Multi-product Trade-off Analysis of E. coli by Multiobjective Flux Balance Analysis

Young-Gyun Oh¹, Dong-Yup Lee¹,², Hongseok Yun¹, Sang Yup Lee¹,² and Sunwon Park¹*

¹Department of Chemical and Biomolecular Engineering and Center for Ultramicrochemical Process Systems, ²Metabolic and Biomolecular Engineering National Research Laboratory Korea Advanced Institute of Science and Technology 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea

Abstract
In this study, we developed a novel multiobjective linear programming (MOLP) strategy based on the noninferior set estimation (NISE) method (Solanki et al., 1993), whereby Pareto solutions for the given set of conflicting objectives and corresponding flux distribution profiles are generated to understand how the internal fluxes are changed in the metabolic system. Furthermore, this MOLP approach was integrated as a new module into the program package, MetaFluxNet, which was developed for metabolic pathway construction and analysis (Lee et al., 2003). As a result, this package enables users to implement the multi-product trade-off analysis as well as the single product optimization. The efficacy and efficiency of the approach were demonstrated by applying it to the in silico E. coli model. Consequently, multiple objectives such as the maximization of succinic acid production and the maximization of NADP were considered simultaneously. The result can provide new insight into the relationship among the measurements, the objective criteria and the possible solutions.

Keywords: Flux balance analysis; Multiobjective linear programming; Multi-product trade-off analysis, Systems biology

1. Introduction
Recent completion of the entire genome sequences of numerous microorganisms and other organisms is setting up a new paradigm of biological research. To fully exploit the genome information, predicting a phenotype from the genome sequences is essential for producing a desired product by genetic modifications. Flux Balance Analysis (FBA) is a prime candidate for characterizing such metabolic genotype of an organism, thereby its metabolic phenotype under particular conditions can be predicted (Schilling et al., 1999). In FBA, the flux distribution is quantified by linear optimization on the basis of the information about the stoichiometry of metabolic reactions and mass balances around the metabolites under pseudo-steady state, or stationary, assumption. Maximizing ATP or metabolite production, minimizing nutrient uptake or redox

* Author to whom correspondence should be addressed: sunwon@kaist.ac.kr
production and predicting time course of growth and by-product secretion have been investigated by FBA.

When analyzing the optimal metabolic flux distribution, there are several important issues in connection with FBA. First, the optimal solution may not correspond to the experimental data. The second issue is to identify all alternative optima if applicable (Edwards et al., 2002). The other important issue is multicriteria optimization dealing with multiple and conflicting objectives. A large scale problem is usually characterized by the presence of many conflicting objectives. Thus it is highly desirable to consider multiobjective optimization, but up to date only a handful of works have been focused in the biology field (Lee et al., 2002; Burgard and Maranas, 2003; Halsall-Whitney et al., 2003; Vera et al., 2003).

For example, consider a simple biochemical reaction network as shown in Figure 1. The network is composed of 10 metabolites and 7 reactions. Of extracellular metabolites, $A_{ex}$, $B_{ex}$ and $C_{ex}$, are regarded as substrates consumed through the system boundary while $D_{ex}$ and $E_{ex}$ are allowed to be secreted or accumulated. All the intermediates are equally constrained and $A$, $B$ and $C$, are converted to $D$ and $E$ by the reaction R4 and R5, respectively.

![Figure 1. Sample reaction network](image)

When fueling of the metabolic network is rendered possible by constrained amounts of $A_{ex}$, $B_{ex}$, and $C_{ex}$ ($\leq 30$, $\leq 12$ and $\leq 20$ mmol/g DW h, respectively), representing limited substrate availability, the maximization of $D_{ex}$ is conflicted with the maximization of $E_{ex}$. If both $D_{ex}$ and $E_{ex}$ are essential products, two questions are encountered: what is the optimal combination of two products, and what is the flux profile at the optimal point? To answer these questions, we propose a novel multiobjective linear programming (MOLP) strategy based on the noninferior set estimation (NISE) method (Solanki et al., 1993).

2. Methodology

Most real-life engineering optimization problems require simultaneous consideration of more than one objective function. In such cases, it is unlikely that the same value of design variables will result in the optimal values for all the objectives. Hence, some trade-off between the objectives is needed to ensure a satisfactory design. The multiobjective optimization problem (MOP) solution is considered to be Pareto-optimal if there are no other solutions that are better in satisfying all of the objectives.
simultaneously. In this case MOP results in finding a set of noninferior solutions called Pareto optimal solutions.

2.1 Multiobjective flux balance model

The flux balance model is set up as follows:

\[ S \cdot v = b \]  

(1)

where \( S \) is the stoichiometric matrix, \( v \) is a vector for reaction rates or fluxes and \( b \) is the vector for the rates of metabolites consumption and excretion. If objective function is single, the flux balance model entails the maximization or minimization of an objective function as given in the following.

Max/Min \[ Z(v) = \sum_{j \in J} c_j v_j \]  

(2)

where \( c_j \) is the weight of reaction \( j \) and \( v_j \) is the element of the flux vector of reaction \( j \). Therefore multiobjective flux balance model can be stated as follows:

Max/Min \[ Z(v) = (Z^1(v), Z^2(v), ..., Z^p(v)) \]  

(3)

Subject to \[ \sum_{j \in J} S_{ij} v_j = b_i \quad \forall i \in I \]  

(4)

\[ l_j \leq \sum_{j \in J} S_{ij} v_j \leq u_j \quad \forall j \in J \]  

(5)

\[ \alpha_j \leq v_j \leq \beta_j \quad \forall i \in E \]  

(6)

where \( \sum_{j \in J} c_j v_j \quad \forall k \in P \)

\( p, P \) = number and set of objectives, respectively

\( I, J \) = set of metabolites and reactions, respectively

\( E \) = set of extracellular metabolites

\( v, v_j \) = flux vector and the flux of reaction \( j \), respectively

\( Z(v) \) = objective vector

\( Z^{k}(v) \) = \( k \)-th objective

\( c_j \) = weight of reaction \( j \) for the formulation of the \( k \)-th objective, \( Z^k \)

\( S_{ij} \) = stoichiometric coefficient of metabolite \( i \) in reaction \( j \)

\( b_i \) = net transport flux of metabolite \( i \)

\( l_i, u_i \) = lower and upper bound for the net transport flux of metabolite \( i \), respectively

\( \alpha_j, \beta_j \) = lower and upper bound for the flux of reaction \( j \), respectively

2.2 Algorithmic procedure for the noninferior set estimation

The techniques of multiobjective programming seek to explore the set of noninferior solutions in the presence of multiple conflicting objectives. To find the noninferior set, we modified the extended noninferior set estimation (eNISE) method (Solanki et al., 1993). A major advantage of the approximation algorithm lies in its ability to generate the approximation by using standard large-scale LP solvers. The procedure can be described as follows:
Step 0: Obtain the noninferior extreme points maximizing the individual objectives.

Step 1: Compute the convex hull of $U$.

$U$ = set of extreme points generated so far

Step 2: Compute the equation of the hyperplane supporting convex hull

$$w'_1Z'_1 + w'_2Z'_2 + \cdots + w'_nZ'_n = d$$

where $w_i = (w'_1, w'_2, \ldots, w'_n)$ denotes the outward normal to the hyperplane.

Step 3: Solve the weighted problem

Maximize $w'_1Z'_1 + w'_2Z'_2 + \cdots + w'_nZ'_n$

Subject to $(Z'_1, Z'_2, \ldots, Z'_n) \in HUZ$

where $HUZ$ = set of objective vectors in the upper bound.

Step 4: Search $\delta$ (distance between hyperplane and generated point)

If $\delta < \varepsilon$ STOP

Else add the new extreme point to $U$ and return Step 1

2.3 MetaFluxNet

We have developed a stand-alone program package, MetaFluxNet which provides an easy and customized way for constructing a metabolic reaction system and performing metabolic flux analysis (MFA). It provides the graphical user interface (GUI) where calculated flux distributions or other computed results are analyzed dynamically as well as interactively (Lee et al., 2003). For multi-product trade-off analysis, we have developed MOLP module, thereby enabling to consider multiple objectives of MFA model simultaneously.

3. Application

The reaction model of the sample network (Fig. 1) is constructed in the MetaFluxNet. Further analysis can be performed for considering the two conflicting objectives such as maximization of $D_{ex}$ and maximization of $E_{ex}$. Resulting in the Pareto curve and flux distribution profiles in Figure 2. In the Pareto curve of Figure 2, trade-off between the two objectives is shown. And flux distributions at the generated point are also shown, from which the change of internal flux is shown at a look.

(a)  
(b)  
(c)  

Figure 2. (a) MOLP module supported in MetaFluxNet, (b) Pareto curve view, (c) flux profile view for the sample network

In this study, the large-size in silico $E. coli$ model was constructed and then the trade-off between succinic acid production and NADP production are investigated by MOLP approach. The metabolic network consists of 315 reactions (128 reversible and 187...
irreversible reactions) and 281 metabolites (263 intermediates and 18 extracellular metabolites). In the metabolic network, glucose, gluconate and sorbitol are examined as carbon substrate while oxygen, NH$_3$ and PI uptake rate are fixed for the FBA.

4. Results & Discussion

The *E. coli* produces several metabolic products by fermentation: acetic acid, ethanol, formic acid, lactic acid, and also succinic acid. Besides succinic acid production, production rates of other organic acids should be considered simultaneously, which is attributable to the reducing power in the network. Through NADH or NADPH, the electrons of relevant metabolites are removed to transfer the reducing power. Therefore trade-off between succinic acid production and NADP production is investigated.

![Figure 3. Pareto curve and flux distributions of the E. coli model. Objectives are to maximize ATP synthesis, succinic acid production and malic acid production.](image)

MetaFluxNet provides a good approximation of the overall shape of the Pareto optimal solutions for two objectives: maximization of succinic acid production and maximization of NADP production (Fig 3). Compared with the weighting method or the constraint method, the computational burden needed for generating the noninferior set is reduced because eNISE method solves MOP with the error bounded by the pre-specified tolerance. The noninferior set has been computed herein by MetaFluxNet with a PC (Intel Pentium IV, 3.06GHz, 1024MB RAM), resulting in flux distributions of the model with a computational time of only 13 seconds. These quickly obtained results assist the metabolic engineers in performing a localized interactive search in the region of interest. Pareto curve (Fig 3a) shows that succinic acid production has not the linear relationship with the NADP production. Therefore results indicate that points 2, 3 and 4 are candidates for making the succinic acid with more NADP. This result is directly used for modifying cellular metabolism and properties of the *E. coli* through the introduction, deletion, and modification of metabolic pathways by using recombinant DNA and other molecular biological tools.

5. Concluding remarks

As the available information about biological system is dramatically accumulated, a new set of tools for analyzing large metabolic reaction networks is required. For that
reason, we have developed MetaFluxNet which is a program package for the management of metabolic reaction information and quantitative metabolic flux analysis. In order to gain more insight into the biochemical network for various objectives, MOLP module is added to MetaFluxNet, which enables user to get Pareto curve and flux distributions without difficulty and to implement multi-product trade-off analysis. Because E. coli is extensively studied to make valuable products, we have investigated in silico E. coli model with two objectives: maximization of succinic acid production and NADP production. The Pareto optimal solutions are obtained by the eNISE method which has the attractive feature of being able to generate an approximation with the deviation from the exact noninferior set within a pre-specified tolerance value. The simultaneous investigation of multiple objectives in E. coli enables metabolic engineer to understand the range of optimal solutions and the trade-offs contained in an approximate representation of the noninferior set, which help decision making and guide to improve the productivity and yield of native products.

References


Acknowledgement

This work was supported by the Korean Systems Biology Research Grant (M10309020000-03B5002-00000) from the Ministry of Science and Technology and also was partially supported by IMT2000 project from the Ministry of Science and Technology and the Ministry of Information and Communication and BK21 project.