17th European Symposium on Computer Aided Process Engineesing – ESCAPE17
V. Plesu and P.S. Agachi (Editors)

Identifying synergistically switching pathways for multi-product strain improvement using multiobjective flux balance analysis

Suresh Selvarasu,[a] Dong-Yup Lee,[a,b*] and Iftekhar A. Karimi[a]

[a]*Deparment of Chemical & Biomolecular Engineering, 4 Engineering Drive 4 National University of Singapore, Singapore 117576*
[b]*Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), 20 Biopolis way, #06-01, Centros, Singapore 138668*
**E-mail: cheld@nus.edu.sg*

Abstract

The current work involves *in silico* analysis of metabolic network of *E. coli*, characterizing its pathways for the production of various industrially important metabolites including pyruvate, acetate, lactate, ethanol and various amino acids. Initially, the correlation among the flux distribution profiles for each objective has been investigated by resorting to a novel multiobjective optimization. Subsequently, we identified the genes which are significant for switching the metabolic pathways from one objective to another. Thus, the present analysis allows us to explore synergism among the metabolic pathways for different combinations of multi-products, providing a new insight into the behavior of the biological system.

Keywords: Metabolic engineering; multiobjective flux balance analysis, gene target identification; strain improvement; synergistic pathway

1. Introduction

Microbial fermentation plays a major role in producing industrially important metabolites which include organic acids, ethanol, amino acids, ethanol, etc [1,

2]. However, even small deviations in culture conditions can affect the performance of these microbes and it is of primary importance in finding approaches to analyze and improve the efficiency of microbial productions. In this regard, *in silico* analysis of microbial systems in conjunction with genome, transcriptome, proteome, and metabolome information allows us to perform systems level analysis of the organism and explore its cellular functions. Such analysis can be also used to achieve the improved strains for overproducing important precursor metabolites [3].

E. coli bacterium is one of the commonly used species for biotechnological productions. Furthermore, its genome is fully sequenced, making it easier to analyze its cell metabolism and physiology [4]. Such information can provide wealth of knowledge in elucidating the functional characteristics of the organism and would be useful in developing new or improved strains for the production of industrially important metabolites. The current work involves *in silico* analysis of metabolic network of *E. coli*, characterizing its pathways for the production of individual components like pyruvate, acetate, lactate, ethanol and other important amino acids and identifying the correlation between the flux distribution profiles for each one of the objectives. A novel multiobjective optimization model has been developed to predict the production of all abovementioned products simultaneously. The flux distributions profile can be further analyzed in identifying genes which are significant for switching the metabolic pathways from one objective to another. Such identified genes can provide an insight into the behaviour of the biological system and also in engineering of the system to develop new or improved strain for overproducing important products [5, 6]. Such analyses would also reveal the possible existence of synergism among the metabolic pathways for the production of different combinations of metabolites.

2. Methodology

Often in biological systems there are situations where more than one objective has to be considered. In such a case, the decision variables will not be the same for different objectives, resulting in trade off among different objective functions. Multiobjective optimization is invaluable for solving such problems as it can identify Pareto optimal solutions satisfying all the objectives.

2.1. Multiobjective flux balance analysis

The metabolic reaction network of *E. coli* can be represented in the form of linear mathematical relations using a stoichiometric matrix. Then flux balance analysis can be used to find the optimal value for cellular objectives. However if the focus is on more than one objective then a multi-objective flux balance analysis is required to identify optimal distribution. In this work, the metabolic network of *E. coli* has been analyzed for the production of metabolites such as

lactate, acetate, pyruvate, organic acids, ethanol, aromatic and other amino acids using multi-objective flux balance analysis technique [6]. There are different methods for solving multiobjective problems. In this work a weighted sum approach has been used. The optimization problem formulated is described below:

$$Max\ Z(v) = \sum_{k \in P} w^k z^k(v) \tag{1}$$

$$Z^k(v) = \sum_{j \in M} c_j^k v_j \tag{2}$$

$$s.t \quad \sum_{j=1}^{M} S_{ij} v_j = 0 \qquad \forall i \in N \tag{3}$$

$$\alpha_j \le v_j \le \beta_j \qquad \forall j \in M \tag{4}$$

Where P = set of objectives
N,M = set of metabolites and reactions in the network
v, v_j = flux vector and metabolic flux of reaction j
$Z(v)$ = objective vector
$Z^k(v)$ = k-th objective
w^k = weight function for different objectives
c^k_j = weight of reaction j for the objective Z^k
S_{ij} = Stoichiometric coefficient of metabolite i in the reaction j
α_j, β_j = lower and upper bounds for the reaction flux j

By choosing suitable weights w^k, priority can be assigned to different objectives and a multi-objective optimization can be carried out. A set of pareto optimal solutions can be generated based on different weights (w^k) assigned.

2.2. Characterization of the metabolic network

Analysis of the metabolic network often yields optimal distribution of reaction fluxes for a given objective function. For a multi-objective problem with different sets of objectives, different sets of flux distributions can be obtained. These flux distributions can be compared and characterized further, to identify important switching genes for different metabolic objectives. Further more, analysis of these flux distributions can reveal the existence of synergism among pathways for different metabolite production and such synergistic pathways can be identified.

3. Results & discussions

The metabolic network of *E. coli* was analyzed for achieving optimal growth phenotypes and also for overproducing individual and multiple byproducts using multiobjective flux balance analysis. The analysis resulted in different sets of metabolic flux distributions for optimal cell growth and different byproduct production. Since the primary objective of the cell is to produce the cell biomass, the flux distributions for the entire individual and multiple byproducts were compared with the flux distribution for cell biomass production. Based on the comparison the metabolic network can be further characterized.

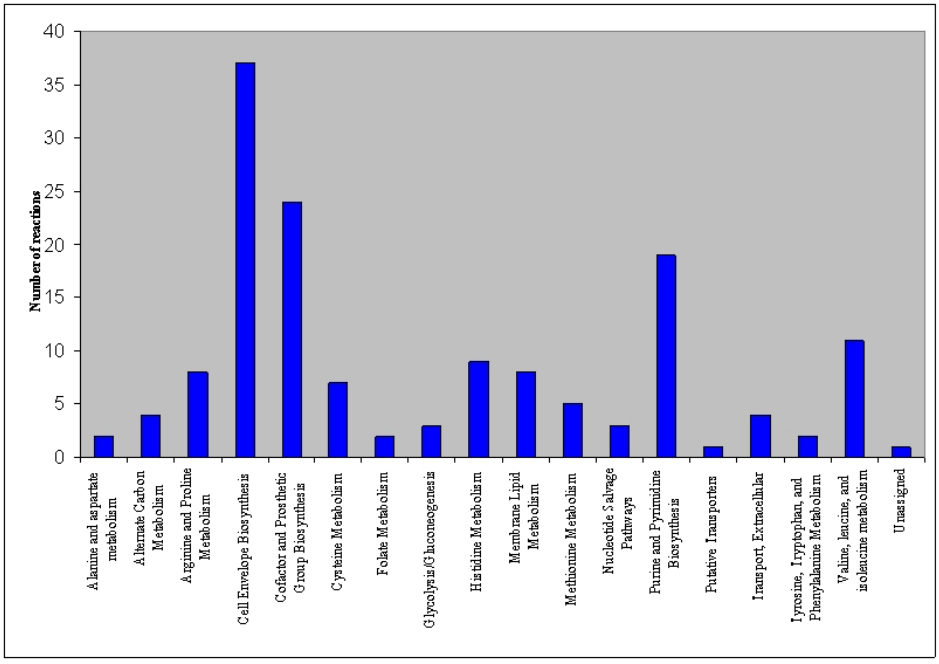

Fig. 1. Distribution of reaction across different metabolism, these reactions participate in all product productions. These reaction sets form set of pathways which are termed as synergistic pathways as these pathways are used in all the cellular objectives

Analysis results revealed a significant variation in flux distributions for different products when the cellular objective was varied from cell biomass to other product production. The shift in cellular objective was characterized by the changes in flux distribution i.e., the flux values changed from a minimum value to a maximum followed by a minimum value when the objective is shifted from byproduct production to cell biomass production. Genes corresponding to these metabolic reaction fluxes were identified and characterized as switching genes. The analysis identified 142 such switching genes for all of the bi-product

production and the number of such switching genes varied from 6 to 59 for each of the individual product production, with the maximum being for all products production and the minimum being for ethanol production. The analysis also identified reactions whose flux remained the same for all byproduct production. Of the 934 metabolic reactions, 150 such reactions were identified from the analysis (Fig 1). The pathways which comprise these reactions are termed synergistic pathways as these pathways are used for the production of all the products.

Comparison of metabolic flux distributions for aromatic amino acids and cell biomass production has been described below as a case study (Fig 2). For this case, when the cellular objective is shifted from aromatic amino acids to cell biomass, significant variation in flux distributions was identified. Further

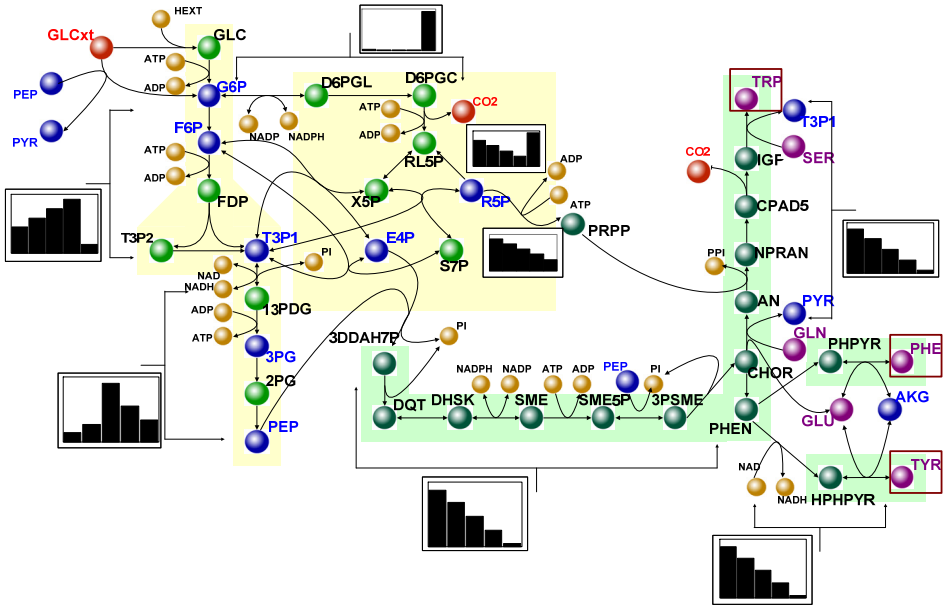

Fig. 2. Screenshot of flux distribution profile for aromatic amino acids production. The change in flux values are highlighted when the objective function slowly changed from aromatic amino acids production to cell biomass

Analysis of these fluxes values identified Pentose Phosphate pathway as important pathway for cell biomass but not for aromatic amino acids production. Similarly glycolytic pathway was found to have switching characteristics, i.e., when both the objectives were considered, a maximum flux distribution across this pathway was identified and reduced flux distributions was observed when only one of the objectives was considered. Hence glycolytic pathway can be considered important for optimal production of both cell

biomass and aromatic amino acids. Similar analyses have been performed for the other products and important switching pathways have also been identified.

4. Concluding remarks and future work

A multi-objective flux balance analysis technique has been used to analyze the metabolic network of *E. coli* for the production of multiple by-products. The analysis has identified synergistic pathways which are used in common for all the product productions. The flux distributions were further analyzed to identify switching genes which have the characteristic behaviour of switching the cell metabolism from production of one product to another. Such characterization can be useful in elucidating the effect of environmental factors on the metabolism of the organism or evolution of organism with respect to external perturbations. This method can also be useful in identifying genes which are important for the optimal production of more than on byproduct. The identified genes can be either over-expressed and strain improvements can be achieved. This analysis can also give insight about the effect of global regulators in the cellular metabolism. The network has also been characterized further by resorting to flux variability analysis which identifies numerical variations in fluxes of individual product production and then comparing the flux variations for different product production (results not shown). However the effect of gene regulation is not considered in this work and this will be considered in our future analyses.

References

1. M. Koffas and G. Stephanopoulos, Curr. Opin. Biotechnol., 16 (2005) 361
2. M. Ikeda, Appl. Microb. Biotechnol., 69 (2006) 615
3. V.F. Wendisch, M. Bott and B.J. Eikmanns, Curr. Opin. Biotechnol., 9 (2006) 268
4. U. Sauer and B.J. Eikmanns, FEMS Microbiol. Rev., 29 (2005) 765
5. S.Y. Lee, D.-Y. Lee and T.Y. Kim, Trends Biotechnol. 23 (2005) 349
6. Y.G. Oh, D.-Y. Lee, H. Yun, S.Y. Lee and S. Park, ESCAPE-14: European Symposium on Computer Aided Process Engineering, Lisbon, Portugal, May 16-19 (2004).