International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:8, No:11, 2014

Integration of Microarray Data into a Genome-Scale Metabolic Model to Study Flux Distribution after Gene Knockout

Mona Heydari, Ehsan Motamedian, Seyed Abbas Shojaosadati

***Abstract*—**Prediction of perturbations after genetic manipulation (especially gene knockout) is one of the important challenges in systems biology. In this paper, a new algorithm is introduced that integrates microarray data into the metabolic model. The algorithm was used to study the change in the cell phenotype after knockout of Gss gene in *Escherichia coli* BW25113. Algorithm implementation indicated that gene deletion resulted in more activation of the metabolic network. Growth yield was more and less regulating gene were identified for mutant in comparison with the wild-type strain.

***Keywords*—**Metabolic network, gene knockout, flux balance analysis, microarray data, integration.

I. Introduction

FLUX balance analysis (FBA) is a mathematical approach for analyzing the flow of metabolites through a metabolic network. This approach selects an objective function and uses linear programming (LP) to find reactions flux in an underdetermined stoichiometric model [1]-[3]. FBA applies various optimization criteria such as maximization of growth rate, minimization of ATP production and maximization of product formation [4]-[6]. This approach has been successfully used to predict different capabilities of cells for example growth, uptake rates, by-product secretion and growing after genetic manipulations [7]-[11]. This approach has been successfully used to predict the effect of different parameters such as temperature, aeration rate and pH [12]-[14] on the metabolism of cells [2], [15], [16].

Also, investigation of intracellular perturbation and genetic manipulation effect on cellular behavior has attracted considerable attention in recent years [3], [5], [17], [18]. Gene deletion analysis is one of the important issues to study mutant phenotypes which could result in more knowledge of metabolism and metabolic engineering strategies for targeted improvement. Gene deletion leads to zero flux for associated reactions and hence, a gene deletion often results in the reduction of the steady-state solution space unless isozymes or equivalent reaction sets present in the network.

Four important approaches FBA, MOMA, lMOMA, and ROOM have been presented to predict the effects of gene deletions on overall network function [19]-[21]. In FBA, LP is carried out the same as before, but in the reduced mutant solution space. In fact, this approach assumes that microorganism optimizes the growth after gene deletion, and the knockout solution is an optimal point in the new feasible solution space. In MOMA, it is supposed that the mutant solution in the new solution space is close to the optimal solution of the wild type. Based on this approach, the mutant is not seeking to optimize its performance against FBA; however, it tries to have the least change and perturbation in the flux distribution compared with wild type. Hence, by using a second order objective function and solving the quadratic programming (QP), the nearest point to the optimal solution of wild type in the new solution space is found. Due to the high computations in MOMA, lMOMA is proposed which uses linear norm instead of the Euclidean norm in objective function [19], [21]. ROOM is another approach to study the cell metabolism after gene deletion which converts LP model to a mixed integer linear programming (MILP) model. In this approach, the aim is the minimization of the total number of significant flux changes from the wild-type flux distribution. The approach has been developed using the idea of Ihmels et al. [22]. According to this idea, metabolic flow in a branch commonly passes through a specific path.

Studies indicate that FBA and ROOM results are very close together especially for prediction of gene lethality. In fact, ROOM is indirectly looking for the optimal growth as well as FBA. Furthermore, Shlomi et al. shows that MOMA is more appropriate for predicting transient growth rates in response to genetic perturbations while ROOM and FBA better predict the final growth rate achieved after the adaptation process [20]. There are some studies on gene deletion analysis using various approaches especially FBA and MOMA. For example, Boghigian et al. [3] simulate the 6-Deoxyerythronolide B production in three common heterologous hosts (*E. coli, B. subtilis,* and *S. cerevisiae*) under a variety of carbon-source and medium compositions. They used MOMA to identify single and double gene knockouts that resulted in increased 6-Deoxyerythronolide B production while maintaining cellular growth. They predict several single and multiple gene knockout mutants to improve growth and 6-Deoxyerythronolide B production. In another work, Hjersted et al. [17] investigated the effect of gene deletion or insertion for growth and ethanol production in the fed-batch culture of S. cerevisiae using FBA.

Previous methods are able to predict gene essentially while

M. Heydari and S. A. Shojaosadati are with the Biotechnology Group, Chemical Engineering Department, TarbiatModares University, Tehran, Iran (e-mail: mona.heidari@modares.ac.ir, shoja_sa@modares.ac.ir).

E. Motamedian is with the Biotechnology Group, Chemical Engineering Department, TarbiatModares University, Tehran, Iran (corresponding author to provide phone: +982182883947; fax: +982182884931; e-mail: motamedian@modares.ac.ir).

their proposed flux distributions may be far from the reality. In this paper, microarray data before and after gene knockout were integrated into a metabolic model using a new algorithm to study the change in flux distribution. The new algorithm is called transcriptionally regulated flux balance analysis (trFBA) and is based on the hypothesis that pattern of genes expression in the growth phase is constant, although it is possible to change the intensity of the gene expression. trFBA extracts a pattern of reaction usage in the growth phase based on a gene expression profile and constrains the upper bounds of reactions. Expression levels are converted to upper bounds by defining the pattern coefficient (PC). Integration of microarray data for *E. coli* BW25113gss$^+$ and gss$^-$ into a metabolic model determined change in flux distribution after gene knockout.

II. Material and Methods

A. Genome-Scale Model and in Silico Simulation

The genome-scale stoichiometric model used in this study for *Escherichia coli* includes 1805 metabolites, 2583 reactions, 1367 genes and three compartments (periplasm, cytoplasm and extracellular) [23]. The model was modified to consider genetic differences between MG1655 and BW25113. Since the araBAD, rhaBAD, and lacZ genes are absent from the BW25113 strain, the associated metabolic reactions were removed [24].

Specific growth and glucose uptake rates were determined for *E. coli* BW25113 growing in aerobic glucose minimal medium cultured in a continuous system [25].

In all *in silico* experiments, biomass formation was the objective function to be maximized. Calculations were made in MATLAB software using the COBRA toolbox. The GLPK (GNU linear programming kit) package was used to solve LP problems.

B. Method

Stoichiometric metabolic models are constructed using a metabolic network and comprise a set of algebraic equations which are usually underdetermined.

Raw microarray data from Chattopadhyay et al.[26] (GSE30679) with three replicates for *E. coli* BW25113 gss$^+$ and gss$^-$ were used. The mRNA level was measured in the log-phase of a minimal medium that includes glucose as the sole carbon source. Expression values were determined using the MAS5 normalization algorithm. Average expression of three replicates was considered to be the expression level of each gene.

The proposed algorithm assigns a value to each reaction based on gene expression. For each reaction that is catalyzed by only one gene, the expression value for the gene whose product catalyzes a reaction is set as the reaction value. For each reaction catalyzed by a complex requiring more than one gene, the reaction value is set to the minimum of expression of requiring genes. Reaction expression vector demonstrates a pattern of reaction usage in metabolism based on expression data.

After transcription, it is possible to have post-transcriptional and post-translational regulation in a cell; thus, the upper bounds instead of the fluxes are limited in the algorithm using reaction expression data. Expression data were converted to the upper bound of reactions using a coefficient named pattern coefficient (PC). The optimal value of PC was determined using experimental data. If the expression of none of the genes associated with a reaction is determined, the upper bound of 1000 is considered for that reaction; if the reaction is reversible, a lower bound of -1000 is considered for it. In the event that more than one reaction is supported by a gene, these reactions are linked to preventing the use of the expression value of a gene multiple times.

III. Results and Discussion

A. Comparison of Reaction Expressions

Fig. 1 shows the distribution of reaction expressions values determined for the 3901 reactions using microarray data of wild-type and Δgss mutant.

It can be seen that reactions expression of wild-type cells tends to lower values. 1280 reactions have values lower than 200 while there are 1444 reactions for wild type. Maximum values of reaction expression were recorded for reactions DBTS (dethiobiotin synthase) from the cofactor and prosthetic group biosynthesis pathway for both strains. Fig. 1 indicates that gss gene knockout commonly resulted in over-expression of metabolic genes. In fact, it can be concluded that genetic perturbations motivated the metabolic genes and further activated the metabolism.

B. Optimal Values of PC and Growth Prediction

To simulate growth using FBA and trFBA, the glucose uptake rate was determined (1.45 (mmol $gDCW^{-1}h^{-1}$) and the growth rate (μ) was optimized by running the model. Implementation of trFBA requires that PC be specified. PC was manually changed to achieve the experimental growth rate (0.093 h^{-1}) for wild-type strain by running the model. The optimal value of 0.0048 was calculated for PC to convert expression values to the upper bound of reactions.

Growth rate predicted by FBA was 0.13 h^{-1}. It can be seen that FBA predicted more growth rate compared with experimental data. This arises from the fact that FBA used the stoichiometric data without limitation for intracellular reactions.

The optimal value of PC calculated for wild-type strain was used to determine the growth rate of Δgss strain. The predicted growth rate was equal to 0.125 h^{-1}that was near growth rate predicted by FBA. While predicted growth yield for wild-type strain was 0.064 g/g, the algorithm predicted that gss gene knockout increase it to 0.86 g/g. in fact, the algorithm based on the microarray data demonstrates that intracellular limitations were decreased when deletion of the gsshas occurred.

C. Change in Flux Distribution

In addition to different growth yield, algorithm predicted that metabolism usage of wild-type and mutant strains are

different. To evaluate the changes in flux distribution, the number of activated and inactivated reactions was compared for two strains and the number of reactions that were active for Δgss, but an increase or decrease in their activity was observed, were presented in Table I.

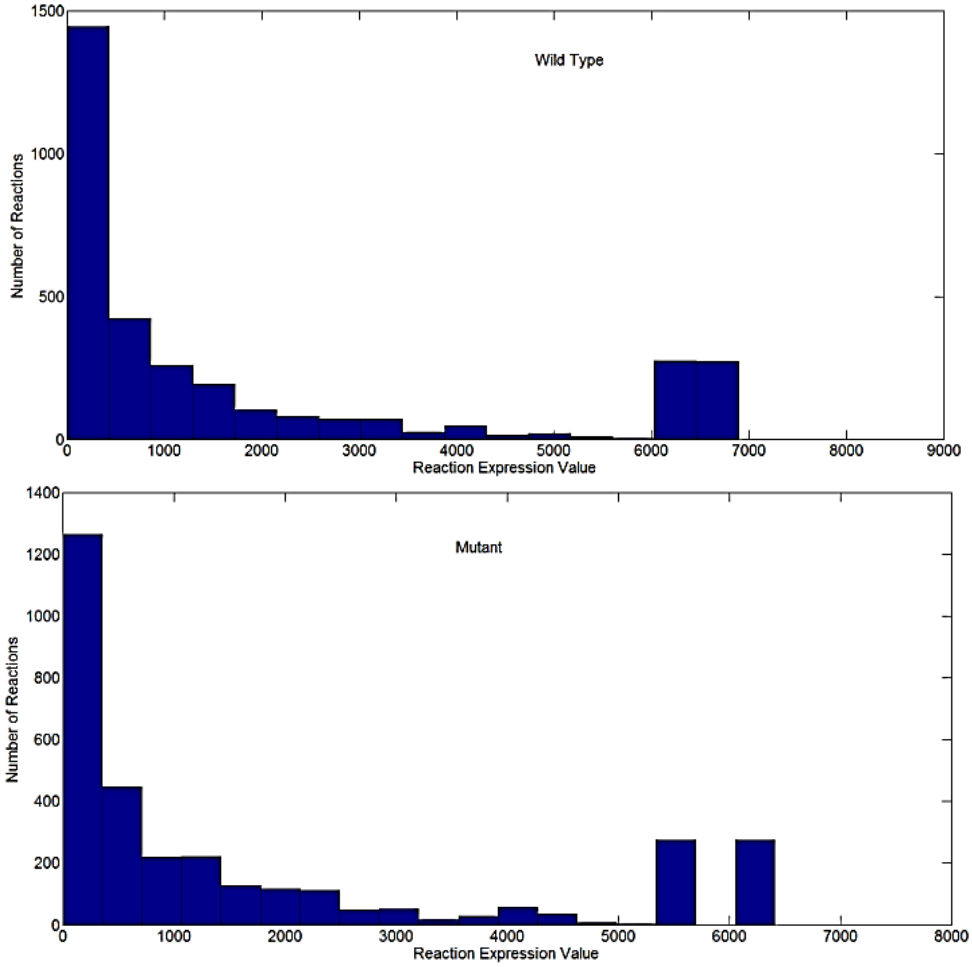

Fig. 1 Distribution of reaction expression values for the reactions in wild-type and mutant strains

TABLE I
NUMBER OF ACTIVATED AND INACTIVATED REACTIONS AND NUMBER OF REACTIONS THAT WERE ACTIVE FOR ΔGSS MUTANT AND SAW AN INCREASE OR DECREASE IN THEIR ACTIVITY

Strain	Yx/s (g/g)	Number of Reactions			
		Activated	Increased	Inactivated	Reduced
WT	0.356	20	172	24	6
Δgss	0.478				

Although the glucose uptake rate was identical, the growth yield was very different. Activated reactions were less than inactivated reactions while reactions within creased fluxes were more than reactions with reduced fluxes. Flux distribution in the mutant was changed, and most of the reactions were activated.

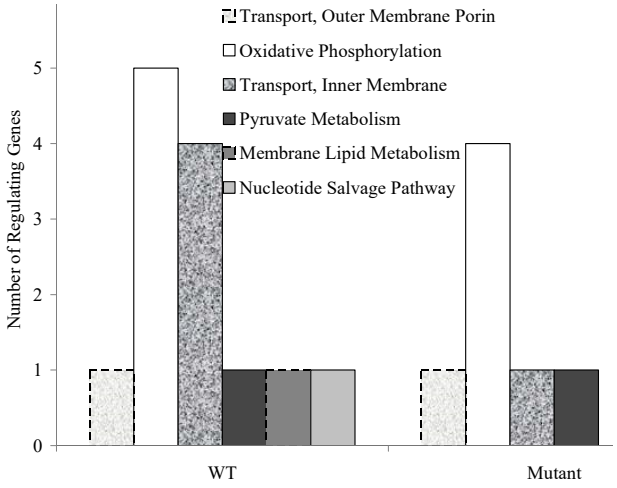

Fig. 2 Metabolic pathways of regulating genes for wild-type and mutant strains

D.Metabolic Pathways of Bottleneck Genes

The genes for which all expressions were used are called bottleneck or regulating genes. These genes limited activity of one or more reactions. 13 and 7 regulating genes were identified for wild type and Δgss strains, respectively. The fewer number of regulating genes confirms lower intracellular limitations of Δgss strain. Regulating genes belong to different metabolic pathways. Fig. 2 shows the metabolic pathways of regulating genes for two wild-type and Δgss strains. It can be seen that various metabolic pathways were limited. Limitation of pathways in wild-type is more than mutant strain. The most limited pathway was oxidative phosphorylation. In fact, the algorithm shows that energy production was the most important limitation for growth.

IV. Conclusion

While a gene was deleted, integration of microarray data into the metabolic model using a new algorithm predicted that growth yield of wild type was lower than a Δgss mutant.

In addition, trFBA predicted that bottleneck genes of wild-type strain were more in amount, and its metabolism is more limited. Therefore, it can be concluded that gene knockout motivated the cell's metabolism to be more activated.

References

[1] K. Raman and N. Chandra, "Flux balance analysis of biological systems: Applications and challenges," *Briefings in Bioinformatics,* vol. 10, no. 4, pp. 435-449, 2009.

[2] F. Llaneras and J. Picó, "Stoichiometric modeling of cell metabolism," *Journal of Bioscience and Bioengineering,* vol. 105, no. 1, pp. 1-11, 2008.

[3] J. S. Edwards, M. Covert, and B. Palsson, "Metabolic modeling of microbes: The flux-balance approach," *Environmental Microbiology,* vol. 4, no. 3, pp. 133-140, 2002.

[4] A. M. Feist and B. O. Palsson, "The biomass objective function," *Current Opinion in Microbiology,* vol. 13, no. 3, pp. 344-349, 2010.

[5] K. J. Kauffman, P. Prakash, and J. S. Edwards, "Advances in flux balance analysis," *Current Opinion in Biotechnology,* vol. 14, no. 5, pp. 491-496, 2003.

[6] A. Varma and B. O. Palsson, "Metabolic flux balancing: Basic concepts, scientific and practical use," *Nature Biotechnology,* vol. 12, no. 10, pp. 994-998, 1994.

[7] J. S. Edwards and B. O. Palsson, "The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 97, no. 10, pp. 5528-5533, 2000.

[8] J. S. Edwards, R. U. Ibarra, and B. O. Palsson, "In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data," *Nature Biotechnology,* vol. 19, no. 2, pp. 125-130, 2001.

[9] I. Family, J. Forster, J. Nielsen, and B. O. Palsson, "*Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 100, no. 23, pp. 13134-13139, 2003.

[10] J. Forster, I. Famili, B. O. Palsson, and J. Nielsen, "Large-scale evaluation of in silico gene deletions in *Saccharomyces cerevisiae*," *OMICS A Journal of Integrative Biology,* vol. 7, no. 2, pp. 193-202, 2003.

[11] R. U. Ibarra, J. S. Edwards, and B. O. Palsson, "*Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth," *Nature,* vol. 420, no. 6912, pp. 186-189, 2002.

[12] K. Kiviharju, U. Moilanen, M. Leisola, and T. Eerikäinen, "A chemostat study of *Streptomyces peucetius* var. caesius N47," *Applied Microbiology and Biotechnology,* vol. 73, no. 6, pp. 1267-1274, 2007.

[13] A. M. Feist, J. C. M. Scholten, B. Ø. Palsson, F. J. Brockman, and T. Ideker, "Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*," *Molecular Systems Biology,* vol. 2, pp. 2006.

[14] V. Satish Kumar, J. G. Ferry, and C. D. Maranas, "Metabolic reconstruction of the archaeon methanogen *Methanosarcina Acetivorans*," *BMC Systems Biology,* vol. 5, pp. 2011.

[15] O. Gonzalez, T. Oberwinkler, L. Mansueto, F. Pfeiffer, E. Mendoza, R. Zimmer, and D. Oesterhelt, "Characterization of growth and metabolism of the haloalkaliphile *Natronomonas pharaonis*," *PLoS Computational Biology,* vol. 6, no. 6, pp. 1-10, 2010.

[16] C. B. Milne, P. J. Kim, J. A. Eddy, and N. D. Price, "Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology," *Biotechnology Journal,* vol. 4, no. 12, pp. 1653-1670, 2009.

[17] J. L. Hjersted, M. A. Henson, and R. Mahadevan, "Genome-scale analysis of *Saccharomyces cerevisiae* metabolism and ethanol production in fed-batch culture," *Biotechnology and Bioengineering,* vol. 97, no. 5, pp. 1190-1204, 2007.

[18] Q. Zhao and H. Kurata, "Genetic modification of flux for flux prediction of mutants," *Bioinformatics,* vol. 25, no. 13, pp. 1702-1708, 2009.

[19] D. Segre, D. Vitkup, and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, no. 23, pp. 15112-15117, 2002.

[20] T. Shlomi, O. Berkman, and E. Ruppin, "Regulatory on/off minimization of metabolic flux changes after genetic perturbations," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, no. 21, pp. 7695-7700, 2005.

[21] M. J. Herrgard, S. S. Fong, and B. O. Palsson, "Identification of genome-scale metabolic network models using experimentally measured flux profiles," *PLoS Computational Biology,* vol. 2, no. 7, pp. 0676-0686, 2006.

[22] J. Ihmels, R. Levy, and N. Barkai, "Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*," *Nature Biotechnology,* vol. 22, no. 1, pp. 86-92, 2004.

[23] J. D. Orth, T. M. Conrad, J. Na, J. A. Lerman, H. Nam, A. M. Feist and B. O. Palsson, "A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism--2011," *Mol Syst Biol,* vol. 7, pp. 535, 2011.

[24] A. R. Joyce, J. L. Reed, A. White, R. Edwards, A. Osterman, T. Baba, H. Mori, S. A. Lesely, B. Ø. Palsson, and S. Agarwalla, "Experimental and computational assessment of conditionally essential genes in *Escherichia coli*," *Journal of Bacteriology,* vol. 188, no. 23, pp. 8259-8271, 2006.

[25] S. A. Nizam and K. Shimizu, "Effects of arcA and arcB genes knockout on the metabolism in *Escherichia coli* under anaerobic and microaerobic conditions," *Biochemical Engineering Journal,* vol. 42, no. 3, pp. 229-236, 2008.

[26] M. K. Chattopadhyay, W. Chen, and H. Tabor, "*Escherichia coli* glutathionylspermidine synthetase/amidase: phylogeny and effect on regulation of gene expression," *FEMS Microbiol Lett,* vol. 338, no. 2, pp. 132-40, 2013.