Complementary elementary mode analysis for large-scale metabolic networks

MD. BAHADUR BADSHA[1] , RYO TSUBOI[1], HIROYUKI KURATA[1,2]

Metabolic pathway analysis facilitates understanding a complex metabolic system and enables prediction of steady-state metabolic flux distributions through elementary mode (EM) analysis. The principal drawback of the ordinary EM analysis is that the number of EMs suffers from a combinatorial explosion, and the use of complete sets of EMs gives rise to problems with scalability when applied to large-scale network models. The current problem is that many organisms still do not provide any specific objective biological function to predict the unknown metabolic fluxes. Since EMs can be described by many scalar products of each EM, the predicted fluxes should be consistent with respect to all of them. To overcome the existing problem, we proposed a fast and efficient EM algorithm named the complementary EMs (cEM). This study opens a new framework for a large-scale metabolic network, which neither requires the initial generation of a full set of EMs nor any objective biological function.

1. Introduction

Network-based metabolic pathway analysis has focused on two approaches, elementary modes (EMs) and extreme pathways. An EM is a minimal set of reactions that can operate in steady state, while the set of extreme pathways is the systemically independent subset of the EMs. The principal drawback of the ordinary EM analysis is that the number of EMs suffers from a combinatorial explosion, and the use of complete sets of EMs gives rise to problems with scalability when applied to large-scale network models. The current problem is that many organisms still do not provide any specific objective biological function to predict the unknown metabolic fluxes. Since EMs can be described by many scalar products of each EM, the predicted fluxes should be consistent with respect to all of them. To overcome the existing problems, we proposed a fast and efficient EM algorithm named the complementary EMs (cEM) analysis.

2. Method

The algorithm of cEM for prediction the flux distributions in a given steady-state metabolic networks is presented in **Fig-1**.

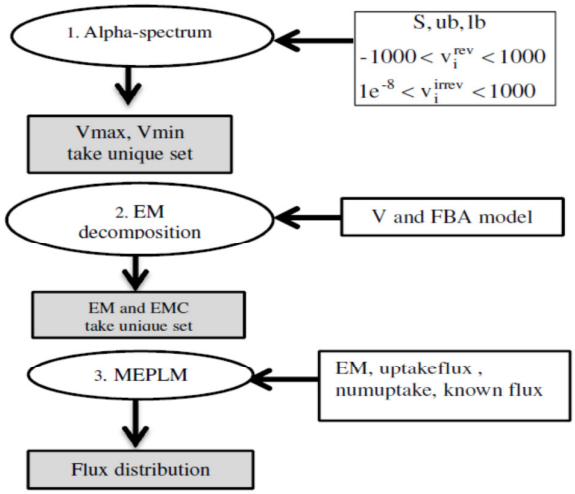

Fig-1: A flow chart of cEM algorithm. The white square boxes are input file; the grey square box is output file. The ovals are algorithms.

2.1 Alpha (α)-spectrum method by FBA

First, we employ the *α*-spectrum [1] method by FBA to determine the possible range of metabolic flux distributions for input of EM decomposition. α-spectrum can be computed even when the flux is partially unknown. The method will obtain the total value of two sets flux distributions and we take the unique set of flux distributions necessary for input of EM decomposition in the next step. Based on α-spectrum the solution of metabolic flux distribution is calculated by maximizing and minimizing each flux in a given steady-state metabolic network as follows:

$$\text{Max and Min} \quad \mathbf{v}_i = (v_1, v_2, ..., v_n)^t \qquad (\text{for } i = 1,2,\ldots,n)$$

$$\text{Subject to}: \mathbf{S}\cdot\mathbf{v} = 0 \qquad \begin{cases} -1000 < v_i^{rev} < 1000 \\ 1e^{-8} < v_i^{irrev} < 1000 \end{cases} \qquad (1)$$

Where, **S**=stoichiometric matrix and **v**=flux vector.

2.2 EM decomposition of metabolic networks

Second, we apply the EM decomposition technique [2], which generates the major EMs responsible for the flux distributions; because computation of the full set of EMs in large-scale metabolic networks still constitutes a challenging issue due to its underlying combinatorial complexity and, in some cases, computationally infeasible for practical purpose. EM decomposition operates by first selecting the reaction with non-zero flux of maximum magnitude from the given unique flux $\boldsymbol{v}$ in the feasible set of optimization problem (1), but it can equally be applied to flux distributions obtained by alternative means. The algorithm then uses mixed integer linear programming (MILP) to find an EM that both contains the selected reaction and is contained in the given distribution. The EM decomposition algorithm takes a flux distribution $\boldsymbol{v}$ as input to obtain major EMs for $\boldsymbol{v}$.

2.3 Maximum entropy principle with Lagrange multipliers (MEPLM)

The EMCs are considered as the contribution that various EMs have on diverse physiological states and it can be estimated by optimization as a particular objective function. The highest

[1] Department of Bioscience and Bioinformatics, Kyushu Institute of Technology 680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan.

[2] Biomedical Informatics R&D Center, Kyushu Institute of Technology 680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan.

EMC, meaning that the related mode is the most relevant for biological interpretation in the given state and 'zero' these modes is not required to describe the present sate. Therefore, the objective function is important for the optimization of EMCs for prediction the flux distribution. But a problem is that many organisms still do not provide any specific objective biological function for estimating the EMCs to prediction the flux distribution relate to the optimum physiological states. Since EMs can be described by many scalar products of each EM, the predicted fluxes should be consistent with respect to all of them.

Third, to overcome this problem, we use the MEPLM [3] as an objective function for estimating the EMCs. MEPLM is convenient in cases where no biological objective function is available and it does not depend on scalar product of each EM. Generally, the flux distribution at steady state can be decomposed onto EMs:

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

$\mathbf{P}_d$ is the sub-matrix of EM matrix $\mathbf{P}$ in which the rows represent the reactions with the determined fluxes and the columns correspond to the elementary modes. $\boldsymbol{\lambda}$ is the EMC vector and $\mathbf{v}_d$ is the flux vector for the determined reactions. The probability of EM is presented as:

$$\rho_j = \frac{1}{v_{substrate\ uptake}} P_{substrate\ uptak,\,j} \cdot \lambda_j \left(\sum_{j=1}^{m} \rho_j = 1 \right) \tag{3}$$

MEPLM is provided by solving the following optimization problem:

$$\text{Maximize} \quad -\sum_{j=1}^{m} \rho_j \log \rho_j$$

$$\text{subject to} \quad \sum_{j=1}^{m} \rho_j = 1 \quad \text{and} \quad \sum_{j=1}^{m} \rho_j x_{d,j} = v_d \quad (d = 1,2,...,nd) \tag{4}$$

where v_d is the dth determined flux; nd is the number of the determined fluxes. MEPLM solve this optimization problem (4) for optimize the EMCs and finally prediction the flux distribution. The prediction error in the flux distributions by ordinary EM and proposed cEM are calculated by:

$$\text{Predection error} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(v_{i,\text{predection}} - v_{i,\exp} \right)^2} \tag{5}$$

3. Results and Discussion

To demonstrate the feasibility of cEM, we compared it with the ordinary EM analysis using simple (results not shown here) and two large-scale metabolic networks of *E.coli*. In model-I, the ordinary EM algorithm performance with generated total 122126 EMs from 156 reactions and 140 metabolites of *E. coli* metabolic network in 500.418 seconds, while the proposed cEM algorithms generated the major 202 EMs, responsible for the flux distributions in only 1.5521 seconds. In model-II, the ordinary EM algorithm performance with generated total 321416 EMs from 157 reactions and 140 metabolites of *E. coli* metabolic network in 2000.728 seconds, while the proposed cEM algorithms generated the major 295 EMs, responsible for the flux distributions in only 1.5580 seconds. The prediction flux results compare with experimental flux [4] shown in **Fig-2** and compare speed and accuracy in **Table-1**.

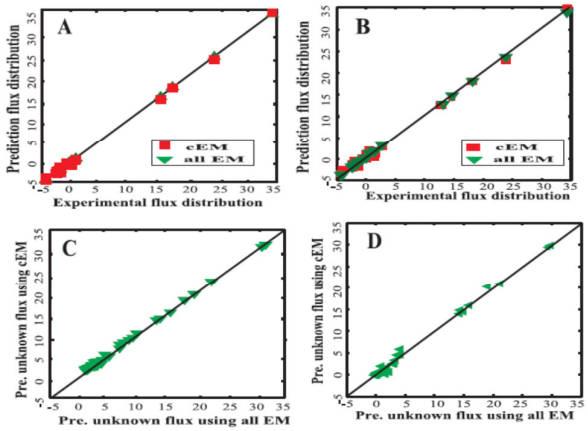

Fig-2: Flux prediction for *E. coli* mutants with experimental flux by both methods (A) model-I, (B) model-II; prediction unknown fluxes (C) model-I and (D) model-II.

Table-1: Accuracy and speed for the prediction of flux distribution by proposed and ordinary method. Calculation accuracy is defined by Eq. (5).

Model	Method	#R	#M	#EM	Time(s)	Accuracy
Model-I	Ordinary EM	156	140	122126	500.418	0.1279
	cEM	156	140	202	1.5521	0.1200
Model-II	Ordinary EM	157	140	321416	2000.728	0.3792
	cEM	157	140	295	1.5580	0.3712

4. Concluding Remarks

The cEM analysis shows similar results and achieve a computational time much shorter than the ordinary EM method but never deteriorated the prediction accuracy to predict the unknown fluxes on large-scale metabolic networks. We anticipate that cEM method exposed new window for large-scale metabolic network, where neither requires the initial generation of a full set of EMs nor any objective biological function, which is memory improvements at the cost of reduced interoperability and maintainability.

Reference

[1]. Wiback SJ, Mahadevan R, Palsson BO: Reconstructing metabolic flux vectors from extreme pathways: defining the alphaspectrum. *J Theor Biol* 2003, **224**:313-324.

[2]. Lp K, Colijn C, Lun DS: Analysis of complex metabolic behavior through pathway decomposition. *BMC Syst Biol* 2011, **5**:91.

[3]. Zhao QY, Kurata H: Use of maximum entropy principle with Lagrange multipliers extends the feasibility of elementary mode analysis. *J Biosci Bioeng* 2010, **110 (2)**: 254–261.

[4]. Siddiquee KA, Arauzo-Bravo MJ, Shimizu K: Effect of a pyruvate kinase (*pykF*-gene) knockout mutation on the control of gene expression and metabolic fluxes in Escherichia coli. *FEBS Lett* 2004, **235**:25–33.