FEBS Letters 583 (2009) 3923–3930

FEBS Letters

journal homepage: www.FEBSLetters.org

Review

Systems analysis of cellular networks under uncertainty

Hans-Michael Kaltenbach, Sotiris Dimopoulos, Jörg Stelling *

Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland
Swiss Institute of Bioinformatics, ETH Zurich, 4058 Basel, Switzerland

ARTICLE INFO

Article history:
Received 15 September 2009
Revised 26 October 2009
Accepted 27 October 2009
Available online 29 October 2009

Edited by Johan Elf

Keywords:
Dynamic model
Uncertainty
Structural network analysis
Chemical reaction network theory
Monotone system
Ensemble modeling

ABSTRACT

Besides the often-quoted complexity of cellular networks, the prevalence of uncertainties about components, interactions, and their quantitative features provides a largely underestimated hallmark of current systems biology. This uncertainty impedes the development of mechanistic mathematical models to achieve a true systems-level understanding. However, there is increasing evidence that theoretical approaches from diverse scientific domains can extract relevant biological knowledge efficiently, even from poorly characterized biological systems. As a common denominator, the methods focus on structural, rather than more detailed, kinetic network properties. A deeper understanding, better scaling, and the ability to combine the approaches pose formidable challenges for future theory developments.

1. Introduction

It is a key concept of systems biology to iteratively combine large-scale experimental analysis with mathematical modeling in order to eventually elucidate how biological systems operate [1]. This concept, per se, is not new – similar approaches have been followed for decades in physiology and theoretical biology [2]. Unprecedented developments of high-throughput, large-scale experimental technologies – such as genomics, proteomics, metabolomics – however, have opened realistic opportunities for system-wide analyses in biology, at least at the cellular level. Mathematical modeling can employ this data to generate qualitative or quantitative predictions and, thereby, evaluate biological hypotheses, suggest new experiments for validation, and ultimately increase our systems-level understanding [3]. Transcriptional regulatory networks provide one example where this combination of experimentation and modeling has been particularly successful [4].

The complexity of cellular systems constitutes an obvious challenge for mathematical modeling in systems biology. For instance, it is unclear how detailed dynamic models of small-scale systems could eventually be scaled to entire cells, or joined with coarser-grained but large-scale models [4]. Uncertainty is another important and less appreciated factor that requires new theory development to increase the power of mathematical models as systems analysis tools. Probabilistic models cope with uncertainty by design, but they often do not respect first principles such as mass conservation, do not cover dynamic processes, and yield only limited insight into mechanistic detail. Hence, it is important to consider the effect of uncertainty on mechanistic models.

Uncertainties fall into two broad categories [5]. Aleatoric uncertainty stems from the inherent randomness in the behavior of the system under examination. In biology, for example, noise in gene expression induces uncertainty in the model output. Since the noise stems from physical principles, this uncertainty cannot be avoided and needs to be addressed by stochastic analysis. This highly active research area is summarized in recent reviews [6–8]. Here, we focus on the second type of uncertainty, epistemic uncertainty, which results from our lack of knowledge on the system. In current mechanistic mathematical models of biological systems, epistemic uncertainty is profoundly present due to practical limitations, such as lack of understanding of the underlying mechanisms, incomplete coverage and measurement errors in various modeling quantities and, most commonly, parameters derived from noisy or incoherent data sets. A recent study of coverage in shotgun proteomics illustrates such technical limitations quantitatively [9] for one type of data required for mathematical modeling.

For future perspectives, consider the example of budding yeast as a best-case scenario. *S. cerevisiae* is arguably the most intensely studied eukaryotic model organism of 'manageable' complexity.

* Corresponding author. Address: Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland. Fax: +41 61 38 73994.
E-mail address: joerg.stelling@bsse.ethz.ch (J. Stelling).

0014-5793/$36.00 © 2009 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2009.10.074

Yet, ten years after sequencing the genome in 1996, roughly 1/6 of the organism's genes still remained un-annotated [10], and only very recently, it was possible to establish growth phenotypes for all individual genes using chemical genomics approaches [11]. Hence, ever accumulating perfect quantitative knowledge on biological systems for epistemic uncertainty to cease currently seems unrealistic.

Incomplete knowledge affects mathematical models based on first principles in different ways. Here, we consider ordinary differential equation (ODE) models derived from mass balances of n individual components in a biochemical network with r reactions. The general form of such an ODE system is:

$$\frac{d\mathbf{c}(t)}{dt} = \mathbf{N}\cdot\mathbf{v}(\mathbf{c}(t),\mathbf{u}(t),\mathbf{k}), \quad \mathbf{c}(t_0)=\mathbf{c}_0 \tag{1.1}$$

with the $n\times 1$ vector of time-dependent concentrations $\mathbf{c}(t)$, the $n\times r$ stoichiometric matrix $\mathbf{N}$, and the $r\times 1$ vector function of reaction rates – or fluxes – $\mathbf{v}(\cdot)$. The fluxes depend on the system state $\mathbf{c}(t)$, on potentially time-varying inputs $\mathbf{u}(t)$, and on kinetic parameters $\mathbf{k}$ such as affinity constants. Finally, $\mathbf{c}_0$ denotes the initial state of the system, for instance, absolute protein concentrations. Uncertainty enters in the form of unknown or poorly estimated parameter values $\mathbf{k}$ and initial conditions $\mathbf{c}_0$. Moreover, missing or incorrect reactions affect the model structures, namely stoichiometries $\mathbf{N}$ and reaction rate laws $\mathbf{v}(\cdot)$. The resulting nested model uncertainties are difficult to handle [12].

To cope with the combination of complexity and uncertainty in biological systems, it is important to realize that mathematical models need to be adapted to the phenomena of interest as well as to the scientific questions they are intended to answer. Or, as a lesson from other complex systems: "Don't model bulldozers with quarks" [13]. In particular, it is not always necessary to specify ODE-based models completely. Using advanced computational methods for exploiting the existing knowledge to the largest possible extent provides a pragmatic approach to gain biological knowledge and to reduce the experimental efforts. However, it poses important theory challenges. Here, we discuss methods that can potentially cope with increasing levels of uncertainty; they originate from different long-term theory developments that start to converge.

2. Structural network analysis

Network structures, especially the stoichiometries of biochemical reactions, are relatively well-characterized and therefore suitable starting points to analyze the largely unknown relationships between structure, function and control in complex cellular networks.

2.1. Concepts

Horn and Jackson were the first to study the effects of stoichiometric coupling in (chemical) reaction networks on their behavior [14]. They realized that – even without knowing parameter values $\mathbf{k}$ and component concentrations $\mathbf{c}$ – the stoichiometry imposes important constraints on network fluxes. Considering only steady states and neglecting the dynamics, Eq. (1.1) simplifies to:

$$\frac{d\mathbf{c}}{dt} = \mathbf{0} \;\Rightarrow\; \mathbf{N}\cdot\mathbf{v} = \mathbf{0} \tag{1.2}$$

Now, the space of feasible fluxes $\mathbf{v}$ is only determined by properties of the stoichiometric matrix $\mathbf{N}$ and potentially other constraints on $\mathbf{v}$ such as reaction reversibilities and capacities as illustrated in Fig. 1. The approach rapidly gained importance for biology after Palsson and colleagues proposed the flux balance analysis (FBA) method in 1992. FBA determines a specific flux distribution in a metabolic network by additionally considering optimality of, for example, biomass production [15]. The method and its extensions have found numerous applications in biology to analyze specific properties and behaviors of genome-scale metabolic networks; see [16,17] for recent reviews. However, it might be even more rewarding to characterize *all* possible behaviors of a network, which is the domain of metabolic pathway analysis.

2.2. Metabolic pathways

Formal metabolic pathways – as opposed to conceptual 'glycolysis', 'TCA cycle', etc. – have well-defined mathematical structures. Two such concepts are elementary flux modes (EFMs) and extreme pathways (EPs); they both originate from Clarke's early work on convex analysis of stoichiometric networks [18,19]. Since EFMs and EPs are closely related, we only consider EFMs in the following. Importantly, EFMs correspond to minimal (that is, non-decomposable) subnetworks that can operate a network at steady state while fulfilling all constraints that are imposed by reaction stoichiometries and reversibilities. They open up a *constructive* inroad for characterizing network behavior: all feasible flux distributions, and only those, are obtained by non-negative (convex) combinations of EFMs (see [20] for technical details). Hence, pathway analysis, in principle, allows one to *comprehensively* investigate the space of all states of (metabolic) networks that are meaningful for the cell.

Early applications of EFM analysis considered small-scale networks such as parts of canonical monosaccharide metabolism, thereby recovering text book pathways as well as proposing new network operation modes [21]. The first larger-scale analysis of *Escherichia coli* central metabolism showed the potential of EFM analysis for characterizing phenotypes and robustness of metabolic networks [22]. Primarily due to increased capabilities for computing EFMs, applications are now becoming possible for genome-scale networks. Recent studies indicated remarkable variability of fluxes within a narrow region of growth optimality [23], and helped uncover optimal, but rather counter-intuitive modes of network operation [24].

Importantly, large-scale analyses showed that focusing on 'core metabolism', while neglecting the peripheral network, can be very misleading [23,25]. This implies important challenges for EFM-based network analysis. First, it is not yet possible to analyze networks larger than ~300 reactions because of the high computational demands for determining EFMs. High-performance computing approaches and parallelization are options for further scaling. In addition, already now, EFM computations result in up to 10^7 pathways for large networks. Even simple statistical analysis such as determining the distribution of path lengths thus becomes non-trivial, and more advanced methods for clustering and other analyses at this scale are needed [23].

2.3. Network structure and regulation

While structural network analysis traditionally focuses on metabolic networks, several theories and computational methods have been proposed to infer (transcriptional) regulatory network structures and behavior from metabolic network structures. This is different from designing hybrid models that abstractly represent *known* genetic control mechanisms, for instance, via Boolean logic to more accurately predict the metabolic phenotype [26]. Reverse-engineering of the control network, at least in microorganisms, appears feasible because the control structures are sparse – only few metabolites directly control each regulator and vice-versa – and hierarchical [27]. In addition, there is increasing evidence that metabolic networks operate in a limited number of dominant functional modes in steady-state. Additional assumptions on optimal

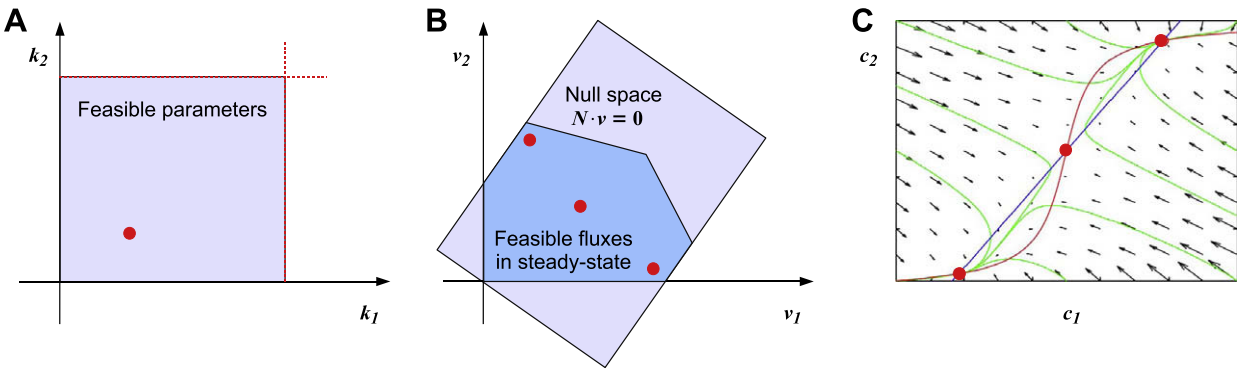

Fig. 1. Relations between spaces in a dynamical system model. In the parameter space (A), constraints such as non-negativity of constants in mass-action kinetics and upper bounds on reaction rates (red lines) make only certain parameter values feasible (shaded area). The flux space (B) permits only sets of steady-state fluxes that are compatible with the stoichiometries (null space, light area), and with reaction reversibilities and capacities (dark area). The system dynamics takes places in the concentration space (state space, C). The vector field of concentration changes for an example system is indicated by arrows and states with zero velocity in one of the directions (nullclines) are shown as red and blue lines, respectively. The system has three steady states (red dots); two steady states are stable and the trajectories (green lines) converge to them with time. The steady states map to three different flux distributions in (B), but to a single set of kinetic constants in (A), as illustrated by the red dots.

control, such as most efficient allocation of resources using a minimal set of regulators could enable predictions of metabolic network regulation. For instance, by comparison of experimental data on gene expression with EFM-derived predictions derived from metabolic network structures alone, the existence of correlations between metabolic fluxes and genetic control has been demonstrated [22].

Other methods assume optimal rejection of perturbations in the sense of a minimal deviation from a given operating point. They use continuous ('minimization of metabolic adjustment', MOMA; [28]) or discrete ('regulatory on/off minimization', ROOM; [29]) distance metrics to quantify the deviation between original and perturbed state. Similarly, sensitivity analysis based on the network structure alone (and using EFMs to construct flux distributions) provided evidence of a close link between network topology and gene expression control [30]. Further developments could not only allow for making specific, experimentally testable predictions, but also for revealing more general design principles of cellular control.

3. Chemical reaction network theory

While the structural network analysis methods rely solely on the network stoichiometry, other methods additionally take into account the algebraic structure of the kinetic rate laws $\mathbf{v}(\cdot)$ and physical constraints such as non-negativity of concentrations $\mathbf{c}$ and parameters $\mathbf{k}$. Thereby, they extend the analysis to dynamic features of a network, despite uncertainties in the kinetic parameters.

3.1. Concepts

Chemical reaction network theory (CRNT) was developed by Horn and Jackson [14,31] in the 1970s, and extended by Feinberg [32,33] in the 1980s. It allows establishing or excluding the existence of multiple steady states for all physically feasible sets of kinetic rate constants in certain very restricted classes of chemical reaction networks captured by mass-action kinetics (see Fig. 1 for an illustration). Most results of CRNT exploit that taking the logarithm of such a mass-action rate law yields a linear equation in the logarithms of the species concentrations.

Given a set of reactions among chemical species, each left- and right-hand side of a reaction is replaced by a new variable, called a 'complex', such that each reaction can be interpreted as one directed edge connecting two complexes in a reaction graph. For example, consider the network given in Fig. 2A. This Michaelis–Menten like network consists of four species E, P, S, C, and $m = 6$ complexes *C0–C5*, namely 0, P, S, E+S, E·S, E+P. Balancing all fluxes in the graph results in a 'complex balancing' [14]; CRNT provides the means to analyze the interplay of complex balancing and steady states on the species level. For this, the 'deficiency' δ of the network describes the degree of linear dependencies among complexes. It can be computed solely from the stoichiometry of the network as $\delta = m - l - s$, where m is the number of complexes, l the number of connected components of the complex graph, and s the rank of the stoichiometric matrix. Our example (Fig. 2A) consists of $l = 2$ linkage classes *L1–L2*, and the stoichiometric matrix has rank $s = 3$. Hence, the network's deficiency is $\delta = 1$.

Key results of CRNT show that for a deficiency zero network, species and complex steady states coincide. For the notion of single and multiple steady states, CRNT introduces so-called compatibility classes which define an equivalence relation among steady states; two steady states are equivalent if their corresponding initial conditions lead to trajectories confined in the same subspace. Under further conditions on the strongly connected components of the complex graph, the existence of at most one steady state per compatibility class can be guaranteed by the Deficiency-1-Theorem for all positive rate constants, provided the deficiency of each connected component is zero or one. If these deficiencies are all equal to zero, the Deficiency-0-Theorem additionally guarantees the existence and asymptotic stability of the steady state [33]. For a different set of regularity conditions on the complex graph, the Deficiency-1-Algorithm applies. It determines if the network can exhibit multiple steady states and subsequently computes one corresponding set of positive rate constants [32]. For the network in Fig. 2A the Deficiency-1-Theorem applies; this establishes the existence of at most one steady state, regardless of the kinetic rate constants. Similarly, CRNT shows potential bistability of the network in Fig. 2B, which is again enzyme catalysis, but with mixed inhibition [34].

3.2. Qualitative dynamics for biology

Despite the theory's elegance and fundamental nature, it has found relatively few applications in (systems) biology. Nevertheless, CRNT is a very good illustration for new theory development induced by biological problems. Feinberg employed examples from biochemistry such as the examples in Fig. 2 to re-formulate central elements of CRNT in terms of the reaction networks' graph structures [34]. The analysis of kinetic proof-reading of T-cell receptors

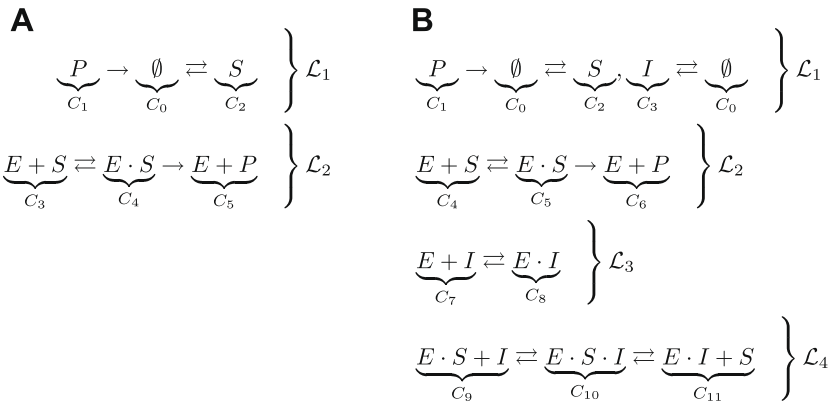

Fig. 2. Chemical reaction networks and multistationarity. (A) Enzyme catalysis following a Michaelis–Menten like scheme with three species, five complexes (*C*) and two connected components (*L*). The substrate is produced with constant rate; both substrate and product are degraded with rate proportional to their concentration. (B) Enzyme catalysis with mixed inhibition consisting of four species, eleven complexes and four connected components. The inhibitor I is additionally produced with constant rate and degraded with rate proportional to its concentration.

(a deficiency-zero network) lead Sontag to extend the theory to certain classes of non-mass action kinetics, and to incorporate robustness analysis, which deals with the network's resilience to perturbations in parameters [35]. More recently, fundamental theorems for classes of biochemical networks that show (near) robust adaptation – as observed in the classical *E. coli* chemotaxis example – have been proposed [36]. Corresponding insights into 'first principles' of biological network functions are clearly needed.

CRNT's main current limitations, however, concern the theory's applicability to broader classes of realistic networks as well as its scalability to larger systems. Two approaches have begun to address these issues, both by sacrificing the generality of CRNT to a certain extent. One approach considers the detailed relations between parameter space and behavior (see Fig. 1 for an illustration). More precisely, formulation of an optimization problem allows to investigate the network dynamics in regions of parameter space with qualitatively different behavior [37]. Similarly, considerations of a systems 'design space' (a compressed version of the parameter space) by Savageau et al. focus on the tolerance of systems in different regions of this space [38].

Another approach aims at decomposing a complex network into smaller subunits, namely EFMs as discussed in the previous section. Each EFM of a network has either deficiency zero or one, such that the Deficiency-1-Algorithm can be applied on each EFM, possibly with relaxed regularity conditions as provided in [39]. If an EFM is found to exhibit multiple steady states, it is often possible to extend this finding to the whole network [40]. On the contrary, multiple steady states cannot be excluded from the fact that each EFM only exhibits a single steady state. With this analysis, CRNT allows discrimination of model variants of a network without knowledge of the kinetic rate constants [41]. Overall, different directions of theory development are being pursued – but we understand their deeper connections only poorly.

4. Monotone systems

The theory of monotone input/output systems, developed recently by Angeli and Sontag [42–44], is founded in engineering control theory. It assumes less about the model structures, which makes it appealing for the analysis of biological systems.

4.1. Concepts

In contrast to CRNT, the theory of monotone systems [45] does not rely on a particular type of kinetic rate law. It linearizes the system of ODEs describing the reactions (Eq. (1.1)) around a steady state $\mathbf{c}^*$ and uses the structure of the Jacobian matrix

$$\mathbf{J} = \mathbf{N} \cdot \left.\frac{\partial \mathbf{v}(\mathbf{c}, \mathbf{u}, \mathbf{k})}{\partial \mathbf{c}}\right|_{\mathbf{c}^*} \tag{1.3}$$

to determine properties of interest. The *i,j*th entry of the Jacobian describes the amount of change in species *i* induced by a unit amount of change in species *j* with a positive or negative entry corresponding to an activation or inhibition, respectively. The Jacobian can be interpreted as an adjacency matrix of a graph with one node per chemical species and connecting edges labeled by −1 or +1 for inhibition or activation, respectively. A system is monotone if there is a vertex labeling {−1,+1} of the species such that the product of each edge label and its adjacent vertex labels results in +1. The existence of such a labeling guarantees that there is no situation in which the increase of a species leads to an activation of another species via one reaction path and an inhibition of the same species via another path. Consider the example network shown in Fig. 3. Assigning activation (+1) to node 1, this activation is propagated to an activation of node 2 and an inhibition of node 3. Then, the response of node 4 on the initial activation of node 1 would depend on the kinetic parameters of the network and, hence, this network is not monotone. However, if the edge 1–2 would be an inhibition,

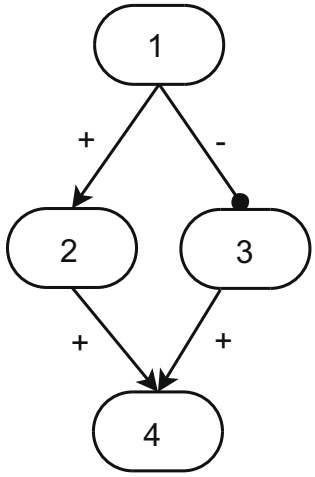

Fig. 3. Example interaction network. Nodes represent components (1–4) and directed links indicate positive (+) or negative (−) interactions between them.

an activation of node 1 would unambiguously lead to an inhibition of node 4 and the network would be monotone.

Sontag and co-workers extended the concept to biological systems with inputs and outputs and examined the system response to constant input signals [44]. The characteristic of a system then describes the steady-state of the output as a function of the constant input. Stimulating a biological system with constant input concentrations of increasing value and measuring the steady-state output concentrations, such a characteristic can be determined by interpolation; an obvious example of a characteristic would be a dose–response curve. The concept of a characteristic also extends to multiple inputs and outputs. Further, analysis of monotone systems can also be applied for revealing multiple steady states.

4.2. Modular network decomposition

As a single chemical species is always a monotone system in itself, a chemical reaction system can always be decomposed into interconnected monotone systems such that the output of one system provides the input to another system. An example of a decomposition is 'pulling out' a negative feedback loop, resulting in a monotone system with the negative feedback connecting its output and input. Sontag and co-workers argue that biological systems are near-monotone, i.e., only few parts have to be pulled out to make the resulting system monotone. Steady states of two connected monotone systems can be determined by the two characteristics alone. Hence, decomposition into monotone subsystems allows characterization of the steady-states of the full system without explicit knowledge of the kinetic rate constants.

Knowing only that the system is monotone, it has been possible to derive the full bifurcation diagram of a simplified MAPK cascade model by using only the characteristic. Similarly, Sontag [44] established the existence of a single, globally attractive, stable steady state in three other examples: a MAPK cascade model with negative feedback, a testosterone model and a simplified model of the Lac operon, by applying a small-gain theorem for monotone input/output systems. In particular, oscillatory behavior could be excluded for the MAPK cascade and the Lac operon feedbacks even for arbitrary delays in the feedback loop. The theory thus provides an important tool for analyzing networks with a high degree of structural uncertainty, as the only knowledge required is the monotonicity of the system and its characteristic, which can both be established without explicit knowledge of the kinetic rate laws and constants.

5. Uncertain dynamics

The theoretical approaches discussed above cover the classes of problems encountered in systems biology, however, only partially. Typically, one is also interested in systems where parameter values ($\mathbf{k}$) and model structures ($\mathbf{N}$, $\mathbf{v}(\cdot)$) are uncertain, even regarding monotonicity for the latter. In addition, the behavior of interest can be a dynamic behavior, for example, the detailed activity profile of a signal transduction pathway. This leads to hard problems of model selection and discrimination [12].

5.1. Concepts

One aspect of model selection concerns the identification of parameters for a model structure that is assumed to be correct. Given a set of experimental data and their variances, model parameters are estimated using one of the optimization algorithms available under a maximum likelihood criterion [46]. Subsequent analyses based on simulation of the parametrized model estimate the likelihood that the model generated the data, the accuracy of the estimated parameters, and other relevant statistical quantities (see [47] for a recent review).

For instance, to address the issue of parametric uncertainty, one can approximate the parameter errors either by combining parameter sensitivities – which measure the change in model output upon parametric perturbations – with experimental measurement errors via the Fisher Information Matrix, or by a quadratic expansion of the estimation error functional involving the Hessian matrix [48]. It is important to note that these methods imply several simplifications (e.g., linearization of the dynamic system for the sensitivity calculation) such that the resulting, symmetric confidence intervals for the parameters only approximate the real distributions.

Alternatively, Monte Carlo methods are the traditional statistical tools for uncertainty quantification [49]. These approaches evaluate the model multiple times by sampling from probability distributions of the model's quantities (such as parameters), responsible for the system's uncertainty. However, except for constraints derived from thermodynamics [12] it is not clear which type of distribution should be used for sampling. Common choices such as the log-normal, normal, or uniform distribution have only weak heuristic justification. A further crucial step involves the selection of a sampling strategy; popular choices are the quasi-Monte Carlo (QMC) method, the Latin hypercube sampling, and various Markov Chain Monte Carlo methods [50]. As illustrated in Fig. 4, simulation results then approximate the effect of uncertainties on model outputs by computing corresponding empirical probability density functions (PDFs). Alternatively, a bootstrap approach [51] can be employed by randomizing the experimental data and re-estimating parameters. However, in large biochemical networks, the repeated simulations together with the high dimension of the parameter space make simple sampling-based approaches extremely computationally demanding and therefore practically infeasible. Hence, even for a known model structure, quantifying the uncertainties associated with a model's parameters involves important trade-offs between computational efficiency and accuracy.

In addition, when the network topology is ill-characterized, it is necessary to rank and discriminate possible models (biological hypotheses) based on their consistency with the experimental observations. The key problem in structure identification is the infeasibility of enumerating all (exponentially many) possible network topologies for large systems when compute-intensive parameter estimation for each dynamic model is needed. Hence, structure identification for dynamic models is conceptually and computationally difficult. It has often been confined to small reaction networks [52] or to aspects of reaction kinetics [53].

5.2. Uncertainty quantification and model selection

Uncertainty analysis plays a crucial role in establishing the credibility of a computational model, since it quantifies the effect of various uncertainties on the model output. Simple sampling-based strategies do not scale well, as discussed above. Therefore, applications of uncertainty quantification in large-scale biological systems are not yet available, but efforts have been made for smaller systems. For example, uncertainty analysis was performed in a subset of parameters of an ODE model of *M. tuberculosis* infection in humans [54]. It was also used for the discrimination of kinetic models by maximizing the overlap between the densities reflecting the model's variability and the experimental data's variability [55] (see Fig. 4C and D for an illustration). Note that various other model selection criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) exist; they utilize maximum likelihood scores as a measure of fit and regularize for model complexity at the same time [56]. In addition, it might be worthwhile to adapt methods from other domains to biology. Probabilistic bounds for an acceptable performance of engineering

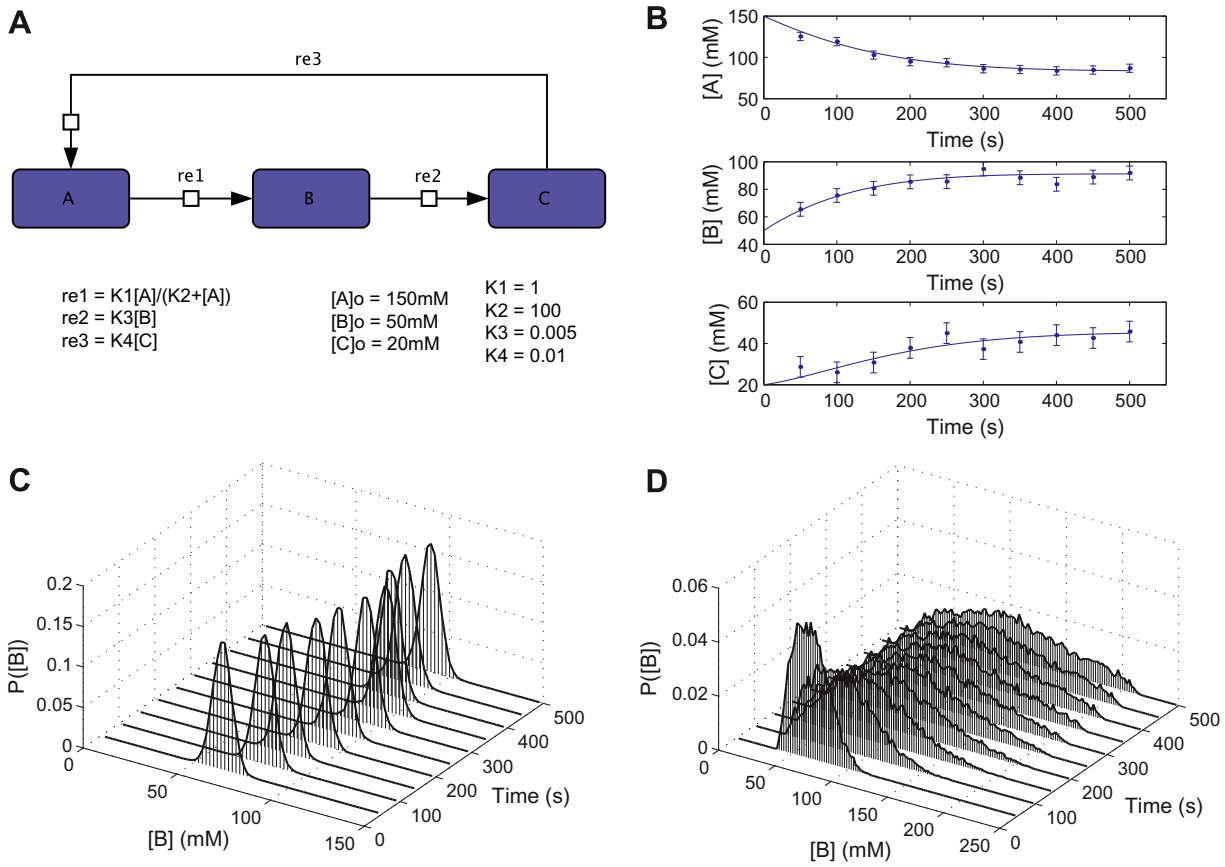

Fig. 4. Parametric uncertainty effect on a simple dynamical system's output. (A) Schematic representation of a simple ODE dynamic system. Initially protein A is converted to B. Then protein B is converted to C, which is finally exerting a feedback to A. Michaelis–Menten and mass-action kinetics are employed for the corresponding conversions. (B) A typical view of the simulation results for the time evolution of the three protein concentrations (solid lines) is shown, together with perturbed synthetic experimental data points (dots) and their standard deviations as typically given by experimentalists. (C) 3D view of the underlying probability density functions (PDFs) of the experimental data points corresponding to protein B, assuming that measurement noise follows a Gaussian distribution. (D) PDFs of the simulated concentration values of protein B, at the time points where experimental data were provided. To assemble each PDF reflecting the variability of protein B at every time point, 10 000 Monte Carlo (MC) runs were performed after uniformly sampling the four-dimensional parametric space.

systems, for example, are provided by concentration-of-measure inequalities [57].

To reduce prediction errors, analyzing ensembles of mathematical models with variations in parameter values has proven powerful when applied to, for instance, protein folding. However, this type of ensemble modeling does not account for uncertainties of model structures. Kauffman was the first to introduce the idea of using constraints on network structures to define ensembles of structurally different models; Boolean model ensembles, for instance, may thereby enable the analysis of 'typical' qualitative dynamics in gene regulatory networks [58]. These ideas have been extended to investigate metabolic networks by local approximations of the kinetics in parameter space and subsequent sampling of Jacobians similar to the form used for monotone systems. This 'structural kinetic modeling' approach, for instance, uncovered dynamic features of the Calvin cycle such as oscillations and multistationarity [59]. Capitalizing on thermodynamics and on experimentally measured steady states to constrain model ensembles, another approach of approximate metabolic network modeling demonstrated the possibility of reducing the candidate ensemble size by iterations with targeted experiments [60]. Finally, a case study for the TOR (Target of Rapamycin) signaling pathway in budding yeast recently demonstrated that a comparative analysis of differently structured models may yield new, fundamental insights into signaling processes also, even with highly uncertain biological mechanisms and few quantitative experimental data. The iterative approach combined a systematic re-casting of (possibly conflicting) biological hypotheses into a model ensemble with highly targeted experimentation [61]. However, it contains heuristic elements – such as the selection of hypotheses to be modeled – that introduce biases and may prohibit scaling to larger systems.

Addressing the hard problem of structure identification, however, has a long tradition in machine learning, and concepts from this domain could be combined with dynamic modeling. Machine learning methods usually rely on large-scale datasets and employ discrete-time grey box models. Probabilistic models such as Bayesian network methods help elucidate causal couplings between the network components by formulating the identification of a network topology as a coherent estimation problem. However, the models do not enable direct mechanistic insight and quantitative predictions, and standard Bayesian models cannot cope with the ubiquitous feedback in cellular networks [62]. In addition, correctly identifying network topologies is not sufficient for predictive mathematical models because – depending on parameter values – the dynamic behavior may differ widely. New paradigms combining probabilistic and deterministic methods are therefore needed. One possible framework might be a coherent probabilistic selection of appropriate mathematical structures of dynamic models.

6. Conclusions and perspectives

Uncertainty about network components, interactions, and their quantitative features is found in all areas of current sys-

tems biology. Although increasing experimental data coverage will improve the situation, we argue that epistemic uncertainty will remain a challenge for mathematical methods that aim at increasing our systems-level understanding in biology. Consequently, it is a key challenge to devise theoretical methods that enable the development and analysis of mechanistic mathematical models and that can cope with uncertainties resulting from limited knowledge.

Although the approaches discussed in this review originate from very different scientific disciplines, they all emphasize the importance of structures of cellular networks. These structural features outweigh the potentially critical fine tuning of rate laws or parameters, which can be attributed to widespread robustness in biology [63]. Notably, the methods' predictive power, in principle, goes well beyond that of the well-established structural analysis of metabolic networks at steady state. Complementary theory developments extend the concepts to other types of biological networks as well as to dynamic network capabilities such as multistationarity.

Notably, many links exist between the approaches discussed in this review – for instance, regarding a modular analysis of complex networks. However, our current understanding of these connections is very limited and needs to be improved. Other challenges concern the scaling and generalization of the methods, such that ultimately biological networks of realistic complexity can be analyzed, for instance, using CRNT or the theory of monotone systems. The most generally applicable concept, ensemble modeling, also is the least mature; to a certain extent, this results from another 'cultural' gap in systems biology – here, the very different traditions in dynamical systems theory and probabilistic approaches - that eventually needs to be overcome. Structure-oriented analysis of biological systems, thus, does not only open broad perspectives for uncovering the organization and functionality of cellular networks, but it also implies challenging problems in the theory domain.

Acknowledgements

We gratefully acknowledge financial support by the EU FP7 Project UniCellSys and by the Swiss Initiative for Systems Biology (SystemsX.ch RTD Project YeastX).

References

[1] Kitano, H. (2002) Systems biology: A brief overview. Science 295 (5560), 1662–1664.
[2] Noble, D. (2008) Claude Bernard, the first systems biologist, and the future of physiology. Exp. Physiol. 93 (1), 16–26.
[3] Kitano, H. (2002) Computational systems biology. Nature 420, 206–210.
[4] Kim, H.D. et al. (2009) Transcriptional regulatory circuits: Predicting numbers from alphabets. Science 325 (5939), 429–432.
[5] Hoffman, F.O. and Hammonds, J.S. (1994) Propagation of uncertainty in risk assessments: The need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. Risk Anal. 14 (5), 707–712.
[6] Shahrezaei, V. and Swain, P.S. (2008) The stochastic nature of biochemical networks. Curr. Opin. Biotechnol. 19 (4), 369–374.
[7] Losick, R. and Desplan, C. (2008) Stochasticity and cell fate. Science 320 (5872), 65–68.
[8] Kaern, M. et al. (2005) Stochasticity in gene expression: From theories to phenotypes. Nat. Rev. Genet. 6 (6), 451–464.
[9] Claassen, M., Aebersold, R. and Buhmann, J.M. (2009) Proteome coverage prediction with infinite Markov models. Bioinformatics 25 (12), i154–i160.
[10] Pena-Castillo, L. and Hughes, T.R. (2007) Why are there still over 1000 uncharacterized yeast genes? Genetics 176 (1), 7–14.
[11] Hillenmeyer, M.E. et al. (2008) The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. Science 320 (5874), 362–365.
[12] Schaber, J., Liebermeister, W. and Klipp, E. (2009) Nested uncertainties in biochemical models. IET Syst. Biol. 3 (1), 1–9.
[13] Goldenfeld, N. and Kadanoff, L.P. (1999) Simple lessons from complexity. Science 284 (5411), 87–89.
[14] Horn, F. and Jackson, R. (1972) General mass action kinetics. Arch. Ration. Mech. Anal. 47 (2), 81–116.
[15] Savinell, J.M. and Palsson, B.O. (1992) Network analysis of intermediary metabolism using linear optimization. Development of mathematical formalism. J. Theor. Biol. 154, 421–454.
[16] Rocha, I., Forster, J. and Nielsen, J. (2008) Design and application of genome-scale reconstructed metabolic models. Methods Mol. Biol. 416, 409–431.
[17] Feist, A.M. et al. (2008) Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol. 7 (2), 129–143.
[18] Papin, J.A. et al. (2004) Comparison of network-based pathway analysis methods. Trends Biotechnol. 22, 400–405.
[19] Clarke, B.L. (1988) Stoichiometric network analysis. Cell Biophys. 12, 237–253.
[20] Klamt, S. and Stelling, J. (2006) Stoichiometric and constraint-based modeling in: System Modeling in Cellular Biology (Szallasi, Z., Stelling, J. and Perival, V., Eds.), pp. 73–96, MIT Press, Cambridge, MA.
[21] Schuster, S., Fell, D.A. and Dandekar, T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat. Biotechnol. 18 (3), 326–332.
[22] Stelling, J. et al. (2002) Metabolic network structure determines key aspects of functionality and regulation. Nature 420 (6912), 190–193.
[23] Terzer, M. and Stelling, J. (2008) Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics 24 (19), 2229–2235.
[24] Teusink, B. et al. (2009) Understanding the adaptive growth strategy of *Lactobacillus plantarum* by in silico optimisation. PLoS Comput. Biol. 5 (6), e1000410.
[25] Kaleta, C., de Figueiredo, L.F. and Schuster, S. (2009) Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. Genome Res. 19 (10), 1872–1883.
[26] Barrett, C.L. et al. (2005) The global transcriptional regulatory network for metabolism in *Escherichia coli* exhibits few dominant functional states. Proc. Natl. Acad. Sci. USA 102 (52), 19103–19108.
[27] Gianchandani, E.P. et al. (2009) Functional states of the genome-scale *Escherichia coli* transcriptional regulatory system. PLoS Comput. Biol. 5 (6), e1000403.
[28] Segre, D., Vitkup, D. and Church, G.M. (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 99 (23), 15112–15117.
[29] Shlomi, T., Berkman, O. and Ruppin, E. (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. USA 102 (21), 7695–7700.
[30] Uhr, M. and Stelling, J. (2008) Structural sensitivity analysis of metabolic networksProc. 17th World Congress, The International Federation of Automatic Control, Seoul, Korea, July 6–11, IFAC.
[31] Horn, F. (1972) Necessary and sufficient conditions for complex balancing in chemical kinetics. Arch. Ration. Mech. Anal. 49 (3), 172–186.
[32] Feinberg, M. (1995) Multiple steady states for chemical reaction networks of deficiency one. Arch. Ration. Mech. Anal. 132, 371–406.
[33] Feinberg, M. (1995) The existence and uniqueness of steady states for a class of chemical reaction networks. Arch. Ration. Mech. Anal. 132, 311–370.
[34] Craciun, G., Tang, Y. and Feinberg, M. (2006) Understanding bistability in complex enzyme-driven reaction networks. Proc. Natl. Acad. Sci. USA 103 (23), 8697–8702.
[35] Sontag, E. (2001) Structure and stability of certain chemical networks and applications to the kinetic proofreading model of T-cell receptor signal transduction. IEEE Trans. Autom. Control 46, 1028–1047.
[36] Shinar, G., Alon, U. and Feinberg, M. (2008) Sensitivity and robustness in chemical reaction networks. SIAM J. Appl. Math. 69 (4), 977–998.
[37] Otero-Muras, I., Banga, J.R. and Alonso, A.A. (2009) Exploring multiplicity conditions in enzymatic reaction networks. Biotechnol. Prog. 25 (3), 619–631.
[38] Savageau, M.A. et al. (2009) Phenotypes and tolerances in the design space of biochemical systems. Proc. Natl. Acad. Sci. USA 106 (16), 6435–6440.
[39] Uhr, M., et al. (2009) Analysis of degenerate chemical reaction networks. In: Proc. Positive Systems: Theory and Applications.
[40] Conradi, C. et al. (2007) Subnetwork analysis reveals dynamic features of complex (bio)chemical networks. Proc. Natl. Acad. Sci. USA 104 (49), 19175–19180.
[41] Conradi, C., et al. (2005) Using chemical reaction network theory to discard a kinetic mechanism hypothesis. In: Proc. 2005 Foundations of Systems Biology in Engineering Conference.
[42] Angeli, D. and Sontag, E.D. (2003) Monotone control systems. IEEE Trans. Automat. Control 48 (10), 1684–1698.
[43] Angeli, D. and Sontag, E.D. (2004) Multi-stability in monotone input/output systems. Syst. Control Lett. 51 (3–4), 185–202.
[44] Sontag, E. (2007) Monotone and near-monotone biochemical networks. Syst. Synth. Biol 1 (2), 59–87.
[45] Smith, H. (1995) Monotone dynamical systems: An introduction to the theory of competitive and cooperative systems, American Mathematical Society (AMS).
[46] Moles, C.G., Mendes, P. and Banga, J.R. (2003) Parameter estimation in biochemical pathways: A comparison of global optimization methods. Genome Res. 13, 2467–2474.
[47] Jaqaman, K. and Danuser, G. (2006) Linking data to models: Data regression. Nat. Rev. Mol. Cell Biol. 7 (11), 813–819.
[48] W.H. Press, S.A.T., W.T. Vetterling, B.P. Flannery, Numerical Recipes 3rd Edition: The Art of Scientific Computing. 3rd ed. 2007: Cambridge University Press.
[49] Liu, J.S. (2008). Monte Carlo Strategies in Scientific Computing, second edSpringer Series in Statistics, xvi, Springer, New York. p. 343.

[50] Lemieux, C. (2009) Monte Carlo and Quasi-Monte Carlo Sampling. Springer Series in Statistics, first ed, Springer.
[51] Joshi, M., Seidel-Morgenstern, A. and Kremling, A. (2006) Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. Metab. Eng. 8 (5), 447–455.
[52] Turkheimer, F.E., Hinz, R. and Cunningham, V.J. (2003) On the undecidability among kinetic models: From model selection to model averaging. J. Cereb. Blood Flow Metab. 23 (4), 490–498.
[53] Wahl, S.A. et al. (2006) Unravelling the regulatory structure of biochemical networks using stimulus response experiments and large-scale model selection. Syst. Biol. (Stevenage) 153 (4), 275–285.
[54] Marino, S. et al. (2008) A methodology for performing global uncertainty and sensitivity analysis in systems biology. J. Theor. Biol. 254 (1), 178–196.
[55] Lorenz, S. et al. (2007) Discrimination of dynamical system models for biological and chemical processes. J. Comput. Chem. 28 (8), 1384–1399.
[56] Johnson, J.B. and Omland, K.S. (2004) Model selection in ecology and evolution. Trends Ecol. Evol. (Amst.) 19 (2), 101–108.
[57] Lucas, L.J., Owhadi, H. and Ortiz, M. (2008) Rigorous verification, validation, uncertainty quantification and certification through concentration-of-measure inequalities. Comput. Methods Appl. Mech. Eng. 197 (51–52), 4591–4609.
[58] Kauffman, S. (2004) A proposal for using the ensemble approach to understand genetic regulatory networks. J. Theor. Biol. 230 (4), 581–590.
[59] Steuer, R. et al. (2006) Structural kinetic modeling of metabolic networks. Proc. Natl. Acad. Sci. USA 103 (32), 11868–11873.
[60] Tran, L.M., Rizk, M.L. and Liao, J.C. (2008) Ensemble modeling of metabolic networks. Biophys. J. 95 (12), 5606–5617.
[61] Küpfer, L. et al. (2007) Ensemble modeling for analysis of cell signaling dynamics. Nat. Biotechnol. 25, 1001–1006.
[62] Markowetz, F. and Spang, R. (2007) Inferring cellular networks – A review. BMC Bioinform. 8 (6), S5.
[63] Stelling, J. et al. (2004) Robustness of cellular functions. Cell 118 (6), 675–685.