Biotechnology Advances 35 (2017) 981–1003

Contents lists available at ScienceDirect

Biotechnology Advances

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

Research review paper

Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks

Pedro A. Saa[a], Lars K. Nielsen[a,b,*]

[a] Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, Brisbane, QLD 4072, Australia
[b] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, DK-2800 Kongens Lyngby, Denmark

ARTICLE INFO

Keywords:
Kinetic models
Metabolic modelling
Enzyme kinetics
Parameter estimation
Metabolic control analysis
Monte Carlo simulation

ABSTRACT

Kinetic models are critical to predict the dynamic behaviour of metabolic networks. Mechanistic kinetic models for large networks remain uncommon due to the difficulty of fitting their parameters. Recent modelling frameworks promise new ways to overcome this obstacle while retaining predictive capabilities. In this review, we present an overview of the relevant mathematical frameworks for kinetic formulation, construction and analysis. Starting with kinetic formalisms, we next review statistical methods for parameter inference, as well as recent computational frameworks applied to the construction and analysis of kinetic models. Finally, we discuss opportunities and limitations hindering the development of larger kinetic reconstructions.

1. Introduction

Mathematical models are essential to broaden our knowledge of metabolic networks. They provide a rational and systematic framework for integrating existing biological knowledge with experimental data, thus enabling appraisal of the complex regulation underpinning the operation of metabolism. During the past decades, several modelling frameworks have been developed for predicting the dynamic behaviour of cellular metabolism supported by the rapid progress in high-throughput omics data generation and advanced metabolic engineering techniques (Chowdhury et al., 2015). The ultimate goal is to integrate these data with mechanistic models to increase our understanding about metabolic networks as well as the information content of the models.

Metabolic network models are described by the set of biochemical reactions mediated by enzymes. Enzymes are proteins whose expression is determined by the genetic program of the cell under specific environmental conditions. The presence of a specific enzyme in the genome implies that a cell has the metabolic capability of the corresponding biotransformation. Annotation and assembly of the repertoire of Gene-Protein-Reactions (GPR) associations from genome sequences and multiple data sources (Fig. 1A) constitutes then a formal representation of the metabolic potential of the cell (Fig. 1B). Subsequent integration of different omics data onto the reconstructed network produces a metabolic model amenable to structural and dynamic analyses (Fig. 1C). Structural analysis relies solely on reaction stoichiometries under steady-state and constitutes the basis of parameter-free Constrained-Based Modelling (CBM) methods (Fig. 1D) (for a detailed review refer to Lewis et al., 2012). These methods have generated fundamental biological insights into the operation of metabolic networks (Ibarra et al., 2002; Schuetz et al., 2012) as well as great advances in biotechnological applications (Lee and Kim, 2015; Yim et al., 2011). By applying various optimization (Burgard et al., 2003; Edwards et al., 2001; Mahadevan and Schilling, 2003) and sampling methods (Almaas et al., 2004; Saa and Nielsen, 2016b), CBM methods enable exploration of the space of feasible metabolic states allowed by the network structure and physiological constraints. The latter methods are however of limited use for the prediction of how metabolic states are achieved, as they lack kinetic information. In contrast, kinetic models of metabolism explicitly describe reaction fluxes as a function of metabolite and enzyme concentrations, enabling dynamic interrogation and quantitative integration of metabolomic, proteomic and transcriptomic data (Fig. 1E).

Metabolic reactions in kinetic models are described by disparate non-linear rate laws, typically involving highly parameterized mathematical expressions. Early modelling efforts proposed different simplified kinetic formalisms to simplify their structure and ease parameter fitting from in vivo data (Hatzimanikatis and Bailey, 1997; Savageau, 1969; Visser and Heijnen, 2003). Although these efforts have yielded valuable insights about the design principles (Savageau et al., 2009) and the dynamic behaviour of different metabolic (Alvarez-Vasquez et al., 2005; Visser et al., 2004), signalling (Vera et al., 2007) and even genetic systems (Atkinson et al., 2003), most of their predictions are inherently limited to the proximity of the chosen operation point.

* Corresponding author at: Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, Brisbane, QLD 4072, Australia.
E-mail addresses: p.saa@uq.edu.au (P.A. Saa), lars.nielsen@uq.edu.au (L.K. Nielsen).

http://dx.doi.org/10.1016/j.biotechadv.2017.09.005
Received 21 March 2017; Received in revised form 30 August 2017; Accepted 10 September 2017
Available online 13 September 2017

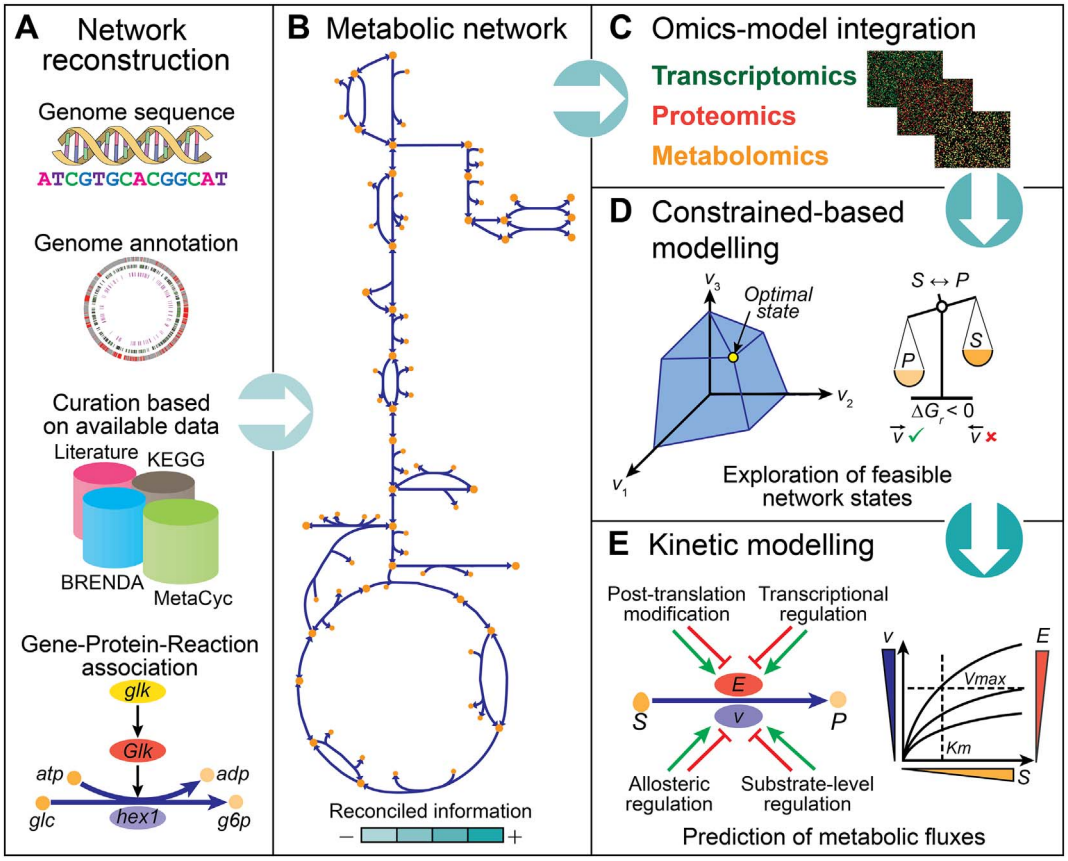

Fig. 1. Model-centric workflow for metabolic networks reconstruction and analysis. A Network reconstruction starts with the annotation of the genome sequence with the encoded metabolic enzymes. Relationships between gene, proteins and reactions are stored in GPR (Gene-Protein-Reaction) associations, enabling rational representation of the biological information flow. Discrepancies and missing information are resolved (where possible) with the support of data from the literature and comprehensive databases (e.g., BRENDA (Placzek et al., 2016), MetaCyc (Caspi et al., 2016), KEGG (Kanehisa et al., 2016)) in a thorough process of manual curation. B The set of curated enzymatic reactions determines the spectrum of metabolic capabilities of the cell. This spectrum is mathematically described by the stoichiometry of the biochemical reactions and defines the topology of the reconstructed network. C Integration of diverse 'omics' data with the metabolic reconstruction enables construction of a metabolic model amenable for rational interrogation and biological discovery. D Structural analysis of the metabolic model is readily achieved using constrained-based modelling methods. Stoichiometric, thermodynamic and kinetic (capacity) constraints defines the space of possible network states which can be readily explored using state-of-the-art optimization and/or sampling methods. This structural analysis is however static, and, it does not quantitatively explain how fluxes are achieved. E Inclusion of kinetic descriptions for the all the enzymes in the model enables prediction of metabolic states as well as dynamic interrogation of the system. The resulting kinetic model can reconcile higher amounts of data; however, it requires substantially more information for its construction.

Furthermore, these models ignore thermodynamic relationships between parameters, often rendering unrealistic behaviours. Conversely, mechanistic-based rate laws are thermodynamically consistent and hold greater prediction power; however, they require a substantial volume of data to fit a multitude of parameters. Approximate mechanistically-inspired formalisms have alleviate these issues to some extent (Ederer and Gilles, 2007; Hofmeyr and Cornish-Bowden, 1997; Liebermeister and Klipp, 2006a; Liebermeister et al., 2010), however they still inevitably suffer from parameter identifiability issues (Heijnen and Verheijen, 2013). Fitting mechanistically-grounded kinetic models using conventional methods has previously been deemed impracticable, as homeostatic control often renders several parameters highly correlated or even outright unidentifiable even in the presence of large amounts of data (Degenring et al., 2004; Hadlich et al., 2009). However, recent Monte Carlo and other simulation-based strategies for kinetic model construction and analysis (Bordbar et al., 2015; Chakrabarti et al., 2013; Saa and Nielsen, 2016a; Steuer et al., 2006; Tran et al., 2008) have shown that satisfactory predictions can be achieved even when many parameters are poorly resolved and/or uncertain. Indeed, scrutiny of most kinetic models has revealed that good predictions do not necessarily require precise parameters (Gutenkunst et al., 2007). As such, modelling frameworks are moving away from precisely fitting coarse grained models, towards building mechanistic models capable of identifying emergent regulatory and dynamic behaviours (Link et al., 2014).

The potential key role of kinetic models in the field of systems biotechnology is certainly undeniable. These models are the only capable of reconciling the multiple layers of omics data, i.e., transcriptomics/proteomics, metabolomics and fluxomics, within a common and coherent mathematical framework. Recent examples of the application of these models include strain design and optimization (Andreozzi et al., 2016a; Khodayari and Maranas, 2016; Savoglidis et al., 2016), identification of drug targets and side effects (Bordbar et al., 2015; Haanstra et al., 2017; Murabito et al., 2011), unravelling key regulatory interactions (Link et al., 2013; Saa and Nielsen, 2016a), to name a few. These case studies constitute a first glance of the potential applications of kinetic models, which justifies the renewed interest of the scientific community in these models. Supported by advanced frameworks for kinetic modelling, kinetic models are increasingly deployed to understand complex metabolic phenotypes.

This review presents a comprehensive overview of mathematical frameworks for kinetic modelling, starting with the relevant formalisms used to describe enzyme-catalysed reactions. We next review relevant classical statistical methods for parameter inference, as well as more recent computational frameworks specific for the analysis of kinetic models. Considering the importance and potential applications of the latter, we focused our attention on these and critically reviewed their main features and capabilities. Finally, we discuss current limitations