BMB
*reports*

# Computational identification of significantly regulated metabolic reactions by integration of data on enzyme activity and gene expression

*Hojung Nam[1], Taewoo Ryu[2], KiYoung Lee[3], Sangwoo Kim[1] & Doheon Lee[1,*]*

[1]Department of Bio and Brain Engineering, KAIST, Deajeon, Korea, [2]Omics Integration Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Korea, [3]Department of Bioengineering, University of California at San Diego, La Jolla, California 92093, USA

**The concentrations and catalytic activities of enzymes control metabolic rates. Previous studies have focused on enzyme concentrations because there are no genome-wide techniques used for the measurement of enzyme activity. We propose a method for evaluating the significance of enzyme activity by integrating metabolic network topologies and genome-wide microarray gene expression profiles. We quantified the enzymatic activity of reactions and report the 388 significant reactions in five perturbation datasets. For the 388 enzymatic reactions, we identified 70 that were significantly regulated (P-value < 0.001). Thirty-one of these reactions were part of anaerobic metabolism, 23 were part of low-pH aerobic metabolism, 8 were part of high-pH anaerobic metabolism, 3 were part of low-pH aerobic reactions, and 5 were part of high-pH anaerobic metabolism. [BMB reports 2008; 41(8): 609-614]**

## INTRODUCTION

Two main processes regulate metabolic rates: the amounts of enzymes and their catalytic activities (1-3). The amount of an enzyme depends on its rate of synthesis and degradation. For example, in *E. coli*, lactose induces a 50-fold increase in the amount of β-galactosidase within several minutes, and this greatly increases lactose catabolism. Enzyme activity is often controlled by small metabolites. For example, tryptophan is the effector molecule for allosteric enzymes such as TrpE and TrpD.

The development of high-throughput technologies provides information on biological processes through the integration of data from genomics, proteomics, and metabolomics (4). Integration of these high-throughput data allows for an understanding of the responses of cells to specific conditions via systems biology (4-6). Previous studies have integrated genetic regulation and metabolism in attempts to understand the mechanisms of metabolic regulation. These studies can be grouped into three classes. The first approach identifies new relationships between metabolites and genes. Ideker *et al*. (7) identified associations between metabolites and transcriptional factors by observing gene and protein expression profiles in the presence of different carbon sources, and Yeang *et al*. (8) inferred links between metabolites and transcriptional factors by using a joint model of genetic regulation and metabolic reactions. The second approach identifies new features of metabolic pathways by analyzing the expression patterns of genes that encode enzymes. Ihmels *et al*. (9, 10) characterized the transcriptional regulation of metabolic pathways by examining the expression patterns of enzymes along the topology of metabolic pathways. Patil *et al*. (3) analyzed the active sub-networks and reporter metabolites, the relative enzyme genes of which show the most significant transcriptional changes. In the third approach, new methodologies are developed to analyze metabolic reactions. Covert *et al*. (6, 11, 12) combined genetic enzyme regulation and metabolic flux modeling to validate flux modes with biomass growth of cells. Cakir *et al*. (13) identified reporter reactions by integrating the transcriptome, the metabolome, and the topology of metabolic pathways.

Most previous studies of the mechanisms of metabolic regulation, however, have only dealt with a restricted level of metabolic regulation. They typically consider only enzyme levels, not the control of enzyme activity (14-16). This is because no high-throughput technology is available for the measurement of enzyme activity (14). One way to infer the activity of an enzyme is to measure the concentration levels of metabolites that serve as regulators (17, 18). Unfortunately, it is not yet possible to identify all metabolites in a genome-wide metabolic network (13). Patil *et al*. proposed a method that assigns "scores" to metabolites based on correlations between transcription levels and metabolic rates (3) . Each score represents the statistically significant change of a metabolite following perturbation by using the expression of neighboring genes that are topologically adjacent to the metabolite. This method successfully identified key metabolites and provided useful information about metabolic changes under specific experimental conditions.

In the present study, we propose a method for scoring the en-

*Corresponding author. Tel: 82-42-869-4316; Fax: 82-42-869-8680; E-mail: dhlee@biosoft.kaist.ac.kr

**Received 3 March 2008, Accepted 15 April 2008**

**Keywords:** Enzyme activity, Gene expression, Metabolic reaction, Metabolism regulation

zyme activity of reactions in order to identify significantly regulated reactions. We extended scoring functions based on the method of Patil *et al*. (3). We assume that the catalytic activity of a reaction will be altered if the changes in metabolites that regulate the reaction are significant. For example, 6-phosphofructokinase 1 is allosterically inhibited by ATP, allosterically activated by AMP, and inhibited by glucagon through the repression of synthesis. Therefore, if a change is made to these regulators, the enzymatic reaction of 6-phosphofructokinase 1 would also change. Fig. 1 provides an overview of our method. First, the significance of a change in the level of genes is extracted by applying *t*-tests to microarray data. This method maps the scores of the altered gene levels to the enzymes in a metabolic pathway. Second, an integrated metabolic network is generated by combining enzymatic activity regulation information with metabolic pathways. Third, the regulation scores of reactions (via control of enzyme activity) are measured by using the levels of significance of the changes in the gene expressions and the integrated network. Finally, based on the regulation scores, significantly regulated reactions are identified. We further evaluate the results of enzyme

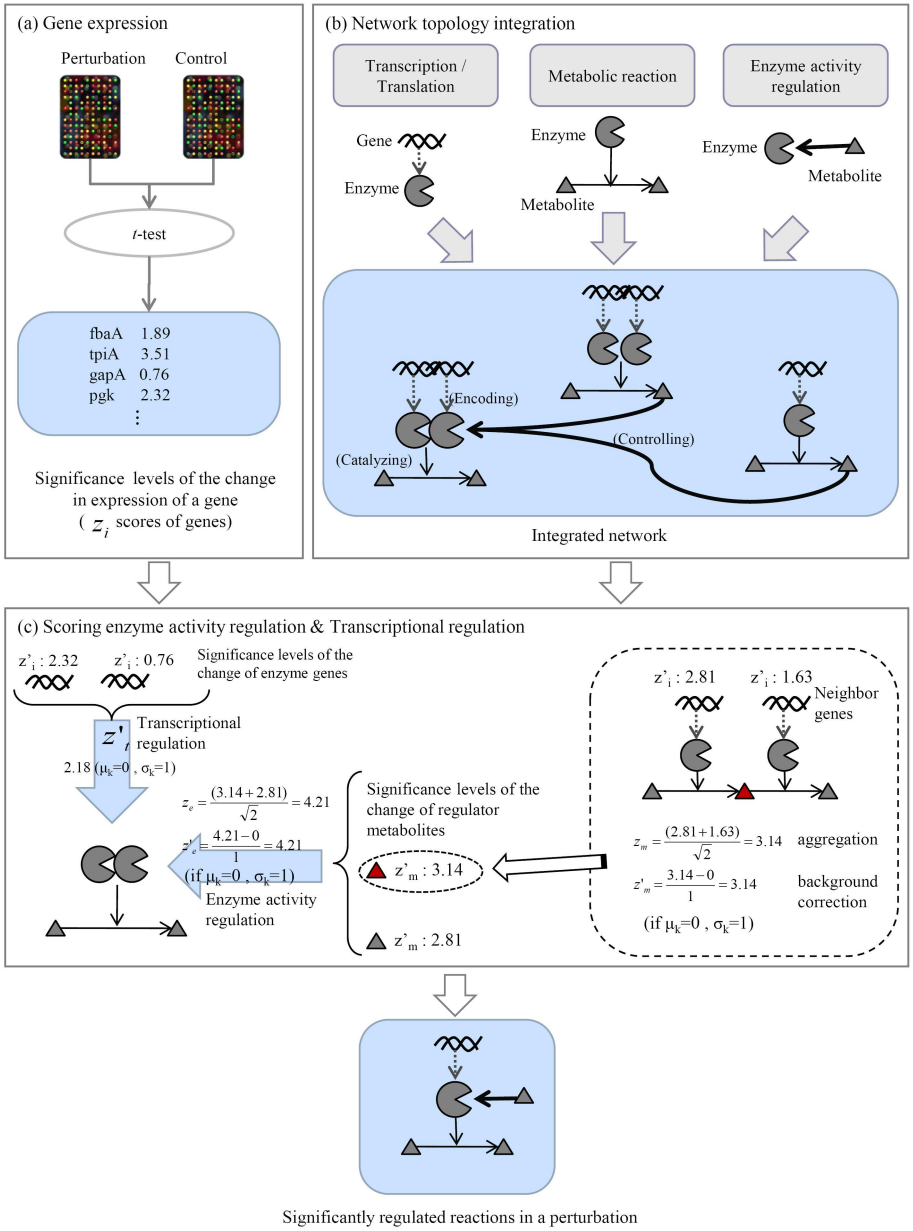


**Fig. 1.** Overview of our methodology. (a) Gene expression. The standardized z-scores of genes are obtained by *t*-test analysis to represent significant changes in the levels of gene expression. (b) Network integration. We merged information from three biological processes to construct an integrated network: i) transcription / translation, ii) metabolic reaction, and iii) enzyme activity regulation. (c) Scoring of the two types of regulation. To calculate the significance score ($z'_e$) of enzyme activity for a reaction, all significance scores ($z'_m$) of regulator metabolites were aggregated, and z-scores of neighboring genes were aggregated for calculation of $z'_m$ scores as described in the Methods section. In the case of transcriptional regulation ($z'_t$), all of the significance levels ($z_i$) of genes that encoded an enzyme were aggregated.

activity scores with transcriptional regulation scores in order to observe the regulatory effects of the two different regulation mechanisms. We tested the proposed method by using publicly available microarray data of *E. coli* K-12 strains that were grown under various conditions (see Supplementary Table 1).

## RESULTS AND DISCUSSION

We applied our methodology to five perturbation conditions (see Method and Supplementary Table 1). Table 1 summarizes the number of reactions that are significantly regulated by enzyme activity regulation ($z'_e$) and transcriptional regulation, ($z'_t$)with P-value < 0.001 for each perturbation.

Among the reactions that were significantly regulated by the control of enzyme activity, we identified reactions in which our results agreed with the previous case studies (6, 19-25) by looking for perturbations that had high significance scores. Following anaerobic perturbation, we identified significant reactions that are related to the central metabolism. Following pH perturbations, we identified reactions that are related to alkalinization or acidification.

The significant reactions that are relevant to central metabolism following anaerobic perturbation are listed in Supplementary Table 2. For example, the citrate (si)-synthase reaction (citsyn-rxn) is a significantly regulated reaction following anaerobic perturbation ($z'_e$ : 5.16, P-value : 1.18E-07). This enzyme is part of the tricarboxylic acid (TCA) cycle and is important in anaerobic respiration (6, 19, 20). Our result agrees with those of previous studies that have identified central respiration metabolism reactions that are affected by anaerobic perturbation (6, 19, 20).

Following pH perturbation in the presence or absence of oxygen, we identified reactions that are related to alkalinization or acidification. These include reactions of the TCA cycle, which consume acids, and hydrogenase/dehydrogenase enzymes, which interconvert hydrogen ions with hydrogen gas. Following pH perturbation in the presence or absence of oxygen, several enzymes with known importance for the catabolism of sugars and amino acids show a pH dependence that minimizes acid production at low external pH and maximizes acid production at high pH. This maintains cellular pH near 7.6 (21-26). Supplementary Table 3 lists the alkalinization- or acidification-related reactions that are significantly regulated by enzyme activity following four different perturbations.

Among these reactions, 26 perturbation-specific reactions are acid-relevant. Supplementary Table 4 shows the four classes of acid-relevant reactions (P-value < 0.001, $z'_e$ > 3.0902). We further considered all reactions in which $z'_e$ > 3.0902 following at least one pH perturbation, and divided these reactions into four classes: low-pH, aerobic (LA, Class 1); high-pH, aerobic (HA, Class 2); low-pH, anaerobic (LAN, Class 3); and high-pH, anaerobic (HAN, Class 4).

As shown in Supplementary Table 4, different types of perturbation yield different regulatory enzymes. In particular, reactions in the central metabolism are widely known to have a complex dependence on pH and oxygen (27). Thus, we tested the effects of pH and oxygen perturbation on the regulation of transcription and enzyme activity in enzymes in glycolysis and the TCA cycle. Fig. 2 shows the $z'_e$ and $z'_t$ scores of each reaction in the central metabolism following the five types of perturbation. This Figure shows how enzymatic reactions of the central metabolic pathway are finely regulated by the control of translation and enzyme activity following specific perturbations (P-value < 0.01, $z'_e$ > 2.3263).

**Table 1.** Number of significantly regulated reactions (P-value < 0.001) by enzyme activity regulation and transcriptional regulation following each of the five perturbations

| Perturbation type | No. of significant reactions by enzyme activity regulation[a] ($z'_e$) | No. of significant reactions by transcriptional regulation[a] ($z'_t$) |
|---|---|---|
| Anaerobic | 31 / 388 (FDR < 2%) | 16 / 1265 (FDR < 7%) |
| Low-pH (aerobic) | 23 / 388 (FDR < 2%) | 25 / 1265 (FDR < 6%) |
| High-pH (aerobic) | 8 / 388 (FDR < 4%) | 45 / 1265 (FDR < 3%) |
| Low-pH (anaerobic) | 3 / 388 (FDR < 9%) | 12 / 1265 (FDR < 11%) |
| High-pH (anaerobic) | 5 / 388 (FDR < 5%) | 20 / 1265 (FDR < 7%) |

[a]The numerators and denominators in columns 2 and 3 indicate the numbers of identified reactions and of candidate reactions, respectively. False discovery rate (FDR) indicates the expected percent of false predictions among all predictions.

## CONCLUSIONS

In this work, we proposed a method for scoring the enzyme activity of reactions in order to identify significantly regulated reactions following perturbation of the cellular environment. For this work, we used metabolic network topology information and microarray gene expression profiles of *E. coli* K-12 strain. Our method quantified the enzyme activity of reactions and identified significance scores for 388 reactions following five different types of perturbation. Among these 388 reactions, we identified 70 significantly regulated reactions (P-value < 0.001). Our results were consistent with previous case studies that examined the effects of metabolic perturbations on regulation. Moreover, we identified 26 reactions (P-value < 0.001) that show responses to the following perturbations: low-pH, aerobic; high-pH, aerobic; low-pH, anaerobic; and high-pH, anaerobic.

By identifying significantly regulated reactions following metabolic perturbation, we identified the reactions that were regulated in order to maintain cellular homeostasis following perturbation.

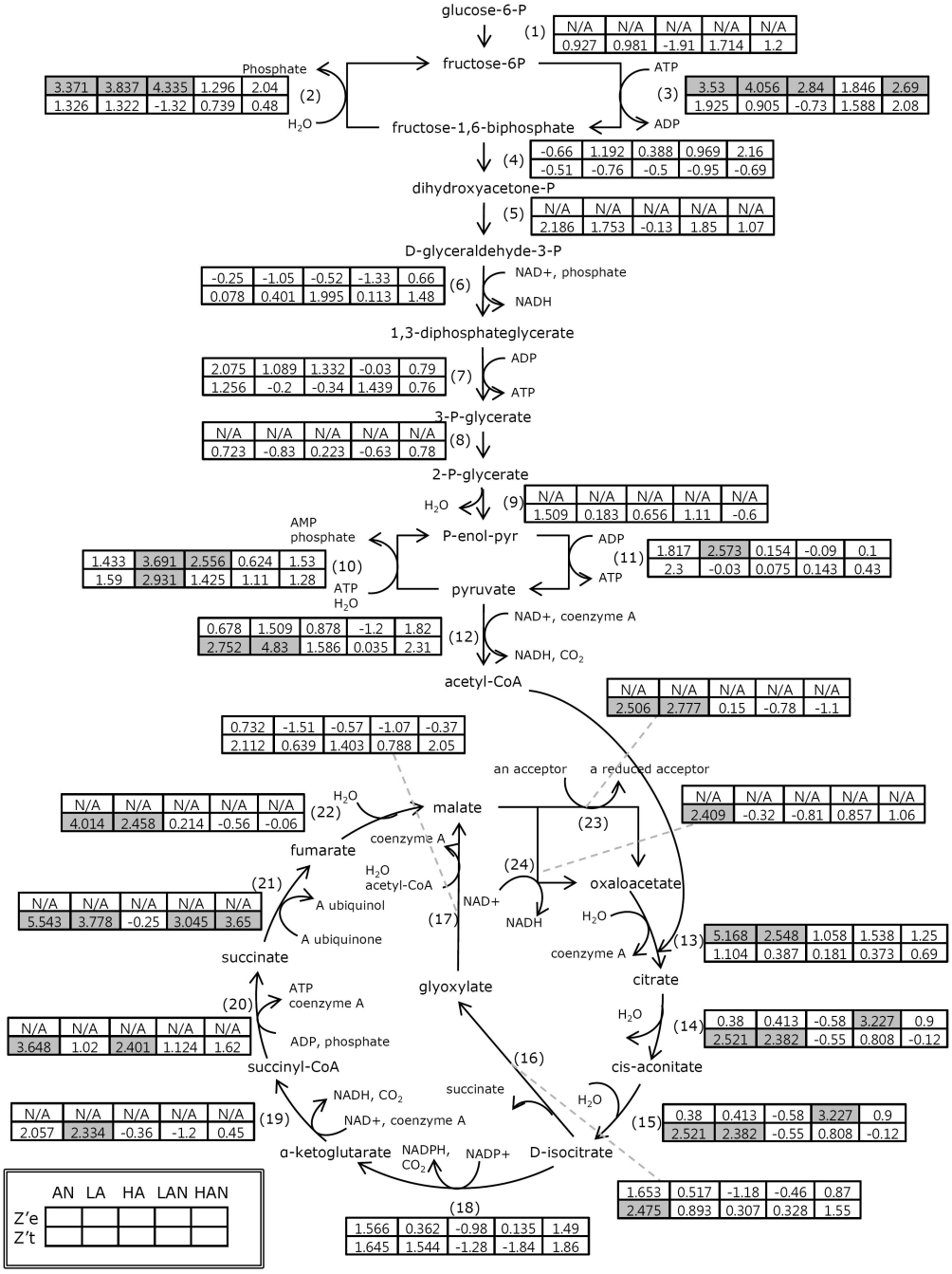


**Fig. 2.** $z'_e$ and $z'_t$ scores of 24 reactions of the central metabolic pathway (glycolysis ~ TCA) following five different perturbations. Significant $z'_e$ and $z'_t$ scores (P-value < 0.01, $z'_e$ > 2.3263) are shaded grey. If a reaction yielded no known information on enzyme activity regulation, the score is marked as N/A. Each 2 × 5 table summarizes the significance scores of the corresponding reaction. The first row of the table lists $z'_e$ scores (enzyme activity regulation) of a reaction following perturbation, and the second row lists $z'_t$ scores (transcriptional regulation) of a reaction for each of the five types of perturbation (AN: anaerobic; LA: low-pH, aerobic; HA: high-pH, aerobic; LAN: low-pH, anaerobic; HAN: high-pH, anaerobic). The number adjacent to each table indicates the reaction ID number from the Ecocyc database. (1) pglucisom-rxn, (2) f16bdephos-rxn, (3) 6pfructphos-rxn, (4) f16aldolase-rxn, (5) triosepisomerization-rxn, (6) gapoxnphosphn-rxn, (7) phosglyphos-rxn, (8) 3pgarearr-rxn, (9) 2pgadehydrat-rxn, (10) pepsynth-rxn, (11) pepdephos-rxn, (12) pyruvdeh-rxn, (13) citsyn-rxn, (14) aconitatedehydr-rxn, (15) aconitate-hydr-rxn, (16) isocit-cleav-rxn, (17) malsyn-rxn, (18) isocitdeh-rxn, (19) 2oxoglutaratedeh-rxn, (20) succcoasyn-rxn, (21) succinate-dehydrogenase-(ubiquinone)-rxn, (22) fumhydr-rxn, (23) malate-dehydrogenase-(acceptor)-rxn, (24) malate-deh-rxn.

## MATERIALS AND METHODS

### Data sets

In this work, we used three publicly available microarray data sets of *E. coli* K-12 that were measured with the Affymetrix GeneChip. The first dataset was an anaerobic perturbation dataset from an experiment in which the bacterium were grown in anaerobic conditions and aerobic conditions were controlled (6). The second dataset was a pH perturbation dataset for bacteria grown at pH 5.0, pH 7.0, and pH 8.7 under aerobic conditions (21). The third dataset was a pH perturbation dataset from an experiment in which the bacterium were grown at pH 5.7, pH 7.0, and pH 8.5 under anaerobic conditions (27). With these three microarray experiments, we analyzed five perturbations and compared the results with those of controls (see Supplementary Table 1).

### *T*-test and z-score

As shown in Fig. 1a, P-values are calculated by applying Student's *t*-test in order to evaluate a particular type of perturbation (e.g. aerobic vs. anaerobic). By using the inverse normal cumulative distribution function ($\theta^{-1}$), the P-value ($P_i$) of each gene ($gene_i$) is converted to a z-score ($z'_t$), which follows a standard normal distribution. A higher z-score signifies more significantly altered gene expression following a specific type of perturbation.

$$z_i = \theta^{-1}(1 - p_i)$$

### Network topology integration

It is necessary to integrate information on transcriptional regulation, enzyme activity regulation, and metabolic reaction rates in order to better understand the overall regulation mechanisms of metabolism. Thus, we merged three datasets to construct an integrated metabolic network (Fig. 1b). We collected information on transcriptional regulation, enzyme regulation, and metabolic reactions from the Ecocyc database (28), which describes the genome and biochemical machinery of *E. coli* K-12 strain. Note that the Ecocyc database offers information about the subunit enzymes as well as the isozymes of a reaction. Thus, in this study, the term 'enzymes' includes a set of subunits and isozymes. Fig. 1b shows a topologically integrated network. In this integrated network, we used 388 distinct reactions in which regulator metabolites regulate enzyme activity, and then scored the significance of each enzyme activity.

### Scoring regulation effects

Fig. 1c shows the conceptual diagram of the proposed scoring scheme for enzyme activity.

*Significance score for enzyme activity regulation* ($z'_e$). To determine the score of the regulation of the activity of an enzyme, we used the significance scores of regulator metabolites that were calculated from the z-scores of the neighboring enzyme genes of the regulator metabolites. A neighboring enzyme gene is defined as a gene that encodes an enzyme that shares at least one common metabolite in a catalyzing reaction (3). Formally stated, the aggregated score, $z_m$, of a metabolite having k neighboring genes is defined as:

$$z_m = \frac{1}{\sqrt{k}}\sum z_i$$

To minimize the potential of false-positive predictions, we normalized each $z_m$ score to a $z'_m$ score by correcting for the background distribution based on the mean ($\mu_{k_m}$) and standard deviation ($\sigma_{k_m}$) derived from $z_m$ scores of 5000 random sets of *k* genes.

$$z'_m = \frac{(z_m - \mu_{k_m})}{\sigma_{k_m}}$$

The $z'_m$ (normalized score) indicates the significance level of metabolite response to a perturbation. Using each $z'_m$, we calculated the significance score ($z_e$) of the enzyme activity of a reaction by aggregating associated $z'_m$ scores:

$$z_e = \frac{1}{\sqrt{k}}\sum z'_m$$

Analogous to the scoring scheme for metabolites, we normalized the $z_e$ score using a random sampling method:

$$z'_e = \frac{(z_e - \mu_{k_e})}{\sigma_{k_e}}$$

$z'_e$ scores identify the reactions with regulator metabolites that show significant changes following a specific perturbation.

*Significance score of transcriptional regulation* ($z'_t$). For our calculations, there can either be a set of isozymes, or an enzyme complex can consist of subunits. Here, we considered each isozyme and subunit as a distinct unit. Thus, we aggregated the z-scores of all distinct enzyme genes that encode the enzymes for a reaction in order to determine the significance score of each transcriptional regulation mechanism. The score of transcriptional regulation ($z'_t$) to a reaction with k genes is defined as:

$$z_t = \frac{1}{\sqrt{k}}\sum z_i$$

Similarly, we normalized $z_t$ using a random sampling method. The normalized $z'_t$ is calculated as:

$$z'_t = \frac{(z_t - \mu_{k_t})}{\sigma_{k_t}}$$

### Acknowledgements

This work was supported by the Ministry of Knowledge Economy, Republic of Korea, under the ITRC support program supervised by the IITA (IITA-2008-C1090-0801-0001). This work was also

supported by the Korean Systems Biology Research Grant (2007-00556) from the Ministry of Education, Science and Technology. We thank CHUNG Moon Soul Center of BioInformation and BioElectronics for providing research facilities.

## REFERENCES

1. Friedrich, C. G. (1998) Physiology and genetics of sulfur-oxidizing bacteria. *Adv. Microb. Physiol.* **39**, 235-289.
2. Smith, E. and Morowitz, H. J. (2004) Universality in intermediary metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 13168-13173.
3. Patil, K. R. and Nielsen, J. (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2685-2689.
4. Joyce, A. R. and Palsson, B. O. (2006) The model organism as a system: integrating 'omics' data sets. *Nat. Rev. Mol. Cell. Biol.* **7**, 198-210.
5. Herrgard, M. J., Lee, B. S., Portnoy, V. and Palsson, B. O. (2006) Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome. Res.* **16**, 627-635.
6. Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J. and Palsson, B. O. (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**, 92-96.
7. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R. and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929-934.
8. Yeang, C. H. and Vingron, M. (2006) A joint model of regulatory and metabolic networks. *BMC Bioinformatics.* **7**, 332.
9. Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y. and Barkai, N. (2002) Revealing modular organization in the yeast transcriptional network. *Nat. Genet.* **31**, 370-377.
10. Ihmels, J., Levy, R. and Barkai, N. (2004) Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **22**, 86-92.
11. Covert, M. W. and Palsson, B. O. (2002) Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **277**, 28058-28064.
12. Covert, M. W. and Palsson, B. O. (2003) Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J. Theor. Biol.* **221**, 309-325.
13. Cakir, T., Patil, K. R., Onsan, Z., Ulgen, K. O., Kirdar, B. and Nielsen, J. (2006) Integration of metabolome data with metabolic networks reveals reporter reactions. *Mol. Syst. Biol.* **2**, 50.
14. Gutteridge, A., Kanehisa, M. and Goto, S. (2007) Regulation of metabolic networks by small molecule metabolites. *BMC Bioinformatics.* **8**, 88.
15. Choi, M. M., Kim, E. A., Choi, S. Y., Kim, T. U., Cho, S. W. and Yang, S. J. (2007) Inhibitory properties of nerve-specific human glutamate dehydrogenase isozyme by chloroquine. *J. Biochem. Mol. Biol.* **40**, 1077-1082.
16. Chen, M., Wei, H., Cao, J., Liu, R., Wang, Y. and Zheng, C. (2007) Expression of Bacillus subtilis proBA genes and reduction of feedback inhibition of proline synthesis increases proline production and confers osmotolerance in transgenic *Arabidopsis*. *J. Biochem. Mol. Biol.* **40**, 396-403.
17. Ito, J., Cox, E. C. and Yanofsky, C. (1969) Anthranilate synthetase, an enzyme specified by the tryptophan operon of *Escherichia coli*: purification and characterization of component I. *J. Bacteriol.* **97**, 725-733.
18. Vlahos, C. J. and Dekker, E. E. (1990) Active-site residues of 2-keto-4-hydroxyglutarate aldolase from *Escherichia coli*. Bromopyruvate inactivation and labeling of glutamate 45. *J. Biol. Chem.* **265**, 20384-20389.
19. Baldoma, L. and Aguilar, J. (1988) Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: aerobic-anaerobic regulation of L-lactaldehyde dissimilation. *J. Bacteriol.* **170**, 416-421.
20. Green, J. and Paget, M. S. (2004) BACTERIAL REDOX SENSORS. *Nature Reviews Microbiology.* **2**, 954-966.
21. Maurer, L. M., Yohannes, E., Bondurant, S. S., Radmacher, M. and Slonczewski, J. L. (2005) pH regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* K-12. *J. Bacteriol.* **187**, 304-319.
22. Yohannes, E., Barnhart, D. M. and Slonczewski, J. L. (2004) pH-dependent catabolic protein expression during anaerobic growth of *Escherichia coli* K-12. *J. Bacteriol.* **186**, 192-199.
23. Stancik, L. M., Stancik, D. M., Schmidt, B., Barnhart, D. M., Yoncheva, Y. N. and Slonczewski, J. L. (2002) pH-dependent expression of periplasmic proteins and amino acid catabolism in *Escherichia coli*. *J. Bacteriol.* **184**, 4246-4258.
24. Blankenhorn, D., Phillips, J. and Slonczewski, J. L. (1999) Acid- and base-induced proteins during aerobic and anaerobic growth of *Escherichia coli* revealed by two-dimensional gel electrophoresis. *J. Bacteriol.* **181**, 2209-2216.
25. Foster, J. W. (2004) *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nat. Rev. Microbiol.* **2**, 898-907.
26. Slonczewski, J. L. a. J. W. F. (1996) pH-related genes and survival at extreme pH. *ASM Press.* 1539-1552.
27. Hayes, E. T., Wilks, J. C., Sanfilippo, P., Yohannes, E., Tate, D. P., Jones, B. D., Radmacher, M. D., BonDurant, S. S. and Slonczewski, J. L. (2006) Oxygen limitation modulates pH regulation of catabolism and hydrogenases, multidrug transporters, and envelope composition in *Escherichia coli* K-12. *BMC Microbiol.* **6**, 89.
28. Keseler, I. M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I. T., Peralta-Gil, M. and Karp, P. D. (2005) EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.* **33**, D334-D337.