Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

The discrepancy between data for and expectations on metabolic models: How to match experiments and computational efforts to arrive at quantitative predictions?

Katja Tummler and Edda Klipp

Abstract

Understanding the regulation of metabolism in time and space is critical for many biological problems, be it the growth of tumors or the adaptation of the gut microbiome to diet. However, the need for quantitative and dynamic understanding and the effort to gain the appropriate data usable in computational models diverge dramatically. Nowadays, metabolism on a genome scale is primarily studied with methods that refer to steady states and conclusions to dynamics and quantitative aspects are only made in an indirect way. There are theoretical concepts that could in principle deliver dynamic and quantitative descriptions, such as ordinary differential equation systems employing tailored rate laws for enzymatic reactions, but they dramatically lack information about the required parameter values and intracellular concentrations as well as computationally feasible parameterization methods.

Addresses

Theoretical Biophysics, Humboldt Universität zu Berlin, Invalidenstraße 42, 10115 Berlin, Germany

Corresponding author: Klipp, Edda (edda.klipp@rz.hu-berlin.de)

Current Opinion in Systems Biology 2017, ■:1–6

This review comes from a themed issue on **Regulatory and metabolic networks (2018)**

Edited by **Bas Teusink** and **Uwe Sauer**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.coisb.2017.11.003

Introduction

The duality of metabolic modeling

Metabolic modeling efforts roughly fall into two categories: Small-scale kinetic models and large to genome-scale constraint based models. Kinetic models mechanistically describe the dependencies of metabolic flux on enzyme levels, enzyme activities, allosteric, post-translational or competitive regulation, substrate and product concentrations as well as thermodynamics. This depth of information allows for quantitative predictions but comes at a high cost for both modeling and simulation as well as the acquisition of suitable and sufficient experimental data for model calibration. Therefore, the coverage of such models is often limited to a few reactions or pathways.

On the other end of the spectrum, genome-scale constraint based models lack mechanistic detail but include up to the entirety of known metabolic reactions of a species. Based on the constraints imposed by the network stoichiometry and (experimentally derived) flux bounds and by applying the steady state mass balance assumption, such models can predict feasible flux distributions and have many functional applications. As the formalism reduces the system in question to a linear description, simulation and optimization are computationally cheap. Due to the models' maximum coverage, systemic features of metabolism become evident in their analysis, but on the downside, owing to the lack of detailed kinetic descriptions of functional dependencies, mechanistic predictions are impossible.

Only kinetics can tell

Why is it important to take kinetic modeling to the large scale? Kinetic models, as is evident from many small-scale studies, are able to describe cellular processes mechanistically and hence have a high predictive power also outside of the condition space used for calibration (e.g in human hepatic glucose metabolism [1] or different nutritional conditions in yeast [2] and others). However, the cellular context is crucial for many, including the very central metabolic systems. Only evaluating metabolism as a whole reveals systemic network effects, essential for understanding cellular functionality. While the intrinsic steady state mass balance constraints are well described and very insightful in FBA models, cellular regulation and catalysis largely happens on transient timescales and is dependent on agent concentrations, their affinities, the transient action of long and short-range feedbacks and many others. Kinetic models are able to capture these essential non-linearities [3] which are crucial for complex systemic behavior.

Therefore, expanding kinetic descriptions to the genome scale is the next frontier in modeling metabolic systems. Many valuable approaches for individual aspects of this challenges have been developed recently. Below we will highlight the most recent studies but also

point out the abundant difficulties in establishing large-scale kinetic models.

Recent steps on the way to large scale kinetic models

Starting big: FBA based methods

One way to approach a genome-scale kinetic model of metabolism is to start with a well curated metabolic reconstruction and fill it with life — in the form of mechanistic dependencies on reactants, cofactors, enzymes or regulators, paving the way to a full kinetic description of the network. Initially, dependencies on metabolite concentrations were included in a simplified way via the laws of thermodynamics (e.g. tFBA [4,5]), later by nonlinear flux constraints estimated from measured fluxes or known kinetic rate laws (DFBA [6,7]). Recent extensions of this latter method make use of better data (e.g. time-resolved metabolomics or transcriptomics in TREM flux [8], MetDFBA [9], or uFBA [10]), but are still limited to smaller networks by high computational costs and sparse data availability. Furthermore, their strong dependency on either matching experimental data or rough simplification of rate laws prohibits flux predictions outside the measured regimen. In a similar way, a number of approaches aims at finding flux distributions that best comply with a number of different omics datasets, partly using simplified kinetic rate laws (e.g. iOMA [11]) or sampling approaches (e.g. ORACLE framework [12]). Generally, the included data do not cover the entire network, such that the models are either reduced to the set of experimentally well covered reactions (e.g. MetDFBA) or the experimentally inaccessible regions of the network are described simplistically (e.g. by simple mass balance in iOMA).

The drawback of such FBA derived methods is their strong dependency on the steady state assumption, which categorically excludes representing mechanistic, nonlinear transitions between states. The approaches often include linearization and are limited to predictions of distinct steady states. This can be very valuable, for example in the analysis of clinical patient data, but lacks mechanistic predictive power, as most information on kinetic parameters is encoded in transient changes of metabolite concentrations.

Starting small: expanding & combining kinetic models

Expanding existing small-scale kinetic models or applying established methods for kinetic modeling on the pathway scale to larger systems is a complementary approach. The inherent model complexity is thereby tackled in different ways. Already published and parameterized models can be combined (e.g. Ref. [13]) or individually readjusted to new average cell, single cell or patient specific data to arrive at specific models [14]. On the interface of the two situations, modelers try to describe the active, i.e. flux carrying or enzyme expressing, parts of a genome-scale model entirely with the help of measured data and standardized rate laws [15,16].

As for all kinetic metabolic models, the limiting factor of such approaches is data availability, for internal concentrations as well as for type and parameterization of the kinetic laws. Often the resulting models also have a very high computational cost. The established methods might therefore also only be feasible for the sparse data and hence small or simplified systems they were developed in a related, hybrid approach would be to express only the well understood part of a metabolic network in kinetic equations and retain the purely stoichiometric nature of the remaining metabolism, entirely or in a lumped fashion based on genome scale reconstructions [17], coupled to the cellular growth rate. While adhering closely to the resolution of data and knowledge, such methods would sacrifice predictability of long range effects as well as of feedbacks from remote parts of the network.

Starting from the data: Model aided omics studies

Experimental techniques have for a long time conquered the genome scale, with ever improving metabolomics lagging slightly behind measurements of proteomes and transcriptomes. As a further approach towards genome scale kinetic models, some metabolomics studies employed tailored mathematical modeling for data analysis and interpretation. The used models are often small but can draw mechanistic and biologically meaningful conclusions, such as identifying metabolite triggered gene expression [18], detecting regulations on the individual reaction level [19] or explaining prominent observations in the large datasets [20].

What prevents genome scale kinetic models?

Despite all these well-formed ideas on large-scale kinetic modeling, the actual number of working examples is small. What are the main obstacles? In the following section we will highlight 4 major points hampering model construction, simulation, calibration and reusability.

The "lack" of data

Modelers might appear slightly greedy by declaring a need for better data in the light of drastically improving and accumulating omics datasets. However, also the kind of data matters and data suitable for kinetic modeling has to fulfill special requirements.

Firstly, quantitativeness is crucial. Metabolite measurements are constantly improving w.r.t temporal resolution (e.g. real-time metabolome profiling [20]),