다목적 선형 계획법을 이용한 대사 회로 분석

오영균, 이동엽, 이상엽, 박선원
대전광역시 유성구 구성동 373-1 한국과학기술원 생명화학공학과

Multiobjective Linear Programming in Flux Balance Analysis

Young-gyun Oh, Dong-Yup Lee, Sang Yup Lee, Sunwon Park
Department of Chemical and Biomolecular Engineering,
Korea Advanced Institute of Science and Technology, 373-1
Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea

Introduction

The interpretation and prediction of metabolic flux distributions require mathematical modeling and computer simulation, and there exists a long history of quantitative metabolic modeling (Bailey, J.E., 1998). To deal with the lack of kinetic information, an alternative approach has been used to study feasible and optimal metabolic flux destributions: flux balance analysis (FBA). FBA provides a solution space that contains all possible steady-state distributions satisfing the applied constraints. Subject to the imposed constraints, optimal metabolic flux distributions can be determined from the set of all allowable flux distributions using linear programming (LP). To identify an optimal solution, it is necessary to establish an objective function. A number of different objective functions have been used for metabolic analysis. These include the following: 1) maximize ATP production, 2) minimize nutrient uptake, 3) minimize redox production, 4) maximize metabolite production, 5) maximize biomass and metabolite production (Edward *et al.*, 1999). In this study, multiobjective linear programming is used to investigate the quantitative relationship between the underlying objectives.

System Description

In this study, an *in silico* representation of *E. coli* metabolism has been constructed. The metabolic network consists of 198 reversible and 644 irreversible reactions and 436 metabolites compiled primarily from the literature (Edward et al., 2001; Pramanik and Keasling, 1997) and databases like EcoCyc (http://ecocyc.org/). The stoichiometric coefficients for each metabolic reaction are used to form the stoichiometric matrix S. In the metabolic network, the only carbon source is glucose, and oxygen uptake rate is also fixed for the FBA.

Methods

1. Flux balance analysis (FBA)

FBA can be used to identify particular behaviors within the allowable solution space, such as behaviors which produce the highest possible growth rate or production of a particular metabolite.

Generally, for a metabolic network all transient material balances can be represented by a single matrix equation,

$$\frac{dX}{dt} = S \cdot v - b \qquad (1)$$

where X is an n-dimensional vector of metabolite amounts per cell, S is the stoichiometric

$n \times m$ matrix, where m is the number of metabolites and n is the number of reactions in the network. The vector v represents all fluxes, in the metabolic network, including the internal fluxes, transport fluxes, and the growth flux. And b is the vector for consumption and secretion rates of metabolites and for biosynthetic requirements for cellular macromolecules. Eliminating the time derivative in equation 1 yields

$$S \cdot v = b \qquad (2)$$

In addition to the mass balance constraints, constraints on the magnitude of each individual metabolic flux is also required.

$$\alpha_i \leq v_i \leq \beta_i \qquad (3)$$

The linear inequality constraints are used to enforce the reversibility/irreversibility of metabolic reactions and the maximal metabolic fluxes in the transport reactions.

A particular metabolic flux distribution within the feasible set is found by using linear programming. A commercially available LP solving method based on the revised simplex algorithm with LU decomposition was used. In this study LP identified a solution that maximize a particular metabolic objective, and was formulated as shown. Maximize Z, where

$$Z = \sum c_i \cdot v_i \qquad (4)$$

The vector c was used to select a linear combination of metabolic fluxes to include in the objective function.

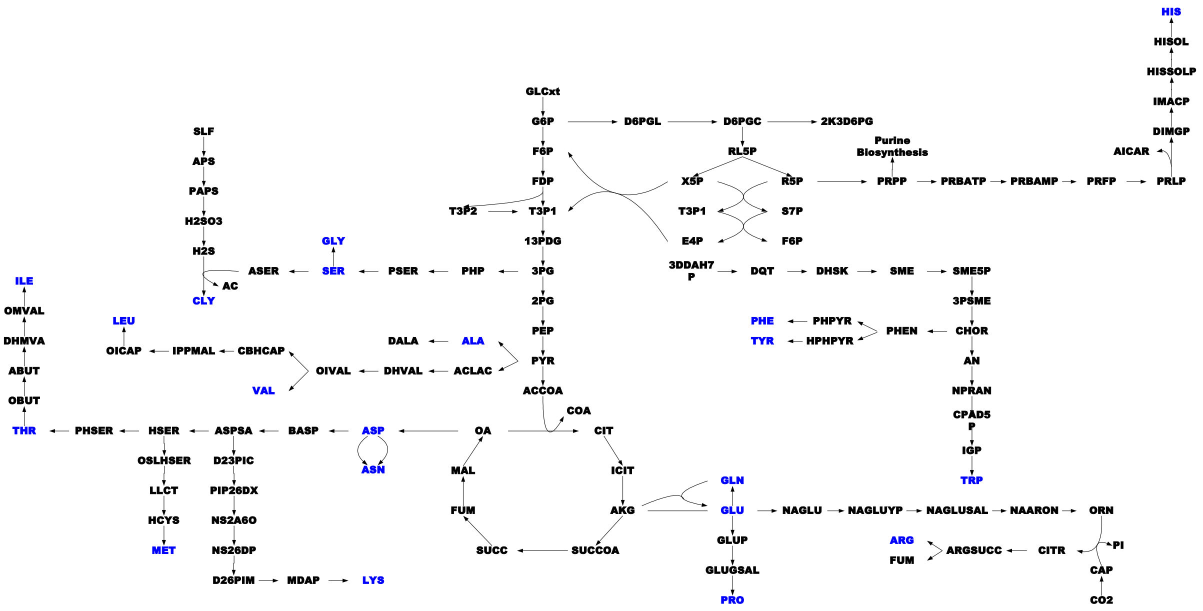


Fig 1. Overview of the metabolic pathways in E. Coli.

2. Multiobjective linear programming

The techniques of multiobjective programming seek to explore the set of noninferior solutions in the presence of multiple conflicting objectives.

A multiobjective linear programming problem is stated as follows:

$$\text{Maximize} \qquad Z(x) = (Z^1(x),\, Z^2(x),\, \ldots,\, Z^p(x)) \qquad (5)$$

$$\text{Subject to} \qquad \sum_{j=1}^{n} a_{ij} x_j \leq b_i \qquad \text{for } i=1,2,\ldots,m \qquad (6)$$

$$x_j \geq 0 \qquad \text{for } j=1,2,...,m \qquad (7)$$

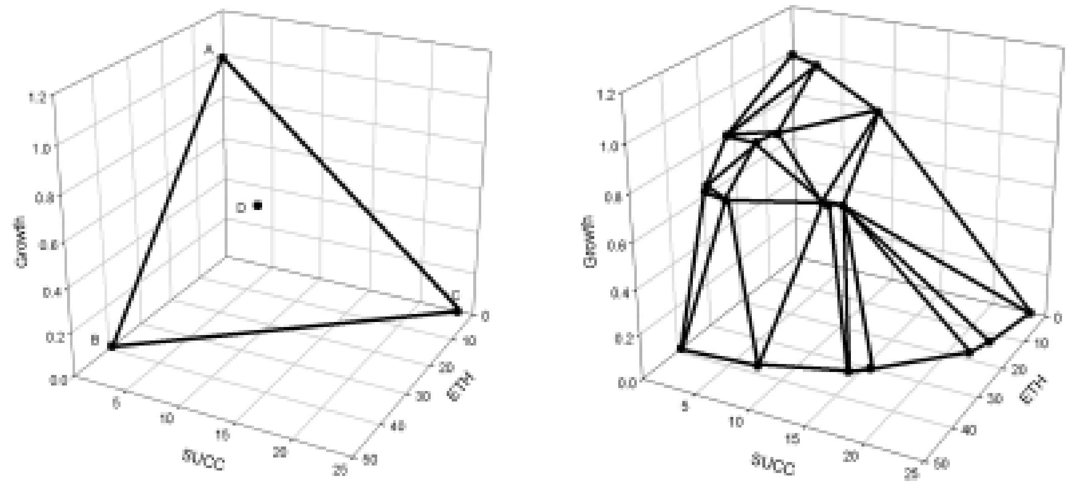
where $$Z^k(x) = \sum_{j=1}^{n} c_j^k x_j \qquad \text{for } k=1,2,...,p \qquad (8)$$

p=number of objectives,
m=number of constraints,
n=number of decision variables,
x=decision vector,
x_j=j-th decision variable,
$Z(x)$=objective vector,
$Z^k(x)$=k-th objective,
and c_j^k=coefficient of x_j in the k-th objective, $Z^k(x)$

The extended NonInferior Set Estimation (extended NISE) method (Solanki, R.S. et al., 1993) for multiobjective linear programing problems has the attractive feature of being able to generate an approximation with the deviation from the exact noninferior set being within a prospecified tolerance value. This method is composed of two parts which are generating the extreme points in the approximation and generating the facets in the approximation.

Result

The metabolic network can be examined using FBA. All the metabolic flux vectors attainable by a reconstructed metabolic network are mathematically confined in the region known as the feasible region in the objective space, and the optimal solution to the LP problem will then lie on an extreme point.



(a) Feasible region of first iteration (b) the feasible region in the objective space

Fig. 2 The results of multiobjective linear programming

In this research, we consider three objective functions. They include maximization of growth rate, succinate production, and ethanol production. According to the extended NISE method, at first LP is used to search extreme points maximizing each objective. As shown in Fig 2. (a), points A, B, and C are generated as results of each LP. The next point (D) is obtained by solving a weighted problem with weights derived from the slope of face ABC. Then points A, B, C, and D represent the extreme points forming the current approximation. The faces in the convex hull of A, B, C, and D are ABC, ABD, ACD, and BCD. Following the procedure of extended NISE, next points are generated by solving the weighted problem whose weights correspond to the outward normal of each face.

The procedure to generate the face in the approximation is based on the assumption that new point is contained in the set of extreme points generated previously. After five iterations, we can obtain the results which is composed of 22 faces. (Fig 2. (b)) Beacause the sixth iteration produces only two new points which were very close to the already generated extreme points, the iteration is stopped at the fifth.

Conclusion

A steady-state, flux-based model was developed to study the distribution of mass and energy fluxes through the *E. coli* metabolic reaction network. Using multiobjective linear programming, the quantitative relationship among the growth rate, the succinate porduction rate, and the ethanol production rate was obtained in the form of 3D-graph which contains the feasible region in the objective space and extreme points generated by extended NISE method. This approach is used to study the feasible metabolic behavior considering the systemic constraints simultaneously. This work should be considered as a step forward for the integrative analysis of bioinformatic databases to analyze, understand and ultimately predict cellular behavior.

Acknowledgement

This work was supported by IMT2000 project from the Ministry of Science and Technology and the Ministry of Information and Communication and also was partially supported by BK21 project and Center for Ultramicrochemical Process Systems sponsored by KOSEF.

References

[1] Bailey, J.E., "Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities", *Biotechnol. Prog.,* vol. 14 no. 1, pp. 8-20, 1998

[2] Edwards, J.S., Ibarra, R.U., and Palsson, B.O., "*In silico* predictions of Escherichia coli metabolic capabilities are consistent with experimental data", *Nat. Biotechnol.* vol. 19, pp. 125-130, 2001

[3] Edwards, J.S., Palsson, B.O., "The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities", *Proc. Natl. Acad. Sci. USA* 97, pp. 5528-5533, 2000

[4] Pramanik, J., and Keasling, J.D., "Stoichiometric Model of *Escherichia coli* Metabolism: Incorporation of Growth-Rate Dependent Biomass Composition and Mechanistic Energy Requirements", *Biotechnol. Bioengin.,* vol. 56, no. 4, pp. 398-421, 1997

[5] Solanki, R.S., Appino, P.A., Cohon, J.L., "Approximating the noninferior set in multiobjective linear programming problems", *European Journal of Operaional Research* 68, pp. 356-373, 1993