Metabolic Engineering 13 (2011) 125–137

Contents lists available at ScienceDirect

Metabolic Engineering

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

A flexible state-space approach for the modeling of metabolic networks I: Development of mathematical methods

Adam C. Baughman [a,b], Susan T. Sharfstein [a,b,c,1], Lealon L. Martin [a,b,*]

[a] Department of Chemical & Biological Engineering, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180, USA
[b] Center for Biotechnology & Interdisciplinary Studies, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180, USA
[c] Department of Biology, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180, USA

ARTICLE INFO

Available online 14 December 2010

Keywords:
Metabolic modeling
Network topology
Global optimization

ABSTRACT

We introduce a novel, flexible, optimization-based mathematical framework for the modeling of arbitrarily complex metabolic networks: topological metabolic analysis (TMA). The framework is adapted from state-space approaches used by Manousiouthakis and co-workers for the representation of complex heat- and mass-exchanger networks. We offer a thorough discussion of the mathematics and general theory underlying the framework, and discuss certain mathematical advantages of our modeling representation in comparison with other commonly used techniques (MFA and FBA). We employ a novel aggregate objective function for use with our basic constraint model, including a generalized least-squares term (for fitting available experimental measurements) and a linear design term (for representing biological or engineering goals). Using a case-study taken from recent literature (McKinlay et al., 2007), we demonstrate (among other benefits) the ability of this objective to identify alternate distinct-yet-equally optimal solutions for a given modeling problem. We also show that these solutions, obtained using only external metabolite uptake and secretion measurements, provide useful biological insights and compare favorably with solutions obtained on the basis of ^{13}C isotope-tracing data.

1. Introduction

The need for quantitative and robust metabolic modeling approaches continues to be driven by an ever-expanding family of high-value industrial applications for both prokaryotic and eukaryotic organisms. Among these applications are the use of microbial, plant-cell, and mammalian-cell cultures in the pharmaceutical industry for the biological manufacture of proteins and other active therapeutic compounds (Dinnis and James, 2005; Chiba and Jigami, 2007; Hamilton and Gerngross, 2007; Jain and Kumar, 2008; Pscheidt and Glieder, 2008; Boghigian et al., 2010), and the use of microbial-cell cultures to generate biogas (Igoni et al., 2008; Cantrell et al., 2008), hydrogen gas (Liu and Fang, 2003; Datar et al., 2007; Porwal et al., 2008), bioethanol (Hatzimanikatis et al., 1998; Otero et al., 2007), and bioplastic materials (Verlinden et al., 2007; Munoz and Riley, 2008). In each case, the network of metabolic reactions within the chosen organism is responsible for executing the desired chemical synthesis.

To improve these processes, it is therefore desirable to engineer these biological networks just as one would for non-biological chemical process networks. As a result of evolutionary pressures, metabolic networks are intricate and robust systems capable of mitigating the overactivity or underactivity of many different reactions. Consequently, modifications to these networks many times do not elicit the desired engineering outcome (Bailey et al., 2002). Detailed and quantitative methods for modeling metabolic networks of arbitrary size and complexity are useful tools to help elucidate complex network responses, thereby helping to guide more effective engineering.

Many approaches currently in use for the modeling of metabolic networks are based upon the stoichiometric balancing method advanced in the early 1990s by Palsson and co-workers (Savinell and Palsson, 1992b,a; Varma and Palsson 1994b, a). Such models are generally expressed in the following form:

$$\underline{\underline{s}}\,\underline{\mathcal{R}} = \underline{k}. \tag{1}$$

The $\mathcal{N} \times \mathcal{M}$ matrix $\underline{\underline{s}}$ stores stoichiometric coefficients describing a system of $\mathcal{M}$ metabolic reactions involving $\mathcal{N}$ metabolites. The $\mathcal{M} \times 1$ vector $\underline{\mathcal{R}}$ describes the overall reaction rates (fluxes) through each of the $\mathcal{M}$ reactions. The $\mathcal{N} \times 1$ vector $\underline{k}$ describes the net rate at which each of the $\mathcal{N}$ metabolites enter or leave the network.

* Corresponding author at: Department of Chemical Engineering, Cockrell School of Engineering, University of Texas at Austin, 1 University Station, Austin, TX 78712, USA.
E-mail address: llmartin@che.utexas.edu (L.L. Martin).
[1] Current address: College of Nanoscale Science and Engineering, University at Albany, 257 Fuller Road, Albany, NY 12203, USA.

1096-7176/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymben.2010.12.002

Nomenclature

Mathematical symbols

$\mathcal{M}$	the total number of reactions in the metabolic network
$\mathcal{N}$	the total number of distinguishable metabolites in the metabolic network
$\underline{\underline{s}}$	the $\mathcal{N} \times \mathcal{M}$ matrix of stoichiometric coefficients $s_{i,j}$
$s_{i,j}$	the stoichiometric coefficient for metabolite i in reaction j
$\underline{\mathcal{R}}$	the $\mathcal{M} \times 1$ vector of overall reaction rates $\mathcal{R}_j$
$\mathcal{R}_j$	the overall rate of reaction j
$\underline{\underline{\mathcal{R}}}$	the $\mathcal{M} \times \mathcal{N}$ matrix of metabolite-specific reaction rates $\mathcal{R}_{i,j}$
$\mathcal{R}_{i,j}$	the rate at which metabolite i is consumed or generated in reaction j. Also equal to $s_{i,j}\mathcal{R}_j$
$\underline{k}$	the $\mathcal{N} \times 1$ vector of net metabolite uptake or secretion rates
$\underline{\mathcal{U}}_i$	the vector of flow rates for the input/uptake streams of family/metabolite i
$\mathcal{U}_i$	the net uptake rate of metabolite i
$\underline{\alpha}_i$	the vector of intensive qualities for the input/uptake streams of family/metabolite i
$\underline{\mathcal{Y}}_i$	the vector of flow rates for the output/secretion streams of family/metabolite i
$\mathcal{Y}_i$	the net secretion rate of metabolite i
$\underline{\beta}_i$	the vector of intensive qualities for the output/secretion streams of family/metabolite i
$\{\underline{D}\}$	the set of parameter vectors $\underline{D}_j$
$\underline{\underline{\sigma}}$	the matrix of flow rates for the substrate streams
$\sigma_{i,j}$	the total flow rate of metabolite i into reaction j
$\underline{\underline{\gamma}}$	the matrix of intensive qualities for the substrate streams
$\underline{\underline{\pi}}$	the matrix of flow rates for the product streams
$\pi_{i,j}$	the total flow rate of metabolite i exiting reaction j
$\underline{\underline{\delta}}$	the matrix of intensive qualities for the product streams
Φ	the set of all network metabolites
Φ_U	the set of all network metabolites permitted to enter the network (i.e. are consumed by the network)
Φ_Y	the set of all network metabolites permitted to leave the network (i.e. are secreted by the network)
Γ	the set of all network reactions
$\underline{w}_i$	the vector of all flow rates $w_{i,j}$
$w_{i,j}$	the flow of metabolite i directly to reaction j from an external source
$\underline{x}_i$	the vector of all flow rates $x_{i,j}$
$x_{i,j}$	the flow of metabolite i directly to an external sink from reaction j
$\underline{\underline{\tau}}^i$	the matrix of all flows $\tau_{i,j,k}$
$\tau_{i,j,k}$	the flow of metabolite i directly from reaction j to reaction k
b_i	the flow of metabolite i directly from an external source to an external sink
m	the total number of constraints in the network model
n	the total number of variables in the network model
$\underline{q}$	the $n \times 1$ vector of all TMA model variables
$\mathcal{Z}$	the number of individual fitting goals within the generalized least-squares objective term
$\underline{\underline{\mu}}$	the $\mathcal{Z} \times \mathcal{Z}$ matrix whose zth diagonal element is μ_z
μ_z	the arbitrary weight applied to the zth fitting goal
$\underline{\underline{g}}$	the $\mathcal{Z} \times n$ matrix of vectors $\underline{g}_z$
$\underline{g}_z$	the zth $1 \times n$ vector such that $\underline{g}_z^T\underline{q} = d_z$
$\underline{d}$	the $\mathcal{Z} \times 1$ vector of experimental measurement values d_z
d_z	the zth experimental measurement value

Metabolite abbreviations

Ace	acetate
AcCoA	acetyl-coenzyme-A
ATP	adenosine-triphosphate
Ala	alanine
Asp	aspartate
CO_2	carbon-dioxide
Cys	cysteine
E4P	erythrose-4-phosphate
EtOH	ethanol
F6P	fructose-6-phosphate
For	formate
Fum	fumarate
Glc	glucose
G3P	glyceraldehyde-3-phosphate
Glxt	glyoxylate
Gly	glycine
His	histidine
Ile	isoleucine
Leu	leucine
Lys	lysine
Mal	malate
Met	methionine
NADH	reduced-nicotinamide-adenine-dinucleotide
NADPH	reduced-nicotanamide-adenine-dinucleotide-phosphate
OXA	oxaloacetate
Phe	phenylalanine
Pyr	pyruvate
PEP	phospho-enol-pyruvate
R5P	ribose-5-phosphate
S7P	sedoheptulose-7-phosphate
Ser	serine
Suc	succinate
Thr	threonine
Val	valine

Implicit in this formulation is a so-called "pseudo-steady-state" assumption, which posits that the rate at which intracellular concentrations of metabolites increase or decrease is quite slow relative to the rate at which metabolites are consumed or generated in metabolic reactions. While this assumption has generally not been rigorously tested, there is some support for its validity.

In the absence of stresses or perturbations to the cell, it has been shown that intracellular concentrations for many metabolites change at a rate comparable to that of overall cell growth (i.e. biomass accumulation) (Hans et al., 2003; Hoque et al., 2005; Chrysanthopoulos et al., 2010). Even for rapidly growing organisms, the cell growth rate is an order of magnitude slower than the rate of many metabolic reactions. Moreover, it has been recently shown that the majority of intracellular metabolites in *Escherichia coli* are present in concentrations well above the known or assumed Michaelis constants (K_m) of more than 300 metabolic enzymes (Bennett et al., 2009), suggesting that many metabolic reactions typically operate at their maximum (or saturated) rates.

Because metabolites routinely participate in multiple reactions within the same network, $\underline{\underline{s}}$ will usually have many more columns than rows, and the system of equations described by Eq. (1) will usually be underdetermined. Networks incorporating branched,