Metabolic Engineering 39 (2017) 117–127

Contents lists available at ScienceDirect

Metabolic Engineering

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

Original Research Article

Thermodynamics-based Metabolite Sensitivity Analysis in metabolic networks

A. Kiparissides[a,b], V. Hatzimanikatis[a,*]

[a] Laboratory of Computational Systems Biotechnology, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
[b] Department of Biochemical Engineering, University College London, WC1E 6BT, London, UK

ARTICLE INFO

Keywords:
TMFA
TFBA
GSA
Metabolic Modeling
FBA
DoE

ABSTRACT

The increasing availability of large metabolomics datasets enhances the need for computational methodologies that can organize the data in a way that can lead to the inference of meaningful relationships. Knowledge of the metabolic state of a cell and how it responds to various stimuli and extracellular conditions can offer significant insight in the regulatory functions and how to manipulate them. Constraint based methods, such as Flux Balance Analysis (FBA) and Thermodynamics-based flux analysis (TFA), are commonly used to estimate the flow of metabolites through genome-wide metabolic networks, making it possible to identify the ranges of flux values that are consistent with the studied physiological and thermodynamic conditions. However, unless key intracellular fluxes and metabolite concentrations are known, constraint-based models lead to underdetermined problem formulations. This lack of information propagates as uncertainty in the estimation of fluxes and basic reaction properties such as the determination of reaction directionalities. Therefore, knowledge of which metabolites, if measured, would contribute the most to reducing this uncertainty can significantly improve our ability to define the internal state of the cell. In the present work we combine constraint based modeling, Design of Experiments (DoE) and Global Sensitivity Analysis (GSA) into the Thermodynamics-based Metabolite Sensitivity Analysis (TMSA) method. TMSA ranks metabolites comprising a metabolic network based on their ability to constrain the gamut of possible solutions to a limited, thermodynamically consistent set of internal states. TMSA is modular and can be applied to a single reaction, a metabolic pathway or an entire metabolic network. This is, to our knowledge, the first attempt to use metabolic modeling in order to provide a significance ranking of metabolites to guide experimental measurements.

1. Introduction

The development of new modeling methods (Miskovic and Hatzimanikatis, 2010) has enabled the formulation of genome-scale kinetic models which are consistent with stoichiometric, thermodynamic and physiological constraints (Chakrabarti et al., 2013; Stanford et al., 2013; Miskovic et al., 2015; Ataman and Hatzimanikatis, 2015). Constraint based methods, such as Flux Balance Analysis (FBA), are commonly used to estimate the flow of metabolites through such metabolic networks, making it possible to identify the ranges of flux values that are consistent with the studied conditions (Palsson, 2006; O'Brien et al., 2013; Orth et al., 2010). In the absence of explicit experimental information for key intracellular reactions, a situation commonly encountered when considering a new organism or strain, fluxes are allowed to vary between physiologically relevant bounds, often derived from literature, which leads to underdetermined problem formulations. The lack of explicit experimental information propagates as uncertainty in the estimated Gibbs free energy of reaction, $\Delta_r G'$, and consequently in the determination of reaction directionalities which in turn constrain the flux profiles and the thermodynamically feasible concentration ranges (Ataman and Hatzimanikatis, 2015; Soh and Hatzimanikatis, 2014). This results in the existence of numerous alternative internal flux distributions that satisfy the model's constraints and can achieve the same optimum (Orth et al., 2010; Lee et al., 2006; Soh et al., 2012). The size of the resulting solution space, comprising the gamut of alternate optimal flux distributions, can be viewed as a representation of the uncertainty in predicting an exact intracellular state relevant to the studied physiology (Binns et al., 2015; Reed, 2012). Therefore knowledge of which metabolites, if measured, would contribute the most to reducing this uncertainty can significantly improve our ability to define the internal state of a cell.

The space of steady state flux solutions has attracted significant scientific interest in a number of studies that either attempted to attribute biological significance to characteristics of the solution space

* Corresponding author.
E-mail address: vassily.hatzimanikatis@epfl.ch (V. Hatzimanikatis).

http://dx.doi.org/10.1016/j.ymben.2016.11.006
Received 17 March 2015; Received in revised form 7 November 2016; Accepted 10 November 2016
Available online 12 November 2016

(Kelk et al., 2012; Trinh et al., 2009) or attempted to achieve more accurate flux distributions by incorporating additional physicochemical information in the form of constraints based on thermodynamic properties (Henry et al., 2007), molecular crowding considerations (Vazquez and Oltvai, 2011) and carbon labeling information (Chen et al., 2011; Masakapalli et al., 2010; Quek et al., 2010). However the problem of identifying network components with a significant contribution towards the observed uncertainty remains far from trivial (Hadlich et al., 2011). Approaches that have been employed to study the sensitivity of FBA outputs include one-factor-a-time numerical perturbation simulations (Edwards and Palsson, 2000; Mahadevan et al., 2002; Segre et al., 2002) or derivative based local measures (Klier, 2012; Hoppe et al., 2007), which almost exclusively study the effect of the model constituents on the value of the objective function (Chen et al., 2011; Masakapalli et al., 2010; Hadlich et al., 2011). Whilst the proposed approaches have provided valuable qualitative information, they have not been able to provide a functional correlation with metabolites that exert strong influence on the qualitative and quantitative behavior of a particular reaction (or a subset of reactions).

Thermodynamics-based flux analysis (TFA) (Ataman and Hatzimanikatis, 2015), also referred to as thermodynamics-based metabolic flux analysis (TMFA) (Henry et al., 2007) and thermodynamics-based flux balance analysis (TFBA) (Soh and Hatzimanikatis, 2014; Soh et al., 2012) was introduced by Hatzimanikatis and colleagues and allows integration of metabolomics data into FBA models (Ataman and Hatzimanikatis, 2015; Soh et al., 2012). Apart from enforcing compliance with the second law of thermodynamics, TFA effectively reduces the size of the solution space by selecting only the thermodynamically feasible sets of the following biochemical and thermodynamic variables: reaction directionalities, net fluxes through a reaction, $\Delta_r G'$ and metabolite concentrations (activities). The permissible ranges of these biochemical variables can be explored within TFA using linear optimization in a methodology defined as Thermodynamic Variability Analysis (TVA) (Soh and Hatzimanikatis, 2014; Henry et al., 2007). In TVA, the ranges for the activity of each metabolite, the $\Delta_r G'$ and the net flux of each reaction are estimated through their minimization and maximization subject to the thermodynamic and mass balance constraints. Consequently we can estimate whether a reaction is thermodynamically reversible under the studied physiological conditions and whether it operates near or far from thermodynamic equilibrium.

Knowledge of the metabolite concentrations for all the products and substrates of a reaction allows the exact estimation of its thermodynamic properties such as $\Delta_r G'$ and displacement from thermodynamic equilibrium (denoted as Γ (Miskovic and Hatzimanikatis, 2011)). The uncertainty in the exact values for metabolite concentrations, either due to experimental variability or due to lack of experimental data, propagates through the thermodynamic constraints introduced by TFA to uncertainty in the thermodynamic properties of the reactions comprising a metabolic network. In the present work we propose a new framework that combines TFA, Design of Experiments (DoE) and Global Sensitivity Analysis (GSA) methods in order to determine and quantify how variability in metabolite concentrations propagates to uncertainty in the thermodynamic properties of any reaction within a metabolic network. TMSA ranks the various metabolites according to their ability to reduce variability in the thermodynamic properties of reactions, and thus it enables the targeted reduction of uncertainty (Soh and Hatzimanikatis, 2014; Henry et al., 2007). The metabolites with the largest contribution to the observed uncertainty can be considered as high priority targets for experimental analysis. Consequently, TMSA can fully exploit and even upgrade the value of metabolomics data, an important, genome-scale source of experimental information (Lee et al., 2006). Moreover, quantification of the propagation of uncertainty from metabolite concentrations to systemic properties can be used for the design of metabolic engineering strategies and the formulation of hypotheses in many problems related to metabolism.

2. Methods

Below we present an overview of the basic concepts of the proposed methodology using a simple example before presenting each step in greater detail. Let us consider the case of a reversible reaction involving two substrates, which react towards two products:

$$A + B \leftrightarrow C + D \tag{1}$$

The $\Delta_r G'$ of this reaction can be estimated (Henry et al., 2007) as:

$$\Delta_r G' = \{\Delta_f G'_C + \Delta_f G'_D - \Delta_f G'_A - \Delta_f G'_B\} + RTln(C{\cdot}D) - RTln(A{\cdot}B) \tag{2}$$

where ($\Delta_f G'_i$) denotes the formation energy of the participating species, R is the universal gas constant and T is the absolute temperature. The formation energy terms ($\Delta_f G'_i$) can be retrieved from relevant literature or estimated using appropriate computational methods (Jankowski et al., 2008). Thus the only unknowns left in Eq. (2) are the concentrations of the four metabolite species participating in the reaction (A, B, C and D). We want to examine whether it is necessary to measure all four concentrations in order to define the $\Delta_r G'$ for reaction (1) within an acceptable uncertainty range. Additionally, if we can achieve the desired accuracy with fewer measurements, we want to know the order in which the participating metabolites should be measured.

Intuitively one would start by testing how different concentrations of the four participating metabolite species (A, B, C and D) affect the $\Delta_r G'$ of reaction (1). This concept can be formalized by considering it in an *uncertainty analysis* setting. In this context, the concentrations of the metabolite species can be considered as uncertain input factors (X) with a predefined range while the thermodynamic properties of reaction (1) will constitute the studied output variables (R_T). If D_T defines the variance observed in the studied output, we are interested in estimating the fractional contributions of the variability of metabolites i towards the variance of the observed output (D_T), which are defined as *fractional variances* d_A, d_B, d_C and d_D, such that:

$$D_T = d_A + d_B + d_C + d_D. \tag{3}$$

Therefore the experimental quantification of metabolite species with higher fractional variance, d_i, can be prioritized over metabolites with a lesser effect.

For single reactions or small subsets of reactions (where the total number of metabolites is ~ < 4) the effects of metabolite concentrations on the studied outputs can be assessed through a small set of randomized or heuristic trial and error model evaluations. However for larger subsets of reactions, pathways or genome scale models the effects of metabolite uncertainty, i.e. the fractional variances of the metabolites, propagate in a non-linear manner. Therefore we need a robust model analysis method able to identify, quantify and analyze such network-level, non-linear effects. Sensitivity Analysis, and in particular Global Sensitivity Analysis (GSA), methods are frequently the methods of choice when studying complex non-linear models due to their ability to differentiate between first and higher order (non-linear) sensitivity (Saltelli, 2008; Chan et al., 1997). Current state-of-the-art Global Sensitivity Analysis (GSA) methods are not directly applicable in a TFA setting due to difficulties in the implementation of commonly used numerical techniques, such as the estimation of high dimensional integrals through their Monte Carlo approximation (see Supplemental Material). In order to circumvent this limitation we developed our method based on the methodology introduced by Kiparissides et al. (2014) that employs Design of (*in silico*) Experiments (DoE) for the estimation of first and higher-order sensitivity measures.

Instead of measuring the effect of one-metabolite-at-a-time we use an experimental design to study how different combinations of concentration values affect the thermodynamic properties of reaction (1).