Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Improving prediction fidelity of cellular metabolism with kinetic descriptions

Anupam Chowdhury[1], Ali Khodayari[1] and Costas D Maranas

Several modeling frameworks for describing and redirecting cellular metabolism have been developed keeping pace with the rapid development in high-throughput data generation and advances in metabolic engineering techniques. The incorporation of kinetic information within stoichiometry-only modeling techniques offers potential advantages for improved phenotype prediction and consequently more precise computational strain design. In addition to substrate-level kinetic regulatory information, the integration of a number of additional layers of regulation at the transcription, translation, and post-translation levels is sought after by many research groups. However, the practical integration of these complex biological processes into a unified framework amenable to design remains an ongoing challenge.

Address
Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, United States

Corresponding author: Maranas, Costas D (costas@psu.edu)
[1] Joint first authors.

Current Opinion in Food Science 2015, **36**:57–64

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

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

http://dx.doi.org/10.1016/j.copbio.2015.08.011

Introduction

The central goal of metabolic engineering is to harness the cellular machinery of microorganisms and eukaryotic cell lines for converting low-value feedstock to economically viable biochemical products and biofuels. While in some cases productivity or yield limitations arise for reasons outside the scope of metabolic models [1] (e.g., toxicity, strain stability, among others), in many cases the cause is within the range of metabolic modeling. For example, product yield may be limited due to enzyme promiscuity, product inhibition, enzyme de-activation, depleted precursor metabolite pools or insufficient enzyme levels. Metabolic models are increasingly becoming more complex expanding beyond simple stoichiometric descriptions of metabolic networks and gene-protein-reaction (GPR) relationships. Kinetic information is being used to postulate a mechanistic description of cellular function from gene transcription to translation and post-translational events [2]. In addition to providing a dynamic view of metabolism, kinetic descriptors offer the opportunity of linking together several layers of regulatory events into an integrated framework. Kinetic model development, however, is still plagued by a number of challenges chief among which are paucity of kinetic parameters, non-universality of rate laws, and implementation of regulatory events. These challenges become even more pressing when kinetic models are tasked with predictions at a genome-scale.

In this review, we concentrate on the recent advances in the construction of kinetic models and kinetics-aided procedures for improved understanding and redirection of cellular metabolism. Incorporation of regulatory events at transcription, translation, post-translational, and substrate-levels with existing formalisms are described along with their associated challenges. Finally, we discuss state-of-the-art strain design protocols for metabolic engineering that incorporate available kinetic description for improving their prediction accuracy.

Kinetic model development

Kinetic model development requires knowledge of stoichiometry, kinetic rate laws, and fluxomic/metabolomic data to support parameter estimation. Earlier efforts relying on *in vitro* derived lumped kinetics led to inconsistencies of model predictions for *in vivo* phenotypes [3]. A number of linear approximate kinetic laws (e.g., log-lin [4], lin-log [5], Taylor series approximation [6], thermokinetic [7], and convenience rate laws [8]) were put forth instead to standardize the kinetic expressions and provide a tractable parameterization workflow. However, the prediction accuracy of these models was inherently limited to the proximity of the reference phenotype as parameterization did not, in general, include data for multiple phenotypes under genetic/environmental perturbations and only a limited number of substrate-level regulatory loops were included [9]. A promising way forward has been to deconstruct complex kinetic expressions into their elementary kinetic steps [10••] while retaining all known regulation. First, metabolic reactions and their substrate-level regulatory interactions are decomposed into their elementary steps followed by the derivation of mass action kinetics for each step. This reduces the nonlinearity of the kinetic description to only bilinear terms without losing any mechanistic detail at the expense of a larger model size. Next, model variables (e.g., elementary kinetic parameters [10••] or elasticity [11]) are

expressed in terms of elementary reaction parameters (i.e., enzyme fractions and reactions reversibility), both of which are bounded between zero and one. This representation provides a natural scaling for parameter sampling from the entire kinetic space consistent with thermodynamic principles [10••]. For example, the Ensemble Modeling (EM) paradigm was developed by uniformly sampling multi-dimensional space of elementary kinetics to construct an ensemble of kinetic models using a reference flux dataset [10••]. In order to reduce the number of model variables during parameter sampling, metabolite concentrations are generally scaled by their reference strain values.

Given a kinetic model scaffold and sampling procedure, the remaining challenge is the availability of metabolomic and fluxomic data for a variety of phenotypes subject to environmental (e.g., aerobic vs. anaerobic), carbon substrate, and genetic perturbations. In an earlier effort, an optimization-based parameter estimation technique was proposed to parameterize EM formalism by integrating multiple mutant strain data under aerobic conditions in *Escherichia coli* [12••]. The developed model led to accurate predictions for aerobic succinate overproduction, but prediction quality was poor when the model was queried under anaerobic conditions [13]. This was because model parameterization was carried out in the absence of anaerobic flux data. This confirms that kinetic model prediction reliability can only be expected if the kinetic model is parameterized using information from perturbations (either genetic or environmental) that approximate the queried metabolic phenotype. At the same time, all relevant substrate-level regulatory interactions must be *a priori* included in the model to allow for their parameterization.

Metabolic regulation

Microbial organisms leverage intertwined layers of regulation to make optimal use of metabolic resources in response to genetic/environmental stimuli. These regulatory layers of transcription, translation, post-translational modification, and substrate-level regulatory interaction act by altering the capacity of metabolic enzymes. Transcriptional and translational regulations mostly control allocation of enzyme resources, while post-translational regulation mainly modulates enzyme kinetics. Substrate-level regulation also provides rapid responses through passive flux regulation by enzyme saturation and thermodynamics. Despite many advances, a gap still exists in effectively linking transcription and translation levels with cellular metabolism. The challenge is twofold: (a) identification of the structure of the underlying regulatory network [14••] and (b) integration of the identified network with the rest of metabolic events through detailed mechanistic descriptions [15].

Regulatory network reconstruction

Environmental and/or metabolic changes modulate gene expression through signal transduction and/or feedback/feedforward regulation (e.g., flux sensing [16]) (Figure 1a–d). There have been several system biology efforts aimed at elucidating transcriptional and translational regulatory networks using forward-engineering [17] and reverse-engineering [14••] approaches. These approaches, however, generally recover the totality of all potential interactions of which only a small subset becomes active under any given condition [18].

A comprehensive assessment of transcriptional regulation in yeast [19] and *E. coli* [20] was carried out by metabolic flux elucidation upon deletion of transcription factors (one at a time). The results revealed a sparse network of active transcriptional factors that control metabolic flux distribution for any given condition. In another effort, the Environmental and Gene Regulatory Influence Network (EGRIN2.0) software tool was developed to systematically identify condition-specific gene regulatory network by combining a biclustering technique (cMonkey) with a regression-based approach (Inferelator) [18]. Unlike transcriptional regulation, the global assessment of translational regulation remains largely uncharted. Riboswitches [21] (Figure 1e) and small noncoding RNAs (sRNAs) [22] (Figure 1f) are two major classes of translational regulators that act by affecting mRNAs translation through binding to small metabolites and basepairing to the target mRNA, respectively. While the events underpinning the control of translation initiation have been identified, many efforts are still underway to elucidate their interplay with the translation-initiation factors and ribosomal subunits [23]. Developing global techniques to systematically identify condition-specific translation regulatory networks at a genome-scale remains an open challenge.

Once the protein is translated, post-translational modification may alter protein catalytic capability through covalent (e.g., (de)phosphorylation) (Figure 1g) and non-covalent modifications (e.g., allostery) (Figure 1h) [24]. Advances in mass-spectrometry based proteomics have greatly facilitated a comprehensive discovery and quantification of covalent protein modifications [25]. Several efforts have also been made towards the systematic inference of allosteric interactions by combining dynamic metabolomics with ensemble modeling [26]. The substrate-level protein regulation (e.g., (un)competitive inhibition) (Figure 1i–j), however, is still limited to *in vitro* activity assays.

Integration of regulatory events with metabolism

Several efforts have been made towards integration of transcriptional regulation into stoichiometry-based models driven by advances in regulatory network reconstructions and availability of DNA microarray and RNAseq datasets. Early studies used Boolean operators by assuming a positive correlation between metabolic flux and gene expression level [27]. However, experimental studies revealed that negative correlations often arise [16].