. 326-332, Mar. 2000.

[45] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E.D. Gilles, "Metabolic Network Structure Determines Key Aspects of Functionality and Regulation," *Nature,* vol. 420, no. 6912, pp. 190-193, Nov. 2002.

[46] J.A. Papin, J. Stelling, N.D. Price, S. Klamt, S. Schuster, and B.O. Palsson, "Comparison of Network-Based Pathway Analysis Methods," *Trends in Biotechnology,* vol. 22, no. 8, pp. 400-405, Aug. 2004.

[47] M.W. Covert and B.O. Palsson, "Constraints-Based Models: Regulation of Gene Expression Reduces the Steady-State Solution Space," *J. Theoretical Biology,* vol. 221, no. 3, pp. 309-325, Apr. 2003.

[48] M. Terzer and J. Stelling, "Elementary Flux Modes - State-of-the-Art implementation and Scope of Application," *BMC Systems Biology,* vol. 1, Suppl 1:P2, 2007.

[49] C.T. Trinh, A. Wlaschin, and F. Srienc, "Elementary Mode Analysis: A Useful Metabolic Pathway Analysis Tool for Characterizing Cellular Metabolism," *Applied Microbiology and Biotechnology,* vol. 81, no. 5, pp. 813-826, Jan. 2009.

[50] F. Llaneras and J. Pico, "Stoichiometric Modelling of Cell Metabolism," *J. Bioscience and Bioeng.,* vol. 105, no. 1, pp. 1-11, Jan. 2008.

[51] Z. Szallasi, J. Stelling, and V. Periwal, *System Modeling in Cell Biology : From Concepts to Nuts and Bolts.* MIT Press, 2006.

[52] N.D. Price, J.A. Papin, C.H. Schilling, and B.O. Palsson, "Genome-Scale Microbial in Silico Models: The Constraints-Based Approach," *Trends in Biotechnology,* vol. 21, no. 4, pp. 162-169, Apr. 2003.

[53] M.W. Covert, C.H. Schilling, and B. Palsson, "Regulation of Gene Expression in Flux Balance Models of Metabolism," *J. Theoretical Biology,* vol. 213, no. 1, pp. 73-88, Nov. 2001.

[54] A. Varma and B.O. Palsson, "Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type Escherichia-Coli W3110," *Applied and Environmental Microbiology,* vol. 60, no. 10, pp. 3724-3731, Oct. 1994.

[55] M.W. Covert and B.O. Palsson, "Transcriptional Regulation in Constraints-Based Metabolic Models of Escherichia Coli," *J. Biological Chemistry,* vol. 277, no. 31, pp. 28058-28064, Aug. 2002.

[56] M.W. Covert, E.M. Knight, J.L. Reed, M.J. Herrgard, and B.O. Palsson, "Integrating High-Throughput and Computational Data Elucidates Bacterial Networks," *Nature,* vol. 429, no. 6987, pp. 92-96, May 2004.

[57] M.J. Herrgard, B.S. Lee, V. Portnoy, and B.O. Palsson, "Integrated Analysis of Regulatory and Metabolic Networks Reveals Novel Regulatory Mechanisms in Saccharomyces Cerevisiae," *Genome Research,* vol. 16, no. 5, pp. 627-635, May 2006.

[58] T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin, "A Genome-Scale Computational Study of the Interplay between Transcriptional Regulation and Metabolism," *Molecular Systems Biology,* vol. 3, p. 101, 2007.

[59] S. Rossell, C.C. Van Der Weijden, A. Lindenbergh, A. van Tuijl, C. Francke, B.M. Bakker, and H.V. Westerhoff, "Unraveling the Complexity of Flux Regulation: A New Method Demonstrated for Nutrient Starvation in Saccharomyces Cerevisiae," *Proc Nat'l Academy of Sciences USA,* vol. 103, no. 7, pp. 2166-2171, Feb. 2006.

[60] M.W. Covert, N. Xiao, T.J. Chen, and J.R. Karr, "Integrating Metabolic, Transcriptional Regulatory and Signal Transduction Models in Escherichia Coli," *Bioinformatics,* vol. 24, no. 18, pp. 2044-2050, Sept. 2008.

[61] A. Kremling, K. Bettenbrock, and E.D. Gilles, "Analysis of Global Control of Escherichia Coli Carbohydrate Uptake," *BMC Systems Biology,* vol. 1, p. 42, 2007.

[62] J.M. Lee, E.P. Gianchandani, J.A. Eddy, and J.A. Papin, "Dynamic Analysis of Integrated Signaling, Metabolic, and Regulatory Networks," *Plos Computational Biology,* vol. 4, no. 5, article no. e1000086, May 2008.

[63] S. Hohmann, "Osmotic Stress Signaling and Osmoadaptation in Yeasts," *Microbiology and Molecular Biology Rev.,* vol. 66, no. 2, pp. 300-372, June 2002.

[64] W. Fontana and L.W. Buss, "The Arrival of the Fittest": toward a Theory of Biological Organization," *Bull. Math. Biology,* vol. 56, pp. 1-64, 1994.

[65] P. Dittrich and P. Speroni di Fenizio, "Chemical Organization Theory," *Bull. Math. Biology,* vol. 69, no. 4, pp. 1199-1231, 2007.

[66] C. Kaleta, F. Centler, and P. Dittrich, "Analyzing Molecular Reaction Networks," *Molecular Biotechnology,* vol. 34, no. 2, pp. 117-123, Oct. 2006.

[67] F. Centler and P. Dittrich, "Chemical Organizations in Atmospheric Photochemistries—A New Method to Analyze Chemical Reaction Networks," *Planetary and Space Science,* vol. 55, no. 4, pp. 413-428, Mar. 2007.

[68] N. Matsumaru, F. Centler, P.S. Di Fenizio, and P. Dittrich, "Chemical Organization Theory Applied to Virus Dynamics," *it—Information Technology,* vol. 48, no. 3, pp. 154-160, 2006.

[69] F. Centler, P.S. di Fenizio, N. Matsumaru, and P. Dittrich, "Chemical Organizations in the Central Sugar Metabolism of Escherichia Coli," *Math. Modeling of Biological Systems, Vol I,* pp. 105-119, vol. 379, 2007.

[70] C. Kaleta, F. Centler, P.S. di Fenizio, and P. Dittrich, "Phenotype Prediction in Regulated Metabolic Networks," *BMC Systems Biology,* vol. 2, article no. 37, Apr. 2008.

[71] F. Centler, C. Kaleta, P.S. di Fenizio, and P. Dittrich, "Computing Chemical Organizations in Biological Networks," *Bioinformatics,* vol. 24, no. 14, pp. 1611-1618, July 2008.

[72] J.A. Asenjo, P. Ramirez, I. Rapaport, J. Aracena, E. Goles, and B.A. Andrews, "A Discrete Mathematical Model Applied to Genetic Regulation and Metabolic Networks," *J. Microbiology and Biotechnology,* vol. 17, no. 3, pp. 496-510, Mar. 2007.

[73] I. Koch and M. Heiner, "Petri Nets," *Analysis of Biological Networks,* B.H. Junker and F. Schreiber, eds., pp. 139-180, John Wiley & Sons, 2008.

[74] C. Chaouiya, "Petri Net Modelling of Biological Networks," *Briefings in Bioinformatics,* vol. 8, no. 4, pp. 210-219, July 2007.

[75] W. Reisig, *Petri Nets: An Introduction.* Springer, 1985.

[76] I. Koch, B.H. Junker, and M. Heiner, "Application of Petri Net Theory for Modelling and Validation of the Sucrose Breakdown Pathway in the Potato Tuber," *Bioinformatics,* vol. 21, no. 7, pp. 1219-1226, Apr. 2005.

[77] D.W. Ding and L.N. Li, "Modeling and Analyzing the Metabolism of Riboflavin Production Using Petri Nets," *J. Biological Systems,* vol. 17, no. 3, pp. 479-490, Sept. 2009.

[78] M. Heiner, I. Koch, and R. Will, "Model Validation of Biological Pathways Using Petri Nets—Demonstrated for Apoptosis," *Biosystems,* vol. 75, nos. 1-3, pp. 15-28, July 2004.

[79] A. Sackmann, M. Heiner, and I. Koch, "Application of Petri Net Based Analysis Techniques to Signal Transduction Pathways," *BMC Bioinformatics,* vol. 7, article no. 482, Nov. 2006.

[80] R. Srivastava, M.S. Peterson, and W.E. Bentley, "Stochastic Kinetic Analysis of the Escherichia Coli Stress Circuit Using Sigma(32)-Targeted Antisense," *Biotechnology and Bioeng.,* vol. 75, no. 1, pp. 120-129, Oct. 2001.

[81] L.J. Steggles, R. Banks, and A. Wipat, "Modelling and Analysing Genetic Networks: From Boolean Networks to Petri Nets," *Proc. Conf. Computational Methods in Systems Biology,* pp. 127-141, 2006.

[82] C. Chaouiya, E. Remy, P. Ruet, and D. Thieffry, "Qualitative Modelling of Genetic Networks: From Logical Regulatory Graphs to Standard Petri Nets," *Proc. Int'l Conf. Applications and Theory of Petri Nets '04,* pp. 137-156, 2004.

[83] C. Chaouiya, E. Remy, and D. Thieffry, "Qualitative Petri Net Modelling of Genetic Networks," *Trans. Computational Systems Biology VI,* vol. 4220, pp. 95-112, 2006.

[84] E. Simao, E. Remy, D. Thieffry, and C. Chaouiya, "Qualitative Modelling of Regulated Metabolic Pathways: Application to the Tryptophan Biosynthesis in E.Coli," *Bioinformatics,* vol. 21, pp. 190-196, Sept. 2005.

[85] H. Genrich, R. Kueffner, and K. Voss, "Executable Petri Net Models for the Analysis of Metabolic Pathways," *Int'l J. Software Tools for Technology Transfer,* vol. 3, pp. 394-404, 2001.

[86] C. Li, Q.W. Ge, M. Nakata, H. Matsuno, and S. Miyano, "Modelling and Simulation of Signal Transductions in an Apoptosis Pathway by Using Timed Petri Nets," *J Biosciences,* vol. 32, no. 1, pp. 113-127, Jan. 2007.

[87] H. Matsuno, R. Murakami, R. Yamane, N. Yamasaki, S. Fujita, H. Yoshimori, and S. Miyano, "Boundary Formation by Notch Signaling in Drosophila Multicellular Systems: Experimental Observations and Gene Network Modeling by Genomic Object Net," *Proc. Pac Symp. Biocomputing,* pp. 152-163, 2003.

[88] E. Grafahrend-Belau, F. Schreiber, M. Heiner, A. Sackmann, B.H. Junker, S. Grunwald, A. Speer, K. Winder, and I. Koch, "Modularization of Biochemical Networks Based on Classification of Petri Net T-Invariants," *BMC Bioinformatics,* vol. 9, article no. 90, Feb. 2008.

[89] M. Heiner and I. Koch, "Petri Net Based Model Validation in Systems Biology," *Proc. Applications and Theory of Petri Nets '04,* vol. 3099, pp. 216-237, 2004.

[90] J. Heath, M. Kwiatkowska, G. Norman, D. Parker, and O. Tymchyshyn, "Probabilistic Model Checking of Complex Biological Pathways," *Theoretical Computer Science,* vol. 391, no. 3, pp. 239-257, Feb. 2008.

[91] J.S. Ferreira, R. Lozano, S. Mondie, and A. Friboulet, "Bifurcation Analysis of a Biochemical Network," *Proc. Positive Systems,* vol. 341, pp. 279-286, 2006.
[92] E. Klipp, *Systems Biology in Practice : Concepts, Implementation and Application.* Wiley-VCH, 2005.
[93] S. Goutelle, M. Maurin, F. Rougier, X. Barbaut, L. Bourguignon, M. Ducher, and P. Maire, "The Hill Equation: A Review of Its Capabilities in Pharmacological Modelling," *Fundamental and Clinical Pharmacology,* vol. 22, no. 6, pp. 633-648, Dec. 2008.
[94] M.A. Savageau, *Biochemical Systems Analysis : A Study of Function and Design in Molecular Biology.* Addison-Wesley Pub. Co., 1976.
[95] E.O. Voit, *Computational Analysis of Biochemical Systems : A Practical Guide for Biochemists and Molecular Biologists.* Cambridge Univ. Press, 2000.
[96] V. Hatzimanikatis and J.E. Bailey, "Effects of Spatiotemporal Variations on Metabolic Control: Approximate Analysis Using (Log)Linear Kinetic Models," *Biotechnology and Bioeng.,* vol. 54, no. 2, pp. 91-104, Apr. 1997.
[97] J.J. Heijnen, "Approximative Kinetic Formats Used in Metabolic Network Modeling," *Biotechnol Bioeng,* vol. 91, no. 5, pp. 534-545, Sept. 2005.
[98] A. Sorribas, B. Hernandez-Bermejo, E. Vilaprinyo, and R. Alves, "Cooperativity and Saturation in Biochemical Networks: A Saturable Formalism Using Taylor Series Approximations," *Biotechnology Bioeng,* vol. 97, no. 5, pp. 1259-1277, Aug. 2007.
[99] C.G. Moles, P. Mendes, and J.R. Banga, "Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods," *Genome Research,* vol. 13, no. 11, pp. 2467-2474, Nov. 2003.
[100] A. Babloyan and M. Sanglier, "Chemical Instabilities of All-or-None Type in Beta-Galactosidase Induction and Active Transport," *Federation of the Soc. of Biochemistry Letters,* vol. 23, no. 3, pp. 364-366, 1972.
[101] M. Santillan and M.C. Mackey, "Influence of Catabolite Repression and Inducer Exclusion on the Bistable Behavior of the Lac Operon," *Biophysical J.,* vol. 86, no. 3, pp. 1282-1292, Mar. 2004.
[102] M.J.A. van Hoek and P. Hogeweg, "In Silico Evolved Lac Operons Exhibit Bistability for Artificial Inducers, but Not for Lactose," *Biophysical J.,* vol. 91, no. 8, pp. 2833-2843, Oct. 2006.
[103] M. van Hoek and P. Hogeweg, "The Effect of Stochasticity on the Lac Operon: An Evolutionary Perspective," *Plos Computational Biology,* vol. 3, no. 6, pp. 1071-1082, June 2007.
[104] E. Klipp, B. Nordlander, R. Kruger, P. Gennemark, and S. Hohmann, "Integrative Model of the Response of Yeast to Osmotic Shock," *Nature Biotechnology,* vol. 23, no. 8, pp. 975-982, Aug. 2005.
[105] O. Demir and I.A. Kurnaz, "An Integrated Model of Glucose and Galactose Metabolism Regulated by the Gal Genetic Switch," *Computational Biology and Chemistry,* vol. 30, no. 3, pp. 179-192, June 2006.
[106] J. Lygeros, K.H. Johansson, S.N. Simic, J. Zhang, and S.S. Sastry, "Dynamical Properties of Hybrid Automata," *IEEE Trans. Automatic Control,* vol. 48, no. 1, pp. 2-17, Jan. 2003.
[107] P. Lincoln and A. Tiwari, "Symbolic Systems Biology: Hybrid Modeling and Analysis of Biological Networks," *Proc. Hybrid Systems: Computation and Control,* vol. 2993, pp. 660-672, 2004.
[108] K.H. Cho, K.H. Johansson, and O. Wolkenhauer, "A Hybrid Systems Framework for Cellular Processes," *Biosystems,* vol. 80, no. 3, pp. 273-282, June 2005.
[109] R. Rosen, "Some Realizations of (M, R)-Systems and Their Interpretation," *Bull. Math. Biophysics,* vol. 33, no. 3, pp. 303-319, 1971.
[110] J.L. Casti, "The Theory of Metabolism-Repair Systems," *Applied Math. and Computation,* vol. 28, no. 2, pp. 113-154, Nov. 1988.
[111] M. Belta, P. Finin, L.C.G.J.M. Habets, A.M. Halasz, M. Imielinski, R.V. Kumar, and H. Rubin, "Understanding the Bacterial Stringent Response Using Reachability Analysis of Hybrid Systems," *Proc. Conf. Hybrid Systems: Computation and Control,* vol. 2993, pp. 111-125, 2004.
[112] A. Halasz, V. Kumar, M. Imielinski, C. Belta, O. Sokolsky, S. Pathak, and H. Rubin, "Analysis of Lactose Metabolism in E.Coli Using Reachability Analysis of Hybrid Systems," *IET Systems Biology,* vol. 1, no. 2, pp. 130-148, Mar. 2007.
[113] H. Matsuno, Y. Tanaka, H. Aoshima, A. Doi, M. Matsui, and S. Miyano, "Biopathways Representation and Simulation on Hybrid Functional Petri Net," *In Silico Biology,* vol. 3, no. 3, pp. 389-404, 2003.
[114] M. Chen and R. Hofestadt, "Quantitative Petri Net Model of Gene Regulated Metabolic Networks in the Cell," *In Silico Biology,* vol. 3, no. 3, pp. 347-365, 2003.
[115] W. Marwan, A. Sujatha, and C. Starostzik, "Reconstructing the Regulatory Network Controlling Commitment and Sporulation in Physarum Polycephalum Based on Hierarchical Petri Net Modelling and Simulation," *J. Theoretical Biology,* vol. 236, no. 4, pp. 349-365, Oct. 2005.
[116] S. Troncale, F. Tahi, D. Campard, J.P. Vannier, and J. Guespin, "Modeling and Simulation with Hybrid Functional Petri Nets of the Role of Interleukin-6 in Human Early Haematopoiesis," *Proc. Pac Symp. Biocomputing,* pp. 427-438, 2006.
[117] J. Wu and E. Voit, "Hybrid Modeling in Biochemical Systems Theory by Means of Functional Petri Nets," *J. Bioinformatics Computer Biology,* vol. 7, no. 1, pp. 107-134, Feb. 2009.
[118] J. Wu and E. Voit, "Integrative Biological Systems Modeling: Challenges and Opportunities," *Frontiers Computer Science in China,* vol. 3, no. 1, pp. 92-100, 2009.
[119] P.S. Swain, M.B. Elowitz, and E.D. Siggia, "Intrinsic and Extrinsic Contributions to Stochasticity in Gene Expression," *Proc. Nat'l Academy of Sciences USA,* vol. 99, no. 20, pp. 12795-12800, Oct. 2002.
[120] S. Kar, W.T. Baumann, M.R. Paul, and J.J. Tyson, "Exploring the Roles of Noise in the Eukaryotic Cell Cycle," *Proc. Nat'l Academy of Sciences USA,* vol. 106, no. 16, pp. 6471-6476, Apr. 2009.
[121] A. Arkin, J. Ross, and H.H. McAdams, "Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage Lambda-Infected Escherichia Coli Cells," *Genetics,* vol. 149, no. 4, pp. 1633-1648, Aug. 1998.
[122] D. Gonze and A. Goldbeter, "Circadian Rhythms and Molecular Noise," *Chaos,* vol. 16, no. 2, article no. 026110, June 2006.
[123] D. Schultz, E. Ben Jacob, J.N. Onuchic, and P.G. Wolynes, "Molecular Level Stochastic Model for Competence Cycles in Bacillus Subtilis," *Proc. Nat'l Academy of Sciences USA,* vol. 104, no. 45, pp. 17582-17587, Nov. 2007.
[124] L.S. Weinberger, J.C. Burnett, J.E. Toettcher, A.P. Arkin, and D.V. Schaffer, "Stochastic Gene Expression in a Lentiviral Positive-Feedback Loop: Hiv-1 Tat Fluctuations Drive Phenotypic Diversity," *Cell,* vol. 122, no. 2, pp. 169-182, July 2005.
[125] V. Prasad and K.V. Venkatesh, "Stochastic Analysis of the Gal Genetic Switch in Saccharomyces Cerevisiae: Modeling and Experiments Reveal Hierarchy in Glucose Repression," *BMC Systems Biology,* vol. 2, article no. 97, Nov. 2008.
[126] D.T. Gillespie, "General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical-Reactions," *J. Computational Physics,* vol. 22, no. 4, pp. 403-434, 1976.
[127] D.T. Gillespie, "The Chemical Langevin Equation," *J. Chemical Physics,* vol. 113, no. 1, pp. 297-306, July 2000.
[128] M. Rathinam, L.R. Petzold, Y. Cao, and D.T. Gillespie, "Stiffness in Stochastic Chemically Reacting Systems: The Implicit Tau-Leaping Method," *J. Chemical Physics,* vol. 119, no. 24, pp. 12784-12794, Dec. 2003.
[129] A. Chatterjee, K. Mayawala, J.S. Edwards, and D.G. Vlachos, "Time Accelerated Monte Carlo Simulations of Biological Networks Using the Binomial Tau-Leap Method," *Bioinformatics,* vol. 21, no. 9, pp. 2136-2137, May 2005.
[130] A. Chatterjee and D.G. Vlachos, "Multiscale Spatial Monte Carlo Simulations: Multigriding, Computational Singular Perturbation, and Hierarchical Stochastic Closures," *J. Chemical Physics,* vol. 124, no. 6, article no. 64110, Feb. 2006.
[131] C.S. Gillespie, "Moment-Closure Approximations for Mass-Action Models," *IET Systems Biology,* vol. 3, no. 1, pp. 52-58, Jan. 2009.
[132] B. Munsky and M. Khammash, "The Finite State Projection Algorithm for the Solution of the Chemical Master Equation," *J. Chemical Physics,* vol. 124, no. 4, article no. 044104, Jan. 2006.
[133] B. Munsky and M. Khammash, "The Finite State Projection Approach for the Analysis of Stochastic Noise in Gene Networks," *IEEE Trans. Circuits and Systems I-Regular Papers,* pp. 201-214, Jan. 2008.
[134] I.C. Chou and E.O. Voit, "Recent Developments in Parameter Estimation and Structure Identification of Biochemical and Genomic Systems," *Math. Bioscience,* vol. 219, no. 2, pp. 57-83, June 2009.
[135] S. Audoly, G. Bellu, L. D'Angio, M.P. Saccomani, and C. Cobelli, "Global Identifiability of Nonlinear Models of Biological Systems," *IEEE Trans Biomedical Eng,* vol. 48, no. 1, pp. 55-65, Jan. 2001.

[136] L. Ljung and S.T. Glad, "On Global Identifiability for Arbitrary Model Parameterizations," *Automatica,* vol. 30, no. 2, pp. 265-276, 1994.
[137] M.P. Saccomani, S. Audoly, G. Bellu, and L. D'Angio, "Parameter Identifiability of Nonlinear Biological Systems," *Positive Systems,* pp. 87-93, 2004.
[138] S. Hengl, C. Kreutz, J. Timmer, and T. Maiwald, "Data-Based Identifiability Analysis of Non-Linear Dynamical Models," *Bioinformatics,* vol. 23, no. 19, pp. 2612-8, Oct. 2007.
[139] L. Pu, J. Hu, and B. Chen, "Information Theoretical Approach to Identification of Hybrid Systems," *Proc. Int'l Workshop Hybrid Systems: Computation and Control,* pp. 650-653, 2008.
[140] F. Lauer and G. Bloch, "Switched and Piecewise Nonlinear Hybrid System Identification," *Proc. Int'l Workshop Hybrid Systems: Computation and Control,* pp. 330-343, 2008.
[141] N. Nandola and S. Bhartiya, "Hybrid System Identification Using a Structural Approach and Its Model Based Control: An Experimental Validation," *Nonlinear Analysis: Hybrid Systems,* vol. 3, no. 2, pp. 87-100, 2009.
[142] J.D. Young, K.L. Henne, J.A. Morgan, A.E. Konopka, and D. Ramkrishna, "Integrating Cybernetic Modeling with Pathway Analysis Provides a Dynamic, Systems-Level Description of Metabolic Control," *Biotechnology and Bioeng.,* vol. 100, no. 3, pp. 542-59, June 2008.
[143] J.I. Kim, J.D. Varner, and D. Ramkrishna, "A Hybrid Model of Anaerobic E. Coli Gjt001: Combination of Elementary Flux Modes and Cybernetic Variables," *Biotechnology Progress,* vol. 24, no. 5, pp. 993-1006, Sept./Oct. 2008.

Nuno Tenazinha received the degree in biological engineering in 2006 from Instituto Superior Técnico, Technical University of Lisbon (IST-UTL) and is working toward his PhD in systems biology at Instituto de Tecnologia Química e Biológica—Universidade Nova de Lisboa (ITQB-UNL) and he completed the first year of the PhD Program in computational biology held by the Calouste Gulbenkian Foundation. He is currently developing his thesis work at INESC-ID (Lisbon, Portugal). His current interests range from the development of mathematical modeling strategies, to understand integrated biological networks, to the use of hybrid systems, and their tools to approach the modeling and analysis of regulated metabolic networks.

Susana Vinga received the mechanical engineering degree in 1999, post-graduate degree in probability and statistics at Instituto Superior Técnico (IST-UTL), and the PhD degree in Bioinformatics in 2005 at ITQB-UNL (Portugal) on biological sequences analysis by alignment-free methods. She is currently a researcher in the KDBIO group at INESC-ID (Lisbon, Portugal) and an invited assistant professor at the Faculty of Medical Sciences at the Universidade Nova de Lisboa (FCM-UNL) and is working in projects in the area of systems biology, in particular the dynamic modeling, analysis and control of metabolic and gene regulatory networks. She was granted in 2010 the Young Researcher Award of the Technical University of Lisbon in the area of Informatics.