Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments

Natal A.W. van Riel

Submitted: 28th August 2006; Received (in revised form): 4th October 2006

Abstract
System biology applies quantitative, mechanistic modelling to study genetic networks, signal transduction pathways and metabolic networks. Mathematical models of biochemical networks can look very different. An important reason is that the purpose and application of a model are essential for the selection of the best mathematical framework. Fundamental aspects of selecting an appropriate modelling framework and a strategy for model building are discussed.

Concepts and methods from system and control theory provide a sound basis for the further development of improved and dedicated computational tools for systems biology. Identification of the network components and rate constants that are most critical to the output behaviour of the system is one of the major problems raised in systems biology. Current approaches and methods of parameter sensitivity analysis and parameter estimation are reviewed. It is shown how these methods can be applied in the design of model-based experiments which iteratively yield models that are decreasingly wrong and increasingly gain predictive power.

Keywords: systems biology; parameter sensitivity analysis; parameter estimation; optimal experiment design; differential equations

INTRODUCTION
Progress in molecular and cell biology has led to the identification of complex biochemical networks involved in the normal functioning of cells, tissues and organs and also defects associated with many diseases. However, most of this information is obtained in cell cultures and animal models. The evaluation of their importance in the human disease context is far from obvious [1]. The advances in experimental techniques provide an opportunity for developing mechanistic mathematical models of biochemical networks, including signal transduction pathways and metabolic networks. Building sound dynamic models of biochemical networks is a key step towards the development of predictive models for cells or whole organisms. Such models can be regarded as the keystones of systems biology [2], ultimately providing scientific explanations of the behaviour of biological systems in health and disease and assisting in the identification of therapeutic targets with the promise of molecular medicine and, on the long run, personalised medicine.

System level understanding of complex processes is common in engineering disciplines and relies heavily on mathematical models. Physiologists have traditionally used a holistic perspective to try to understand how life works by collecting quantifiable information on biological stimuli and responses. However, a systems approach is a rather new, and not generally accepted, way of thinking in molecular...
and cell biology. As currently employed, the term ‘systems biology’ encompasses many different approaches and models, and studies many organisms from bacteria to man [3].

Despite the amount and quality of experimental data increase rapidly, quantitative measurements of many cellular components in time and space are still relatively scarce. Data sparsity in combination with biological complexity impose severe challenges to standard engineering methodologies for modelling, simulation and mathematical analysis of biochemical networks and there is a great need of sound model building methods. Predictive, mechanism-based mathematical models of biochemical networks require quantitative information on reaction rates and molecular concentrations. For most processes, these parameters are not directly accessible in vivo. Existing biochemical data usually originate from different experimental settings, cell types and states of cells and can therefore not be automatically trusted and used for quantitative models. Furthermore, cellular processes are described on different levels of information quality, ranging from mechanistically well-understood interactions to purely hypothetical interactions described in qualitative terms like activation or inhibition. In cases where the mechanistic details are unclear, it is necessary to fill in the gaps by postulating simple mechanisms without having any kinetic parameters available [4]. If the dynamic behaviour of a system is highly dependent on the value of (some of) the parameters, then accurate and reliable quantification of the parameters is essential for the development of predictive models and therefore parameter estimation is an important research topic in systems biology [5–8].

**BOTTOM-UP VERSUS TOP-DOWN**

Approaches to mathematical and computational modelling in biology can be thought of as falling into two philosophical categories [9–11]. The first includes large-scale modelling approaches for which it is assumed that important biological features emerge from their simulation and analysis (inductive science). Complementary to this approach is hypothesis-driven research, traditional in among other physics, which aims for the simplest possible model that captures the key features of a system consistent with the level of available experimental data. This fundamental principle of generating parsimonious hypotheses and models is referred to as ‘Occam’s razor’ [12], which, as Einstein noted, also implies that models should not be too simple.

For successful application of systems biology these philosophies are complementary rather than mutually exclusive. The two paradigms can also be recognised in the two principal approaches to the construction of biochemical network models [13, 14]. In one of his founding papers on systems biology, Kitano referred to a ‘bottom-up’ approach as ‘trying to construct a gene regulatory network from independent experimental data, mostly obtained through a literature search and the rest obtained from experiments designed to provide information about specific aspects of the network of the interest’. A ‘top-down’ approach ‘tries to make use of data gathered using a high-throughput DNA micro-array and other measurement technologies to infer genetic network structures’ [15], which is also referred to as ‘reverse engineering’ [16–19]. A substantial number of methods have been developed to try to solve these complex inverse problems [20–24]. However, the focus of this review is on mechanism-based models of biochemical networks.

It should be noted that in several papers on systems biology the ‘top-down’ and ‘bottom-up’ terms have been used with a different meaning. In drug discovery ‘bottom-up’ refers to going from molecular biology techniques applied to in vitro systems up to clinical trials [1, 3]. This interpretation of bottom-up and top-down is more closely related to their meaning in computational physiology, such as applied in the Physiome project, reflecting the multi-scale hierarchy going from organ systems (top) to the genome (bottom, Figure 1) [11, 25].

Whether an inductive or a parsimonious, hypothesis-driven approach is most suited, and which mathematical framework to use is essentially dependent on the purpose of the model. Therefore, any modelling effort should start with defining this purpose. In brief, the following reasons could warrant the use of mathematical models [10, 26]: (1) to organise disparate information into a coherent, self-consistent whole, (2) to think (and calculate) logically about what components and interactions are important in a complex system, (3) to simulate, predict, and optimise procedures, experiments and therapies, (4) to disprove hypotheses and to define improved hypotheses and (5) to understand the essential features of a system. Disproof of hypotheses that are formalised as mathematical models is based
on the comparison of model-predicted and experimentally measured variables. Thus working toward goal 4 requires developing models and designing experiments in tandem, ensuring that sets of modelled and measured variables can be matched to each other and that experiments are optimally designed to identify model unknowns [10, 11, 27].

Over the past years, many papers have reviewed approaches to the development of computational models of biochemical networks and described the insights gained from models [4, 28–37]. Some cover an extensive overview of many different mathematical frameworks, including Boolean and Bayesian networks and stochastic models [38], others are restricted to a specific class of models, such as so-called hybrid systems, which combine continuous differential equations and discrete events [39]. Like most reviews, we focus on continuous, deterministic descriptions of the dynamics by applying non-linear ordinary differential equation (ODE) models of biochemical pathways. The so-called state-space formulation is most convenient to describe and implement an input–output system as a set of first-order differential equations for a vector of state variables and an algebraic output equation.

Several review, opinion and tutorial papers use rather simple and abstracted examples of signalling pathways to illustrate how mathematical modelling and analysis can help in understanding concepts like emergence, feedforward and feedback loops, stability, robustness (and fragility), adaptation and homeostasis. This might leave the impression that systems biology per se aims for simplification and functional descriptions, while losing sight of the details and complexity of actual reaction networks. Many efforts pursue realistic large-scale complex models, but simplified models show to be valuable in terms of understanding the essential and/or qualitative behaviour of biological systems [10, 34, 40]. A focus on system behaviour of the network does not imply that known molecular details are always to be neglected [41]. Nevertheless, a course-grained level of description is often the only practical approach if molecular details of parts of the biochemical network lack [4, 42].

MODEL BUILDING CYCLE
Arguably, systems theory provides the most formal and most general (not application domain specific) definition of the cycle of building models of...
dynamic systems. The typical model building cycle as considered in the area of system identification [5, 43–45] starts from a goal definition (purpose of the model) and some a priori knowledge (i.e. preliminary data and initial hypotheses) on the basis of which a model framework is chosen and a model structure is proposed. From the available data, parameter estimation is then performed, leading to a first working model. This initial model must be validated with new experiments, which, in most cases, will reveal a number of deficiencies. Thus, a new model structure and/or a new experimental design must be planned and the process is repeated iteratively until the validation step is considered satisfactory.

An underlying assumption of this cycle is that the system inputs can be easily and freely perturbed by the experimenter to reveal a rich ensemble of dynamic responses and that the responses of many of the system variables can be measured accurately. Furthermore, it is assumed that this cycle can be iterated several times without significant extra ‘costs’. These assumptions do not hold for biological systems. For systems biology a somewhat different, more specific workflow is emerging as a practical approach towards mechanism-based models of biochemical networks. After the purpose of the model has been defined, the next step is to define the connection map of the network. For all the components and their interactions parameter information needs to be collected. This involves (basal) concentrations for each component and kinetic rate constants for interactions and enzymatic reactions. Some of this information is available as tabulated numerical values in the literature. However, in many cases, the (in vitro) quantitative information has to be manually extracted from the primary literature. In cases where the mechanistic details are unclear, it is necessary to fill in the gaps with simple mechanisms [1, 4, 42]. Before starting any biological experiment, it is most efficient to adapt the model until it is consistent with most of the heterologous datasets (i.e. using data from different tissues or species). Based on model simulations that are discrepant for (some of) the data, the model can already be iteratively improved without performing own ‘wet’ experiments. Preferably, the literature also provides quantitative physiological ‘endpoints’ that have been measured in an intact system (in vivo, in situ, ex vivo) and can be compared to the corresponding endpoints as predicted by the model.

Despite such endpoints will often be static, this helps to verify the consistency of the model and is important to gain confidence in the predictions. Next, system analysis tools, such as parameter sensitivity analysis, can be used to design experiments to generate optimally informative data for parameter estimation or to validate model predictions. Optimal experimental design (OED) includes how to perturb the system, which components to measure and at which time points to sample (model-based experiments). To develop predictive models of realistically large biochemical networks dozens of components (proteins, metabolites) will have to be measured in time. To use parameter estimation techniques, quantitative and repeatedly measured data on cell responses to physiological stimuli and to pharmaceutical agents are needed [3]. Compared to the high-throughput experiments in functional genomics these are generally medium or low throughput assays. To validate the model, it is used to formulate hypotheses about the system behaviour under different conditions than used to estimate the parameters. Subsequently, these experiments are performed and the measured responses are compared quantitatively with the model predictions. The iterative cycle of prediction and experimental validation progressively strengthens the predictive power of the model. Sensitivity analysis, parameter estimation and optimal experiment design will be discussed in the remainder of this review.

PARAMETER SENSITIVITY ANALYSIS

Mechanism-based models allow to predict the behaviour of the specified system over time and to track its dynamics for each set of fixed system parameters. However, all of the parameters, including rate constants and initial component concentrations, must be experimentally measured or inferred to specify the model. Even for those models with experimentally determined parameters, it is still uncertain whether the particular set of parameters closely approximates the corresponding biological system, because (some of) the kinetic parameters are usually taken or estimated from measurements reported by different laboratories using different in vitro models and conditions. It is therefore important not only to study the dynamical properties governed by the particular kinetic parameters, but also to further investigate the effects of their
perturbations on the overall system [46]. The identification of the network components and rate constants that are most critical to the output behaviour of the system is one of the major problems raised in systems biology [47]. Besides analysis of the effects of parameter uncertainty on model outcome, other applications of parameter sensitivity analysis are: (1) analysing robustness [48, 49], (2) determining modularity of biochemical networks and model reduction [42], (3) identification of control points for, among others, drug target selection [1], (4) discrimination of phenotypes for biomarker selection [50] and (5) experiment design [5, 42, 45, 51, 52]. Robustness allows a system to maintain its functions despite external and internal perturbations [53] and is one of the system properties of biochemical networks that has been widely addressed using computational approaches [49]. With parameter sensitivity analysis it can be investigated if a system is robust in the sense that it is capable to operate reliably when its physical and biochemical parameters vary within their expected ranges, or that its outcome is insensitive to the precise value of the parameters when these are not (accurately) known. Conversely, parameters and components that are very sensitive to parametric variations can introduce fragility into the system. Sensitivity analysis assists in identifying the critical steps in the network that could drive disease development and these control points provide potential therapeutic targets. If, by changing parameters (including basal component concentrations), the model can be made to reproduce different characteristic disease phenotypes and this outcome is sensitive and specific for a certain parameter value, then this parameter might be a novel biomarker.

Local parameter sensitivity analysis (LPSA) has been used as an in silico investigation method to identify critical parameters in dynamic models of signal transduction [8, 54–56] and metabolic pathways [57]. The sensitivity coefficient is defined as:

$$S^M_\theta = \frac{\delta M(y)/M(y)}{\delta \theta/\theta}$$  \hspace{1cm} (1)

where $\theta$ represents the parameter that is varied and $M(y)$ a characteristic of the response of the system output $y$. $\delta M$ is the change in $M$ due to the change $\delta \theta$ in $\theta$. This relative sensitivity coefficient is similar to the control coefficients of metabolic control analysis [58]. Typical response characteristics that have been applied are: sum of squared differences between the output for the reference parameter values and the perturbed system output [45, 46, 59], the area under the curve of the output [8], the amplitude and period of oscillation [56, 57], the deviation from the steady-state values [51] and the output value after a specific amount of time, which could be an endpoint that is clinically used to assess drug efficacy [1, 50].

Local sensitivity analysis allows only one parameter to vary for each simulation and deals with only small perturbations of the reference model, i.e. local sensitivity analysis pertains to a particular point in the parameter space. However, rate constants and concentrations of diverse molecules may vary extensively in an interactive manner among different cellular environments. For these reasons it is more appropriate to explore, in a probabilistic context, possibilities of non-linear effects from simultaneous parameter variations of arbitrary magnitudes by means of a global sensitivity analysis [46]. Only recently global parameter sensitivity analysis (GPSA) methods have been started to be applied to biochemical network models [42, 45, 46, 59]. GPSA has the advantage that it provides a more comprehensive analysis. However, varying all possible combinations of parameters for a wide range of possible values per parameters rapidly becomes computationally infeasible for realistically large biochemical networks. Therefore, all GPSA methods apply a Monte Carlo strategy using sampling to generate random sets of parameter values for simulations [59]. A range of the parameter distributions can often be obtained from the available literature or guided by the experiences of the researchers. Latin hypercube sampling [60] is an efficient method to sample random parameter vectors while guaranteeing that individual parameter ranges are evenly covered [46]. Zhang and Rundell [59] embedded GPSA in a random-search parameter identification routine. The sensitivity index is calculated for each parameter set generated by the Genetic Algorithm during parameter optimisation. The fit of the output of the system with the perturbed parameters to the output for the reference parameter values has commonly been used as sensitivity criterion. A sum of squared errors between the perturbed and reference output is used as objective function to evaluate the fit [42, 45, 46, 59].

Bentele et al. [42] calculated global sensitivities from a weighted average of local sensitivities.
The methods of Cho et al. [45], Zi et al. [46] and Zhang and Rundell [59] classify the sampled parameter sets as either acceptable or unacceptable given a threshold value of the objective function. The distributions of the parameter values associated with acceptable and unacceptable cases are statistically evaluated to rank the sensitivities.

LPSA and GPSA have been compared by applying the methods to the same model. The resulting rankings of the parameters were significantly different [46, 59]. Zi et al. [46] applied LPSA and their GPSA method to a model of the IFN-γ induced JAK-STAT signalling pathway, a stress response pathway related to the immune system, containing 32 state variables and 51 parameters. The response of phosphorylated STAT1 dimers in the nucleus (output) to a step function in IFN-γ (input) was used as the characteristic response for MPSA. Besides LPSA, Zhang and Rundell [59] also compared five GPSA methods: the method of Bentele et al. [42], their own novel method, the partial rank correlation coefficient (PRCC) method [61], Sobol’s method [62] and the Fourier amplitude sensitivity test (FAST) [63]. The latter three have been used in engineering applications [64]. The methods were applied to a model of a T-cell receptor-activated Erk-MAPK signalling pathway (24 state variables and 69 parameters of which 49 were included in the sensitivity analysis). The sensitivities as predicted by the ‘traditional’ GPSA methods (PRCC, Sobol’s method and FAST) were highly similar. Like LPSA, the method of weighted average of local sensitivities [42] produced a very different sensitivity pattern. The results of the Zhang and Rundell method were in between the local approaches and the existing GPSA methods. The method of Zi et al. [46] was not included in this comparison.

PARAMETER ESTIMATION FROM PARTIAL MEASUREMENTS

The number of assessable parameters and therefore the maximum size of biochemical network models have been very limited due to the large amount of experimental data required for high-dimensional parameter estimation problems. The space of possible sets of parameter values grows exponentially with the number of unknown parameters, which severely impairs the search for the globally optimal parameter values (the ‘curse of dimensionality’ [42]). In a few studies, sufficient components of a biochemical network were measured in time such that the parameters of a simplified model could be uniquely and accurately estimated given the data of inputs and outputs (the model was shown to be ‘identifiable’) [6, 8]. Swameye et al. [8] estimated five unknown parameters in a model with four state variables. The model of van Riel and Sontag contained five state variables and seven unknown model parameters [6]. The parameters of both models were estimated by applying the maximum likelihood estimator (MLE) and the accuracy of the estimates was assessed by calculating the variances from the Fisher Information Matrix (FIM). MLE has been widely used for parameter estimation problems [5–7, 43, 52, 65, 66]. It is assumed that the measured data have been obtained from a stochastic process, because of the presence of unmodelled dynamics and measurement noise. The measurement error is assumed to be additive with a Gaussian distribution with zero mean and known variance. This implicates that the estimated parameters are also stochastic quantities with a Gaussian distribution. It can be shown that an unbiased parameter estimate (the expected mean value) can be obtained by minimising a weighted sum of squared model residuals, which is similar to the cost function used in MPSA. The covariance matrix of parameter estimates has the inverse of the FIM as lower bound (the so-called Cramér–Rao bound [44]). The FIM quantifies the accuracy of the estimates. If the weighting matrix in the cost function is selected as the inverse of the data covariance matrix, then the parameter covariance matrix is equal to (FIM)\(^{-1}\). The parameter estimates have a minimum variance and the diagonal elements of (FIM)\(^{-1}\) are approximations of this variance.

The non-linear least squares criterion in combination with gradient-based optimisation methods is available in many software tools and computational languages [67]. One difficulty with parameter estimation in non-linear differential–algebraic equation models is the existence of multiple local minima in the cost function. The traditional gradient-based optimisation methods often fail to arrive at the global optimal solutions. Banga and coworkers [68] have compared several advanced deterministic and stochastic global optimisation methods for parameter identification. The random search methods include methods such as Genetic Algorithms and Simulated Annealing. The stochastic strategies appeared to be the most efficient and reliable to locate the
parameter region containing the global solution, but
global optimality is not guaranteed. Furthermore, all
global optimisation methods suffer from the large
computational burden required, even for moderately
sized problems. (As case study, 36 parameters of a
non-linear biochemical model were estimated using
a PC Pentium-III, which took 40 h.) In a recent
contribution, Banga and co-workers [5] presented
a new optimisation methodology reducing compu-
tation time by one order of magnitude by means
of a hybrid method combining global and local
optimisation methods which increases efficiency
while guaranteeing robustness.

Recently, three different approaches to overcome
the curses of data scarcity and high dimensionality of
the parameter optimisation problem in large-scale,
mechanistic models of biochemical networks were
reported: (1) infer the state trajectories of compo-
nents that have not been measured and subsequently
use these profiles as additional data [52, 69],
(2) include well-known biological features (semi-
quantitative information) of the behaviour of
the system in the parameter estimation algorithm [6, 70,
71] and (3) make use of the modular topology of
many biochemical networks [6, 42].

State regulator approach
Generally, it is not possible to measure all time-
varying components in a genetic, signalling or
metabolic network. Doyle III and coworkers
[52, 69] have decoupled the problem of parameter
estimation from partial measurements into two parts.
First, the available measurements are used to estimate
the profiles of all unmeasured concentrations and
fluxes. In the second part the reconstructed state
trajectories and fluxes are used as additional ‘data’
to estimate the model parameters. In system and
control theory a well-known solution to estimate the
behaviour of unmeasured components given partial
measurements of other system variables is the
Kalman filter. However, a Kalman filter requires an
accurate model of the system and the development
of such models is exactly the aim of systems biology.
Doyle III and coworkers used an extension of
dynamic flux balance analysis (dFBA, [72, 73])
to estimate unmeasured protein and reaction rate
trajectories. The premise of dFBA is that cellular
processes have evolved regulatory structures that
optimally use cellular resources. This translates
into the postulate that the network flows are
managed to minimise internal accumulation
of components while minimising the use of reaction
fluxes. This requirement is analogous to a classic
problem in automatic control, namely, the state
regulator problem (SRP), which penalises deviations
of the state variables as well as large control action.
The SRP estimator calculates an optimal set of state
trajectories via a constrained convex programming
problem in which the available measurements are the
constraints. It is important to note that the reaction
kinetics and therefore the unknown parameters are
not used in the SRP algorithm.

Feature-based constraints
In contrast to the scarce quantitative time-series data,
biochemistry is gifted with enormous amounts of
qualitative observations and semi-quantitative infor-
mation. Recently, several publications have tried to
make use of qualitative features of network beha-
vour for the automatic quantification of model
parameters [6, 70, 71]. It is not trivial to translate
qualitative features into a mathematical function that
can be included in the numerical optimisation
algorithm for parameter estimation. The most
applied approach is to heuristically construct an
objective function by summing penalty functions for
each of the feature constraints [70, 71]. The relative
importance of the different terms is rather arbitrary
and is determined by introducing a weighting factor
for each term, yielding a weighted optimisation
criterion. Amonlirdviman et al. [71] used the patterns
of the hairs on the Drosophila wings as qualitative
features. Locke et al. [70] combined qualitative
features of mRNA profiles extracted from different
publications on the circadian clock circuit in the
plant Arabidopsis thaliana.

Van Riel and coworkers [6, 74] developed an
approach which combines data fitting and feature
constraints in a multi-objective estimation problem.
The data fit objective function was augmented with
a penalty function prohibiting parameter values that
would violate the a priori information. The a priori
information consisted of the steady-state fluxes,
calculated by flux analysis, in a metabolic network
of nitrogen metabolism in the yeast Saccharomyces
cerevisiae. These flux values were subsequently used
as constraints when a dynamic model of the same
network was fitted to time-series data of extracellular
and intracellular metabolites after perturbing
the steady-state cell culture by injecting a nitrogen
substrate [74].
Modular topology

As indicated before, parameter sensitivity analysis can be used to analyse the modularity of biochemical networks. If the network has a modular structure the distributions of the sensitivities calculated with GPSA show distinct and narrow peaks indicating that the sensitivities of the system are highly robust to large variations in most of the parameter values [42]. Because of this modular structure of biochemical networks the inverse problem to estimate model parameters from experimental data can be split in a hierarchically nested set of parameter estimation problems [42]. Clusters can be identified that contain a subset of molecules whose concentrations depend on a subset of parameters only, indicating that these parameters can be estimated locally. Global parameters, those that belong to more than one cluster, are estimated by optimising all clusters. Bentele et al. [42] could reduce the dimensionality of the optimisation problem from 58 unknown parameters to 18 in a model of CD95-induced apoptosis.

Van Riel and Sontag [6] developed a different approach to utilise the modular structure of biochemical networks to reduce the complexity of parameter estimation [6]. If the profiles of some of the proteins or metabolites can be (accurately) measured in time, these signals can be used as inputs during simulation. These measured inputs emerge from the total system and cannot be independently defined in an experiment; therefore Van Riel and Sontag have called these as ‘dependent inputs’. The concept of dependent inputs allows the ‘opening’ of some of the connections of the pathways of interest with the rest of the cell and its environment (Figure 2). The model can be restricted to a smaller part of the network and the unmodelled dynamics are replaced by the measured dependent inputs. This approach of closed-loop system identification has been previously applied in pharmacokinetic compartment modelling to estimate subject specific physiological parameters (e.g. quantification of insulin sensitivity [75]). In systems biology the approach is particularly useful in validation experiments, because model parameters can be fitted to experimental data generated by a reference cell type (‘wild-type’) and then testing this model on data generated by a variation (‘mutant’).

OPTIMAL EXPERIMENTAL DESIGN

For a systems biology approach and especially parameter estimation, the dynamics of cellular interaction networks need to be measured by taking comprehensive, high-resolution quantitative measurements of the intracellular status, such as the concentrations, interactions, modifications and localisations of molecules, and of cellular structures in time and space under various conditions [76]. The experiments to generate such data are complex and expensive, as a consequence of which the time-series available are usually rather short, with few if any replicates. Almost certainly, not all variables one would like to include in a model can be measured. Furthermore, the robustness of many biomolecular systems implies that arbitrary perturbations may result in a minimal response in many of its molecular components. Thus, if the experiments are not appropriately designed, it is very likely that little information of value can be recovered from the experimental measurements for identifying the model parameters. OED is therefore an important research problem in systems biology [7, 51, 77].

Identification of the most sensitive parameters can guide the design of model-based experiments to provide data that are required to estimate those parameters [42, 45]. However, system identification theory offers a more rigorous approach to OED. According to the theory of MLE, the FIM is a measure of the information content of the data and its inverse provides an estimate of the accuracy of the parameter estimates. Therefore, OED can be formulated as an optimisation problem, in which a scalar function of the FIM is maximised to provide ‘rich’ experimental data to increase the accuracy of the parameter estimates [5, 52, 77]. The shape of the input perturbation is one of the important aspects of an optimal design. In general, complex perturbations generate more informative data than simple ones [78]. OED can also be used to optimise the sampling times [7]. Restrictions with respect to the number

![Figure 2](http://bib.oxfordjournals.org/)

**Figure 2:** The concept of the ‘dependent inputs’ makes use of the modular structure of biochemical networks and can be used to reduce the complexity of the model. Input $u$ can be experimentally manipulated and output $y$ is measured. Signal $v$ results from the interaction of module A with module B. If $v$ is also measured it can be treated as a dependent input to simulate a model of A, without requiring a model of B.
of samples that can be taken or the type of input perturbations that can be applied can be taken into account.

CONCLUSION

Systems biology is arguably one of the most influential ideas in life science research today, promising to yield a more global and in-depth understanding of biological systems, furnishing insights into the root causes of complex diseases, contributing to the drug discovery process and facilitating a variety of aspects of health care including treatment evaluation and predictive medicine [79].

Precise estimates of parameters are not always needed to understand certain qualitative features of the dynamic behaviour of biochemical networks. Parameter sensitivity analysis is an important tool to analyse which parameters do need to be accurately known to develop predictive models.

Key Points

- Inductive research, common to omics approaches, and holistic and hypothesis-driven research, typical of physiology and physics, should be applied in parallel for successful systems biology.
- Mechanism-based dynamic models provide a platform to integrate heterologous data into a self-consistent description that can be used for hypothesis-driven research and the design of model-based experiments.
- System and control theory provides many valuable concepts and methods for systems biology. Concepts from control theory are successful in explaining how biochemical networks can be very stable in the face of unwanted perturbations, while simultaneously responding in a highly sensitive and specific manner to desired stimuli. System identification theory provides methods for parameter sensitivity analysis, parameter estimation and the design of optimally informative experiments, which iteratively yield models that are decreasingly wrong and increasingly gain predictive power. The specific requirements of systems biology also stimulate new developments in system theory [80].
- Local methods of parameter sensitivity analysis and parameter estimation are well-established and have been successfully applied. More advanced methods for parameter sensitivity analysis based on Monte Carlo simulations and global optimisation methods based on random search algorithms for parameter estimation problems and optimal experimental design receive increasing attention, but these areas need further development. Currently the methods are tailored towards specific problems to reduce the computational burden.

Acknowledgements

The author wishes to thank Mark Musters for his helpful comments and Ad Damen and Paul van den Bosch for useful discussions.

References


