Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Rites of passage: requirements and standards for building kinetic models of metabolic phenotypes

Ljubisa Miskovic[1,2], Milenko Tokic[1,2], Georgios Fengos[1,2] and Vassily Hatzimanikatis[1,2]

The overarching ambition of kinetic metabolic modeling is to capture the dynamic behavior of metabolism to such an extent that systems and synthetic biology strategies can reliably be tested *in silico*. The lack of kinetic data hampers the development of kinetic models, and most of the current models use *ad hoc* reduced stoichiometry or oversimplified kinetic rate expressions, which may limit their predictive strength. There is a need to introduce the community-level standards that will organize and accelerate the future developments in this area. We introduce here a set of requirements that will ensure the model quality, we examine the current kinetic models with respect to these requirements, and we propose a general workflow for constructing models that satisfy these requirements.

Addresses
[1] Laboratory of Computational Systems Biotechnology (LCSB), Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland
[2] Swiss Institute of Bioinformatics (SIB), CH-1015 Lausanne, Switzerland

Corresponding author: Hatzimanikatis, Vassily (vassily.hatzimanikatis@epfl.ch)

Current Opinion in Biotechnology 2015, **36**:146–153

This review comes from a themed issue on **Pathway engineering**

Edited by **Michael J Betenbaugh** and **William E Bentley**

For a complete overview see the Issue and the Editorial

Available online 2nd September 2015

http://dx.doi.org/10.1016/j.copbio.2015.08.019

Introduction

Mathematical modeling is an essential tool for understanding and explaining complex behavior and properties of living organisms. In recent years, the prevalent frameworks for modeling metabolic pathways were constraint-based approaches that make use of network stoichiometry to characterize the intracellular fluxes at steady state [1–3]. While proving their utility in studies of cellular physiology and metabolic engineering [4,5], the stoichiometric models lack information about metabolic regulation and enzyme kinetics. Therefore, these static descriptions cannot be used for predicting the complex dynamic responses to environmental and genetic perturbations, or, for example, for studying dynamic transitions of the metabolism [6•,7] or oscillatory phenomena [8].

Kinetic models couple dynamics of metabolic concentrations and fluxes to enzyme concentrations and they allow us to take into consideration regulation at the enzyme and post-translational level [9]. Although the potential of kinetic models compared to their stoichiometric counterparts is promising, it comes at a price. Kinetic models are typically built in a bottom-up manner, wherein for each reaction a kinetic rate expression along with corresponding parameter values is required. This results in model structures with large number of parameters. Due to the absence of experimental assays that could provide the required extent of measurements for the rigorous parameterization of these models, researchers incorporate the needed information from different sources: (i) literature; (ii) databases such as Brenda [10]; or (iii) they perform experimental measurements themselves [11,12•,13]. Whenever the model parameters are not experimentally measured, parametric estimation methods [14] or Monte Carlo methods are used [15••,16–19]. In the latter, the parameters are characterized within well-defined bounds that are consistent with the studied conditions and physicochemical laws. The available experimental values of kinetic parameters are often uncertain due to measurement and estimation errors, and variations stemming from different experimental conditions and set-ups [20]. As a consequence, many existing kinetic models are of a limited scope, often with *ad hoc* stoichiometry, they cover one or a few metabolic pathways, and frequently they neglect the whole network dynamics as observed in [9,21].

Recent efforts have been made toward building genome-scale kinetic models [6•,15••,17,22–25]. In the quest for models with a large-scale or genome-scale scope, one must ensure that the increased size and scope is attained without sacrificing the consistency with physicochemical laws and the necessary mechanistic details. As the activity in kinetic modeling is expected to grow intensively in the coming years, there is a need for establishing community-level standards. The objective of this paper is to review the current state of kinetic modeling and propose a set of requirements that every kinetic model should satisfy. We expect that standardized kinetic models will facilitate future community efforts in model building where knowledge from different sources and research groups is incorporated as advocated in [9].

Issues in building kinetic models

When building kinetic models we are given a set of observations and we seek to identify the set of kinetic

parameters that best describe the observations. In metabolic kinetic models we usually start with a set of metabolic fluxes and concentrations, and we assume that the stoichiometry and the thermodynamic properties of the reactions in the metabolic network are known. The basic problem then is to identify kinetic model(s) that consistently describe the experimental observations. We discuss here the main issues in building kinetic models.

Uncertainty

Uncertainty is recognized in the literature as the main challenge in kinetic modeling of biological systems [9,16,20–22,26,27]. The dynamic behavior of metabolism is a result of complex interactions of metabolite concentrations, through kinetics and thermodynamics, and uncertainty in these interactions propagate to the structure and parameters of the kinetic models.

- Uncertainty in kinetic properties of enzymes
  We can distinguish two types of uncertainty in kinetic properties of enzymes: (i) *structural uncertainty* is associated with the missing information concerning kinetic mechanisms; (ii) *quantitative uncertainty* refers to the inconsistency about the values of kinetic parameters in the data [16].
  While the databases that collect and organize the information about kinetic parameters are growing in size [10,28], the available kinetic data are not standardized, and the reported values of kinetic parameters often range within several orders of magnitude. Furthermore, factors that impact the values of kinetic parameters such as temperature or pH are frequently not reported. An additional question is if the values of the kinetic parameters that are quantified *in vitro* and for each enzyme separately can represent well the behavior of a multitude of enzymes interacting in a crowded *in vivo* environment [11,22]
- Uncertainty in metabolic fluxes
  Despite the availability of abundant fluxomics data, the complex topology of metabolic networks prohibits determination of the exact values and directionality of intracellular metabolic fluxes [29,30]. This translates into the existence of multiple alternative flux profiles that are consistent with the measured data but with uncertainty in determining a *unique* flux profile.
- Uncertainty in metabolite concentration levels and thermodynamic properties
  The introduction of thermodynamics-based constraints in the context of flux balance analysis allows integration of metabolomics data through coupling of the directionality of fluxes with metabolite concentrations [17,29–31]. The thermodynamic properties of many reactions are not measured, instead, they are estimated using group contribution methods [32]. These estimates contain both measurement and estimation errors and together with uncertainties in metabolite concentration measurements they can affect the conclusions about cellular physiology.

Size and content of metabolic networks

As the main purpose of the models is the understanding of system-wide properties, we need large models in order to capture the interactions determining the behavior of the system as a whole. The size of a model introduces a trade-off between the accuracy of the models that comes from the description of all possible and important interactions and the number of unknown and uncertain parameters. There are issues to be considered when large-scale and genome-scale kinetic models are constructed.

- Large number of unknown parameters, sloppiness and overfitting
  As the size of the metabolic network increases, the portion of available kinetic parameters is rapidly decreasing. Consequently, a large number of parameters have to be quantified using parameter estimation techniques [14]. However, due to a large number of parameters, the uncertainty in available data, and the intrinsic sloppiness of parametric models in systems biology [33,34] it is impossible to compute unique parameter values. When the number of parameters is large relative to the number of observations, the obtained models tend to describe measurement errors rather than functional relationships within the modeled process (overfitting). As a result, poor predictions are obtained when these models are validated against independent data sets.
- Issues with parameter estimation methods
  Parameter estimation methods use optimization procedures to obtain the values of parameters. Depending on the underlying formulation, network structure and employed optimization technique parameter estimation might become computationally intractable for large metabolic networks [35].
- Issues with Monte Carlo methods
  In Monte Carlo methods, the admissible parameter space is constrained with physicochemical and thermodynamic laws along with the constraints obtained from available measurements, and then a population of alternative parameter sets is drawn from such a reduced solution space [15••,16–19,23,25,36]. Sampling of such space is a computationally daunting task for large metabolic networks. Another important challenge is that an efficient sampling necessitates well-defined bounds on kinetic parameters such as Michaelis constants, and these bounds are rarely known. To address these issues a new tailor-made formulation and a new sampling technique were proposed [22].
- Stiffness of metabolism dynamics
  Large-scale and genome-scale kinetic models of metabolism are stiff systems of ordinary differential equations (ODE) since they span over metabolic reactions with a wide range of rate dynamics. The