Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computers and Chemical Engineering 29 (2005) 459–466
Computers & Chemical Engineering
www.elsevier.com/locate/compchemeng

Closing the loop between feasible flux scenario identification for construct evaluation and resolution of realized fluxes via NMR☆

S. Ghosh[a], T. Zhu[a], I.E. Grossmann[a], M.M. Ataai[b], M.M. Domach[a,*]

[a] *Department of Chemical Engineering, Carnegie Mellon University, 3102A Doherty Hall, Pittsburgh, PA 15213, USA*
[b] *Department of Chemical Engineering, The University of Pittsburgh, USA*

Received 20 October 2003
Available online 30 November 2004

Abstract

Mixed integer linear programming (MILP) and other methods can provide information on product yield horizons and the optimal "target" values of fluxes to achieve in metabolic networks. These methods can also elucidate the metabolite trafficking possibilities in tissues. While methods that enumerate flux scenarios are useful for the "front-end" analysis of metabolic engineering problems or probing the possibilities in tissue physiology, the evaluation of actual data from a tissue or engineered cells is ultimately needed to verify the veracity of theoretical predictions and mechanistic hypotheses. However, using NMR data to identify fluxes can be problematic because a non-linear optimization problem can result that is highly non-convex due to the bi-linear and tri-linear terms present in the isotopomer balance equations. One consequence is that local versus global solutions can be found. To surmount the problem of obtaining local solutions, we have investigated the utility of using a joint problem formulation. It involves using the results of the "front-end" analysis (MILP solutions) to also provide bounds for the data-to-fluxes problem, where using a deterministic global optimization algorithm of the branch-and-bound type solves the latter. The joint formulation produced the correct solutions for a streamlined example and a more realistic problem. An alternate gradient-based algorithm failed in that it produced an incorrect solution. The utility of different NMR analytes for yielding correct fluxes was also investigated via simulation.

Keywords: Metabolic engineering; Flux analysis; Global optimization; NMR; Mixed integer linear programming (MILP)

1. Introduction

The modeling and optimization of metabolic networks have important applications in pharmaceutical and biotechnology research and development. One application in metabolic engineering is to determine the rates at which material and energetic resources flow (i.e., "fluxes" in units of mmol/(g cell h)) within a metabolic network. These flux distributions provide information on cell and product yield horizons (e.g., Majewski & Domach, 1990; Varma & Palsson, 1994). Furthermore, flux analysis enables one to examine the possibilities of optimally redirecting the fluxes towards a desired product. For pharmaceutical research and development, metabolic fluxes provide insights on how raw materials normally traffic within tissue cells. With such baseline information in hand, an improved basis then exists for understanding the "disease state" (e.g., Cohen, 1987). Moreover, because many medicinal compounds directly or ultimately (e.g. via receptor-mediated processes) alter the rate of biochemical reactions, having baseline information on flux distributions is useful for assessing and documenting the impact of metabolic effectors that may alleviate or otherwise alter the "disease state" (e.g., Cohen, Werrmann, & Tota, 1998).

One can use linear programming (LP) methods to identify the flux sets that satisfy an objective (e.g. product maximization or by-product minimization) and constraints such as min-

☆ Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the EPA or the National Science Foundation.

* Corresponding author. Tel.: +1 412 268 2246; fax: +1 412 268 7139.
E-mail address: md0q@andrew.cmu.edu (M.M. Domach).

0098-1354/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compchemeng.2004.08.027

imum or fixed amount of ATP must be produced for energy (e.g., Phalakornkule et al., 2001). When a specific objective is not used, these methods can also identify all feasible flux scenarios that, for example, satisfy stoichiometric balances and physiological constraints (e.g., Varma & Palsson, 1994). LP-based and other solution enumeration methods are now easier to apply due to the advent of "friendly" user interfaces that allow one to generate and edit a model using a "picture" drawn on a computer screen as opposed to programming (e.g., Zhu et al., 2003).

While LP-based and other methods are useful for the "front-end" evaluation and/or design phases of metabolic engineering or probing the possibilities in tissue physiology, the evaluation of actual data from a tissue or engineered cells is ultimately needed to verify the veracity of theoretical predictions and mechanistic hypotheses. In a sense, for metabolic engineering the data-to-fluxes step closes an information feedback "loop" in a design problem where preanalysis suggests what trafficking scenario should be attained and in some cases, what mutations are needed to yield the desired end point. An example of the latter is a zero flux in an optimal flux set implies that an engineered cell should loose an enzyme activity via gene deletion, an insertional mutation, or other means. An information "loop" exists because if the data suggests that the expected fluxes are actually realized, then the confidence is elevated that the system model and assumptions are correct. Correctness, in turn, not only solidifies the conclusions; the intellectual property and patent prospects are also strengthened when commercialization is the goal. If instead a disparity results between theoretical predictions and fluxes, this information can be used for model refinement and assumption re-evaluation.

Either mass or ^{13}C NMR spectroscopy is used to measure the isotopic abundance of metabolites derived from a labeled precursor (e.g. 1-^{13}C glucose). For ^{13}C NMR spectroscopy, the fine structure of an NMR spectrum is primarily determined by the fractional ^{13}C enrichment of the isotopomers of the signal molecule(s) chosen, and, as such, the spectrum is directly related to the intracellular reaction fluxes within the metabolic network.

This paper addresses some of the challenges associated with solving the "inverse problem", i.e. closing the information "loop," which entails determining the fluxes from NMR data. As will be reviewed, the problem corresponds to a non-linear optimization problem, which is highly non-convex due to the existence of bi-linear and tri-linear terms present in the isotopomer balance equations. Therefore, conventional gradient-based algorithms are likely to converge to local solutions (Horst & Tuy, 1996). We have attempted to address this problem with two linked steps. First, MILP or depth first search (DFS) solutions to a metabolic network problem are obtained beforehand to bound the domain of each net intracellular reaction rate (Lee, Phalakornkule, Ataai, Domach, & Grossmann, 2000; Zhu et al., 2003). The solutions for yield evaluation and/or flux optimization can also be easily obtained from this formulation simply by incorporating the appropriate objective function. Then, using the information on flux bounds, we have implemented a deterministic global optimization algorithm of the branch-and-bound type, BARON (Sahinidis & Tawarmalani, 2002) to circumvent the problem of local convergence. Thus, solving the flux trafficking problem can assist with "front-end" evaluation as well as closing the information "loop".

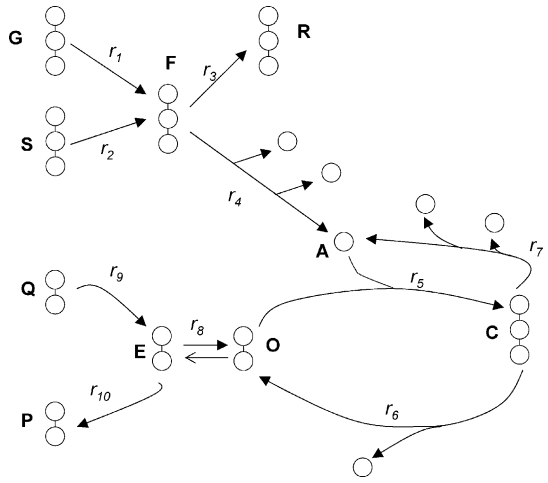

Fig. 1. Abbreviated pathway model adopted from Forbes et al. (2001).

We proceed by first using a simplified example to illustrate the problem and the joint MILP and inverse formulations. We consider a hypothetical experiment involving the network shown in Fig. 1 (adapted from Forbes, Clark, & Blanch, 2001). A recursive mixed integer linear programming (MILP) approach (Lee et al., 2000) has been adopted to first determine the different feasible flux trafficking alternatives within the cell. These solutions are essentially the alternate optima of the LP obtained through flux balance and measured rate constraints. These flux solutions provide suitable bounds for attempting to solve the non-linear, inverse problem of determining the correct fluxes from NMR experimental data. As will be shown, the NMR computations are an adaptation of the isotopomer mapping matrix method described by Schmidt, Carlsen, Nielsen, and Villadsen (1997) and Schmidt, Nielsen, and Villadsen (1999). We then present an application of the formulation to solving a larger NMR data-to-fluxes problem that involves *E. coli* metabolism. The problem entails using NMR data to identify the fluxes in an *E. coli* mutant that has had pyruvate kinase activity deleted as a strategy for eliminating acid production (Goel, Lee, Domach, & Ataai, 1995; Lee et al., 2000).

2. A streamlined example

2.1. Overview of network and data available

The network structure is shown in Fig. 1, where r represents a net flux (e.g., mmol/(g cell h)). To determine the fluxes, one can typically make use of the two types of mea-