Biochemical Engineering Journal 90 (2014) 121–130

Contents lists available at ScienceDirect

Biochemical Engineering Journal

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

Regular Article

Complementary elementary modes for fast and efficient analysis of metabolic networks

Md. Bahadur Badsha [a], Ryo Tsuboi [a], Hiroyuki Kurata [a,b,*]

[a] Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan
[b] Biomedical Informatics R&D Center, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan

ARTICLE INFO

Article history:
Received 19 January 2014
Received in revised form 30 May 2014
Accepted 31 May 2014
Available online 6 June 2014

Keywords:
Bioreactors
Elementary mode
Metabolic network
Metabolite over production
Microbial growth
Modelling

ABSTRACT

Metabolic pathway analysis facilitates understanding or designing a complex metabolic system and enables prediction of steady-state metabolic flux distributions. A serious problem of elementary mode (EM) or extreme pathway (Expa) analysis is that the computational time increases exponentially with an increase in network sizes, which makes the computation of the EMs/Expas expensive and infeasible for large-scale networks. To overcome such problems, we proposed a fast and efficient algorithm named complementary EM (cEM) analysis. To achieve the computational time improvement, we employ the EM decomposition method that explores EMs or linear combinations of them which are responsible for the metabolic flux distributions. Flux balance analysis (FBA) is used to determine possible ranges of metabolic flux distributions as the input data necessary for the EM decomposition method. The maximum entropy principle (MEP) is employed as an objective function for estimating the coefficients of cEMs. To demonstrate the feasibility of cEM analysis, we compared it with EM/Expa analysis by using two medium-scale metabolic networks of *Escherichia coli* and a genome-scale metabolic network of head and neck cancer cells. The cEM analysis greatly reduces the computational time and memory cost, exposing a new window for large-scale metabolic network analysis.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Systems biotechnology is an approach to develop comprehensive and ultimately predictive models of how components of a biological system reproduce its observed behavior. Mathematical modeling was proven successful when applied to relatively small-scale systems, while applications to large-scale models are being challenged by the technical advances that generate high-dimensional and high-throughput data [1]. Since cellular metabolic and regulatory networks are often large and complex, the construction and analysis of their computational models can be useful for identifying physiological states and evaluating the effects of network perturbations on desired phenotypes. Recently, genome-scale computational models have gained increasing prominence and importance; capturing stoichiometric models with thermodynamic constraints have been published for over 30 organisms [2] ranging from relatively simple prokaryotes such as *Escherichia coli*, to complex eukaryotes such as *Homo sapiens* [3,4].

Metabolic pathways are complex in every living cell, where a coherent set of enzymes catalyze various biochemical reactions [5–7]. Pathway-based analysis generally employs a constraint-based modeling approach [8], e.g., FBA that uses a stoichiometric matrix and an objective function to define a network's allowable solution space. The target flux capacity is provided by optimizing a specific objective function such as cell growth, energy, biomass, adenosine triphosphate (ATP) production or metabolite synthesis [9,10].

Metabolic pathway analysis has focused on two approaches, elementary modes (EMs) [11] and extreme pathways (Expas) [12]. EMs are a minimal set of reactions that can operate in a steady state, while Expa analysis contains one additional constraint to make all Expas systematically independent. EM analysis allows one to systematically enumerate all independent minimal pathways that are stoichiometrically and thermodynamically feasible and to offer great opportunities for studying functional and structural properties of metabolic pathways [13–15]. The EM-based enzyme control flux (ECF) and genetic modification of flux (GMF) are very effective in correlating transcriptome or proteome data to their

* Corresponding author at: Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan.
Tel.: +81 948 29 7828; fax: +81 948 29 7828.
E-mail addresses: m791703b@bio.kyutech.ac.jp, mbahadur_stat_ru@yahoo.com (Md.B. Badsha), m673216r@bio.kyutech.ac.jp (R. Tsuboi), kurata@bio.kyutech.ac.jp (H. Kurata).

http://dx.doi.org/10.1016/j.bej.2014.05.022
1369-703X/© 2014 Elsevier B.V. All rights reserved.

associated metabolic network architecture or flux distributions [16–19]. A serious problem of EM analysis is that the computational time increases exponentially with an increase in network sizes, which makes the computation of the EMs very expensive and infeasible for large-scale networks [20–23]. For example, central metabolism of *E. coli* model with 112 reactions has more than two million EMs. When possible substrates are extended, the number of EMs increases to more than 26 million [24]. Thus, the huge computation time and memory storage are required to enumerate all the EMs of large-scale or genome-scale metabolic networks. To overcome such problems, distributed memory parallelization and parallel processing have been employed together with compression of the stoichiometric matrix [25,26] or with removal of biologically infeasible solutions [27]. On the other hand, alternative approaches have been presented that do not enumerate the whole set of EMs [28–30]. The EM decomposition method [29] was reported to pick up major EMs or linear combinations of EMs responsible for the flux distributions for metabolic networks, while the entire flux distributions must be input for EM decomposition.

The EM coefficients (EMCs), which indicate the quantitative contribution of their associated EMs, can be estimated by maximizing a particular objective function. While EMs can be described by many possible scalar products of each EM, the predicted fluxes must be consistent with respect to all of them. The linear programming (LP) method is often used, where the maximum biomass and specific metabolite formation are selected as objective functions [31]. Such objective functions relate to the optimum physiological states, but they are not provided for many organisms. The quadratic programming (QP) could optimize EMCs by defining the objective function as the minimal norm of the EMCs, but QP has neither a physical nor a biological background and is still restricted to relatively small-scale networks [32]. A serious problem is that QP depends on scalar products of each EM. Therefore, the QP method may not be valid for optimizing the EMCs [33]. The linear programming denoted as ECFLP (enzyme control flux linear programing) [16] maximizes and minimizes each EMC to represent its available ranges in the same manner of the α-spectrum method [34], averaging all the estimated EMCs. It is practically useful, but it has neither a biological nor a theoretical background. To obtain reliable EMCs, we proposed the MEP algorithm [17,19], which is a universal principle established based on Shannon entropy [35] when insufficient information is available. MEP readily optimizes hundreds of thousands of the EMCs in large-scale networks. MEP is convenient in cases where no biological objective function is available and it does not depend on the scalar product of each EM.

In this paper, we propose complementary elementary mode (cEM) analysis to improve the calculation speed and efficiency of EM/Expa analyses. The cEM analysis consists of the EM decomposition method coupled with FBA and the MEP-based optimization. FBA is used to determine many possible ranges of metabolic flux distributions necessary for the EM decomposition method. MEP optimizes the coefficients of cEMs to predict the flux distributions. To demonstrate the feasibility of cEM analysis, we compared it with EM/Expa analyses by using two medium-scale metabolic networks of *E. coli* and one genome-scale metabolic network model of head and neck cancer cells. The cEM analysis remarkably reduces the computational time and memory cost without requiring the generation of a full set of EMs nor any biological objective function.

2. Materials and methods

The algorithm for cEM analysis in a given steady-state metabolic network is presented as shown in Fig. 1. The cEM analysis consists of three steps: generation of many flux distributions by FBA, cEM extraction by the EM decomposition method, and flux prediction by MEP. To demonstrate the feasibility of cEM analysis, we use a simple example model as shown in Fig. A.1 (Supplementary data A), two medium-scale metabolic networks of *E. coli* [36] and one genome-scale metabolic network model [37]. The mathematical procedure of cEM analysis is intelligibly illustrated (Supplementary data B).

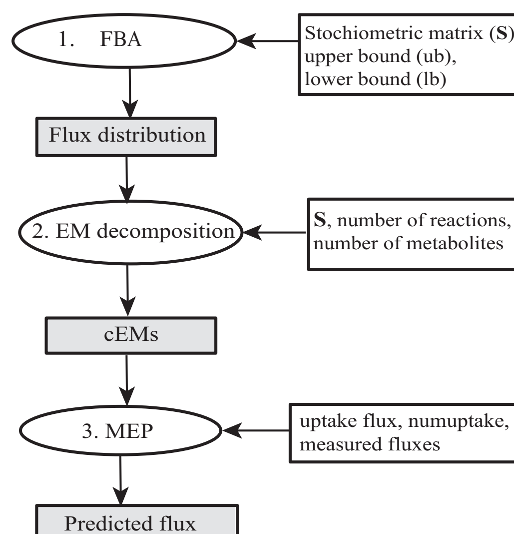

Fig. 1. A flow chart of cEM analysis. The white and gray square boxes are the data. The ovals are the algorithms.

2.1. Elementary mode (EM) analysis

Metabolic networks can be represented by a stoichiometric matrix **S**. The rows and columns of **S** correspond to the metabolites and reactions, respectively. At the steady-state, the flux-balance equation is given by:

$$\mathbf{S} \cdot \mathbf{v} = 0, \tag{1}$$

where $\mathbf{v} = (v_1, v_2, \ldots, v_n)^t$ is the vector whose elements correspond to metabolic fluxes and n is the number of reactions. The set of all possible solutions to Eq. (1) can be described by a set of basis vectors. EM matrix **P** is uniquely determined from the stoichiometric matrix and the flux vector, as follows:

$$\mathbf{v} = \mathbf{P} \cdot \lambda, \tag{2}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)^t$ is the EMC vector and m is the number of EMs. The ingredients of these vectors and matrix are displayed as follows:

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1m} \\ e_{21} & e_{22} & \cdots & e_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nm} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} \tag{3}$$

The ith column for the **P** matrix is the ith EM vector: $\boldsymbol{e}_i = (e_{1i}, e_{2i}, \ldots, e_{ni})^t$. The flux distribution can be also represented as superposition of the EM vectors with non-negative EMCs as follows:

$$\mathbf{v} = \sum_{i=1}^{m} \lambda_i \mathbf{e}_i \tag{4}$$

Expa/cEM analyses were performed in the same manner as EM analysis, where the Expas/cEMs are employed instead of the EMs in Eqs. (2)–(4). Expa analysis splits only the internal reversible reactions into two irreversible reactions while not decomposing reversible exchange reactions. Such an additional constraint makes

Download English Version:

https://daneshyari.com/en/article/3000

Download Persian Version:

https://daneshyari.com/article/3000

Daneshyari.com