Metabolic Engineering 24 (2014) 38–60

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/ymben

Minireview

Kinetic models in industrial biotechnology – Improving cell factory performance

Joachim Almquist [a,b,*], Marija Cvijovic [c,d], Vassily Hatzimanikatis [e], Jens Nielsen [b], Mats Jirstrand [a]

[a] Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Göteborg, Sweden
[b] Systems and Synthetic Biology, Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-412 96 Göteborg, Sweden
[c] Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Göteborg, Sweden
[d] Mathematical Sciences, University of Gothenburg, SE-412 96 Göteborg, Sweden
[e] Laboratory of Computational Systems Biotechnology, Ecole Polytechnique Federale de Lausanne, CH 1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 11 September 2013
Received in revised form
7 March 2014
Accepted 9 March 2014
Available online 16 April 2014

Keywords:
Mathematical modeling
Kinetic models
Dynamic models
Parameter estimation
Cell factory
Industrial biotechnology

ABSTRACT

An increasing number of industrial bioprocesses capitalize on living cells by using them as cell factories that convert sugars into chemicals. These processes range from the production of bulk chemicals in yeasts and bacteria to the synthesis of therapeutic proteins in mammalian cell lines. One of the tools in the continuous search for improved performance of such production systems is the development and application of mathematical models. To be of value for industrial biotechnology, mathematical models should be able to assist in the rational design of cell factory properties or in the production processes in which they are utilized. Kinetic models are particularly suitable towards this end because they are capable of representing the complex biochemistry of cells in a more complete way compared to most other types of models. They can, at least in principle, be used to in detail understand, predict, and evaluate the effects of adding, removing, or modifying molecular components of a cell factory and for supporting the design of the bioreactor or fermentation process. However, several challenges still remain before kinetic modeling will reach the degree of maturity required for routine application in industry. Here we review the current status of kinetic cell factory modeling. Emphasis is on modeling methodology concepts, including model network structure, kinetic rate expressions, parameter estimation, optimization methods, identifiability analysis, model reduction, and model validation, but several applications of kinetic models for the improvement of cell factories are also discussed.

1. Introduction

Throughout the World there is a desire to move towards sustainable production of energy, fuels, materials and chemicals, and biobased production of transportation fuels and chemicals is expected to contribute significantly towards reaching this objective. This has resulted in the advancement of industrial biotechnology, where microbial fermentation is used for the conversion of bio-based feedstocks to fuels and chemicals (Nielsen and Jewett, 2008; Tang and Zhao, 2009; Otero and Nielsen, 2010; Du et al., 2011; Sauer and Mattanovich, 2012). Not only has this resulted in a significant expansion of traditional processes such as bioethanol production, which has increased from 10 billion liters produced in 2010 to 75 billion liters produced in 2012, but it has also resulted in the introduction of novel processes for the production of chemicals that can be used for the production of polymers, e.g. lactic acid that goes into poly-lactate and 1,3 propanediol that goes into Sorona®. With these successes the chemical industry is looking into the development of other processes for the production of platform chemicals that can find application in the manufacturing of solvents and polymers. Traditionally the fermentation industry used naturally producing microorganisms, but today there is a focus on using a few microorganisms, often referred to as platform cell factories, and then engineering their metabolism such that they efficiently can produce the chemical of interest. This engineering process is referred to as metabolic engineering, and it involves the introduction of directed genetic modifications. Due to the complexity of microbial metabolism, both due to the large number of interacting reactions and the complex regulation, there has been an increasing focus on the use of mathematical models for the identification of metabolic

* Corresponding author at: Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Göteborg, Sweden.
E-mail address: joachim.almquist@fcc.chalmers.se (J. Almquist).

http://dx.doi.org/10.1016/j.ymben.2014.03.007
1096-7176/

engineering targets (Patil et al., 2004; Cvijovic et al., 2011; Wiechert and Noack, 2011; Soh et al., 2012).

Industrial biotechnology can benefit from mathematical models by using them to understand, predict, and optimize the properties and behavior of cell factories (Tyo et al., 2010). With valid models, improvement strategies can be discovered and evaluated in silico, saving both time and resources. Popular application of models thus includes using them to suggest targets for metabolic engineering leading to increases in yield, titer, and productivity of a desired product. Since these quantities not only depend on the genetic constitution of cells but to a large extent also on how the cells are utilized, models can additionally play a critical role in the optimization and control of the bioreactor and fermentation processes. Other possible model focus includes expanding the range of cell factory substrates, minimizing the formation of undesired by-products, increasing product quality, and guidance in the choice of cell factory when introducing a novel product.

Many biological processes or systems of importance to biotechnology, such as the metabolism of a cell culture during a fed-batch process, cellular stress responses, or the decision making during the cell cycle, are non-stationary in their nature. These systems are characterized by their dependence on time and the fact that the effect of inputs to the systems depends on the systems history. The most common way of modeling such dynamic systems is to set up mathematical expressions for the rates at which biochemical reactions of the systems are taking place. The reaction kinetics are then used to form mass balance equations which in turn describe the temporal behavior of all biochemical species present in the modeled system. Mathematical models of this type are usually referred to as kinetic models but the literature sometimes tends to use the terms dynamic and kinetic models interchangeably due to their largely overlapping concepts as far as biological models are concerned. Reaction kinetics being the fundamental building block of kinetic models, they are clearly distinguished from the large body of so-called genome-scale metabolic models (GEMs) which mainly focus on the stoichiometry of reactions (Thiele et al., 2009; Sohn et al., 2010; Chung et al., 2010; Österlund et al., 2012). Although kinetic models are frequently being used to describe dynamic behaviors, they are equally important in the study of processes that may be stationary or close to stationary, such as cell metabolism during exponential phase, since they can relate the properties of a (quasi) steady-state to the kinetic properties of the model components.

This review looks at the work-flow and methods for setting up, analyzing, and using kinetic models, focusing on models and modeling methodology with relevance for industrial biotechnology. The paper is divided into three main parts. The first part discusses and describes different aspects of the model building procedure, including defining the model focus, how to set up a model structure, determine parameter values and validate the model. The second part looks at how kinetic models have been used once they are set up. Applications of kinetic cell factory models for improving production, substrate utilization, product quality, and process design are reviewed. In the last part, a number of advantages and challenges of kinetic modeling are listed and some future perspectives of kinetic modeling in biotechnology are discussed. A complete overview of the organization can be found in Table 1. To increase the readability, especially for readers who are not experienced modelers, parts of the material which are of technical or mathematical nature are displayed in special boxes. The models and methods on which this review has been based have been supplied by the partners of SYSINBIO (Systems Biology as a Driver for Industrial Biotechnology, a coordination and support action funded by the European commission within the seventh framework programme) and through a thorough literature review.

Table 1
Organization of this review.

Contents

2. Setting up kinetic models – Modeling framework

The kinetic modeling procedure can be divided into a number of steps which are illustrated in Fig. 1. Since the choices and decisions made at the different steps are dependent both on the objective of the modeling and on the previous steps, the exact details of how a model is set up will be different from case to case. Also, some steps will probably have to be iterated several times before a complete model can be presented (van Riel, 2006). For instance, the model structure will most certainly evolve during the model building process, having new elements added and other removed or changed. Parameter estimation may have to be performed again as new data sets are collected, and different types of analysis on the finished model may lead to new applications that was not initially foreseen. This type of iterative work-flow is not unique for kinetic models of cell factories, but apply for modeling efforts in general (Ljung, 1987). The steps of the kinetic modeling procedure are now described briefly, and then followed by elaboration and in-depth discussions on some of their aspects.

Purpose: The first step of modeling is to define the purpose of the model, an important step as it includes the very reason for setting up a model in the first place. Typical questions are: Why do we model? What do we want to use the model for? What type of behavior should the model be able to explain? The majority of the goals of modeling cell factories are related to understanding and predicting their behavior when perturbing them either internally through genetic modifications, or externally by changing various environmental factors. The model purpose defines the complexity of the modeling problem and will influence all subsequent steps of the modeling procedure.

Network structure: The model network structure is the wiring diagram of the model. It defines the network of interconnected elements that are assumed to be important for the modeling task in question. For instance, it will contain elements such as compartments, concentrations of metabolites, enzymes and transcripts, and reactions (including transport across membranes), including their effectors and stoichiometric coefficients. It also defines the interfaces of the model with the un-modeled exterior.

Box 1–Mass balance equations and model outputs.

Combining the stoichiometric information from the model network structure with the symbolic form of the kinetic rate expressions, mass balance equations with explicitly given kinetics can be set up for all dynamic components of the modeled system. In the deterministic, continuous case, these equations can be written as

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{S} \cdot \mathbf{v}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) \quad (1)$$

and their associated initial conditions are

$$\mathbf{x}(0) = \mathbf{x_0}(\boldsymbol{\theta}). \quad (2)$$

Here, $\mathbf{x}(t)$ denotes an m-dimensional vector of time-dependent state variables, $\mathbf{S}$ a stoichiometric matrix of dimension $m \times n$, and $\mathbf{v}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})$ an n-dimensional vector of reaction rates which are dependent on the state variables, a vector of input variables $\mathbf{u}(t)$, and a set of parameters $\boldsymbol{\theta}$. Eq. (1) sometimes needs to be extended to take volume changes of the respective compartment into account, for example the dilution of intracellular species in growing cells. Additionally, it may be necessary to supplement the ordinary differential equations in Eq. (1) with a set of algebraic equations for certain models. Since the quantities measured in experiments are not necessarily the same as the model state variables, a function $\mathbf{h}(\mathbf{x}(t), \boldsymbol{\theta})$ is also needed to relate $\mathbf{x}(t)$ to a vector of model outputs

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}). \quad (3)$$

Kinetic rate expressions: Having defined the model network structure, the next step in the modeling process is the determination of the mathematical expressions that define the interactions between the different components. The model network structure already delivers information about which elements should take part in the mathematical expressions. Kinetic rate expressions can be derived from actual reaction mechanisms, with different degrees of detail, or be represented by approximate expressions capturing the essential quantitative and qualitative features of a reaction. The complexity of a reaction's kinetics is defined by the scope of the reaction, the scope of the model and the biochemical knowledge about the reactions. Both deterministic and stochastic formulations of the reaction rates may be used.

Model structure: When the network structure and kinetic rate expressions have been determined, the structure of the kinetic model is complete. The model can now be written as a set of mass balance equations with explicitly given kinetic expressions, which determines the time trajectories of the modeled species, and a list of model outputs indicating which parts of the modeled system that are being observed in experiments, see Box 1.

Parameter determination: Next, the numerical values of the parameters appearing in the rate expressions, the initial conditions, and the outputs need to be determined. Parameter values are sometimes established one by one, either from targeted experiments measuring them directly or from other types of a priori information on individual parameter values. In contrast, parameter values can also be determined simultaneously in an inductive way by utilizing the implicit information in measurements of other quantities than the parameters themselves, using parameter estimation methods. If the parameter estimation problem does not have a unique solution, the space of admissible parameter values can be further constrained using physicochemical and thermodynamics laws. Subsequently, from such a reduced space parameter values can be determined by using Monte-Carlo sampling techniques.

Validation: With the parameter values determined, the quality of the model should be assessed. Such model validation can consist of both qualitative reasoning as well as formal statistical testing. In addition to explaining experimental data used for setting up the model, it is common to further validate the model's predictive power based on new sets of experimental data that was not used previously in the modeling process.

Usage: When a model has been established it can be used in a number of different ways to answer the questions for why it was created. This involves various types of what–if analysis that explores different scenarios and investigates the impact of model assumptions. Examples of model usage include analysis of flux control in a pathway, in silico evaluation of metabolic engineering strategies, and design of optimal process conditions.

2.1. Purpose

Building models of biological systems is a way of collecting, organizing, and representing knowledge and hypotheses. The models can be thought of as formalized descriptions of what is known expressed by precise mathematical statements. They can be used for a variety of purposes including hypothesis testing, understanding how different components of a system work together to achieve some function or behavior, and learning about system components which are hard to access experimentally. Most importantly in the context of industrial biotechnology, they can be used for making predictions about the effects of genetic engineering, e.g. deleting or overexpressing a metabolic enzyme, and for optimizing the design and conditions of bioreactor or fermentation processes, e.g. determining the details of a fed-batch feeding strategy.

A common goal for many cell factory production processes, especially those for low-value products, is the desire to increase either yield, titer, or productivity, or combinations thereof. As a consequence, these quantities are ultimately what models should aim to describe and they are defined in Box 2. Which quantity is most relevant for a particular process is determined by a large number of factors such as the value and market size of the product, the substrate availability and cost, and the downstream processing. Although the models presented in this review do not always work directly with the above quantities, the models are usually describing aspects of cells and production processes that at least indirectly affect them and they should therefore always be kept in the back of the mind.

Essentially, any kinetic model whose purpose is to describe some aspects of the cellular machinery, or of the production process, that may impact the performance of a cell factory is of interest to biotechnology. Because there are many different types of cell factories and a plenitude of interesting products to be produced by them, the range of purposes and focus of potentially relevant kinetic models is wide. Depending on the problem they may address cellular processes such as metabolism, protein maturation and secretion, signaling, gene regulation, stress

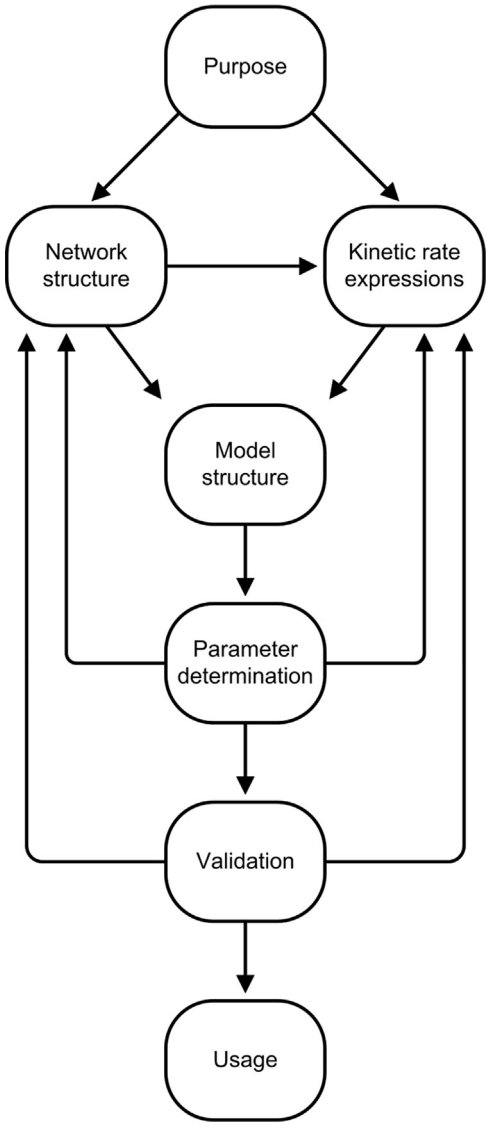

Fig. 1. Illustration of the main steps of the kinetic modeling procedure.

responses, cell cycle progression and apoptosis, as well as external or environmental factors like temperature, pH, osmolarity, product and by-product toxicity, and not least the type and operation mode of the bioreactor or fermentor. When describing the above features of cell factories a model may be specifically designed for a particular application, such as a specific pathway for the production of a special metabolite, or it can describe more general functions of the cell that may be exploited in different applications, such as primary metabolism or the protein synthesis machinery. The diversity in the purposes and scopes of kinetic models in biotechnology is reflected in the wide range of timescales of commonly modeled processes. Fig. 2 shows how important processes such as signaling, the action of metabolic enzymes, gene expression, protein secretion, the cell cycle, and bioreactor processes have characteristic timescales that span and cover almost ten orders of magnitude. Also the size of kinetic models can be very different, ranging from single enzymes (Chauve et al., 2010; Hattersley et al., 2011), to entire pathways (Hynne et al., 2001), to larger models comprising several interacting modules or pathways (Klipp et al., 2005b; Kotte et al., 2010).

2.2. Model structure

Contemporary kinetic modeling is increasingly targeting cells at the molecular level, describing components like genes, enzymes, signaling proteins, and metabolites. From a metabolic engineering perspective this is in principle advantageous since it is at this level that genetic alterations eventually would take place. In a process referred to as a bottom-up or forward modeling, mechanistic descriptions of a system's components are integrated to form a description of the system as a whole (Bruggeman and Westerhoff, 2007). The central idea of this approach is that the behavior of a system emerges from the interaction of its components, and, importantly, that the behavior can be calculated if the properties of the components and their interactions have been characterized in sufficient detail. In principle the bottom-up concept can also be applied to merge already existing models of cellular sub-systems into larger models (Klipp et al., 2005b; Snoep et al., 2006). As indicated in Fig. 1 a kinetic model consists of a network structure, a corresponding set of rate expressions, and their associated parameter values. Knowledge of all three parts is needed to form a complete model.

Box 2–Production process quantities.

If we let the time dependent functions $x(t), p(t)$, and $c(t)$ denote the biomass concentration, the specific productivity, and the specific substrate consumption, respectively, of a cell factory production process with a duration time T, the accumulative yield can be defined as

$$\frac{\int_0^T x(t)p(t)\,dt}{\int_0^T x(t)c(t)\,dt}, \tag{4}$$

the titer as

$$\int_0^T x(t)p(t)\,dt, \tag{5}$$

and the productivity as

$$\frac{1}{T}\int_0^T x(t)p(t)\,dt. \tag{6}$$

Note that the T in the expression of the productivity might itself be a parameter for optimization. For models that only consider situations where $p(t)$ and $c(t)$ are approximatively constant, such as for a continuous cultivation or perhaps for a population of cells growing in exponential phase, the yield can instead be quantified by p/c and the titer and productivity can both be replaced by looking at the specific productivity p if only a particular profile of $x(t)$ is considered.

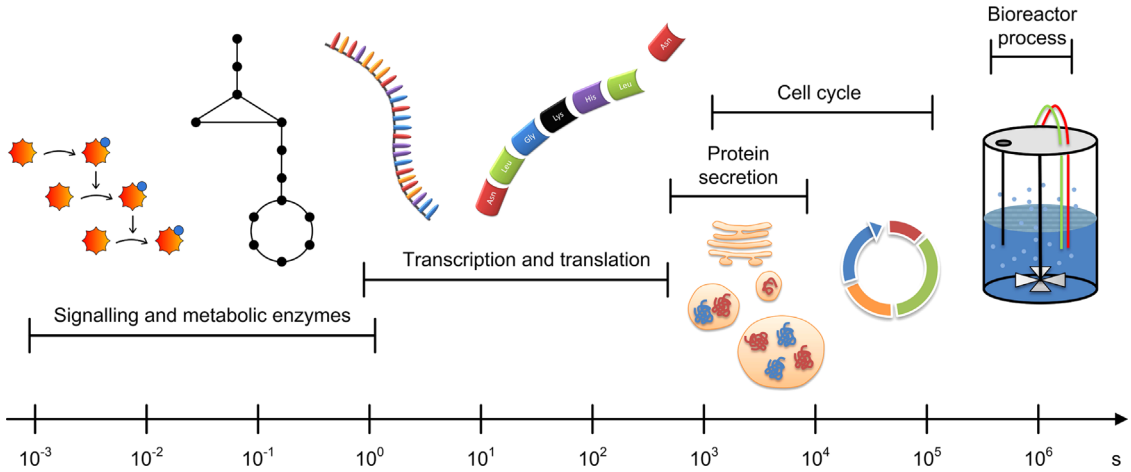

Fig. 2. Characteristic timescales for signaling, the action of metabolic enzymes, gene expression, protein secretion, the cell cycle, and bioreactor production processes.

Determination of the network structure and the symbolic structure of rate expressions in kinetic models are usually done according to the bottom-up approach (but exceptions exist, see for instance Mettetal and Muzzey, 2008). It is dependent on experimental studies characterizing the properties of the individual components appearing in the model, information that is collected directly from the literature or compiled in databases. For some systems the components have been characterized in such detail that the bottom-up approach can be applied in its entirety (Bruggeman et al., 2005), also including the determination of all parameter values. However, it is common that some or all of the parameters are unknown and instead determined indirectly from system-level measurements of other quantities using parameter estimation methods, a strategy sometimes referred to as a top-down or inverse approach.

While the biochemistry and biophysics underlying the decision-making when setting up the model structure is in some cases well understood, this is generally far from true (Kaltenbach et al., 2009). Undoubtedly, network structures and rate expressions will be set up in incomplete or even incorrect ways. It may thus seem logical trying to infer the model structure from system-level data, in the same vein as the inverse problem of parameter estimation, but because of the countless possibilities of network structures and symbolic forms of rate expressions, a top-down approach is not feasible for this part of the model building process. One strategy for handling uncertainty in the model structure is to work with an ensemble of models with different structures. This approach has for instance been employed in a study of the TOR signaling pathway in *Saccharomyces cerevisiae* (Kuepfer et al., 2007). Other efforts have focused on the development of computational tools that support the handling of such model families (Haunschild et al., 2005). The problem can also in part be tackled by using different kinds of flexible standardized kinetic rate expressions that can display a large range of kinetic behaviors depending on their parameter values. In this way part of the structural identification problem can be turned into a parameter estimation problem (Chou and Voit, 2009; Srinath and Gunawan, 2010). In another variant of the bottom-up approach, addressing the issue of determining a suitable network structure, Hildebrandt et al. (2008) proposed a strategy where mechanistic modeling on the molecular level is combined with an incremental adding of model components in a systematic way. Starting from a basic backbone model, the effect of each added component can be evaluated to gain insight into its contribution to the overall behavior of the system. The authors of that study used the procedure to construct a model for optimizing the production of single-chain antibody fragment in *S. cerevisiae*, focusing on the chaperon binding protein and the foldase protein disulfide isomerase.

In contrast to the molecular level model structures, coarse-grained, lumped descriptions of biological systems and their parts are sometimes employed instead. Setting up models with less complex network structures can be a good way of capturing known higher-level mechanisms, such as the activity of a complete pathway, even though not all molecular mechanisms are understood. This is especially true for models of protein production and the protein secretion machinery where many details are still unidentified. For example, Wiseman et al. (2007) used a simplified treatment of the endoplasmic reticulum pathways for protein folding, degradation, and export to study their contributions to protein homeostasis and protein export efficiency. Similarly, the intricate details of the pathways of the unfolded protein response (Curtu and Diedrichs, 2010) were condensed into a minimal model featuring the basic mechanisms (Trusina et al., 2008; Trusina and Tang, 2010). Despite the simplified treatment the model could provide insight into the function of this homeostatic-restoring system, in particular in addressing the differences between yeast and mammalian cells and the role of translation attenuation. An even simpler, but nevertheless very useful, model of recombinant protein secretion in *Pichia pastoris* was presented by Pfeffer et al. (2011). This model is unique in that it was able to quantify the degree of intracellular protein degradation under production like conditions. A study addressing sustained oscillations in continuous yeast cultures is yet another example of successful modeling using a relatively simple model structure (Heinzle et al., 1982). At the extreme end of simple network structures there are of course also the so-called unstructured models which only use a single state variable to describe the cell biomass in addition to a few state variables accounting for extracellular substrates and products (Menezes et al., 1994; Portner and Schäfer, 1996; Carlsen et al., 1997; Ensari and Lim, 2003; Sarkar and Modak, 2003; Liu and Wu, 2008; Yüzgeç et al., 2009). Such models are for the most part just phenomenological representations of what is empirically observed. An exception is a type of unstructured kinetic models that are derived using prior knowledge of intracellular reactions; based on a stoichiometric description of a metabolic network, a set of macroscopic reactions connecting the extracellular substrates and products are determined by decomposing the network into its elementary flux modes (Provost and Bastin, 2004; Haag et al., 2005; Provost et al., 2006; Teixeira et al., 2007; Dorka et al., 2009; Zamorano et al., 2013).

A drawback of all the less detailed network structure approaches mentioned above is the missing or complicated links between entities of the model and the actual molecular entities inside the cell. These links are particularly important if the model is to be used for identification of explicit targets for strain improvement by genetic engineering. However, depending on the purpose of modeling, a model with a simpler structure may still be useful. It can for instance foster a better general understanding of the system behavior or give

insights of the system that can be used as a starting point for further detailed modeling. A less detailed model can also be used for making predictions without explicit reference to the underlying, un-modeled reactions at the molecular level. For instance, to accurately simulate the concentration profiles of substrate, product, and biomass during a fermentation, which may be valuable for process design, a simple unstructured model may be sufficient. Thus, in situations where a simple model structure is believed to meet the requirements of the modeling purpose, nonessential details should be avoided since they will only make the modeling process unnecessary cumbersome.

2.2.1. *Representation of network structure*

The goal of the model network structure is the collection of all necessary and available biological information that will be converted into a mathematical representation. However, the network structure also serves as a basis for discussion between biologists and engineers, physicists or mathematicians. The graphical representation is therefore an important aspect of the model network structure. An accurate and standardized visual language facilitates the communication between researchers, especially for those with different backgrounds, and it rationalizes the interchange of models and biological knowledge, reducing the risk of misunderstandings and ambiguity. The Systems Biology Graphical Notation (SBGN) (Le Novère et al., 2009) was developed by members of the systems biology community to address these issues and is now emerging as a standard for graphical notation. The use of SBGN in biochemical modeling was recently reviewed by Jansson and Jirstrand (2010). Tools for visualization of model simulation results, arranged in the form of a network structure map, have also been developed (Oldiges et al., 2006; Noack et al., 2007).

2.2.2. *Kinetic rate expressions*

The kinetic rate expressions are the symbolic expressions that describe the reactions and interactions between the elements of the network structure. Determination of the numerical values of the parameters occurring in them are discussed later. A fundamental type of reaction kinetics is the so-called law of mass action. It states that the reaction rate is proportional to the concentrations of the reactants, or the reactant for a unimolecular reaction, and it is frequently used as a description for elementary reactions (reactions with one step). Kinetics of multi-step reactions, such as those of enzymes and transporters, can be derived by combining the mass action kinetics of their elementary reactions (Goryanin and Demin, 2009). The resulting dynamical systems are usually simplified based on time-scale considerations (Klipp et al., 2005a; Almquist et al., 2010a), or on symmetries, such as the commonly used assumption of identical and independent behavior of ion channel subunits (Almquist et al., 2010b). Typically the simplification is done to the point where the internal dynamics of the reaction process is lost, and the description has reduced to an explicit function of the reactants and any effectors. The reduction also means that many of the parameters appearing in the final rate expressions are aggregates of elementary reaction parameters and therefore do not always have the same type of biochemical interpretability. An example of a well-known rate expression derived from elementary reactions is the Michaelis–Menten kinetics. It is obtained by separation of slow and fast dynamics and it is usually used to describe enzyme kinetics where the concentration of substrate is much higher than the concentration of the enzyme. A thorough treatment of the Michaelis–Menten approximation and its connection to the underlying dynamic system of elementary reactions was recently presented by Chen et al. (2010).

Determination of kinetic rate expressions is complicated by the fact that mechanisms of enzymes, transporters, and other complex biochemical reactions are often unknown (Costa et al., 2011). In those cases where reaction mechanisms have been derived through careful experimental studies, detailed modeling of the different reaction steps can produce rate expressions with complicated symbolic forms and large numbers of associated parameters (Goryanin and Demin, 2009), making subsequent model analysis and parameter determination difficult tasks. It must also not be forgotten that all kinetic rate expressions, no matter how comprehensive in their details, are just models. They have limitations in their applicability, they may be incomplete, or even incorrect. For instance, the experimental conditions under which a rate expression was established may differ from those of the living cell being modeled, making the kinetics inappropriate. In addition, reaction rates will to different degrees of extent depend on variables that were not considered in the derivation, such as pH, temperature, ionic strength, or the cooperative effect of enzyme effectors.

Box 3–Approximative kinetic formats.

Generalized mass action (*GMA*) describes reactions by power law kinetics with non-integer exponents (Savageau, 1976). GMA allows an analytical steady-state solution to be calculated for linear pathways.

S-systems also use power laws kinetics but here the individual reaction rates are aggregated into two reactions for every mass balanced biochemical species (Savageau, 1976). This approximation makes analytical solutions of steady-states possible also for branched pathways, but at the risk of introducing large errors and unrealistic results in certain situations (Heijnen, 2005).

Log-linear kinetics approximates reaction rates with a linear expression of logarithmic dependencies on reactants and effectors (Hatzimanikatis and Bailey, 1996, 1997). However, the enzyme concentration appears among the linear terms and the reaction rate is thus not proportional to the enzyme concentrations, something that is generally observed.

Lin-log kinetics (Visser and Heijnen, 2003; Heijnen, 2005) is also a linear expression of logarithms but with the difference that the enzyme concentration is a multiplicative factor to this linear sum, giving a reaction rate that is proportional to enzyme concentration. Like the power law approximations of GMA and S-systems, the log-linear and lin-log approaches enable analytic solutions of steady states. However, unlike the scale-free power laws, their concentration elasticities go towards zero for high concentrations, which is in agreement with the downward concave behavior of most enzymes' kinetics (Heijnen, 2005).

Convenience kinetics is a generalization of Michaelis–Menten kinetics that covers arbitrary reaction stoichiometrics (Liebermeister and Klipp, 2006a). It can be derived from a non-ordered enzyme mechanism under the assumption of rapid equilibrium between the enzyme and its substrates and products. The convenience kinetics differs from the above rate laws in that it is saturable and can handle concentrations that are equal or close to zero, the latter situation being known to cause problems for kinetics containing logarithmic functions (Wang et al., 2007; del Rosario et al., 2008). It has also been described how to avoid violating the laws of thermodynamics by using thermodynamically independent system parameters (Liebermeister and Klipp, 2006a).

Modular rate laws is a family of different rate laws which were presented with an emphasis on thermodynamical correctness (Liebermeister et al., 2010).

Box 4–Stochastic kinetics.

Models with stochastic reaction kinetics can be based on either discrete or continuous state spaces. In a discrete stochastic model, the state of the system corresponds to the exact numbers of different types of molecules. Since it is impossible to predict the individual reactions changing the state of the system, the system must instead be described by the probability of being in each possible state. Knowing the transition probabilities between states, referred to as the reaction propensity, the time evolution of the probabilities for the different states can be described by a differential equation known as the master equation. Because of the large number of possible states even for the most simple biochemical systems it is not feasible to solve the master equation in most practical applications. What can be done, however, are (repeated) realizations of the stochastic process described by the master equation using the stochastic simulation algorithm (Gillespie's algorithm) (Gillespie, 1976), or extensions of it such as tau-leaping (Gillespie, 2001).

Another strategy to deal with the discrete stochastic process of the master equation is to approximate it by a continuous stochastic process. This is typically done by the use of stochastic differential equations known as Langevin equations, enabling simulations that are more efficient (Higham, 2001; Adalsteinsson et al., 2004). Although Langevin equations can be rigorously derived to approximate the discrete stochastic process described by the master equation (Gillespie, 2000; Lang et al., 2009), they can also be used to introduce randomness to an ordinary differential equation in an ad hoc manner (Hasty et al., 2000; Ghosh et al., 2012). The continuous process described by a Langevin equation can also be expressed by the corresponding deterministic partial differential equation for the dynamics of the probability distribution, the Fokker–Planck or Kolmogorov forward equation (Jazwinski, 1970; Gillespie, 2000; van Kampen, 2007).

2.2.3. *Approximate kinetic rate expressions*

Since most kinetic rate expressions are unknown, and because of the complexity and unreliability of those who are claimed to be known, a number of different approximative kinetic rate expressions have been suggested as alternatives. These rate expressions have in common that their symbolic structures are intended to be simple but yet flexible enough to describe many types of reaction kinetics. They aim for a small number of parameters to facilitate parameter determination, and some of them are designed to have good analytical properties or to guarantee correct parameterization from a thermodynamical point of view. Because of their standardized formats they simplify the modeling-building process, also encouraging automatic construction of kinetic models (Liebermeister and Klipp, 2006a; Borger et al., 2007; Adiamah et al., 2010; Liebermeister et al., 2010). Some of the approximative rate expressions used in kinetic modeling (generalized mass action, S-systems, log-linear, lin-log, convenience kinetics, and modular rate laws) are briefly described in Box 3.

The use of approximative rate expressions have been compared both to other approximative rate expressions as well as to traditional mechanistic formulations of reaction kinetics in a number of modeling studies. For example, a lin-log model (Visser et al., 2004) was derived based on a already established mechanistic model of the central carbon metabolism in *Escherichia coli* (Chassagnole et al., 2002), and was found to give similar simulation results despite its simpler structure and fewer parameters. In three parallel models of sphingolipid metabolism in yeast (Alvarez-Vasquez et al., 2004), the power law formats, GMA and S-systems, were compared to Michaelis–Menten kinetics. It was found that the models behaved similarly both with respect to steady states and dynamics responses. The performance of GMA, convenience kinetics, and Michaelis–Menten kinetics was compared in a number of model variants describing the biosynthesis of valine and leucine in *Corynebacterium glutamicum* (Dräger et al., 2009). Hybrid models consisting of both approximative kinetics and mechanistic kinetics have also been evaluated and concluded to be suitable approaches (Bulik et al., 2009; Dräger et al., 2009; Costa et al., 2010).

2.2.4. *Stochastic kinetics*

A deterministic formulation of reaction kinetics will gradually lose its validity as the number of reacting molecules becomes small. As a rule of thumb, there should be at least 10^2–10^3 molecules per reactant (Chen et al., 2010) when describing reactions with deterministic models. Metabolic reactions, the most commonly modeled aspects of cell factories, typically fulfill the requirements for deterministic modeling. However, low numbers of reacting molecules and stochastic behavior can occur in for instance signaling (Wang et al., 2006), gene expression (Paulsson, 2004), and protein secretion (Love et al., 2010), processes potentially relevant in cell factory applications. Modeling of these and other processes is therefore in some cases best done using stochastic approaches that take the randomness of biochemical reactions into account (Ullah and Wolkenhauer, 2010). Such simulations have for instance been used for models of *S. cerevisiae* to study the *GAL* network (Ramsey et al., 2006), and the Ras/cAMP/PKA signaling pathway (Cazzaniga et al., 2008) including the nucleocytoplasmic oscillations of the downstream transcription factor Msn2 (Gonze et al., 2008). For more details on stochastic kinetics see Box 4.

2.3. *Parameter determination*

Parameters in kinetic models are essentially determined in two different ways; either one at a time, considering the different components and processes of the model individually, or by collectively calibrating the parameters to make the model fit measurements of the intact system. The two approaches are often combined by setting some parameters to previously known or measured values while simultaneously fitting the remaining ones (Zi et al., 2010).

Following the first approach, there are studies where the model building process has been complemented by experimental work aiming to measure parameter values directly (Teusink et al., 2000), but more commonly parameters are set to values already reported in the literature (Alvarez-Vasquez et al., 2004). These values can sometimes be found in databases compiling experimental information on kinetic parameters (Kanehisa and Goto, 2000; Rojas et al., 2007; Schomburg and Schomburg, 2010; Scheer et al., 2011). A serious problem with this approach is that it usually means that parameter values will have to be collected from different sources, involving different experimental conditions, different physiological states of the cells, different strains, or even different organisms (Costa et al., 2011). Notably, it is also common that such parameter values are derived from in vitro measurements, where conditions may differ drastically to those of in vivo systems (Minton, 2001, 2006), an approach which has been shown to have shortcomings even if great care is taken (Teusink et al., 2000). The above issues are being tackled by the development of standardized experimental systems imitating in vivo conditions for specific organisms or cell types (van Eunen et al., 2010). Sometimes model parameters are determined in even less accurate ways, for instance according to rule of thumb-like considerations such as using generic rate constants for protein–protein associations or by educated guessing of enzyme K_m values (Hoefnagel et al., 2002). Finally, there are many parameters whose values cannot be determined directly due to the limitations of experimental techniques.

The alternative to determining parameters one by one is to collectively calibrate the parameters to make the model reproduce experimental measurements of other quantities than the parameter themselves. This way of indirectly determining parameters is referred to as parameter estimation (but also as system identification, model fitting, or model calibration). The parameter estimation problem can be seen either as the geometrical problem of minimizing the distance between the model output and the corresponding experimental data, or it can be interpreted statistically as the problem of maximizing the likelihood of observing the data given a model that takes the experimental uncertainty into account. It can be shown that these views on parameter estimation are related. Specifically, when the geometrical approach uses a (weighted) sum of squares as the distance measure it is equivalent to when the statistical approach model measurement errors as additive, independent, and normally distributed. For more details on how the parameter estimation problem is formulated see Box 5. Some of the challenges of parameter estimation include large qualitative and quantitative uncertainties faced in biological systems, and parameter estimation for large-scale models. In these cases, it is common that multiple sets of parameter values can make the model reproduce the measurements. When the lack of sufficient information in experimental data results in a population rather than in a unique set of parameter values, an alternative to conventional parameter estimation methods might be more appropriate (Miskovic and Hatzimanikatis, 2010; Soh et al., 2012; Chakrabarti et al., 2013). In this approach, the space of admissible parameter values is first reduced by applying physicochemical and thermodynamic constraints integrated with available measurements.

Box 5–Formulating the parameter estimation problem.

The parameter estimation problem can be formulated as the following minimization problem. Consider N measured data points, $\mathscr{D}_N = d_1, , d_N$, taken at time points $t_1, , t_N$, which are described by a scalar-valued model output, $y(t)$ (at the expense of a little more notation the line of though easily extends to the case with vector-valued outputs, see for instance Raue et al., 2009). Now an objective function $V(\boldsymbol{\theta})$ can be defined for some distance measure of the vector of residuals, $[d_1 - y(t_1,\boldsymbol{\theta}), , d_N - y(t_N,\boldsymbol{\theta})]$. For instance, using a weighted sum of squares as a measure of the distance, the objective function, $V_{SS}(\boldsymbol{\theta})$, becomes

$$V_{SS}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{(d_i - y(t_i,\boldsymbol{\theta}))^2}{\sigma_i^2} \tag{7}$$

where σ_i^2 is the weight for the ith data point. The parameter estimate, $\hat{\boldsymbol{\theta}}$, is then the set of parameters that minimizes $V_{SS}(\boldsymbol{\theta})$

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} V_{SS}(\boldsymbol{\theta}). \tag{8}$$

The parameter estimation problem can also been seen from a statistical view point, treating experimental observations as realizations of random variables (Ljung, 1987). If the model is assumed to be a perfect description of the system, the deviation of each observed data point, d_i, from the model prediction, $y(t_i)$, must originate from a measurement error, ϵ_i, here assumed to be of additive nature

$$d_i = y(t_i) + \epsilon_i. \tag{9}$$

By changing the model of the outputs in Eq. (3) to

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \epsilon, \tag{10}$$

the observed data can at any time point be seen as a deterministic part, as previously, plus the realization of the random numbers in the vector $\boldsymbol{\epsilon}$. If the measurement errors are assumed to be independent and normally distributed, with zero mean and variance σ_i^2 for the ith data point (again considering a scalar-valued model output), the likelihood of observing $\mathscr{D}_N$ given $\boldsymbol{\theta}$, $\mathscr{L}(\boldsymbol{\theta})$, can be written as

$$\mathscr{L}(\boldsymbol{\theta}) = c \prod_{i=1}^{N} \exp\left[-\frac{(d_i - y(t_i,\boldsymbol{\theta}))^2}{2\sigma_i^2}\right] \tag{11}$$

where c is a constant not affecting the optimum of the likelihood function. The parameter vector $\hat{\boldsymbol{\theta}}$ that maximizes $\mathscr{L}(\boldsymbol{\theta})$ is called the maximum likelihood estimate. Using the fact that the logarithm is a strictly monotonically increasing function, the problem of maximizing $\mathscr{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ can be replaced with the problem of minimizing the negative logarithm of the likelihood function

$$-2 \ln \mathscr{L}(\boldsymbol{\theta}) = -2 \ln c + \sum_{i=1}^{N} \frac{(d_i - y(t_i,\boldsymbol{\theta}))^2}{\sigma_i^2}, \tag{12}$$

making the optimization problem equivalent to the sum of squares minimization described in Eq. (8). Therefore, the geometrical approach using a weighted sum of squares as discrepancy measure will coincide with the statistical approach if measurement errors are independent and normally distributed. More generally, any conceivable model of the measurement error like the one used here will correspond to some kind of distance measure of the vector of residuals.

The likelihood function above describes the probability of observing the data $\mathscr{D}_N$ given the parameters $\boldsymbol{\theta}$. It is also possible to treat the parameters themselves as random variables (Ljung, 1987; Secrier et al., 2009). Using Bayes' rule, the probability density function for the parameters given the data, $p(\boldsymbol{\theta}|\mathscr{D}_N)$, or the posterior, can be written as

$$p(\boldsymbol{\theta}|\mathscr{D}_N) = \frac{p(\mathscr{D}_N|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathscr{D}_N)} \propto p(\mathscr{D}_N|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{13}$$

and the parameter set maximizing $p(\boldsymbol{\theta}|\mathscr{D}_N)$ is called the maximum a posteriori estimate. The posterior distribution is a combination of the likelihood (of observing $\mathscr{D}_N$ given the parameters) and any prior knowledge of the parameters. Prior knowledge could for instance come from typical distributions of similar parameters, or from previous estimates which did not include the data used for the likelihood. If there is no prior information about parameter values, i.e., the prior is a uniform distribution whose logarithm adds nothing but a constant to the objective function, the maximum a posteriori estimate is reduced to the maximum likelihood estimate. The Bayesian approach with maximum a posteriori estimation has for example been applied to a model of the threonine synthesis pathway (Liebermeister and Klipp, 2006b).

Then, the reduced solution space is sampled using Monte-Carlo techniques to extract a population of alternative sets of parameter values.

2.3.1. Computing the estimate

When an objective function describing a model's ability to reproduce the experimental data have been formulated – be it a likelihood function based on a probabilistic model of model prediction errors, or some other function – the parameter estimate is obtained by locating its optimum. This is accomplished by different ways of iteratively searching through the parameter space, usually taking constraints on admissible parameter values into account, and a large number of different optimization algorithms have been designed for this task, see Box 6. However, the problem is complicated by the fact that most models of biological systems contain nonlinearities and many of these models have large number of parameters to be estimated. A high dimensional parameter space in combination with strong nonlinearities can result in complexly shaped objective functions with many local optima. Such multimodality makes it hard to assess whether the global solution to the optimization problem has been located or if only a local optimum has been found. Adding further to the problem are the often vast and relatively flat parts of the parameter space, which only shows a weak response in the objective function (Transtrum et al., 2010) and consequently may delay the convergence of the search. As the objective function is not given as an explicit function of the model parameters, its values for a certain parameter vector must be determined by solving the model equations. Every iterate of the optimization algorithm therefore requires one or more evaluations of the model equations and the majority of time spent on computing the estimate is typically used for integrating ODEs (Chou and Voit, 2009). The main challenges when optimizing the objective function are thus to locate the global optimum, and doing this in reasonable time.

2.3.2. Identifiability analysis and experimental design

An important but sometimes overlooked aspect of parameter estimation is the level of confidence in the obtained estimates and whether it is possible at all to uniquely assign values to the parameters (Cedersund, 2006; Gutenkunst et al., 2007b; Ashyraliyev et al., 2009; Roper et al., 2010; Raue et al., 2011; Erguler and Stumpf, 2011; Meshkat et al., 2011; Hattersley et al., 2011). To accurately estimate parameters requires a balance between the information content in the experimental data and the complexity level of the model. However, it is widely acknowledged that kinetic models often are overparameterized and too complex in their structures in relation to available quantitative data (Nikerel et al., 2006, 2009; Schmidt et al., 2008; Sunnåker et al., 2010; Schaber and Klipp, 2011). Some models have intrinsic symmetries that allow transformations of state variables and parameters in a way that does not change the model output. Such redundant parameterization leads to a likelihood function that instead of a unique minimum has a completely flat valley, meaning that there

Box 6–Optimization.

Two main categories of optimization methods can be distinguished, so-called local and global methods. Local methods require some kind of initialization of parameters, a position in the parameter space from where to start the optimization. This parameter set can come from in vitro measurements of reaction kinetics or other kinds of estimates, perhaps reported in the literature, but may also require guessing. The initial parameter set is then improved by repeated application of the optimization algorithm. Many local methods determine their direction of search in the parameter space based on the gradient and Hessian of the objective function at the present point in parameter space (Nocedal and Wright, 1999). The Newton method uses the exact Hessian, but quasi-Newton methods approximating the computationally costly Hessian using gradients, like the SR1 or BFGS algorithms, are more commonly used. For least squares problems, which are the most common in biochemical modeling, the Hessian approximation of the Gauss–Newton and Levenberg–Marquardt (Marquardt, 1963) methods are especially appropriate (Nocedal and Wright, 1999). The gradient of the objective function needed by these methods are typically computed by finite difference approximations. However, numerical solutions of the model equations using adaptive step length ODE solvers are known to introduce "quantification errors" to the objective function, making it non-smooth on small scales (Bohlin, 2006; Carlsson and Nordheim, 2011). The finite difference approximation may thus become an unreliable description of the gradient and gradient-based methods can as a consequence experience difficulties. To overcome such problems the gradient can instead be determined by integration of the so-called sensitivity equations (Ljung and Glad, 1994a; Skaar, 2008; Carlsson and Nordheim, 2011). Another strategy of handling issues with non-smooth objective functions is the use of non-gradient based methods like the Nelder–Mead method (Nelder and Mead, 1965), the Hooke–Jeeves method (Hooke and Jeeves, 1961), or the principal axis method (Brent, 1973). Although such methods are robust and easy to implement, they generally have much slower convergence in terms of the number of objective function evaluations.

Since the objective function typically has several local optima the choice of initial values is crucial for finding the global optimum using local methods. The inefficiency of local methods in finding the global optimum (Mendes and Kell, 1998; Moles et al., 2003) has spurred the development of global optimization methods that search the parameter space more comprehensively. A common drawback with these algorithms is a slower rate of convergence. Some of the popular global methods include simulated annealing (Kirkpatrick et al., 1983; Nikolaev, 2010), a large number of different genetic and evolutionary algorithms (Sarkar and Modak, 2003; Yüzgeç et al., 2009; Chou and Voit, 2009; Ashyraliyev et al., 2009), and particle swarms (Kennedy and Eberhart, 1995), and their performance has been compared in several studies (Moles et al., 2003; Dräger et al., 2009; Baker et al., 2010).

Most successful is the combination of local and global search methods. Such hybrid methods benefit both from the global methods' ability to explore the parameter space and from the faster convergence rate of the local methods once close to a (local) optimum. As an example, the results obtained by Moles et al. (2003) using the global SRES method (Runarsson and Yao, 2000, 2005) were substantially improved by different combinations with local methods (Rodriguez-Fernandez et al., 2006b), and further strengthened by a systematic strategy for when to switch from the global to the local method (Balsa-Canto et al., 2008). Even more promising results have been obtained with a hybrid approach based on a scatter search metaheuristic (Rodriguez-Fernandez et al., 2006a). An enhanced version of the scatter search (Egea et al., 2010) has also been shown to benefit from a cooperative parallelization (Balsa-Canto et al., 2012), as illustrated in a comparison with a non-cooperative parallelization of the algorithm on the parameter estimation problem of the 193 parameter *E. coli* model by Kotte et al. (2010).

Several of the local, global, and hybrid methods mentioned above are available through modeling software tools like SBML-PET (Zi and Klipp, 2006), the Systems Biology Toolbox (Schmidt and Jirstrand, 2006; Schmidt, 2007), COPASI (Hoops et al., 2006; Mendes et al., 2009), PottersWheel (Maiwald and Timmer, 2008), and AMIGO (Balsa-Canto and Banga, 2011).

are several parameter sets that are equally likely to have produced the measured data. Models of this type are said to be structurally unidentifiable (Bellman and Åstrom, 1970; Pohjanpalo, 1978). It should be emphasized that this property is only dependent on the model structure itself, including the set of measured model outputs and the known input variables, but not on the quality or quantity of data used for estimation. The analysis of structural identifiability can therefore be done a priori, meaning that neither experimental data, nor a certain parameterization, is required.

Structural identifiability is a necessary condition for an unambiguous estimation of parameters. It may however not be sufficient because it can happen that even though the likelihood function has a unique minimum for some parameter set, the surroundings of this minimum could be very flat. Consequently there may be other parameter sets with potentially very different values that are almost as likely. Such diverse parameter sets yielding very similar outputs have for example been observed in a model of the methionine cycle dynamics (Piazza et al., 2008) and in a model of monoclonal antibody production in Chinese hamster ovary (CHO) cells (McLeod et al., 2011). This situation is referred to as a lack of practical identifiability. Unlike structural identifiability, this property does depend on the amount, quality, and time points of experimental observations. Methods for determining practical identifiability also require that a parameter estimate has already been obtained, and can therefore not be applied a priori. A review of methods for identifiability analysis is found in Box 7.

When estimating model parameters from experimental data, decisions have to be made about what kind of experiments to perform. It is rare that all state variables can be measured and typically there are several quantities appearing in the model for which experimental methods exist but come at a high cost in terms of time- or resource-consumption. In these situations, identifiability analysis can be a useful tool to guide the experimental design. For instance, the structural identifiability of a model depends on the set of model outputs but it is not only interesting to know whether a particular set of measured outputs renders the model identifiable but it is also of great interest to learn which potential sets of outputs that have to be measured in order to ensure structural identifiability. Addressing this question, an algorithm was developed in the group of Jirstrand and colleagues (Anguelova et al., 2012) that a priori finds so-called minimal output sets, which are sets of outputs that when measured results in an identifiable model. The algorithm has been implemented in Mathematica (Wolfram Research, Inc., Champaign, USA) and used successfully in the analysis of models with over 50 parameters (Anguelova et al., 2012). Since methods that only determine structural identifiability will not be able to detect practical identifiability, they can never be used to prove the feasibility of a certain experimental design. Rather, because approaches like the

Box 7–Identifiability analysis.

One algorithm for determining structural identifiability has been presented by Sedoglavic (2002) which is particularly interesting. Unlike previous efforts (Vajda et al., 1989; Audoly et al., 2001; Margaria et al., 2001) this method does not suffer from the limitation of only being applicable to smaller systems. In fact, a recent implementation of the algorithm, which was also extended to handle parameterized initial conditions, has been successfully applied to models with a size of about 100 state variables and 100 parameters using a standard desktop computer (Karlsson et al., 2012). The results obtained by Sedoglavic are so far unfortunately not disseminated in the biological modeling community, one of the reasons perhaps being the use of the related term observability instead of the, in the biological field, more common term identifiability. It should be noted that this method, and all other methods based on the so-called rank-test, are testing for so-called local structural identifiability. Thus, these methods will identify redundant parameterizations that correspond to completely flat and continuous regions in the likelihood function but there may still be an enumerable set of non-neighboring single points in the parameter space, also resulting in identical model output, which are not detected by this analysis. One situation, where multiple parameter sets are possible and where local structural identifiability analysis might be insufficient, is when measuring one or more components of a pathway containing an upstream reaction which is catalyzed by two or more isoenzymes whose concentrations and activities are not explicitly measured. If the different enzymes are described by the same type of model structure, permutations of concentrations and kinetic parameters for the set of isoenzymes results in models with identical output. The models themselves are however not identical because the different parameter sets have different implications when interpreting the properties and functions of the actual enzymes and their corresponding genes. Methods for the analysis of global structural identifiability exist (Ljung and Glad, 1994b; Bellu et al., 2007) but are typically only applicable to smaller systems with just a few state variables and parameters (Roper et al., 2010), or systems with a particular structure (Saccomani et al., 2010), and therefore so far of lesser interest in the analysis of most models addressed in this review. A notable exception is the successful application of the generating series approach to a medium-sized model of the NFκB regulatory module (Chis et al., 2011). Though, potential issues with non-identifiability in the global sense could be eliminated if there is a priori knowledge about parameter values that can be used as a starting guess when computing the estimate or to discard an incorrect solution to the parameter identification problem.

A simple way of evaluating how accurately parameters can be identified in practice is to look at the standard parameter confidence intervals determined from a quadratic approximation of the log-likelihood function around its optimum. However, due to the frequent combination of limited amounts of experimental data and model outputs that depend non-linearly on the parameters, this type of confidence intervals can be unsuitable (Raue et al., 2011; Schaber and Klipp, 2011). Another way of assessing the accuracy of the parameter estimates is to use exact confidence intervals determined by a threshold level in the likelihood. A method to calculate such likelihood-based confidence intervals based on the profile likelihood was recently proposed (Raue et al., 2009, 2010). Here, all parameter directions of the likelihood function are explored by moving along the negative and positive directions of each parameter while minimizing the likelihood function with respect to the remaining parameters (which means that one studies the projection of the likelihood onto a specific likelihood-parameter axis plane). The confidence intervals are determined by the points where these likelihood profiles cross over a certain threshold, and the confidence levels are determined by the level of that threshold. If the profile likelihood for a parameter never reaches the threshold in either the negative or positive direction, or in neither, the confidence interval of this parameter extends infinitely in at least one direction. According to this approach, parameters with unbounded confidence intervals are defined as non-identifiable. This definition would make no sense for confidence intervals determined from the likelihood curvature at the point of the estimate, since these are always finite (with the exception of a completely flat likelihood resulting from a structural non-identifiable parameter). Profiling can also be applied to posterior distributions (Raue et al., 2013).

minimal output sets do not require any wet-lab efforts at all, the appropriate use of such structural identifiability analysis is to beforehand disprove any experimental design that is bound to fail in identifying the parameter values, and to give well-founded suggestions of which additional quantities that have to be measured to resolve the identifiability issues. Insights obtained in this way can potentially save a lot of valuable laboratory resources. The analysis of practical identifiability will on the other hand require an existing set of measurements, but it can not only determine which parameters that are impossible to estimate uniquely but also those that are too poorly constrained. This type of analysis is therefore able to confirm if a given set of measurements really is sufficient for parameter identification in practice. If this is not the case, and additional measurements are required, practical identifiability analysis can be used to improve the experimental design in more specific ways than methods like the minimal output sets, for instance by indicating certain time points at which the measurement of a particular quantity is most efficient in (further) constraining a parameter value (Raue et al., 2009, 2010). Thus, structural and practical identifiability analysis fulfills different needs and can be said to have complementary roles when used for experimental design.

2.3.3. Model reduction

It was shown above that identifiability analysis can guide the experimental design so that the correct type and amount of data required for system identification is collected. Another way of achieving the balance between model and data is to decrease the complexity of the model by different model reduction techniques. These techniques aim at simplifying models to reach an appropriate level of detail for experimental validation (Klipp et al., 2005a), and if done properly the reduced model retains the essential properties of the original model. Model reduction can also be performed on models where the parameters have already been identified and whose applicability has been validated. In these cases the purpose of the reduction is to facilitate the understanding of essential structures and mechanisms of the model and to decrease the computational burden of simulation and analysis. Methods for model reduction are discussed in Box 8.

In addition to the more formal methods mentioned in Box 8, a lot of model reduction is often done by the modeler already when setting up the network structure and formulating the rate expressions. For instance, different post-translationally modified versions of a protein might de described by a single lumped state variable, concentrations of co-factors might be excluded as state variables and consequently not considered in the rate expressions of reactions in which they participate, reactions which are thought to be marginally relevant for the problem at hand might be left out from the model, known rate expression might be simplified and described by approximate kinetic formats as explained previously, and quantities that are changing slowly in the characteristic time scale of the model, such as the synthesis and degradation of enzymes during a much faster metabolic process, may be considered frozen and hence set constant. Decisions like these are usually dependent on a combination of the purpose of the model, the modelers experience and intuition, and prior knowledge of the modeled system.

2.4. Validation

Before a model is ready to be used its quality should be established. This is done not only by evaluating the model's ability to explain the experimental data used for parameter estimation but also by comparing some of its predictions to new data that was not used earlier in the model building process (Ljung, 1987). If a priori information is available on values of parameters with a biophysical interpretation, these should be compared to the estimated values as a feasibility check. Additionally, other aspects of the model, such as the predictions of unobserved state variables,

Box 8–Model reduction.

Two popular categories of model reduction methods are the ones based on time-scale separation and lumping. The time-scale separation approach is based on defining a time-scale of interest and neglecting changes in state variables that occur on slower time-scales and approximating state variables and processes associated with faster times-scales using the quasi-steady-sate and the quasi-equilibrium approaches (Klipp et al., 2005a; Nikerel et al., 2009). Thus, the dynamics of some state variables will be replaced by either constants or algebraic relations. If the time-scales of the reactions in a system are not known, several reduced versions of a model may be considered (Almquist et al., 2010a) or further assumptions could be made (Almquist et al., 2010b). Lumping, on the other hand, transforms the original state variables to a set of new state variables in a lower dimensional state space (Okino and Mavrovouniotis, 1998). The choice of which state variables to lump together is frequently based on time-scale considerations, which results in groups of quickly equilibrating state variables being completely eliminated and replaced by a new state variable. One example of model reduction through lumping can be found in a study of secondary metabolism pathways in potato (Heinzle et al., 2007). Here, the steady-state assumptions which were used to motivate the lumping of different metabolites were derived from experimental work. Even though model reduction through lumping and time-scale separation often overlap, this is not always the case. Examples of time-scale separation not involving lumping include setting slowly varying variables to constant values, and examples of lumping not involving time-scale separation include mean concentration models of cellular compartments, i.e., reaction–diffusion equations represented without the spatial dimension. Other model reduction techniques include sensitivity analysis (Degenring et al., 2004; Danø et al., 2006; Schmidt et al., 2008) and balanced truncation (Liebermeister et al., 2005). The previously mentioned profile likelihood approach to practical and structural identifiability analysis can also be used for model reduction (Raue et al., 2009, 2010, 2011).

In most models with relevance for biotechnology the model components, such as state variables, their rates of change, and parameter values, have precise physical meaning. A successful model reduction should therefore not only preserve the input–output relations, which may be sufficient in other disciplines where models are used, but also preserve the interpretation of model components (Cedersund, 2006). These ideas are central in a recently developed method that reduce models by lumping (Sunnåker et al., 2010). Based on the approximation that state variables involved in fast reactions are in quasi-steady-state, interconnected groups of such quickly adjusting states are identified and lumped together. The distribution among the original states of a lump is determined analytically by so-called fraction parameters. These parameters can be used to retrieve the details of the original model, which is known as back-translation, thereby allowing better biochemical interpretation of analysis and simulations done with the reduced model. The method has also been extended to be able to handle nonlinear models and was successfully applied to a model of glucose transport in *S. cerevisiae* (Sunnåker et al., 2011).

may be interrogated with respect to their biological plausibility. Quality controls like the above are referred to as model validation. Strictly speaking, however, a model can never be validated. It may explain all experimental data generated sofar but it can never be proven to correctly account for future experiments. What is meant by validation is rather that the model has withstood repeated attempts to falsify or invalidate it. The rationale here is that the more experiments that have been successfully explained by the model, and the more reasonable it is with respect to a priori information about the biological system, the more it can be trusted to correctly predict future experiments. If a model fails to pass the validation step, researchers need to revise their model by suitable iteration of the modeling steps outlined in Fig. 1.

The ability of a model to explain experimental data is frequently judged by visual inspection of the respective time-series (Heinzle et al., 2007; Li et al., 2011; Cintolesi et al., 2012) or by qualitative comparison of model characteristics (Gonzalez et al., 2001). A qualitative comparison may for instance involve an investigation of whether the model can produce certain observed behaviors such as oscillations, homeostasis, or switching. Such analysis is sometimes actually performed before parameters have been formally determined, typically using some initial estimate of the parameter values, which might result in models being discarded already at this point. While these less rigorous assessments may be a good first step of the validation procedure there are also formal statistical tests for determining the quality of a model, see Box 9. Regardless of the outcome of statistical tests and formal methods of validation, it should not be forgotten that these are best used as support for decisions made by the modeler (Cedersund and Roll, 2009) and that the ultimate validation is whether the model can fulfill the purpose for why it was created in the first place (Ljung, 1987).

Sometimes validation is done by qualitatively different types of data than what was used for model identification. For instance, the biological system can be measured under new external conditions (Shinto et al., 2007; Oshiro et al., 2009), resulting in a different operating point, new types of input schemes (such as steps, pulses, periodic pules, or staircases) may be used (Klipp et al., 2005b; Zi et al., 2010), data can be collected on previously unmeasured molecular species, and validation experiments can be conducted on modified versions of the original system, i.e. mutants, where enzymes or other components are inactive, constitutively active, or have been underexpressed, overexpressed, or completely deleted (Alvarez-Vasquez et al., 2005; Klipp et al., 2005b; Wang et al., 2006; Zi et al., 2010; Cintolesi et al., 2012). When models can successfully explain such new data, it is a strong indication that the mechanistic principles and assumptions behind the model are sound.

3. Using kinetic models

From a biotechnology perspective, a complete and validated model according to the steps outlined previously is usually not in itself the ultimate goal of modeling. The real value of a model lies instead in using it to predict, evaluate, and explore different scenarios or assumptions involving the modeled system and its surrounding environment. An established model should thus foremost be seen as a tool that can be used to answer questions about the cell factory and it should be used as a complement or alternative to performing actual experiments in the lab.

3.1. Improving production

A major question which has been attempted to be answered using kinetic models is how to rationally design directed metabolic engineering strategies that will improve a cell factory's ability to produce a desired product. This requires models that can predict the behavior of the cell in response to genetic alterations like gene deletion or overexpression. One way of using kinetic models to identify suitable targets is to perform a local parameter sensitivity analysis. A more thorough treatment of the problem involves simulating larger changes in the levels of enzymes and other components.

3.1.1. Local parameter sensitivity analysis

The aim of a local parameter sensitivity analysis is to determine the degree of change of some model property like a flux, a concentration, or a more complex quantity such as the area under the curve of some state variable, in response to a change in the model parameters. As the parameters may represent quantities that can be manipulated by genetic engineering, such as enzyme concentrations, the analysis provides predictive links between potential targets and their effect on the cell factory behavior. Since a local analysis only considers small or even infinitesimal perturbations around a point in parameter space, it is not indented to mimic any actual changes in, for example, an enzyme concentration. However, a

Box 9–Validation.

Model validation is typically done by analyzing the deviation between the measured data and the model outputs, $\epsilon_i = d_i - y(t_i)$. For a model to be good these residuals should be sufficiently small and uncorrelated. First of all, if the parameters have been collectively estimated, the model should be able to satisfactorily describe this 'training' data. For instance, the size of the residuals can be tested by a χ^2 test (Jaqaman and Danuser, 2006; Cedersund and Roll, 2009) and the correlation of residuals can be tested by a run test or a whiteness test (Cedersund and Roll, 2009). Secondly, the residual analysis should be performed also with new data that were not previously used. This is done to assure that a good fit is not just because a too complex model has been over-fitted to the particular data points of the estimation set. Validating a model with fresh data means that unless new data can be collected after parameter estimation, some data has to be saved. This can be a problem if there is not much data to begin with. A common approach to this situation is the use of resampling methods (Molinaro et al., 2005) where the model validation procedure is repeated and averaged over different partitions of the original data into training and validation sets. One such method is *k*-fold cross-validation, which has for instance been used in modeling of the TOR pathway (Kuepfer et al., 2007).

Model validation sometimes also involves comparison between competing models describing the same biological system, to see which one is 'most valid' (Schaber et al., 2012). Two common criteria used to find the most suitable model include the Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978), and two common tests that also address the statistical significance of model discrimination are the likelihood ratio test (Kreutz and Timmer, 2009; Cedersund and Roll, 2009) and the F-test (Jaqaman and Danuser, 2006; Cedersund and Roll, 2009). Other approaches to model discrimination, which included the dependence of model discrimination on experimental design, have been explored in studies on formate dehydrogenase production in *Candida boidinii* (Takors et al., 1997) of and L-valine production in *C. glutamicum* (Brik Ternbach et al., 2005).

local parameter sensitivity analysis is easy to perform, gives a concise and transparent output, and despite its limitations it does have some predictive power allowing the results to be used as guidelines for identifying reasonable metabolic engineering targets.

A popular application of kinetic models is a special type of sensitivity analysis called metabolic control analysis (MCA) (Fell, 1992; Nielsen, 1998; Visser and Heijnen, 2002), the basis of which was already developed in the seventies (Kacser and Burns, 1973; Heinrich and Rapoport, 1974). It is concerned with the problem of quantifying how the control of steady-state flux is distributed among the enzyme-catalyzed reactions of a pathway. Two of the central quantities in MCA are the elasticity coefficients and the flux control coefficients (FCCs), defined in Box 10. They are both measures of sensitivity that have been scaled to obtain dimensionless numbers.

Normally there are many non-zero FCCs, meaning that there is no single rate-limiting enzyme but that the control of a flux instead is distributed over several reactions. However, it is likely that some reactions have larger values of their FCCs than others, indicating that these reactions are the ones primarily controlling the flux. The enzymes of those reactions may consequently be promising targets for successful metabolic engineering of the pathway. Given a kinetic model the FCCs can readily be calculated directly from its steady-state(s). The steady-state can be obtained either by simulations asymptotically approaching it, or by analytical or numerical solutions of the model equations. Alternatively, the FCCs may for linear pathways be determined indirectly from the summation and connectivity theorems using elasticity coefficients derived from the individual reaction rates of the model.

Sensitivity analysis in the form of MCA has been applied to a variety of kinetic models describing many different kinds of cell factories and types of products. It has for example been used to determine suitable genetic targets for improved production of lysine in *C. glutamicum* (Hua et al., 2000). This study found that lysine production was primarily controlled by the enzymes aspartokinase and lysine permease. The outcome of the analysis was verified experimentally by overexpression of aspartokinase, resulting in a significant increase in lysine production. However, the lysine flux did not increase as much as would be expected from the sensitivity analysis, suggesting that model predictions of this type are best used as supporting guidelines and that they never should be taken as indisputable facts. Further verification of the model's predictive capability was obtained by overexpression of the low flux-control enzyme dihydrodipicolinate, which only had a very limited effect on the production rate. Recently, Cintolesi et al. (2012) applied MCA to a model of ethanol production from glycerol in *E. coli*. Their analysis suggested that the control of the glycerol fermentation was almost exclusively shared between glycerol dehydrogenase and dihydroxyacetone kinase. The validity of this prediction was confirmed by the 2.4-fold increase in glycerol to ethanol flux observed when simultaneously overexpressing both enzymes. It was additionally seen that overexpression of other enzymes involved in glycerol metabolism, but whose flux control coefficients were close to zero, did not lead to increased rates of glycerol consumption and ethanol synthesis. The use of MCA is not limited to fluxes of metabolites but can be applied to the steady-state flux of any chemical entity. For instance, Gonzalez et al. (2001) used MCA to study monoclonal antibody synthesis in eukaryotic cells. They came to the conclusion that control of antibody production is shared between different steps of the synthesis pathway and that this division depends on the extracellular conditions and the physiological state of the cell. Their predictions were shown to compare qualitatively well with previously published experiments. Other examples of MCA applied to kinetic models include glycerol synthesis in *S. cerevisiae* (Cronwright et al., 2002), valine production in *C. glutamicum* (Magnus et al., 2009), the central carbon metabolism (Chassagnole et al., 2002) and production of threonine (Chassagnole et al., 2001), tryptophan (Schmid et al., 2004), and serine (Nikolaev, 2010) in *E. coli*, L-cysteine production in *Pseudomonas* sp. (Huai et al., 2009), production of lactic acid (Oh et al., 2011) and compounds of the acetolactate branch (Hoefnagel et al., 2002) in *Lactococcus lactis*, and the penicillin biosynthetic pathway in *Penicillium chrysogenum* (Theilgaard and Nielsen, 1999). Except for the work by (Hoefnagel et al.), which is further discussed in the next subsection, the model predictions of those studies were not tested by actually constructing the correspondingly modified

Box 10–Metabolic control analysis.

Consider a pathway, possibly containing branching points, consisting of metabolites x_i, and reactions rates v_j which are catalyzed by enzymes with concentrations e_j. The elasticity coefficients (ECs) are then defined as

$$\epsilon^j_{x_i} = \frac{x_i}{v_j}\frac{\partial v_j}{\partial x_i}, \tag{14}$$

which means that for each reaction of the pathway there is a set of ECs measuring its sensitivity to the concentrations of the different metabolites. Each EC is a property of an individual enzyme and is therefore independent of the activity of the other enzymes in the pathway. For any steady-state flux J in the pathway the flux control coefficients (FCCs) are defined as

$$C^J_j = \frac{e_j}{J}\frac{\partial J}{\partial e_j}. \tag{15}$$

They quantify the degree of control exerted by the different enzymes on a steady-state flux of the pathway as a whole. This means that an FCC for one of the enzymes can depend on the properties of the other enzymes, and the FCCs are therefore system properties. The ECs and the FCCs are related by the summation theorem

$$\sum_j C^J_j = 1 \tag{16}$$

which states that the sum of all FCCs is 1, and by the connectivity theorem

$$\sum_j C^J_j \epsilon^j_{x_i} = 0 \tag{17}$$

which states that for each metabolite, the sum of the product of the FCCs and the ECs with respect to that metabolite is zero. The full details of MCA comprise additional sensitivity coefficients which are related through similar theorems. Thus, MCA is not just a sensitivity analysis but also a theoretical framework that formally describes the connection between properties of a system and its components.

microbes. There are also computational studies where MCA has been combined with parameter sampling approaches in order to examine the effect of parameter uncertainty (Pritchard and Kell, 2002; Wang et al., 2004; Wang and Hatzimanikatis, 2006a, 2006b; Miskovic and Hatzimanikatis, 2010).

Other types of local parameter sensitivity analysis are also abundantly represented in the literature. For example, Oshiro et al. (2009) determined the impact of parameter perturbations in a kinetic model describing the dynamics of lactic acid production in xylose fermenting *L. lactis*. Based on their results, the enzymes around the pyruvate node were proposed as targets for genetic manipulation. Some of these authors had also previously carried out a similarly designed sensitivity analysis of acetone–butanol–ethanol production in *Clostridium acetobutylicum* (strain N1-4, formerly known as *Clostridium saccharoperbutylacetonicum*) (Shinto et al., 2007). Here, several genetic engineering strategies for increased butanol production were suggested, including decreasing the activity of CoA transferase for butyrate and increasing the activity of the reverse pathway of butyrate production. This model (Shinto et al., 2007) was later improved by Li et al. (2011) who arrived at similar conclusions in their sensitivity analysis. In addition to looking at single parameters, their analysis also considered all combinations of parameter pairs. Though, as the size of parameter perturbations was small, the combined effects of simultaneously changing two parameters always equaled the sum of the separate parameter effects and no nonlinear crossover effects were thus found. Unfortunately did neither the *L. lactis* study nor the *Clostridium* studies genetically implement the proposed strategies. In recent work on CHO cells, McLeod et al. (2011) used sensitivity analysis to investigate which cellular process that controlled the production of a recombinant monoclonal antibody. The sensitivity analysis was repeated to specifically target different days of a two week fed-batch process. Unlike the modeling study by Gonzalez et al. (2001) it was found that control was divided almost exclusively between transcription, degradation, and translation of mRNA, and that this control structure did not change appreciably during the different phases of culturing. The authors consequently suggested that genetic engineering strategies for their system should focus on these processes, but the validity of their predictions was not tested experimentally.

3.1.2. Simulating larger changes

The theory behind MCA and other local sensitivity approaches is based on small perturbations of the parameter values and the resulting sensitivities are normally only valid in the vicinity of the nominal parameter values. Realistic cases of genetic manipulation will on the other hand likely involve larger changes in the levels of gene product concentrations. The extent to which the results of a local analysis of the model can be extrapolated to larger perturbations differs from case to case and cannot generally be determined (Visser et al., 2004; Schmid et al., 2004; Nikolaev, 2010). However, if a kinetic model has been formulated there is usually no reason for limiting the model analysis to local parameter sensitivities. Just as the control coefficients of MCA can be calculated directly from model simulations, the model can in principle be used for simulating any kind of perturbation of its components. By simulating more extensive changes to models, metabolic engineering scenarios can be explored in more realistic ways. In this respect, such approaches are more powerful compared to the traditionally used sensitivity analysis like MCA, and the predictions made have the potential to be much more accurate. Though, performing simulations that involve large changes in the model parameters, or even changes to the model structure, may require that the model has good predictive power not only for the specific physiological setting for which it was developed but also for other operating points, something that cannot generally be assumed to be true. Therefore, the more extensive the perturbations to the model are, the more careful one should be when interpreting the results.

One example of model-based analysis of actual metabolic engineering strategies was provided by Hoefnagel et al. (2002). Based on MCA-derived candidate targets for increasing the production of acetoin and diacetyl in *L. lactis*, they proceeded with simulations of larger changes in the concentrations of two enzymes. First, a mutant with a lactate dehydrogenase deletion was simulated. This did indeed lead not only to a substantial flux towards the acetolactate synthase branch but also to a reduction in glycolytic flux, indicating potential problems with growth rates for such a strain. Then, a 40-fold overexpression of NADH-oxidase was simulated also resulting in some of the flux being diverted into the acetolactate branch. Finally, a simulation combining the two modifications was performed and it predicted that 92% of the flux through the pyruvate node would go into the desired direction, and that the glycolytic flux would be less affected. This fraction should be compared to a negligible 0.1% measured in the wild-type strain. The model prediction was tested, and at least to a certain degree confirmed, by an experiment which showed that 75% of the pyruvate ended up as acetoin in a strain where lactate dehydrogenase had been knocked out and NADH-oxidase was overexpressed. In another purely computational study, Chen et al. (2012) developed two separate kinetic models of glycolysis and the pentose phosphate pathway in *S. cerevisiae* and CHO cells. The authors then used the yeast model to analyze the impact of metabolic engineering targeting the production of dihydroxyacetone phosphate. Specifically, a deletion of the enzyme triose phosphate isomerase was simulated by setting its activity to zero. The rate of dihydroxyacetone phosphate production, and its yield on glucose, for this in silico deletion mutant was subsequently determined under different glucose uptake rates. Yet other studies have simulated the effects of realistically sized perturbations in the central metabolism of *E. coli* (Usuda et al., 2010; Kadir et al., 2010), comparing their results to experimental data.

3.1.3. Optimization problems

Even though the analysis of a specific metabolic engineering strategy is relatively easy to implement in silico given a kinetic model, there are still at least in theory infinitely many possible strategies to consider (assuming a continuum of expression levels) and it may be unclear which particular ones to try out in simulations. To overcome this difficulty, scenarios involving perturbations to model parameters are sometimes formulated as optimization problems (Hatzimanikatis et al., 1996a, 1996b; Mendes and Kell, 1998; Chang and Sahinidis, 2005; Pozo et al., 2011). A typical objective function to be optimized would be the rate of formation of the desired product and the optimization procedure may moreover be subject to constraints regarding the maximum changes in levels of enzymes and metabolites. All methods that can be used to compute the parameter estimates, described in Box 6, are typically applicable also for these problems. Using the output from kinetic models to set up optimization problems is perhaps the most rigorous and ambitious way of approaching the search for metabolic engineering targets.

Different optimization approaches to determining appropriate levels of metabolic enzymes have been used in a number of purely simulation-based studies for various aspects of microbial metabolism, including the production of ethanol in *S. cerevisiae* (Polisetty et al., 2008), citric acid in *Aspergillus niger* (Alvarez-Vasquez et al., 2000; Polisetty et al., 2008), and of serine (Visser et al., 2004; Vital-Lopez et al., 2006; Nikolaev, 2010), tryptophan (Marín-Sanguino and Torres, 2000; Schmid et al., 2004), and L-(-)-carnitine (Alvarez-

Vasquez et al., 2002) in *E. coli*. The simultaneous production of serine and tryptophan in *E. coli* has also been considered using a multi-objective optimization strategy (Lee et al., 2010).

3.2. Improving substrate utilization

An interesting prospect in the development of competitive cell factories is the expansion of their range of substrates. One example is the improved use of lignocellulosic biomass for ethanol production made possible by the introduction of genes for xylose utilization in *S. cerevisiae*. To assist in the evaluation of directed genetic engineering efforts towards improved efficiency in catabolism of this pentose sugar, Parachin et al. (2011) used a kinetic modeling approach for analyzing two different scenarios. One model was set up to represent a strain in which extracellular xylose reaches the pentose phosphate pathway through membrane transport followed by conversions by the enzymes xylose reductase, xylitol dehydrogenase, and xylulokinase. This model also featured a membrane transport reaction for the excretion of the intermediate metabolite xylitol. Additionally, another model was constructed for an alternative pathway comprising xylose isomerase and xylulokinase. In both these models the effects of a 10-fold overexpression as well as a severe knockdown (a 10-fold decrease in activity) of the different enzymes were examined. Simulating overexpression of xylose reductase in the first catabolic pathway slightly not only increased the ability to consume xylose but it also led to an increased excretion of xylitol. Conversely, not only the knockdown of xylose transport capacity decreased the xylitol formation but also the xylose consumption. Changing the activity of xylitol dehydrogenase, either by overexpression or knockdown, only had a marginal effect on model simulations. The best outcome was observed when xylulose kinase was overexpressed. This resulted in a slight increase in xylose consumption combined with a dramatic decrease in excretion of xylitol. Similarly, for the pathway using xylose isomerase, the model analysis suggested overexpression of xylulose kinase to be the best alternative for improving the utilization of xylose. The predictions of the models were in essence validated experimentally by aerobic and anaerobic cultivation of the correspondingly engineered yeast strains.

Other kinetic modeling contributions aiming for improved utilization of substrate includes enhancement of glucose uptake in *E. coli* (Visser et al., 2004; Nishio et al., 2008; Nikolaev, 2010) and application of the aforementioned MCA to the catabolism of L-arabinose (de Groot et al., 2005) and xylose (Prathumpai et al., 2003) in *Aspergillus nidulans* and *Aspergillus niger*. Of those studies, only Nishio et al. (2008) proceeded to validate their predictions in experimental follow-ups.

3.3. Improving product quality

For complex products such as glycoproteins, the quality of the product may be subject to improvement by genetic manipulations. The patterns of glycosylation have impact on in vivo activity, immunogenicity, and product half-life and their importance has encouraged the development of glycoengineered yeasts (Hamilton and Gerngross, 2007; Ye et al., 2011; Nett et al., 2012), serving as an alternative to production of human-like glycoprotein in animal cells. Kinetic modeling has been employed to describe glycosylation in mammalian cells (Krambeck and Betenbaugh, 2005; Hossler et al., 2007). Expanding on an earlier model (Umaña and Bailey, 1997), Krambeck and Betenbaugh (2005) set up a model describing the non-linear kinetics of the enzymes involved in N-linked glycosylation in CHO cells. They were able to use their model to simulate how the glycosylation profile changes when the concentration of maturing protein in the Golgi increases. Specifically, they investigated a scenario where the concentration was increased 4-fold in order to represent a hypothetical cell line with an increased specific productivity. The results of their simulations showed that the distribution of different glycoforms changes in response to the increased productivity, indicating a potential problem with reduced product quality. The authors then explored in silico the possibility of restoring the original glycosylation pattern in the high producer by means of changing the levels of glycosylation enzymes and the availability of uridine diphosphate *N*-acetylglucosamine. By just adjusting the level of a single enzyme, *N*-acetyllactoseaminide α-2,3-sialyltransferase, more than half of the deviation could be reverted. Other solutions, involving changes in several targets, that almost completely restores the glycan distribution were also proposed based on the analysis. Models such as this are clearly interesting tools for making predictions of how to preserve correct glycosylation in high producing cells, but possibly also for how to engineer new glycan patterns.

3.4. Improving process design

In addition to predicting the effects of internal perturbations to a cell factory, kinetic models are also useful for predicting their behavior in response to various external conditions. Understanding the interplay between the cell and its environment is valuable since it can be used for improving the fermentation or bioreactor process. To describe the complete production process, a model which can reproduce cellular properties such as the rate of growth, substrate consumption, and product formation, is combined with a model of the bioreactor in which the cells are cultivated. Bioreactor models are usually set up as quite simple dynamical systems based on mass balances of substrates, products, biomass, and viable cells, normally assuming ideal mixing (however, highly complex models also exists Lapin et al., 2010). Considering the bioreactor as part of the modeled system is necessary for calculating the quantities discussed in the section on model purpose, like productivity and final titer. Not only are both the time trajectories for biomass concentration and the specific rates of consumption and production need for their determination but the dynamics of these variable are usually dependent on one another (Maurer et al., 2006; Douma et al., 2010) and their dynamics must be dealt with simultaneously.

To be useful for the design or optimization of fermentation or bioreactor processes, the models of the cell metabolism need only be predictive in a input–output sense. As long as this is the case it does not matter whether they are mechanistically correct representations of intracellular biochemistry or just empirical models. Because of the challenges of setting up mechanistic models on the molecular level, the production processes have traditionally been, and commonly still are, modeled with either unstructured kinetic models or by other simplified model designs (DiMasi and Swartz, 1995). Such kinetic models have been used to describe both continuous and fed-batch cultivations. For continuous cultures, modeling has for instance been used to study growth and metabolism of mammalian cells (DiMasi and Swartz, 1995) and the effect of oxygen uptake on L-lysine production in *Corynebacterium lactofermentum* (Ensari and Lim, 2003). Models have also been used for optimizing operating conditions such as the dilution rate in order to maximize production of protein in *S. cerevisiae* (Carlsen et al., 1997) and L-(-)-carnitine in *E. coli* (Alvarez-Vasquez et al., 2002). The model-based predictions in both of these studies turned out to agree very well with experiments. In industry many processes are run in fed-batch mode and kinetic models of fed-batch processes have for example been used to study penicillin fermentation in *P. chrysogenum* (Menezes et al., 1994) and production of proteins in mammalian cell lines like baby hamster kidney (Teixeira et al., 2007), murine hybridoma (Dorka et al., 2009), and

CHO (Xing et al., 2010). Several investigators have also used similar models for the optimization of fed-batch feeding profiles (Sarkar and Modak, 2003), for instance to maximize astaxanthin production in *Xanthophyllomyces dendrorhous* (Liu and Wu, 2008) or protein production in *E. coli* (Levisauskas et al., 2003) and *P. pastoris* (Maurer et al., 2006), and for maximizing production of biomass while at the same time minimizing ethanol formation in a *S. cerevisiae* fermentation (Yüzgeç et al., 2009). The predicted optimal cultivation strategies in all of these studies were shown to compare well with validation experiments and did indeed lead to significantly improved fed batch processes. As well as using kinetic process models for optimizing operational strategies in advance, kinetic models are also potentially useful for online control (Chae et al., 2000; Teixeira et al., 2007; Yüzgeç et al., 2009).

Although models with simpler structures have proven useful in many cases, models of production processes and cell cultivation with increasing mechanistic details of intracellular reactions are now starting to appear (Bettenbrock et al., 2006; Shinto et al., 2007; Oshiro et al., 2009; Kadir et al., 2010; Li et al., 2011; Nolan and Lee, 2011). In fact, kinetic models of substantial complexity, which have been successful in describing how the metabolic state of cells varies with the external conditions, have recently been presented for both *E. coli* (Kotte et al., 2010; Usuda et al., 2010) and *S. cerevisiae* (Moisset et al., 2012). In addition to detailed representations of primary metabolic reaction networks, these models include genetic regulation of enzyme concentrations. One of the advantages of using more detailed models of the production process is that it allows the synergistic effects of metabolic engineering and process conditions to be evaluated, something which was recently explored in a mainly computational study addressing CHO cell metabolism (Nolan and Lee, 2012).

4. Advantages, challenges and perspectives

4.1. Advantages

The general strength of the kinetic modeling approach is that it quantitatively takes into account the factors that determine the rate of reactions. Compared to the modeling paradigm of the constraint-based stoichiometric models, which mainly is focusing on which reactions that can occur and the proportions of their reactants and products, kinetic models also define when and to what extent reactions take place. For an enzymatic reaction, for instance, not only can the effects of substrate and product concentrations be incorporated into the kinetic rate expression, but also the effects of co-factors, activators, inhibitors, and other modulators of enzyme activity. The ability of kinetic models to incorporate detailed information about reactions gives them a number of advantageous properties. Though, it must be emphasized that the advantages listed below partially reflect the potential capabilities of kinetics models, and not necessarily what is routinely achieved for all kinetic models.

The principles of kinetic modeling are applicable for all parts of the cell as well as the extracellular environment. Thus, a kinetic model can not only describe the rates of several interlinked enzyme-catalyzed reactions and the corresponding dynamics of the interconverted metabolites, but it may also include additional layers accounting for the rates of synthesis and degradation of transcripts and enzymes, as well as the rates of reactions involved in various sensing mechanisms and signal transduction. The many different levels of control, regulation and coordination of biochemistry are essential features of living cells (Heinemann and Sauer, 2010) and a modeling framework with a broad applicability is clearly an advantage if one desires to study the integration of different cellular processes. A kinetic model of the cell factory is furthermore easily embedded in a dynamic model of the bioreactor process itself. GEMs, on the other hand, are less flexible and work best for modeling fluxes of metabolites.

Kinetic models can assist in understanding the complex behaviors of biological systems. Although the qualitative behavior may be intuitive, such as end-product inhibition in a linear pathway, understanding both the qualitative and quantitative aspects of how system behavior emerges from the properties of its components and their interactions is generally not trivial. In fact, even really small molecular circuits with just a few components are capable of producing non-intuitive dynamic behaviors such as adaptation, homeostasis, irreversible switching and oscillation (Tyson et al., 2003). Modeling behavior like these requires a kinetic approach and is beyond the scope of GEMs. One interesting example of how kinetic modeling has provided insight into the emergence of complex behaviors is the model of metabolic adaption in *E. coli* (Kotte et al., 2010). Here, a kinetic formulation of the reactions of the central metabolism, including their transcriptional and translational regulation, was shown to be capable of reproducing system-level metabolic adjustments through a mechanism termed distributed sensing of intracellular metabolic fluxes. This can be compared to the incorporation of Boolean rules for known gene regulation in GEMs (Herrgård et al., 2006). Since the regulatory information is explicitly hard-wired into the model, this strategy can never offer the same explanatory power in terms of actual molecular mechanisms.

Kinetic modeling can turn understanding of how cell factories work into predictions about how to improve them. When models have been set up linking relevant aspects on the system-level with the properties of the system components, they become valuable for predicting and optimizing the performance of cell factories. Ideally, the model components represent things that can be manipulated such as expression levels or process parameters, but also when model components are more abstract there may be general predictions achievable that still are useful. Kinetic models are unique in that predictions and optimizations are quantitative and can be very detailed, going beyond the regime of gene addition and deletion typically identified from flux balance analysis of GEMs. Thus, if such details are desired, kinetic models are conceptually superior to GEMs which instead are better suited for pathway-oriented problems involving prediction of the steady-state flux-capabilities of metabolic networks.

4.2. Challenges

A number of challenges must be addressed and overcome for biotechnology to capitalize from the advantages of kinetic modeling. The overall challenge lies in producing predictive models of high quality that really can make a difference for improving cell factory performance. Although this review has presented a number of studies where models have been used for predicting metabolic engineering targets, some of which have been experimentally verified, we are still far away from having kinetic models that are sufficiently good to be used for in silico design of industrially competitive cell factories. In this respect, kinetic models have not reached the same degree of maturity and industrial applicability as the much more successful GEMs.

The difficulty of producing high quality predictive models is that it requires a lot of detailed information about the system that one wishes to model. If too little information is available, the strengths and advantages of the kinetic modeling approach cannot be realized. Unfortunately kinetic modeling efforts frequently suffer from incomplete and uncertain knowledge of the underlying biochemistry with respect to both network structures, kinetic

rate expressions, and parameter values (Schaber et al., 2009; Kaltenbach et al., 2009; Soh et al., 2012). Thanks to the many reconstructions of genome-scale metabolic networks, the pathways and stoichiometry of metabolic reactions are often mapped out quite well, but regulatory mechanisms, both at the level of enzyme–metabolite interaction and at the transcriptional and translational level, are usually characterized to a lesser extent. For systems involving other types of reactions, such as protein secretion networks or signal transduction, knowledge of components and interaction is usually scarce, making the formulation of the network structure a challenging task (Schaber and Klipp, 2011). Little is also known about the exact mechanisms of the majority of reactions, meaning that the structure of kinetic rate expressions is also mostly unknown. Though, if a network structure as defined here (including qualitative knowledge about reaction modifiers) can be formulated, this issue can to some degree be circumvented using approximative kinetic rate expressions. The challenge of determining suitable structures for kinetic models should not be underestimated. While the literature often emphasizes the lack of quantitative information regarding parameter values, the lack of qualitative information based on which the model structure is set up may prove to be an even more difficult problem, at least when modeling certain parts of the cellular biochemistry. Nevertheless, the limited information about parameter values also deserves a lot of attention. Already in smaller models there are typically lots of parameters with unknown values and determining them is indeed a challenge. Even in those cases where parameter values are claimed to be known as the result of studying individual components, models that agree with experimental data of system properties do not automatically follow. A well-known example illustrating this point is the study of glycolysis in *S. cerevisiae* by Teusink et al. (2000). Here, kinetic parameters were determined experimentally under standard conditions in vitro for most of the glycolytic enzymes. When the individual enzyme kinetics were pieced together to form a model of the entire pathway, the model predictions deviated substantially from the in vivo behavior in some parts. This and other examples suggest that the accumulated uncertainty introduced by in vitro measurements, differences in experimental protocols, using data from different organisms, etc., make the resulting models questionable. If possible, modelers should try to make a transition from the bottom-up philosophy of determining parameter values, and instead collectively estimate them using in vivo data with the same scope as that of the system being modeled. As shown in this review, rigorous mathematical frameworks have been established for this task and there are several available methods for solving the resulting optimization problems. There are also established methodologies within identifiability analysis and model reduction which will help in achieving well-posed estimation problems.

Producing the right kind of data is critical for parameter estimation in kinetic models. Ideally, methods from identifiability analysis and experimental design should assist in laying down the directions for what data to collect, rather than uncritically basing these decisions common practice or on intuition. Performing relevant analysis and simulations before even a single experiment has been performed can potentially save both time and resources, and lead to better models. Of course, it may then turn out that the construction of a particular model is best done with the aid of data that is currently not routinely produced. To estimate parameter in larger scale models it is for instance expected that high-throughput time-series data will be a crucial factor. Thus, kinetic modeling can act as a driver for the development of new experimental techniques as well as a better use of existing ones.

The time it takes to set up kinetic models must be reduced. Since modeling projects can be very different in their scopes and purposes it is hard to find a recipe that fits all scenarios and as a result the modeling procedure often becomes rather slow, typically involving a lot of manual work and case-to-case considerations by the modeler. One part of the solution towards a faster modeling cycle may be for the kinetic modeling community to continue to strive for a higher degree of standardization and automization. This is important not only for the representation and implementation of models, but foremost for the methods and workflows used to set them up. In this respect, valuable insights may come from looking at workflows for setting up GEMs (although the methods of course are different) where a substantial number of models have been produced in relatively short time.

If the routine generation of highly predictive kinetic models would become a reality, this will in turn pose new challenges for molecular biologists. As the predictions of metabolic engineering strategies derived from such models might be quantitatively very precise, an equally high precision in their implementation may potentially be needed to materialize the full potential of those predictions. This may require a precision in molecular biology methods that is currently not achievable, for instance such as very finely tuned expression levels or precise alteration of the catalytic properties of an engineered enzyme. In fact, the lack of such precision is already today preventing an exact implementation of the detailed results from the previously mentioned computational studies on optimal levels of metabolic enzymes.

4.3. Perspectives

It is not unrealistic to envision a future scenario where industrially relevant strategies for cell factory improvements based on classical methods gradually become exhausted or obsolete, and where the design is successively replaced by model-driven methods (Otero and Nielsen, 2010; Miskovic and Hatzimanikatis, 2010; Cvijovic et al., 2011). The most mature mathematical models of today, the GEMs, are the obvious candidates for this transition and they have already generated valuable results (Bro et al., 2006; Lee et al., 2006; Asadollahi et al., 2009; Becker et al., 2011; Neuner and Heinzle, 2011; Park et al., 2011). However, as the lower-hanging fruits of computational strategies are collected, the stoichiometric models will eventually also run into problems of predicting new targets. In this long-term perspective, kinetic models may well become a strong driving force for advancing the industrial application of cell factories. Two of the future aspects that are likely to be important for moving kinetic modeling forward are how their size and coverage can be increased and how they should deal with the previously mentioned limitations and uncertainty in the information needed to set them up.

In the future we will need to start producing large-scale kinetic models. The organization of the different biochemical reactions and pathways of the cell is characterized by a high degree of interconnectivity, for example through common precursor, energy, and redox metabolites. Because of this, changes in one part of the network of reactions may have unexpected consequences for other parts, rendering a global system perspective necessary. In the light of this complexity, one of the reasons for the successful application of stoichiometric models for predicting metabolic engineering targets is the fact that they can be set up on the genome-scale. Their aim for completeness means that they are re-usable for many kinds of problems and their popularity has even encouraged community consensus reconstructions of metabolic networks for *S. cerevisiae* (Herrgård et al., 2008). In addition to an extensive coverage of metabolism, recent work on GEMs is taking a genome-scale perspective also on the transcriptional and translational machinery (Thiele et al., 2009) as well as on protein secretion pathways (Feizi et al., 2013). The GEMs clearly have the advantage of being suitable for large scales, but they ultimately lack the

details required for a full characterization of the cell. Existing kinetic models, on the other hand, are usually set up on a small or at most medium scale. They are often built under rather specific assumptions which make them less adaptable for re-use in new situations and their lack of standard impedes the possibilities of merging smaller models into larger ones. Moving towards large-scale kinetic models, especially for the most important platform cell factories, will hopefully allow for better predictions and widen the possible model applications. Although the routine construction and use of genome-scale kinetic models definitely lies many years ahead, there are however already some emerging efforts towards the formulation of kinetic models with a more complete coverage. Various modeling methodologies have been proposed aiming to move large-scale modeling from stoichiometric constraint-based approaches to the kinetic domain (Famili et al., 2005; Smallbone et al., 2007; Jamshidi and Palsson, 2008; Ao et al., 2008; Adiamah et al., 2010; Smallbone et al., 2010), but their usefulness for cell factory improvements remains to be proven. Important results have been achieved for consistent reduction of metabolic networks (Soh, 2013), which may contribute towards genome-scale kinetic models through facilitating intermediate large-scale steppingstones (Chakrabarti et al., 2013). The idea of a community of modelers that together drives the development of large-scale kinetic models is potentially also interesting. Clearly, it is very difficult at this stage to predict which particular parts of the kinetic modeling procedure will be most crucial for eventually achieving kinetic genome-scale models. If it at all is possible given our current capabilities, it will likely involve a combination of the different topics covered within this review. In parallel with these developments, the constraint-based approaches have been modified to account for the dynamics of fluxes by sequential solutions of different steady states. These rather popular methods of so-called dynamic flux balance analysis (Mahadevan et al., 2002; Lee et al., 2008; Oddone et al., 2009) are however not addressing the kinetics of reactions, and are therefore still limited in this sense. Though, the concepts of dynamic flux balance might become useful in hybrid strategies where stoichiometric models and kinetic models are combined.

The awareness of the limited and uncertain information available for setting up kinetic models, especially large models, should lead to a critical examination of modeling strategies. One way in which modelers are facing up to the challenge posed by uncertainty is the inclusion of uncertainty itself as a part of the models. As shown in this review there is an increasing trend of publications addressing the uncertainty of both structure (Chou and Voit, 2009; Kaltenbach et al., 2009; Schaber et al., 2012), and of parameters, both directly (Pritchard and Kell, 2002; Liebermeister and Klipp, 2005; Piazza et al., 2008; Contador et al., 2009; Kotte and Heinemann, 2009; McLeod et al., 2011) and indirectly through sampling of enzyme state spaces (Wang et al., 2004; Wang and Hatzimanikatis, 2006a, 2006b; Miskovic and Hatzimanikatis, 2010, 2011). The presence of nested uncertainties of model structure and parameter values has also been emphasized (Schaber et al., 2009; Schaber and Klipp, 2011). Taking an even more fundamental approach to the uncertainty of molecular cell biology, researchers should also continue to investigate the prospect of accounting for uncertainty of parameters, reaction rates, and networks, using kinetic models based on stochastic differential equations, something which has been successful in other fields of biological modeling (Berglund et al., 2012). Related approaches can additionally be used to account for the uncertainty and variation between individual cells in a population (Almquist et al., 2008). Fueled by novel experimental techniques such as single cell metabolomics (Heinemann and Zenobi, 2011) and single cell level cultivation (Grünberger et al., 2012), cellular heterogeneity is a topic of growing interest (Lidstrom and Konopka, 2010) but its implications for cell factory design are largely uncharted territory. Based on these trends, further development of kinetic modeling strategies that can handle uncertainty will likely be one essential ingredient for generating better cell factory models in the future. As part of such efforts it should be particularly important to elucidate the impact of uncertainty on the predictive power of models Gutenkunst et al. (2007a,b).

Acknowledgments

This work has been supported by the EU Framework VII project SYSINBIO – Systems Biology as a Driver for Industrial Biotechnology (call FP7-KBBE-2007-1).

References

Adalsteinsson, D., McMillen, D., Elston, T.C., 2004. Biochemical network stochastic simulator (BioNetS): software for stochastic modeling of biochemical networks. BMC Bioinf. 5, 24, http://dx.doi.org/10.1186/1471-2105-5-24.

Adiamah, D.A., Handl, J., Schwartz, J.M., 2010. Streamlining the construction of large-scale dynamic models using generic kinetic equations. Bioinformatics 26, 1324–1331, http://dx.doi.org/10.1093/bioinformatics/btq136.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control AC-19, 716–723.

Almquist, J., Lang, P., Prätzel-Wolters, D., Deitmer, J.W., Jirstrand, M., Becker, H.M., 2010a. A kinetic model of the monocarboxylate transporter MCT1 and its interaction with carbonic anhydrase II. J. Comput. Sci. Syst. Biol. 3, 107–116.

Almquist, J., Sunnåker, M., Hagmar, J., Kvarnstrom, M., Jirstrand, M., 2008. System identification from spatiotemporal cell population data. In: The 9th International Conference on Systems Biology.

Almquist, J., Wallman, M., Jacobson, I., Jirstrand, M., 2010b. Modeling the effect of Kv1.5 block on the canine action potential. Biophys. J. 99, 2726–2736, http://dx.doi.org/10.1016/j.bpj.2010.08.062.

Alvarez-Vasquez, F., Cánovas, M., Iborra, J.L., Torres, N.V., 2002. Modeling, optimization and experimental assessment of continuous L-(-)-carnitine production by *Escherichia coli* cultures. Biotechnol. Bioeng. 80, 794–805, http://dx.doi.org/10.1002/bit.10436.

Alvarez-Vasquez, F., González-Alcón, C., Torres, N.V., 2000. Metabolism of citric acid production by *Aspergillus niger*: model definition steady-state analysis and constrained optimization of citric acid production rate. Biotechnol. Bioeng. 70, 82–108.

Alvarez-Vasquez, F., Sims, K.J., Cowart, L.A., Okamoto, Y., Voit, E.O., Hannun, Y.A., 2005. Simulation and validation of modelled sphingolipid metabolism in *Saccharomyces cerevisiae*. Nature 433, 425–430, http://dx.doi.org/10.1038/nature03232.

Alvarez-Vasquez, F., Sims, K.J., Hannun, Y.A., Voit, E.O., 2004. Integration of kinetic information on yeast sphingolipid metabolism in dynamical pathway models. J. Theor. Biol. 226, 265–291.

Anguelova, M., Karlsson, J., Jirstrand, M., 2012. Minimal output sets for identifiability. Math. Biosci. 239, 139–153, http://dx.doi.org/10.1016/j.mbs.2012.04.005.

Ao, P., Lee, L.W., Lidstrom, M.E., Yin, L., Zhu, X., 2008. Towards kinetic modeling of global metabolic networks: *Methylobacterium extorquens* AM1 growth as validation. Sheng Wu Gong Cheng Xue Bao 24, 980–994.

Asadollahi, M.A., Maury, J., Patil, K.R., Schalk, M., Clark, A., Nielsen, J., 2009. Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering. Metab. Eng. 11, 328–334, http://dx.doi.org/10.1016/j.ymben.2009.07.001.

Ashyraliyev, M., Fomekong-Nanfack, Y., Kaandorp, J.A., Blom, J.G., 2009. Systems biology: parameter estimation for biochemical models. FEBS J. 276, 886–902, http://dx.doi.org/10.1111/j.1742-4658.2008.06844.x.

Audoly, S., Bellu, G., D'Angiò, L., Saccomani, M.P., Cobelli, C., 2001. Global identifiability of nonlinear models of biological systems. IEEE Trans. Biomed. Eng. 48, 55–65, http://dx.doi.org/10.1109/10.900248.

Baker, S.M., Schallau, K., Junker, B.H., 2010. Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models. J. Integr. Bioinform. 7, http://dx.doi.org/10.2390/biecoll-jib-2010-133.

Balsa-Canto, E., Banga, J.R., 2011. AMIGO, a toolbox for advanced model identification in systems biology using global optimization. Bioinformatics 27, 2311–2313, http://dx.doi.org/10.1093/bioinformatics/btr370.

Balsa-Canto, E., Banga, J.R., Egea, J.A., Fernandez-Villaverde, A., de Hijas-Liste, G.M., 2012. Global optimization in systems biology: stochastic methods and their applications. Adv. Exp. Med. Biol. 736, 409–424, http://dx.doi.org/10.1007/978-1-4419-7210-1_24.

Balsa-Canto, E., Peifer, M., Banga, J.R., Timmer, J., Fleck, C., 2008. Hybrid optimization method with general switching strategy for parameter estimation. BMC Syst. Biol. 2, 26, http://dx.doi.org/10.1186/1752-0509-2-26.

Becker, J., Zelder, O., Häfner, S., Schroder, H., Wittmann, C., 2011. From zero to hero-design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. Metab. Eng. 13, 159–168, http://dx.doi.org/10.1016/j.ymben.2011.01.003.

Bellman, R., Åstrom, K.J., 1970. On structural identifiability. Math. Biosci. 7, 329–339.

Bellu, G., Saccomani, M.P., Audoly, S., D'Angiò, L., 2007. DAISY: a new software tool to test global identifiability of biological and physiological systems. Comput.

Methods Progr. Biomed. 88, 52–61, http://dx.doi.org/10.1016/j.cmpb.2007.07.002.
Berglund, M., Sunnåker, M., Adiels, M., Jirstrand, M., Wennberg, B., 2012. Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations. Math. Med. Biol. 29, 361–384, http://dx.doi.org/10.1093/imammb/dqr021.
Bettenbrock, K., Fischer, S., Kremling, A., Jahreis, K., Sauter, T., Gilles, E.D., 2006. A quantitative approach to catabolite repression in *Escherichia coli*. J. Biol. Chem. 281, 2578–2584, http://dx.doi.org/10.1074/jbc.M508090200.
Bohlin, T., 2006. Practical Grey-Box Process Identification: Theory and Applications. Springer-Verlag, London.
Borger, S., Liebermeister, W., Uhlendorf, J., Klipp, E., 2007. Automatically generated model of a metabolic network. Genome Inf. 18, 215–224.
Brent, R.P., 1973. Algorithms for Minimization without Derivatives. Prentice-Hall, Englewood Cliffs, NJ.
Brik Ternbach, M., Bollman, C., Wandrey, C., Takors, R., 2005. Application of model discriminating experimental design for modeling and development of a fermentative fed-batch L-valine production process. Biotechnol. Bioeng. 91, 356–368, http://dx.doi.org/10.1002/bit.20504.
Bro, C., Regenberg, B., Forster, J., Nielsen, J., 2006. In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. Metab. Eng. 8, 102–111, http://dx.doi.org/10.1016/j.ymben.2005.09.007.
Bruggeman, F.J., Boogerd, F.C., Westerhoff, H.V., 2005. The multifarious short-term regulation of ammonium assimilation of *Escherichia coli*: dissection using an *in silico* replica. FEBS J. 272, 1965–1985, http://dx.doi.org/10.1111/j.1742-4658.2005.04626.x.
Bruggeman, F.J., Westerhoff, H.V., 2007. The nature of systems biology. Trends Microbiol. 15, 45–50, http://dx.doi.org/10.1016/j.tim.2006.11.003.
Bulik, S., Grimbs, S., Huthmacher, C., Selbig, J., Holzhütter, H.G., 2009. Kinetic hybrid models composed of mechanistic and simplified enzymatic rate laws—a promising method for speeding up the kinetic modelling of complex metabolic networks. FEBS J. 276, 410–424.
Carlsen, M., Jochumsen, K.V., Emborg, C., Nielsen, J., 1997. Modeling the growth and proteinase A production in continuous cultures of recombinant *Saccharomyces cerevisiae*. Biotechnol. Bioeng. 55, 447–454 http://dx.doi.org/3.0.CO;2-C.
Carlsson, J., Nordheim, C., 2011. A Parameter Estimation Method for Continuous Time Dynamical Systems Based on the Unscented Kalman Filter and Maximum Likelihood (Master's thesis). Chalmers university of technology.
Cazzaniga, P., Pescini, D., Besozzi, D., Mauri, G., Colombo, S., Martegani, E., 2008. Modeling and stochastic simulation of the Ras/cAMP/PKA pathway in the yeast *Saccharomyces cerevisiae* evidences a key regulatory function for intracellular guanine nucleotides pools. J. Biotechnol. 133, 377–385, http://dx.doi.org/10.1016/j.jbiotec.2007.09.019.
Cedersund, G., 2006. Core-Box Modelling (Ph.D. thesis). Chalmers University of Technology.
Cedersund, G., Roll, J., 2009. Systems biology: model based evaluation and comparison of potential explanations for given biological data. FEBS J. 276, 903–922, http://dx.doi.org/10.1111/j.1742-4658.2008.06845.x.
Chae, H.J., Delisa, M.P., Cha, H.J., Weigand, W.A., Rao, G., Bentley, W.E., 2000. Framework for online optimization of recombinant protein expression in high-cell-density *Escherichia coli* cultures using GFP-fusion monitoring. Biotechnol. Bioeng. 69, 275–285.
Chakrabarti, A., Miskovic, L., Soh, K.C., Hatzimanikatis, V., 2013. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. Biotechnol. J. 8, 1043–1057, http://dx.doi.org/10.1002/biot.201300091.
Chang, Y., Sahinidis, N.V., 2005. Optimization of metabolic pathways under stability considerations. Comput. Chem. Eng. 29, 467–479.
Chassagnole, C., Fell, D.A., Raïs, B., Kudla, B., Mazat, J.P., 2001. Control of the threonine-synthesis pathway in *Escherichia coli*: a theoretical and experimental approach. Biochem. J. 356, 433–444.
Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M., 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnol. Bioeng. 79, 53–73.
Chauve, M., Mathis, H., Huc, D., Casanave, D., Monot, F., Ferreira, N.L., 2010. Comparative kinetic analysis of two fungal beta-glucosidases. Biotechnol. Biofuels 3, 3, http://dx.doi.org/10.1186/1754-6834-3-3.
Chen, N., Koumpouras, G.C., Polizzi, K.M., Kontoravdi, C., 2012. Genome-based kinetic modeling of cytosolic glucose metabolism in industrially relevant cell lines: *Saccharomyces cerevisiae* and Chinese hamster ovary cells. Bioprocess. Biosyst. Eng. 35, 1023–1033, http://dx.doi.org/10.1007/s00449-012-0687-3.
Chen, W.W., Niepel, M., Sorger, P.K., 2010. Classic and contemporary approaches to modeling biochemical reactions. Genes Dev. 24, 1861–1875, http://dx.doi.org/10.1101/gad.1945410.
Chis, O.T., Banga, J.R., Balsa-Canto, E., 2011. Structural identifiability of systems biology models: a critical comparison of methods. PLoS One 6, e27755, http://dx.doi.org/10.1371/journal.pone.0027755.
Chou, I.C., Voit, E.O., 2009. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math. Biosci. 219, 57–83, http://dx.doi.org/10.1016/j.mbs.2009.03.002.
Chung, B.K., Selvarasu, S., Andrea, C., Ryu, J., Lee, H., Ahn, J., Lee, H., Lee, D.Y., 2010. Genome-scale metabolic reconstruction and *in silico* analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. Microb. Cell Fact. 9, 50, http://dx.doi.org/10.1186/1475-2859-9-50.
Cintolesi, A., Clomburg, J.M., Rigou, V., Zygourakis, K., Gonzalez, R., 2012. Quantitative analysis of the fermentative metabolism of glycerol in *Escherichia coli*. Biotechnol. Bioeng. 109, 187–198, http://dx.doi.org/10.1002/bit.23309.
Contador, C.A., Rizk, M.L., Asenjo, J.A., Liao, J.C., 2009. Ensemble modeling for strain development of L-lysine-producing *Escherichia coli*. Metab. Eng. 11, 221–233, http://dx.doi.org/10.1016/j.ymben.2009.04.002.
Costa, R.S., Machado, D., Rocha, I., Ferreira, E.C., 2010. Hybrid dynamic modeling of *Escherichia coli* central metabolic network combining Michaelis–Menten and approximate kinetic equations. Biosystems 100, 150–157, http://dx.doi.org/10.1016/j.biosystems.2010.03.001.
Costa, R.S., Machado, D., Rocha, I., Ferreira, E.C., 2011. Critical perspective on the consequences of the limited availability of kinetic data in metabolic dynamic modelling. IET Syst. Biol. 5, 157–163, http://dx.doi.org/10.1049/iet-syb.2009.0058.
Cronwright, G.R., Rohwer, J.M., Prior, B.A., 2002. Metabolic control analysis of glycerol synthesis in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 68, 4448–4456.
Curtu, R., Diedrichs, D., 2010. Small-scale modeling approach and circuit wiring of the unfolded protein response in mammalian cells. Adv. Exp. Med. Biol. 680, 261–274, http://dx.doi.org/10.1007/978-1-4419-5913-3_30.
Cvijovic, M., Bordel, S., Nielsen, J., 2011. Mathematical models of cell factories: moving towards the core of industrial biotechnology. Microb. Biotechnol. 4, 572–584, http://dx.doi.org/10.1111/j.1751-7915.2010.00233.x.
Danø, S., Madsen, M.F., Schmidt, H., Cedersund, G., 2006. Reduction of a biochemical model with preservation of its basic dynamic properties. FEBS J. 273, 4862–-4877, http://dx.doi.org/10.1111/j.1742-4658.2006.05485.x.
de Groot, M.J.L., Prathumpai, W., Visser, J., Ruijter, G.J.G., 2005. Metabolic control analysis of *Aspergillus niger* L-arabinose catabolism. Biotechnol. Prog. 21, 1610–1616, http://dx.doi.org/10.1021/bp050189o.
Degenring, D., Froemel, C., Dikta, G., Takors, R., 2004. Sensitivity analysis for the reduction of complex metabolism models. J. Process Control 14, 729–745.
del Rosario, R.C.H., Mendoza, E., Voit, E.O., 2008. Challenges in lin-log modelling of glycolysis in *Lactococcus lactis*. IET Syst. Biol. 2, 136–149, http://dx.doi.org/10.1049/iet-syb:20070030.
DiMasi, D., Swartz, R.W., 1995. An energetically structured model of mammalian cell metabolism. 1. Model development and application to steady-state hybridoma cell growth in continuous culture. Biotechnol. Prog. 11, 664–676, http://dx.doi.org/10.1021/bp00036a010.
Dorka, P., Fischer, C., Budman, H., Scharer, J.M., 2009. Metabolic flux-based modeling of mAb production during batch and fed-batch operations. Bioprocess Biosyst. Eng. 32, 183–196, http://dx.doi.org/10.1007/s00449-008-0236-2.
Douma, R.D., Verheijen, P.J.T., de Laat, W.T.A.M., Heijnen, J.J., van Gulik, W.M., 2010. Dynamic gene expression regulation model for growth and penicillin production in *Penicillium chrysogenum*. Biotechnol. Bioeng. 106, 608–618, http://dx.doi.org/10.1002/bit.22689.
Dräger, A., Kronfeld, M., Ziller, M.J., Supper, J., Planatscher, H., Magnus, J.B., Oldiges, M., Kohlbacher, O., Zell, A., 2009. Modeling metabolic networks in *C. Glutamicum*: a comparison of rate laws in combination with various parameter optimization strategies. BMC Syst. Biol. 3, 5, http://dx.doi.org/10.1186/1752-0509-3-5.
Du, J., Shao, Z., Zhao, H., 2011. Engineering microbial factories for synthesis of value-added products. J. Ind. Microbiol. Biotechnol. 38, 873–890, http://dx.doi.org/10.1007/s10295-011-0970-3.
Egea, J.A., Martí, R., Banga, J.R., 2010. An evolutionary method for complex-process optimization. Comput. Oper. Res. 37, 315–324.
Ensari, S., Lim, H.C., 2003. Kinetics of L-lysine fermentation: a continuous culture model incorporating oxygen uptake rate. Appl. Microbiol. Biotechnol. 62, 35–40, http://dx.doi.org/10.1007/s00253-003-1266-y.
Erguler, K., Stumpf, M.P.H., 2011. Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. Mol. Biosyst. 7, 1593–1602, http://dx.doi.org/10.1039/c0mb00107d.
van Eunen, K., Bouwman, J., Daran-Lapujade, P., Postmus, J., Canelas, A.B., Mensonides, F.I.C., Orij, R., 2010. Measuring enzyme activities under standardized in vivo-like conditions for systems biology. FEBS J. 277, 749–760, http://dx.doi.org/10.1111/j.1742-4658.2009.07524.x.
Famili, I., Mahadevan, R., Palsson, B.O., 2005. *k*-Cone analysis: determining all candidate values for kinetic parameters on a network scale. Biophys. J. 88, 1616–1625, http://dx.doi.org/10.1529/biophysj.104.050385.
Feizi, A., Österlund, T., Petranovic, D., Bordel, S., Nielsen, J., 2013. Genome-scale modeling of the protein secretory machinery in yeast. PLoS One 8, e63284, http://dx.doi.org/10.1371/journal.pone.0063284.
Fell, D.A., 1992. Metabolic control analysis: a survey of its theoretical and experimental development. Biochem. J. 286 (Pt 2), 313–330.
Ghosh, S., Banerjee, S., Bose, I., 2012. Emergent bistability: effects of additive and multiplicative noise. Eur. Phys. J. E: Soft Matter 35, 11, http://dx.doi.org/10.1140/epje/i2012-12011-4.
Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical-reactions. J. Comput. Phys. 22, 403–434.
Gillespie, D.T., 2000. The chemical Langevin equation. J. Chem. Phys. 113, 297–306.
Gillespie, D.T., 2001. Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. 115, 1716–1733.
Gonzalez, R., Andrews, B.A., Asenjo, J.A., 2001. Metabolic control analysis of monoclonal antibody synthesis. Biotechnol. Prog. 17, 217–226, http://dx.doi.org/10.1021/bp000165b.

Gonze, D., Jacquet, M., Goldbeter, A., 2008. Stochastic modelling of nucleocytoplasmic oscillations of the transcription factor msn2 in yeast. J. R. Soc. Interface 5 (Suppl. 1), S95–109, http://dx.doi.org/10.1098/rsif.2008.0141.focus.
Goryanin, I., Demin, O., 2009. Kinetic Modelling in Systems Biology. Chapman and Hall/CRC, Boca Raton, FL.
Grünberger, A., Paczia, N., Probst, C., Schendzielorz, G., Eggeling, L., Noack, S., Wiechert, W., Kohlheyer, D., 2012. A disposable picolitre bioreactor for cultivation and investigation of industrially relevant bacteria on the single cell level. Lab Chip 12, 2060–2068, http://dx.doi.org/10.1039/c2lc40156h.
Gutenkunst, R.N., Casey, F.P., Waterfall, J.J., Myers, C.R., Sethna, J.P., 2007a. Extracting falsifiable predictions from sloppy models. Ann. N.Y. Acad. Sci. 1115, 203–211, http://dx.doi.org/10.1196/annals.1407.003.
Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., Sethna, J.P., 2007b. Universally sloppy parameter sensitivities in systems biology models. PLoS Comput. Biol. 3, 1871–1878, http://dx.doi.org/10.1371/journal.pcbi.0030189.
Haag, J.E., Vande Wouwer, A., Bogaerts, P., 2005. Dynamic modeling of complex biological systems: a link between metabolic and macroscopic description. Math. Biosci. 193, 25–49, http://dx.doi.org/10.1016/j.mbs.2004.11.007.
Hamilton, S.R., Gerngross, T.U., 2007. Glycosylation engineering in yeast: the advent of fully humanized yeast. Curr. Opin. Biotechnol. 18, 387–392, http://dx.doi.org/10.1016/j.copbio.2007.09.001.
Hasty, J., Pradines, J., Dolnik, M., Collins, J.J., 2000. Noise-based switches and amplifiers for gene expression. Proc. Natl. Acad. Sci. U.S.A. 97, 2075–2080, http://dx.doi.org/10.1073/pnas.040411297.
Hattersley, J.G., Pérez-Velázquez, J., Chappell, M.J., Bearup, D., Roper, D., Dowson, C., Bugg, T., Evans, N.D., 2011. Indistinguishability and identifiability of kinetic models for the MurC reaction in peptidoglycan biosynthesis. Comput. Methods Progr. Biomed. 104, 70–80, http://dx.doi.org/10.1016/j.cmpb.2010.07.009.
Hatzimanikatis, V., Bailey, J.E., 1996. MCA has more to say. J. Theor. Biol. 182, 233–242, http://dx.doi.org/10.1006/jtbi.1996.0160.
Hatzimanikatis, V., Bailey, J.E., 1997. Effects of spatiotemporal variations on metabolic control: approximate analysis using (log)linear kinetic models. Biotechnol. Bioeng. 54, 91–104 http://dx.doi.org/3.0.CO;2-Q.
Hatzimanikatis, V., Floudas, C.A., Bailey, J.E., 1996a. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. AIChE J. 42, 1277–1292.
Hatzimanikatis, V., Floudas, C.A., Bailey, J.E., 1996b. Optimization of regulatory architectures in metabolic reaction networks. Biotechnol. Bioeng. 52, 485–500 http://dx.doi.org/3.0.CO;2-L.
Haunschild, M.D., Freisleben, B., Takors, R., Wiechert, W., 2005. Investigating the dynamic behavior of biochemical networks using model families. Bioinformatics 21, 1617–1625, http://dx.doi.org/10.1093/bioinformatics/bti225.
Heijnen, J.J., 2005. Approximative kinetic formats used in metabolic network modeling. Biotechnol. Bioeng. 91, 534–545, http://dx.doi.org/10.1002/bit.20558.
Heinemann, M., Sauer, U., 2010. Systems biology of microbial metabolism. Curr. Opin. Microbiol. 13, 337–343, http://dx.doi.org/10.1016/j.mib.2010.02.005.
Heinemann, M., Zenobi, R., 2011. Single cell metabolomics. Curr. Opin. Biotechnol. 22, 26–31, http://dx.doi.org/10.1016/j.copbio.2010.09.008.
Heinrich, R., Rapoport, T.A., 1974. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. Eur. J. Biochem. 42, 89–95.
Heinzle, E., Dunn, I.J., Furukawa, K., Tanner, R.D., 1982. Modelling of sustained oscillations observed in continuous culture of *Saccharomyces cerevisiae*. In: Modelling and Control of Biotechnical Processes.
Heinzle, E., Matsuda, F., Miyagawa, H., Wakasa, K., Nishioka, T., 2007. Estimation of metabolic fluxes, expression levels and metabolite dynamics of a secondary metabolic pathway in potato using label pulse-feeding experiments combined with kinetic network modelling and simulation. Plant J. 50, 176–187, http://dx.doi.org/10.1111/j.1365-313X.2007.03037.x.
Herrgård, M.J., Lee, B.S., Portnoy, V., Palsson, B.Ø., 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. Genome Res. 16, 627–635, http://dx.doi.org/10.1101/gr.4083206.
Herrgård, M.J., Swainston, N., Dobson, P., Dunn, W.B., Arga, K.Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M.L., Oliveira, A.P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasić, I., Weichart, D., Brent, R., Broomhead, D.S., Westerhoff, H.V., Kirdar, B., Penttilä, M., Klipp, E., 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat. Biotechnol. 26, 1155–1160, http://dx.doi.org/10.1038/nbt1492.
Higham, D.J., 2001. An algorithmic introduction to numerical simulation of stochastic differential equations. SIAM Rev. 43, 525–546.
Hildebrandt, S., Raden, D., Petzold, L., Robinson, A.S., Doyle, F.J., 2008. A top-down approach to mechanistic biological modeling: application to the single-chain antibody folding pathway. Biophys. J. 95, 3535–3558, http://dx.doi.org/10.1529/biophysj.107.125039.
Hoefnagel, M.H.N., Starrenburg, M.J.C., Martens, D.E., Hugenholtz, J., Kleerebezem, M., Van Swam, I.I., Bongers, R., Westerhoff, H.V., Snoep, J.L., 2002. Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. Microbiology 148, 1003–1013.
Hooke, R., Jeeves, T.A., 1961. Direct search solution of numerical and statistical problems. J. Assoc. Comput. Mach. 8, 212–229.
Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U., 2006. COPASI—a COmplex PAthway SImulator. Bioinformatics 22, 3067–3074, http://dx.doi.org/10.1093/bioinformatics/btl485.
Hossler, P., Mulukutla, B.C., Hu, W.S., 2007. Systems analysis of N-glycan processing in mammalian cells. PLoS One 2, e713, http://dx.doi.org/10.1371/journal.pone.0000713.
Hua, Q., Yang, C., Shimizu, K., 2000. Metabolic control analysis for lysine synthesis using *Corynebacterium glutamicum* and experimental verification. J. Biosci. Bioeng. 90, 184–192.
Huai, L., Chen, N., Yang, W., Bai, G., 2009. Metabolic control analysis of L-cysteine producing strain TS1138 of *Pseudomonas* sp. Biochemistry 74, 288–292.
Hynne, F., Danø, S., Sørensen, P.G., 2001. Full-scale model of glycolysis in *Saccharomyces cerevisiae*. Biophys. Chem. 94, 121–163.
Jamshidi, N., Palsson, B.Ø., 2008. Formulating genome-scale kinetic models in the post-genome era. Mol. Syst. Biol. 4, 171, http://dx.doi.org/10.1038/msb.2008.8.
Jansson, A., Jirstrand, M., 2010. Biochemical modeling with systems biology graphical notation. Drug Discov. Today 15, 365–370, http://dx.doi.org/10.1016/j.drudis.2010.02.012.
Jaqaman, K., Danuser, G., 2006. Linking data to models: data regression. Nat. Rev. Mol. Cell Biol. 7, 813–819, http://dx.doi.org/10.1038/nrm2030.
Jazwinski, A.H., 1970. Stochastic Processes and Filtering Theory. Academic Press, New York.
Kacser, H., Burns, J.A., 1973. The control of flux. Symp. Soc. Exp. Biol. 27, 65–104.
Kadir, T.A.A., Mannan, A.A., Kierzek, A.M., McFadden, J., Shimizu, K., 2010. Modeling and simulation of the main metabolism in *Escherichia coli* and its several single-gene knockout mutants with experimental verification. Microb. Cell Fact. 9, 88, http://dx.doi.org/10.1186/1475-2859-9-88.
Kaltenbach, H.M., Dimopoulos, S., Stelling, J., 2009. Systems analysis of cellular networks under uncertainty. FEBS Lett. 583, 3923–3930, http://dx.doi.org/10.1016/j.febslet.2009.10.074.
van Kampen, N.G., 2007. Stochastic Processes in Physics and Chemistry. North Holland, Amsterdam.
Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30.
Karlsson, J., Anguelova, M., Jirstrand, M., 2012. An efficient method for structural identifiability analysis of large dynamic systems. In: 16th IFAC Symposium on System Identification.
Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Networks.
Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220, 671–680, http://dx.doi.org/10.1126/science.220.4598.671.
Klipp, E., Herwig, R., Kowald, A., Wierling, C., Lehrach, H., 2005a. Systems Biology in Practice. Wiley-VCH, Berlin.
Klipp, E., Nordlander, B., Krüger, R., Gennemark, P., Hohmann, S., 2005b. Integrative model of the response of yeast to osmotic shock. Nat. Biotechnol. 23, 975–982, http://dx.doi.org/10.1038/nbt1114.
Kotte, O., Heinemann, M., 2009. A divide-and-conquer approach to analyze underdetermined biochemical models. Bioinformatics 25, 519–525, http://dx.doi.org/10.1093/bioinformatics/btp004.
Kotte, O., Zaugg, J.B., Heinemann, M., 2010. Bacterial adaptation through distributed sensing of metabolic fluxes. Mol. Syst. Biol. 6, 355, http://dx.doi.org/10.1038/msb.2010.10.
Krambeck, F.J., Betenbaugh, M.J., 2005. A mathematical model of n-linked glycosylation. Biotechnol. Bioeng. 92, 711–728, http://dx.doi.org/10.1002/bit.20645.
Kreutz, C., Timmer, J., 2009. Systems biology: experimental design. FEBS J. 276, 923–942, http://dx.doi.org/10.1111/j.1742-4658.2008.06843.x.
Kuepfer, L., Peter, M., Sauer, U., Stelling, J., 2007. Ensemble modeling for analysis of cell signaling dynamics. Nat. Biotechnol. 25, 1001–1006, http://dx.doi.org/10.1038/nbt1330.
Lang, M., Waldherr, S., Allgower, F., 2009. Amplitude distribution of stochastic oscillations in biochemical networks due to intrinsic noise. PMC Biophys. 2, 10, http://dx.doi.org/10.1186/1757-5036-2-10.
Lapin, A., Klann, M., Reuss, M., 2010. Multi-scale spatio-temporal modeling: lifelines of microorganisms in bioreactors and tracking molecules in cells. Adv. Biochem. Eng. Biotechnol. 121, 23–43, http://dx.doi.org/10.1007/10_2009_53.
Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M.I., Wimalaratne, S.M., Bergman, F.T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y., Villéger, A., Boyd, S.E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, T.C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D.B., Sander, C., Sauro, H., Snoep, J.L., Kohn, K., Kitano, H., 2009. The systems biology graphical notation. Nat. Biotechnol. 27, 735–741, http://dx.doi.org/10.1038/nbt.1558.
Lee, F.C., Pandu Rangaiah, G., Lee, D.Y., 2010. Modeling and optimization of a multi-product biosynthesis factory for multiple objectives. Metab. Eng. 12, 251–267, http://dx.doi.org/10.1016/j.ymben.2009.12.003.
Lee, J.M., Lee, J.M., Gianchandani, E.P., Eddy, J.A., Papin, J.A., 2008. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. PLoS Comput. Biol. 4, e1000086, http://dx.doi.org/10.1371/journal.pcbi.1000086.
Lee, S.J., Song, H., Lee, S.Y., 2006. Genome-based metabolic engineering of mannheimia succiniciproducens for succinic acid production. Appl. Environ. Microbiol. 72, 1939–1948, http://dx.doi.org/10.1128/AEM.72.3.1939-1948.2006.
Levisauskas, D., Galvanauskas, V., Henrich, S., Wilhelm, K., Volk, N., Lübbert, A., 2003. Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*. Bioprocess Biosyst. Eng. 25, 255–262, http://dx.doi.org/10.1007/s00449-002-0305-x.
Li, R.D., Li, Y.Y., Lu, L.Y., Ren, C., Li, Y.X., Liu, L., 2011. An improved kinetic model for the acetone–butanol–ethanol pathway of *Clostridium acetobutylicum* and

model-based perturbation analysis. BMC Syst. Biol. 5 (Suppl. 1), S12, http://dx.doi.org/10.1186/1752-0509-5-S1-S12.
Lidstrom, M.E., Konopka, M.C., 2010. The role of physiological heterogeneity in microbial population behavior. Nat. Chem. Biol. 6, 705–712, http://dx.doi.org/10.1038/nchembio.436.
Liebermeister, W., Baur, U., Klipp, E., 2005. Biochemical network models simplified by balanced truncation. FEBS J. 272, 4034–4043, http://dx.doi.org/10.1111/j.1742-4658.2005.04780.x.
Liebermeister, W., Klipp, E., 2005. Biochemical networks with uncertain parameters. Syst. Biol. (Stevenage) 152, 97–107.
Liebermeister, W., Klipp, E., 2006a. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. Theor. Biol. Med. Model. 3, 41, http://dx.doi.org/10.1186/1742-4682-3-41.
Liebermeister, W., Klipp, E., 2006b. Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data. Theor. Biol. Med. Model. 3, 42, http://dx.doi.org/10.1186/1742-4682-3-42.
Liebermeister, W., Uhlendorf, J., Klipp, E., 2010. Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. Bioinformatics 26, 1528–1534, http://dx.doi.org/10.1093/bioinformatics/btq141.
Liu, Y.S., Wu, J.Y., 2008. Modeling of *Xanthophyllomyces dendrorhous* growth on glucose and overflow metabolism in batch and fed-batch cultures for astaxanthin production. Biotechnol. Bioeng. 101, 996–1004, http://dx.doi.org/10.1002/bit.21978.
Ljung, L., 1987. System Identification: Theory for the User. Prentice-Hall, Englewood Cliffs, NJ.
Ljung, L., Glad, T., 1994a. Modeling of Dynamic Systems. Prentice-Hall, Englewood Cliffs, NJ.
Ljung, L., Glad, T., 1994b. On global identifiability for arbitrary model parametrizations. Automatica 30, 265–276.
Love, K.R., Panagiotou, V., Jiang, B., Stadheim, T.A., Love, J.C., 2010. Integrated single-cell analysis shows *Pichia pastoris* secretes protein stochastically. Biotechnol. Bioeng. 106, 319–325, http://dx.doi.org/10.1002/bit.22688.
Magnus, J.B., Oldiges, M., Takors, R., 2009. The identification of enzyme targets for the optimization of a valine producing *Corynebacterium glutamicum* strain using a kinetic model. Biotechnol. Prog. 25, 754–762, http://dx.doi.org/10.1002/btpr.184.
Mahadevan, R., Edwards, J.S., Doyle, F.J., 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys. J. 83, 1331–1340, http://dx.doi.org/10.1016/S0006-3495(02)73903-9.
Maiwald, T., Timmer, J., 2008. Dynamical modeling and multi-experiment fitting with PottersWheel. Bioinformatics 24, 2037–2043, http://dx.doi.org/10.1093/bioinformatics/btn350.
Margaria, G., Riccomagno, E., Chappell, M.J., Wynn, H.P., 2001. Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences. Math. Biosci. 174, 1–26.
Marín-Sanguino, A., Torres, N.V., 2000. Optimization of tryptophan production in bacteria. Design of a strategy for genetic manipulation of the tryptophan operon for tryptophan flux maximization. Biotechnol. Prog. 16, 133–145, http://dx.doi.org/10.1021/bp990144l.
Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. 11, 431–441.
Maurer, M., Kühleitner, M., Gasser, B., Mattanovich, D., 2006. Versatile modeling and optimization of fed batch processes for the production of secreted heterologous proteins with *Pichia pastoris*. Microb. Cell Fact. 5, 37, http://dx.doi.org/10.1186/1475-2859-5-37.
McLeod, J., O'Callaghan, P.M., Pybus, L.P., Wilkinson, S.J., Root, T., Racher, A.J., James, D.C., 2011. An empirical modeling platform to evaluate the relative control discrete CHO cell synthetic processes exert over recombinant monoclonal antibody production process titer. Biotechnol. Bioeng. 108, 2193–2204, http://dx.doi.org/10.1002/bit.23146.
Mendes, P., Hoops, S., Sahle, S., Gauges, R., Dada, J., Kummer, U., 2009. Computational modeling of biochemical networks using COPASI. Methods Mol. Biol. 500, 17–59, http://dx.doi.org/10.1007/978-1-59745-525-1_2.
Mendes, P., Kell, D., 1998. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics 14, 869–883.
Menezes, J.C., Alves, S.S., Lemos, J.M., de Azevedo, S.F., 1994. Mathematical modelling of industrial pilot-plant penicillin-G fed-batch fermentations. J. Chem. Technol. Biotechnol. 61, 123–138, http://dx.doi.org/10.1002/jctb.280610207.
Meshkat, N., Anderson, C., Distefano 3rd, J.J., 2011. Finding identifiable parameter combinations in nonlinear ODE models and the rational reparameterization of their input–output equations. Math. Biosci. 233, 19–31, http://dx.doi.org/10.1016/j.mbs.2011.06.001.
Mettetal, J.T., Muzzey, D., Gómez-Uribe, C., van Oudenaarden, A., 2008. The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*. Science 319, 482–484, http://dx.doi.org/10.1126/science.1151582.
Minton, A.P., 2001. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. J. Biol. Chem. 276, 10577–10580, http://dx.doi.org/10.1074/jbc.R100005200.
Minton, A.P., 2006. How can biochemical reactions within cells differ from those in test tubes? J. Cell Sci. 119, 2863–2869, http://dx.doi.org/10.1242/jcs.03063.
Miskovic, L., Hatzimanikatis, V., 2010. Production of biofuels and biochemicals: in need of an ORACLE. Trends Biotechnol. 28, 391–397, http://dx.doi.org/10.1016/j.tibtech.2010.05.003.
Miskovic, L., Hatzimanikatis, V., 2011. Modeling of uncertainties in biochemical reactions. Biotechnol. Bioeng. 108, 413–423, http://dx.doi.org/10.1002/bit.22932.
Moisset, P., Vaisman, D., Cintolesi, A., Urrutia, J., Rapaport, I., Andrews, B.A., Asenjo, J.A., 2012. Continuous modeling of metabolic networks with gene regulation in yeast and in vivo determination of rate parameters. Biotechnol. Bioeng. 109, 2325–2339, http://dx.doi.org/10.1002/bit.24503.
Moles, C.G., Mendes, P., Banga, J.R., 2003. Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res. 13, 2467–2474, http://dx.doi.org/10.1101/gr.1262503.
Molinaro, A.M., Simon, R., Pfeiffer, R.M., 2005. Prediction error estimation: a comparison of resampling methods. Bioinformatics 21, 3301–3307, http://dx.doi.org/10.1093/bioinformatics/bti499.
Nelder, J., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7, 308–313.
Nett, J.H., Gomathinayagam, S., Hamilton, S.R., Gong, B., Davidson, R.C., Du, M., Hopkins, D., Mitchell, T., Mallem, M.R., Nylen, A., Shaikh, S.S., Sharkey, N., Barnard, G.C., Copeland, V., Liu, L., Evers, R., Li, Y., Gray, P.M., Lingham, R.B., Visco, D., Forrest, G., DeMartino, J., Linden, T., Potgieter, T.I., Wildt, S., Stadheim, T.A., d'Anjou, M., Li, H., Sethuraman, N., 2012. Optimization of erythropoietin production with controlled glycosylation-PEGylated erythropoietin produced in glycoengineered *Pichia pastoris*. J. Biotechnol. 157, 198–206, http://dx.doi.org/10.1016/j.jbiotec.2011.11.002.
Neuner, A., Heinzle, E., 2011. Mixed glucose and lactate uptake by *Corynebacterium glutamicum* through metabolic engineering. Biotechnol. J. 6, 318–329, http://dx.doi.org/10.1002/biot.201000307.
Nielsen, J., 1998. Metabolic engineering: techniques for analysis of targets for genetic manipulations. Biotechnol. Bioeng. 58, 125–132.
Nielsen, J., Jewett, M.C., 2008. Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. FEMS Yeast Res. 8, 122–131, http://dx.doi.org/10.1111/j.1567-1364.2007.00302.x.
Nikerel, I.E., van Winden, W.A., van Gulik, W.M., Heijnen, J.J., 2006. A method for estimation of elasticities in metabolic networks using steady state and dynamic metabolomics data and linlog kinetics. BMC Bioinf. 7, 540, http://dx.doi.org/10.1186/1471-2105-7-540.
Nikerel, I.E., van Winden, W.A., Verheijen, P.J.T., Heijnen, J.J., 2009. Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics. Metab. Eng. 11, 20–30, http://dx.doi.org/10.1016/j.ymben.2008.07.004.
Nikolaev, E.V., 2010. The elucidation of metabolic pathways and their improvements using stable optimization of large-scale kinetic models of cellular systems. Metab. Eng. 12, 26–38, http://dx.doi.org/10.1016/j.ymben.2009.08.010.
Nishio, Y., Usuda, Y., Matsui, K., Kurata, H., 2008. Computer-aided rational design of the phosphotransferase system for enhanced glucose uptake in *Escherichia coli*. Mol. Syst. Biol. 4, 160, http://dx.doi.org/10.1038/msb4100201.
Noack, S., Wahl, A., Qeli, E., Wiechert, W., 2007. Visualizing regulatory interactions in metabolic networks. BMC Biol. 5, 46, http://dx.doi.org/10.1186/1741-7007-5-46.
Nocedal, J., Wright, S.J., 1999. Numerical Optimization. Springer-Verlag, New York, NY.
Nolan, R.P., Lee, K., 2011. Dynamic model of CHO cell metabolism. Metab. Eng. 13, 108–124, http://dx.doi.org/10.1016/j.ymben.2010.09.003.
Nolan, R.P., Lee, K., 2012. Dynamic model for CHO cell engineering. J. Biotechnol. 158, 24–33, http://dx.doi.org/10.1016/j.jbiotec.2012.01.009.
Österlund, T., Nookaew, I., Nielsen, J., 2012. Fifteen years of large scale metabolic modeling of yeast: developments and impacts. Biotechnol. Adv. 30, 979–988, http://dx.doi.org/10.1016/j.biotechadv.2011.07.021.
Oddone, G.M., Mills, D.A., Block, D.E., 2009. A dynamic, genome-scale flux model of *Lactococcus lactis* to increase specific recombinant protein expression. Metab. Eng. 11, 367–381, http://dx.doi.org/10.1016/j.ymben.2009.07.007.
Oh, E., Lu, M., Park, C., Park, C., Oh, H.B., Lee, S.Y., Lee, J., 2011. Dynamic modeling of lactic acid fermentation metabolism with *Lactococcus lactis*. J. Microbiol. Biotechnol. 21, 162–169.
Okino, M.S., Mavrovouniotis, M.L., 1998. Simplification of mathematical models of chemical reaction systems. Chem. Rev. 98, 391–408.
Oldiges, M., Noack, S., Wahl, A., Qeli, E., Freisleben, B., Wiechert, W., 2006. From enzyme kinetics to metabolic network modeling—visualization tool for enhanced kinetic analysis of biochemical network models. Eng. Life Sci. 6, 155–162.
Oshiro, M., Shinto, H., Tashiro, Y., Miwa, N., Sekiguchi, T., Okamoto, M., Ishizaki, A., Sonomoto, K., 2009. Kinetic modeling and sensitivity analysis of xylose metabolism in *Lactococcus lactis* IO-1. J. Biosci. Bioeng. 108, 376–384, http://dx.doi.org/10.1016/j.jbiosc.2009.05.003.
Otero, J.M., Nielsen, J., 2010. Industrial systems biology. Biotechnol. Bioeng. 105, 439–460, http://dx.doi.org/10.1002/bit.22592.
Parachin, N.S., Bergdahl, B., van Niel, E.W.J., Gorwa-Grauslund, M.F., 2011. Kinetic modelling reveals current limitations in the production of ethanol from xylose by recombinant *Saccharomyces cerevisiae*. Metab. Eng. 13, 508–517, http://dx.doi.org/10.1016/j.ymben.2011.05.005.
Park, J.H., Kim, T.Y., Lee, K.H., Lee, S.Y., 2011. Fed-batch culture of *Escherichia coli* for L-valine production based on in silico flux response analysis. Biotechnol. Bioeng. 108, 934–946, http://dx.doi.org/10.1002/bit.22995.
Patil, K.R., Akesson, M., Nielsen, J., 2004. Use of genome-scale microbial models for metabolic engineering. Curr. Opin. Biotechnol. 15, 64–69, http://dx.doi.org/10.1016/j.copbio.2003.11.003.
Paulsson, J., 2004. Summing up the noise in gene networks. Nature 427, 415–418, http://dx.doi.org/10.1038/nature02257.

Pfeffer, M., Maurer, M., Kollensperger, G., Hann, S., Graf, A.B., Mattanovich, D., 2011. Modeling and measuring intracellular fluxes of secreted recombinant protein in *Pichia pastoris* with a novel 34S labeling procedure. Microb. Cell Fact. 10, 47, http://dx.doi.org/10.1186/1475-2859-10-47.

Piazza, M., Feng, X.J., Rabinowitz, J.D., Rabitz, H., 2008. Diverse metabolic model parameters generate similar methionine cycle dynamics. J. Theor. Biol. 251, 628–639, http://dx.doi.org/10.1016/j.jtbi.2007.12.009.

Pohjanpalo, H., 1978. System identifiability based on the power series expansion of the solution. Math. Biosci. 41, 21–33.

Polisetty, P.K., Gatzke, E.P., Voit, E.O., 2008. Yield optimization of regulated metabolic systems using deterministic branch-and-reduce methods. Biotechnol. Bioeng. 99, 1154–1169, http://dx.doi.org/10.1002/bit.21679.

Pozo, C., Marín-Sanguino, A., Alves, R., Guillén-Gosálbez, G., Jiménez, L., Sorribas, A., 2011. Steady-state global optimization of metabolic non-linear dynamic models through recasting into power-law canonical models. BMC Syst. Biol. 5, 137, http://dx.doi.org/10.1186/1752-0509-5-137.

Prathumpai, W., Gabelgaard, J.B., Wanchanthuek, P., van de Vondervoort, P.J.I., de Groot, M.J.L., McIntyre, M., Nielsen, J., 2003. Metabolic control analysis of xylose catabolism in *Aspergillus*. Biotechnol. Prog. 19, 1136–1141, http://dx.doi.org/10.1021/bp034020r.

Pritchard, L., Kell, D.B., 2002. Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis. Eur. J. Biochem. 269, 3894–3904.

Provost, A., Bastin, G., 2004. Dynamic metabolic modelling under the balanced growth condition. J. Process. Control 14, 717–728.

Provost, A., Bastin, G., Agathos, S.N., Schneider, Y.J., 2006. Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells. Bioprocess. Biosyst. Eng. 29, 349–366, http://dx.doi.org/10.1007/s00449-006-0083-y.

Portner, R., Schäfer, T., 1996. Modelling hybridoma cell growth and metabolism—a comparison of selected models and data. J. Biotechnol. 49, 119–135.

Ramsey, S.A., Smith, J.J., Orrell, D., Marelli, M., Petersen, T.W., de Atauri, P., Bolouri, H., Aitchison, J.D., 2006. Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. Nat. Genet. 38, 1082–1087, http://dx.doi.org/10.1038/ng1869.

Raue, A., Becker, V., Klingmüller, U., Timmer, J., 2010. Identifiability and observability analysis for experimental design in nonlinear dynamical models. Chaos 20, 045105, http://dx.doi.org/10.1063/1.3528102.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 1923–1929, http://dx.doi.org/10.1093/bioinformatics/btp358.

Raue, A., Kreutz, C., Maiwald, T., Klingmuller, U., Timmer, J., 2011. Addressing parameter identifiability by model-based experimentation. IET Syst. Biol. 5, 120–130, http://dx.doi.org/10.1049/iet-syb.2010.0061.

Raue, A., Kreutz, C., Theis, F.J., Timmer, J., 2013. Joining forces of bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. Philos. Trans. A. Math. Phys. Eng. Sci. 371, 20110544, http://dx.doi.org/10.1098/rsta.2011.0544.

Rodriguez-Fernandez, M., Egea, J.A., Banga, J.R., 2006a. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. BMC Bioinf. 7, 483, http://dx.doi.org/10.1186/1471-2105-7-483.

Rodriguez-Fernandez, M., Mendes, P., Banga, J.R., 2006b. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. Biosystems 83, 248–265, http://dx.doi.org/10.1016/j.biosystems.2005.06.016.

Rojas, I., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Wittig, U., 2007. Storing and annotating of kinetic data. In Silico Biol. 7, S37–S44.

Roper, R.T., Saccomani, M.P., Vicini, P., 2010. Cellular signaling identifiability analysis: a case study. J. Theor. Biol. 264, 528–537, http://dx.doi.org/10.1016/j.jtbi.2010.02.029.

Runarsson, T.P., Yao, X., 2000. Stochastic ranking for constrained evolutionary optimization. IEEE Trans. Evol. Comput. 4, 284–294.

Runarsson, T.P., Yao, X., 2005. Search biases in constrained evolutionary optimization. IEEE Trans. Syst. Man Cybern.—Part C 35, 233–243.

Saccomani, M.P., Audoly, S., Bellu, G., D'Angiò, L., 2010. Examples of testing global identifiability of biological and biomedical models with the DAISY software. Comput. Biol. Med. 40, 402–407, http://dx.doi.org/10.1016/j.compbiomed.2010.02.004.

Sarkar, D., Modak, J.M., 2003. Optimisation of fed-batch bioreactors using genetic algorithms. Chem. Eng. Sci. 58, 2283–2296.

Sauer, M., Mattanovich, D., 2012. Construction of microbial cell factories for industrial bioprocesses. J. Chem. Technol. Biotechnol. 87, 445–450.

Savageau, M.A., 1976. Biochemical Systems Analysis. Addison-Wesley Pub. Co., Reading, MA.

Schaber, J., Baltanas, R., Bush, A., Klipp, E., Colman-Lerner, A., 2012. Modelling reveals novel roles of two parallel signalling pathways and homeostatic feedbacks in yeast. Mol. Syst. Biol. 8, 622, http://dx.doi.org/10.1038/msb.2012.53.

Schaber, J., Klipp, E., 2011. Model-based inference of biochemical parameters and dynamic properties of microbial signal transduction networks. Curr. Opin. Biotechnol. 22, 109–116, http://dx.doi.org/10.1016/j.copbio.2010.09.014.

Schaber, J., Liebermeister, W., Klipp, E., 2009. Nested uncertainties in biochemical models. IET Syst. Biol. 3, 1–9, http://dx.doi.org/10.1049/iet-syb:20070042.

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Sohngen, C., Stelzer, M., Thiele, J., Schomburg, D., 2011. BRENDA, the enzyme information system in 2011. Nucleic Acids Res. 39, D670–D676, http://dx.doi.org/10.1093/nar/gkq1089.

Schmid, J.W., Mauch, K., Reuss, M., Gilles, E.D., Kremling, A., 2004. Metabolic design based on a coupled gene expression-metabolic network model of tryptophan production in *Escherichia coli*. Metab. Eng. 6, 364–377, http://dx.doi.org/10.1016/j.ymben.2004.06.003.

Schmidt, H., 2007. Sbaddon: high performance simulation for the systems biology toolbox for matlab. Bioinformatics 23, 646–647, http://dx.doi.org/10.1093/bioinformatics/btl668.

Schmidt, H., Jirstrand, M., 2006. Systems biology toolbox for matlab: a computational platform for research in systems biology. Bioinformatics 22, 514–515, http://dx.doi.org/10.1093/bioinformatics/bti799.

Schmidt, H., Madsen, M.F., Danø, S., Cedersund, G., 2008. Complexity reduction of biochemical rate expressions. Bioinformatics 24, 848–854, http://dx.doi.org/10.1093/bioinformatics/btn035.

Schomburg, D., Schomburg, I., 2010. Enzyme databases. Methods Mol. Biol. 609, 113–128, http://dx.doi.org/10.1007/978-1-60327-241-4_7.

Schwarz, G., 1978. Estimating the dimension of model. Ann. Stat. 6, 461–464.

Secrier, M., Toni, T., Stumpf, M.P.H., 2009. The abc of reverse engineering biological signalling systems. Mol. Biosyst. 5, 1925–1935, http://dx.doi.org/10.1039/b908951a.

Sedoglavic, A., 2002. A probabilistic algorithm to test local algebraic observability in polynomial time. J. Symb. Comput., 735–755.

Shinto, H., Tashiro, Y., Yamashita, M., Kobayashi, G., Sekiguchi, T., Hanai, T., Kuriya, Y., Okamoto, M., Sonomoto, K., 2007. Kinetic modeling and sensitivity analysis of acetone–butanol–ethanol production. J. Biotechnol. 131, 45–56, http://dx.doi.org/10.1016/j.jbiotec.2007.05.005.

Skaar, N., 2008. Parameter Estimation Methods for Continuous Time Dynamical Systems Given Discrete Time Measurements (Master's thesis). Chalmers University of Technology.

Smallbone, K., Simeonidis, E., Broomhead, D.S., Kell, D.B., 2007. Something from nothing: bridging the gap between constraint-based and kinetic modelling. FEBS J. 274, 5576–5585, http://dx.doi.org/10.1111/j.1742-4658.2007.06076.x.

Smallbone, K., Simeonidis, E., Swainston, N., Mendes, P., 2010. Towards a genome-scale kinetic model of cellular metabolism. BMC Syst. Biol. 4, 6, http://dx.doi.org/10.1186/1752-0509-4-6.

Snoep, J.L., Bruggeman, F., Olivier, B.G., Westerhoff, H.V., 2006. Towards building the silicon cell: a modular approach. Biosystems 83, 207–216, http://dx.doi.org/10.1016/j.biosystems.2005.07.006.

Soh, K.C., 2013. Computational Studies on Cellular Bioenergetics (Ph.D. thesis). Thèse École polytechnique fédérale de Lausanne EPFL.

Soh, K.C., Miskovic, L., Hatzimanikatis, V., 2012. From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. FEMS Yeast Res. 12, 129–143, http://dx.doi.org/10.1111/j.1567-1364.2011.00771.x.

Sohn, S.B., Graf, A.B., Kim, T.Y., Gasser, B., Maurer, M., Ferrer, P., Mattanovich, D., Lee, S.Y., 2010. Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production. Biotechnol. J. 5, 705–715, http://dx.doi.org/10.1002/biot.201000078.

Srinath, S., Gunawan, R., 2010. Parameter identifiability of power-law biochemical system models. J. Biotechnol. 149, 132–140, http://dx.doi.org/10.1016/j.jbiotec.2010.02.019.

Sunnåker, M., Schmidt, H., Jirstrand, M., Cedersund, G., 2010. Zooming of states and parameters using a lumping approach including back-translation. BMC Syst. Biol. 4, 28, http://dx.doi.org/10.1186/1752-0509-4-28.

Sunnåker, M., Cedersund, G., Jirstrand, M., 2011. A method for zooming of nonlinear models of biochemical systems. BMC Syst. Biol. 5, 140, http://dx.doi.org/10.1186/1752-0509-5-140.

Takors, R., Wiechert, W., Weuster-Botz, D., 1997. Experimental design for the identification of macrokinetic models and model discrimination. Biotechnol. Bioeng. 56, 564–576 http://dx.doi.org/3.0.CO;2-C.

Tang, W.L., Zhao, H., 2009. Industrial biotechnology: tools and applications. Biotechnol. J. 4, 1725–1739, http://dx.doi.org/10.1002/biot.200900127.

Teixeira, A.P., Alves, C., Alves, P.M., Carrondo, M.J.T., Oliveira, R., 2007. Hybrid elementary flux analysis nonparametric modeling: application for bioprocess control. BMC Bioinf. 8, 30, http://dx.doi.org/10.1186/1471-2105-8-30.

Teusink, B., Passarge, J., Reijenga, C.A., Esgalhado, E., van der Weijden, C.C., Schepper, M., Walsh, M.C., Bakker, B.M., van Dam, K., Westerhoff, H.V., Snoep, J.L., 2000. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. Eur. J. Biochem. 267, 5313–5329.

Theilgaard, H.A., Nielsen, J., 1999. Metabolic control analysis of the penicillin biosynthetic pathway: the influence of the LLD-ACV:bisACV ratio on the flux control. Antonie Van Leeuwenhoek 75, 145–154.

Thiele, I., Jamshidi, N., Fleming, R.M.T., Palsson, B.Ø., 2009. Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. PLoS Comput. Biol. 5, e1000312, http://dx.doi.org/10.1371/journal.pcbi.1000312.

Transtrum, M.K., Machta, B.B., Sethna, J.P., 2010. Why are nonlinear fits to data so challenging? Phys. Rev. Lett. 104, 060201.

Trusina, A., Papa, F.R., Tang, C., 2008. Rationalizing translation attenuation in the network architecture of the unfolded protein response. Proc. Natl. Acad. Sci. U.S.A. 105, 20280–20285, http://dx.doi.org/10.1073/pnas.0803476105.

Trusina, A., Tang, C., 2010. The unfolded protein response and translation attenuation: a modelling approach. Diabetes Obes. Metab. 12 (Suppl. 2), 27–31, http://dx.doi.org/10.1111/j.1463-1326.2010.01274.x.

Tyo, K.E.J., Kocharin, K., Nielsen, J., 2010. Toward design-based engineering of industrial microbes. Curr. Opin. Microbiol. 13, 255–262, http://dx.doi.org/10.1016/j.mib.2010.02.001.

Tyson, J.J., Chen, K.C., Novak, B., 2003. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr. Opin. Cell Biol. 15, 221–231.

Ullah, M., Wolkenhauer, O., 2010. Stochastic approaches in systems biology. Wiley Interdiscip. Rev. Syst. Biol. Med. 2, 385–397, http://dx.doi.org/10.1002/wsbm.78.

Umaña, P., Bailey, J.E., 1997. A mathematical model of N-linked glycoform biosynthesis. Biotechnol. Bioeng. 55, 890–908 http://dx.doi.org/3.0.CO;2-B.

Usuda, Y., Nishio, Y., Iwatani, S., Dien, S.J.V., Imaizumi, A., Shimbo, K., Kageyama, N., Iwahata, D., Miyano, H., Matsui, K., 2010. Dynamic modeling of *Escherichia coli* metabolic and regulatory systems for amino-acid production. J. Biotechnol. 147, 17–30, http://dx.doi.org/10.1016/j.jbiotec.2010.02.018.

Vajda, S., Godfrey, K.R., Rabitz, H., 1989. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. Math. Biosci. 93, 217–248.

van Riel, N.A.W., 2006. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. Br. Bioinf. 7, 364–-374, http://dx.doi.org/10.1093/bib/bbl040.

Visser, D., Heijnen, J.J., 2002. The mathematics of metabolic control analysis revisited. Metab. Eng. 4, 114–123, http://dx.doi.org/10.1006/mben.2001.0216.

Visser, D., Heijnen, J.J., 2003. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. Metab. Eng. 5, 164–176.

Visser, D., Schmid, J.W., Mauch, K., Reuss, M., Heijnen, J.J., 2004. Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics. Metab. Eng. 6, 378–390, http://dx.doi.org/10.1016/j.ymben.2004.07.001.

Vital-Lopez, F.G., Armaou, A., Nikolaev, E.V., Maranas, C.D., 2006. A computational procedure for optimal engineering interventions using kinetic models of metabolism. Biotechnol. Prog. 22, 1507–1517, http://dx.doi.org/10.1021/bp060156o.

Wang, F.S., Ko, C.L., Voit, E.O., 2007. Kinetic modeling using S-systems and lin-log approaches. Biochem. Eng. J. 33, 238–247.

Wang, L., Birol, I., Hatzimanikatis, V., 2004. Metabolic control analysis under uncertainty: framework development and case studies. Biophys. J. 87, 3750–-3763, http://dx.doi.org/10.1529/biophysj.104.048090.

Wang, L., Hatzimanikatis, V., 2006a. Metabolic engineering under uncertainty. II. Analysis of yeast metabolism. Metab. Eng. 8, 142–159, http://dx.doi.org/10.1016/j.ymben.2005.11.002.

Wang, L., Hatzimanikatis, V., 2006b. Metabolic engineering under uncertainty. I. Framework development. Metab. Eng. 8, 133–141, http://dx.doi.org/10.1016/j.ymben.2005.11.003.

Wang, X., Hao, N., Dohlman, H.G., Elston, T.C., 2006. Bistability, stochasticity, and oscillations in the mitogen-activated protein kinase cascade. Biophys. J. 90, 1961–1978, http://dx.doi.org/10.1529/biophysj.105.073874.

Wiechert, W., Noack, S., 2011. Mechanistic pathway modeling for industrial biotechnology: challenging but worthwhile. Curr. Opin. Biotechnol. 22, 604–610, http://dx.doi.org/10.1016/j.copbio.2011.01.001.

Wiseman, R.L., Powers, E.T., Buxbaum, J.N., Kelly, J.W., Balch, W.E., 2007. An adaptable standard for protein export from the endoplasmic reticulum. Cell 131, 809–821, http://dx.doi.org/10.1016/j.cell.2007.10.025.

Xing, Z., Bishop, N., Leister, K., Li, Z.J., 2010. Modeling kinetics of a large-scale fed-batch CHO cell culture by Markov chain Monte Carlo method. Biotechnol. Prog. 26, 208–219, http://dx.doi.org/10.1002/btpr.284.

Ye, J., Ly, J., Watts, K., Hsu, A., Walker, A., McLaughlin, K., Berdichevsky, M., Prinz, B., Sean Kersey, D., d'Anjou, M., Pollard, D., Potgieter, T., 2011. Optimization of a glycoengineered *Pichia pastoris* cultivation process for commercial antibody production. Biotechnol. Prog. 27, 1744–1750, http://dx.doi.org/10.1002/btpr.695.

Yüzgeç, U., Türker, M., Hocalar, A., 2009. On-line evolutionary optimization of an industrial fed-batch yeast fermentation process. ISA Trans. 48, 79–92, http://dx.doi.org/10.1016/j.isatra.2008.09.001.

Zamorano, F., Vande Wouwer, A., Jungers, R.M., Bastin, G., 2013. Dynamic metabolic models of CHO cell cultures through minimal sets of elementary flux modes. J. Biotechnol. 164, 409–422, http://dx.doi.org/10.1016/j.jbiotec.2012.05.005.

Zi, Z., Klipp, E., 2006. SBML-PET: a systems biology markup language-based parameter estimation tool. Bioinformatics 22, 2704–2705, http://dx.doi.org/10.1093/bioinformatics/btl443.

Zi, Z., Liebermeister, W., Klipp, E., 2010. A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in *Saccharomyces cerevisiae*. PLoS One 5, e9522, http://dx.doi.org/10.1371/journal.pone.0009522.