BioSystems 147 (2016) 1–10

Contents lists available at ScienceDirect

BioSystems

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

Review article

Advances in the integration of transcriptional regulatory information into genome-scale metabolic models

R.P. Vivek-Ananth, Areejit Samal *

The Institute of Mathematical Sciences (IMSc), Chennai, India

ARTICLE INFO

Article history:
Received 26 March 2016
Received in revised form 14 May 2016
Accepted 7 June 2016
Available online 7 June 2016

Keywords:
Flux balance analysis
Metabolic networks
Transcriptional regulatory networks
Data integration
Integrated regulatory-metabolic models

ABSTRACT

A major goal of systems biology is to build predictive computational models of cellular metabolism. Availability of complete genome sequences and wealth of legacy biochemical information has led to the reconstruction of genome-scale metabolic networks in the last 15 years for several organisms across the three domains of life. Due to paucity of information on kinetic parameters associated with metabolic reactions, the constraint-based modelling approach, flux balance analysis (FBA), has proved to be a vital alternative to investigate the capabilities of reconstructed metabolic networks. In parallel, advent of high-throughput technologies has led to the generation of massive amounts of *omics* data on transcriptional regulation comprising mRNA transcript levels and genome-wide binding profile of transcriptional regulators. A frontier area in metabolic systems biology has been the development of methods to integrate the available transcriptional regulatory information into constraint-based models of reconstructed metabolic networks in order to increase the predictive capabilities of computational models and understand the regulation of cellular metabolism. Here, we review the existing methods to integrate transcriptional regulatory information into constraint-based models of metabolic networks.

Contents

1. Introduction

Extensive research by biochemists in the last century has resulted in the chemical characterization of thousands of biochemical reactions (Kanehisa and Goto, 2000; Chang et al., 2009). Towards the end of 20th century, complete genome sequences became available for the first time (Fleischmann et al., 1995; Tomb et al., 1997). In the post-genomic era, a focus in systems biology has been the reconstruction of genome-scale metabolic networks using the annotated sequences along with available biochemical, genetic and phenotypic information for organisms (Edwards et al., 2001; Forster et al., 2003; Duarte et al., 2007; Feist et al., 2009; Thiele and Palsson, 2010; Thiele et al., 2013). In the last 15 years, considerable effort has led to the reconstruction of manually curated and high quality genome-scale metabolic networks for more than 50 organisms including humans (Durot et al., 2009; Feist et al., 2009; Thiele and Palsson, 2010). However, the current pace of manual reconstruction of high quality genome-scale metabolic networks lags far behind the sequencing effort, and thus, automated methods have also been developed to aid and accelerate the speed of metabolic network reconstruction process (Henry et al., 2010; Schellenberger

* Corresponding author.
E-mail address: asamal@imsc.res.in (A. Samal).

http://dx.doi.org/10.1016/j.biosystems.2016.06.001
0303-2647/

et al., 2011; Agren et al., 2013). Several studies have demonstrated the utility of these genome-scale metabolic reconstructions for biological discovery and hypothesis generation (Feist and Palsson, 2008; Oberhardt et al., 2009).

Current paucity of information on relevant parameters such as rate constants, enzyme concentrations and metabolite concentrations for most reactions renders kinetic modelling of genome-scale metabolic networks infeasible. In the face of inadequate kinetic information, the constraint-based modelling method, flux balance analysis (FBA), has proved to be a vital alternative to study the capabilities of genome-scale metabolic networks (Varma and Palsson, 1994; Kauffman et al., 2003; Price et al., 2004; Orth et al., 2010; Lewis et al., 2012). In contrast to kinetic models, constraint-based FBA primarily uses stoichiometric information (Heinrich and Schuster, 1996; Schilling et al., 1999; Palsson, 2006) of reactions in a metabolic network to predict the flux of reactions in steady state and biomass synthesis rate in a given environmental condition. Due to its simplicity, constraint-based FBA has become a popular framework to study genotype-phenotype relationships and predict the metabolic response to environmental and genetic perturbations using genome-scale metabolic reconstructions (Feist and Palsson, 2008; Papp et al., 2009; Lewis et al., 2012).

Concurrently, advances in post-genomic high-throughput data collection techniques has led to the generation of vast amounts of *omics* data. High-dimensional multi-omics data provides quantitative information on multitude of cellular components across diverse scales of organization. A major goal of systems biology is to turn the recent explosion of omics data into predictive holistic models of biological systems (Kitano, 2002; Joyce and Palsson, 2006). In this direction, contextualization of omics data within constraint-based FBA models of genome-scale metabolic reconstructions can lead to more accurate models. However, there are several challenges in integration of omics data stemming from the inherent experimental and biological noise in such datasets (Quackenbush, 2004). Nonetheless, several constraint-based methods have been developed to integrate experimental data, especially on transcriptional regulation and gene expression, within the FBA framework to build improved models (Åkesson et al., 2004; Covert et al., 2004; Becker and Palsson, 2008; Blazier and Papin, 2012; Hyduke et al., 2013). In this review, we discuss the existing methods to integrate regulatory information into constraint-based FBA models by broadly classifying them into three different approaches (Covert et al., 2001; Åkesson et al., 2004; Covert et al., 2004; Becker and Palsson, 2008; Chandrasekaran and Price, 2010; Blazier and Papin, 2012; Hyduke et al., 2013; Kim and Reed, 2014). Such methods have already proven successful in building context-specific metabolic models for human tissues and predicting novel drug targets in pathogens (Becker and Palsson, 2008; Folger et al., 2011; Bordbar et al., 2012; Collins et al., 2012).

The review is organized as follows. In the second section, we describe the constraint-based FBA framework. In the third section, we discuss existing methods to integrate omics data within the FBA framework as additional flux constraints to build context-specific metabolic models. In the fourth section, we describe the reconstruction and analysis of integrated regulatory-metabolic models where Boolean transcriptional regulatory networks (TRNs) are incorporated within the FBA framework. In the fifth section, we discuss the need for automated methods to integrate information on regulatory network architecture and expression measurements within metabolic networks to reconstruct integrated regulatory-metabolic models. Note that previous reviews in this area only emphasize on methods that are descriptive in nature which are presented in section 3 of this review. In comparison to previous reviews, we here provide a much more comprehensive overview of the area by also describing in detail the methods which are predictive rather than just descriptive in nature in sections 4 and 5 of this review.

2. Flux balance analysis

Flux balance analysis (FBA) is a constraint-based modelling approach that is widely used to investigate the capabilities of available genome-scale metabolic networks (Varma and Palsson, 1994; Kauffman et al., 2003; Price et al., 2004; Orth et al., 2010; Lewis et al., 2012). FBA primarily uses the information on the list of biochemical reactions in an organism along with the stoichiometric coefficients of involved metabolites to predict the fluxes of all reactions in the metabolic network. Such biochemical information is contained within available organism-specific genome-scale metabolic reconstructions. For any organism, the genome-scale metabolic reconstruction contains information on all known metabolic reactions and genes encoding enzymes catalysing different reactions in the network (Palsson, 2006) (Fig. 1A). Notably, genome-scale metabolic reconstructions for most organisms also include reactions for transport of metabolites across the cell boundary, and a pseudo-reaction capturing the production of biomass in terms of their precursor metabolites (Fig. 1A).

In the FBA framework, the list of reactions along with the stoichiometric coefficients of involved metabolites in a network reconstruction is mathematically represented in the form of a stoichiometric matrix $\mathbf{S}$ of dimensions $m \times n$, where m denotes the number of metabolites and n denotes the number of reactions in the network (Fig. 1B). Entries in each column of the matrix $\mathbf{S}$ give the stoichiometric coefficients of metabolites participating in a particular reaction, where negative coefficients signify consumption of a metabolite, positive coefficients signify production of a metabolite, and zero coefficients signify no participation of a metabolite in the reaction (Fig. 1B). These stoichiometric coefficients of metabolites in various reactions impose constraints on the flow of metabolites in the network (Heinrich and Schuster, 1996; Schilling et al., 1999; Palsson, 2006). Subsequently, the method capitalizes on these stoichiometric constraints and assumes steady state to predict the fluxes of all reactions in the network.

In any metabolic steady state, different metabolites attain a mass balance wherein the rate of production of each metabolite is equal to its rate of consumption, and this leads to the system of mass balance equations given by:

$$\mathbf{S}.\mathbf{v} = 0 \tag{1}$$

where $\boldsymbol{v}$ is the vector of fluxes through all reactions in the network (Fig. 1B). For each metabolite in the network, Eq. (1) gives a linear equation relating fluxes of various reactions in which the metabolite participates (Fig. 1B). Since, the number of metabolites is much less than the number of reactions in genome-scale metabolic networks of most organisms, the number of linear equations is much less than the number of reaction fluxes (unknowns) to be determined. Thus, Eq. (1) typically leads to an under-determined system of linear equations, and a large solution space of allowable fluxes for genome-scale metabolic networks (Fig. 1B and C).

The size of the allowable space can also be reduced by incorporating additional constraints on reaction fluxes. Firstly, certain reactions in the metabolic network are irreversible under physiological conditions, and such thermodynamic constraints (Beard et al., 2002; Orth et al., 2010) can be used to constrain the flux of irreversible reactions. Secondly, the activity of specific enzymes may limit the flux through certain reactions. Thirdly, the availability of nutrients in the growth medium can be used to constrain the fluxes of transport reactions. Note that unlike stoichiometric or mass-balance constraints, these additional constraints represent bounds on reaction fluxes in the metabolic network (Fig. 1B).