The First International Symposium on Optimization and Systems Biology (OSB'07)
Beijing, China, August 8–10, 2007
Copyright © 2007 ORSC & APORC, pp. 143–150

Qualitative Cellular Dynamics Analysis with Gene Regulatory Information and Metabolic Pathways

Hojung Nam Kwang H. Lee Doheon Lee*

Department of Bio and Brain Engineering, Korea Advanced Institute Science and Technology (KAIST), Korea

1 Introducion

The genome of an organism plays a central role in the control of cellular processes such as genetic regulation, metabolic pathway, and signal transduction. Usually, these processes are very complex and closely connected. In a cell environment, cells use a number of mechanisms to regulate their metabolism to adapt to changing environments.

Metabolism is regulated by controlling the amounts of enzymes, the catalytic activities of enzymes (Figure 1). The amount of a particular enzyme depends on both its rate of synthesis and its rate of degradation. The level of most enzymes is adjusted primarily by changing the rate of transcription of the genes encoding them. The catalytic activity of enzymes is controlled in several ways, especially allosteric control is the most important. Several of these regulatory mechanisms are used simultaneously in many biological systems. One of the most intensely studied bacterial pathways is the pathway of tryptophan biosynthesis. The pathway of tryptophan biosynthesis is regulated by feedback inhibition. Tryptophan is the effector molecule for allosteric enzyme. When the end product of the pathway (tryptophan) attaches to enzyme, the enzyme is inactive and can no longer join glutamine and chorismic acid into anthranilate. If tryptophan is disjoined from the enzyme the pathway is resumed, and tryptophan synthesis will continue. Tryptophan biosynthesis is also regulated at a genetic level by the processes of enzyme repression. Tryptophan does so by binding to an allosteric (non-active) site on the tryptophan repressor. This alters the shape of the repressor protein so that the operator blocks the attachment of RNA polymerase to the promoter.

In this reason, to understand the real control principal of metabolism, it is necessary to consider the diverse regulation roles of metabolism. With the development

*Contacts: dhlee@biosoft.kaist.ac.kr

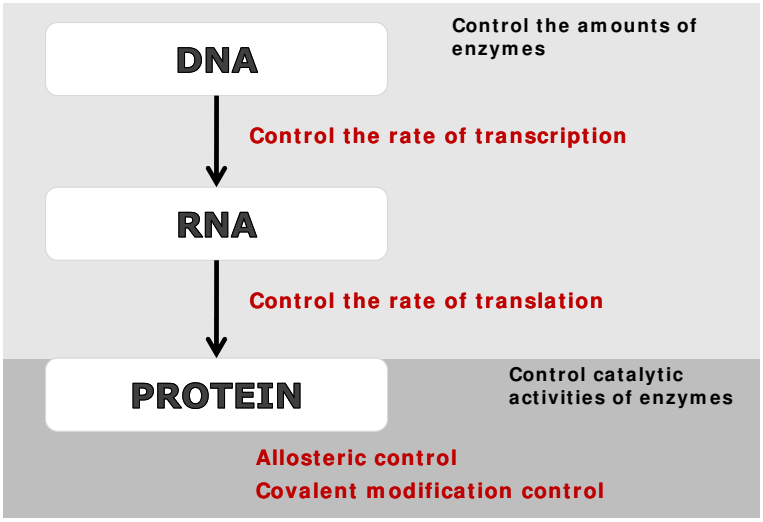

Figure 1: Various types of regulation mechanism

of biological high-throughput technologies allows a chance to understand broad picture of biological process by combining different levels of data. Transcriptome data provide an overview of the global regulation in the metabolism. Integrating genetic transcriptome data and functional metabolism makes it possible to understand cell's response against a specific condition in a concrete manner.

Several works have studied the significant roles of genetic transcriptional regulation in regulation of metabolism in the view point of joint process. (Covert, et al., 2004; Covert and Palsson, 2002; Covert and Palsson, 2003; Ideker, et al., 2001; Yeang and Vingron, 2006) have identified the new links between metabolites and transcriptional factors. (Patil and Nielsen, 2005) analyzed the coregulated subnetworks and reporter metabolites which correspond to specific conditions of microarray experiment. (Ihmels, et al., 2002; Ihmels, et al., 2004) integrated the two system with the aspect of regulation of metabolic enzymes. (Cakir, et al., 2006) identified reporter reactions, which are reactions where there are significant coordinated changes in the level of surrounding metabolites and combined the results with transcriptome data.

However, previous works mainly concentrated on the aspect of transcriptional regulation of metabolic enzymes, it is still insufficient to understand the real mechanism of metabolic regulation. To address this problem, we propose an analyzing method which quantifies the regulation effects of enzyme regulation to find significantly changed reactions in perturbation environments.

In our study, significance scores of metabolites which act as a regulator is quantified based on the z-scores of its neighboring enzyme gene by integrating metabolic pathway topology information and transcriptome data (See method). The basic hy-

pothesis of our work is that the change of concentration of metabolites, and the regulation responses of metabolic reactions can be quantified by using the network topology of metabolism.

2 Method

2.1 Microarray data

We used microarray examination of *Escherichia coli* K-12 strains grown in either aerobic or anaerobic conditions in M9 minimal media supplemented with glucose (Covert, et al., 2004). This data set is originally generated to investigate the changes of global gene expression in E. coli during an oxygen shift. We used samples that are produced under aerobic growth of wild-type strain on M9 media with glucose, anaerobic growth of wild-type strain on M9 media with glucose.

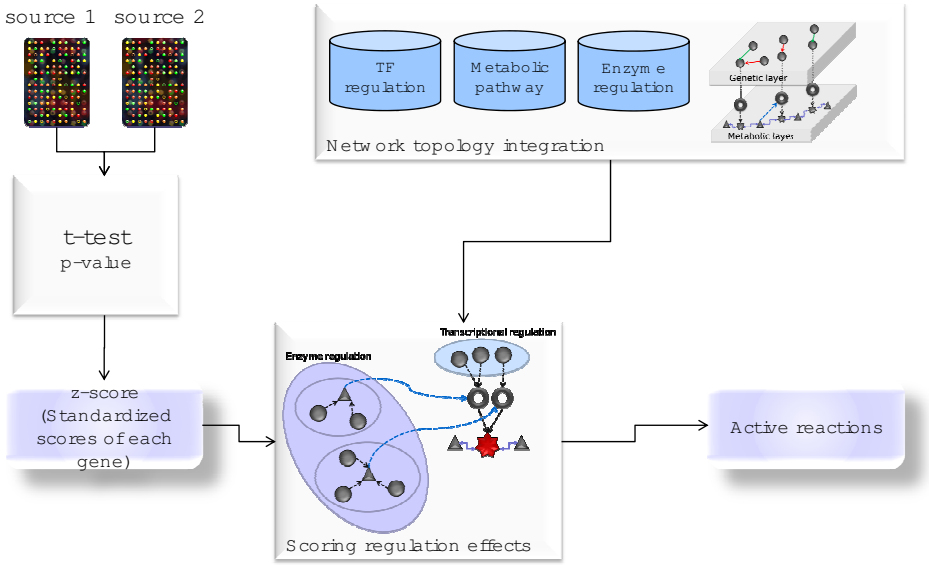

Figure 2: System overview

2.2 Network topology integration

It is necessary to merge transcriptional regulation information, enzyme regulation information and metabolic reaction information to understand overall regulation roles of metabolic pathway. Integrating the network topology of transcriptional regulatory network and metabolic pathway can be accomplished by making connections between enzyme proteins for metabolic reactions and genes which translate those enzyme proteins. Combing enzyme regulation information and metabolic pathway can be finished by adding binding relations as regulators from metabolites to enzymes.

We have collected transcriptional regulation information, enzyme regulation information, metabolic reaction information from the Ecocyc database (Keseler, et al., 2005) which is a bioinformatics database that describes the genome and the biochemical machinery of *E. coli* K-12 MG1655. Through integrating these three different information sets, topologically integrated network like figure 3 can be obtained.

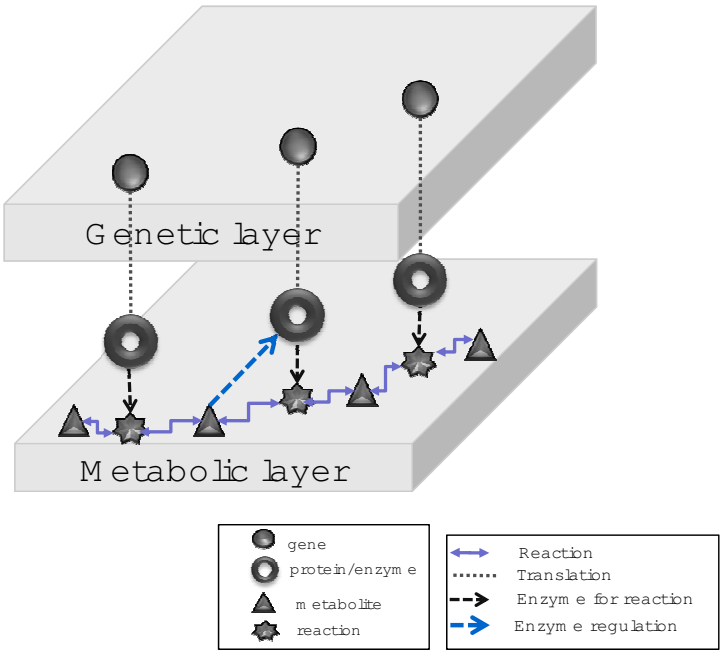

Figure 3: Topologically integrated network

2.3 T-test and z-score

The significance of the change in the expression levels of genes between two different conditions (e.g. aerobic and anaerobic) was calculated by applying student's t-test and gained p values. Every p_i of a $gene_i$ is transformed to a z-score which follows a standard normal distribution by using the inverse normal cumulative distribution (θ^{-1}). A higher z-score stands for more significantly changed gene expression of a gene in a perturbation.

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

Scoring regulation effects

Even if we can know all possible regulation information, not all regulation actions are responding at every moment of cell life or perturbation environments. In

this reason, it is necessary to quantify the regulation effects of two different regulation mechanisms to find significantly changed reactions in perturbation environments. To do quantify such regulation scores among transcriptional regulation and enzyme regulation, we begin by examining z-scores of genes (Figure 4). To identify the significant score of enzyme regulation, significance scores of metabolites which act as a regulator is quantified based on the z-scores of its neighboring enzyme gene (Patil and Nielsen, 2005). An aggregate z-score of metabolite (z_m) of k neighbor enzyme genes:

$$z_m = \frac{1}{\sqrt{k}} \sum z_i, \qquad z'_m = \frac{z_m - \mu_k}{\sigma_k}$$

Because z_m scores are depend on the size of k, z_m scores are needed to be corrected (z'_m) for the background distribution using mean (μ_k) and standard deviation (σ_k) of z-scores of sets of k genes which are obtained by random sampling from the same microarray data set. To gain mean (μ_k) and standard deviation (σ_k) value from random data, we repeated the scoring algorithm 2000 times for each k.

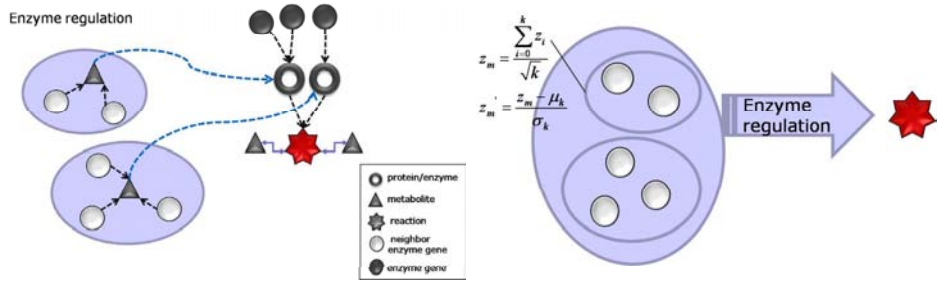

Figure 4: Regulation scoring scheme

3 Results

In this work, we quantify the regulation effects of two different regulation mechanisms to find active (significantly changed) reactions in perturbation environments from microarray experiments of *Escherichia coli* K-12 strains grown in either aerobic or anaerobic conditions and integrated network.

The cellular respiration is the mechanism that cells use to harvest energy and we divide the process into three stages. The first two stages of cellular respiration, glycolysis and the TCA cycle, are energy-releasing processes that break down glucose and other organic fules. Glycolysis begins repiration by breaking glucose into two molecules of a compound called pyruvic acid. The TCA cycle completes the breakdown of glucose by decomposing a derivative of pyruvic acid to carbon dioxide. The third stage of cellular respiration is the electron transport chain. The electrons from NAD(P)H and FADH2 are transferred to the electron transport chain in respiring or-

ganisms, with the formation of ATP. In fermentation cells, the NADH is reoxidized by an organic acceptor that is generated during catabolism.

Aerobic and anaerobic respirations are similar processes. Both depend on an electron transport chain to form a proton gradient that can be used to generate ATP. The chains themselves and their yield of ATP differ somewhat. But the critical difference is the terminal electron acceptor. In aerobic respiration, oxygen accepts electrons and is reduced to water. In anaerobic respiration, another compound is reduced by accepting these electrons. Compounds that can act as a terminal electron acceptor in anaerobic respiration include sulfate, nitrate, fumarate, and trimethylamine oxide.

By applying our method, we obtained the results that are consentaneous with the biological backgrounds which are stated above. Table 1 shows metabolite regulators with significant z-scores (z'_m). Many metabolites (e.g. ATP, ADP, NADH, NAD, SUC, ACETYL-coA, etc) that appear in cellular respiration processes are detected as significantly responding metabolites. (See supplementary data 1 for a full list of z-scores of metabolites.)

Table 1: Metabolite regulators with significant z'_m score ($p < 0.0003$)

Environmental perturbation (aerobic vs. anaerobic)							
Metabolite	# neighbor genes	z-score	p-value	Metabolite	# neighbor genes	z-score	p-value
ATP	431	8.090	0	O_2	48	4.831	0
ADP	360	8.053	0	NO_2	24	4.355	0
Pi	371	7.731	0	isocitrate	4	3.864	0.0001
SUC	34	6.807	0	CIS-ACONITATE	4	3.654	0.0001
NADH	94	6.061	0	NH_3	69	3.639	0.0001
FORMATE	28	6.053	0	L	13	3.567	0.0002
MAL	12	5.872	0	FH_4	13	3.478	0.0003
NAD	100	5.781	0	CH_3	9	3.457	0.0003
CoA	64	5.658	0	Ac-CoA	45	3.446	0.0003

Among 5220 metabolic reactions in Ecocyc, 258 distinct reactions are regulated by one or more metabolite regulators which showed significance change ($p < 0.05$). Figure 5 shows the metabolic pathway names and the numbers which tell how many reactions are involved in that pathway among 258 reactions.

In this paper, we suggest a method to quantify the regulation effects enzyme regulation by metabolite regulator to find significantly changed reactions in perturbation environments of aerobic vs. anaerobic. The results show that metabolites which are related with the cellular respiration are significantly changed and we get the list of significantly regulated pathways which are regulated by metabolite regulators.

Acknowledgement

This work was supported by the Korean Systems Biology Research Grant (2007-00556) from the Ministry of Science and Technology. We would like to thank CHUNG

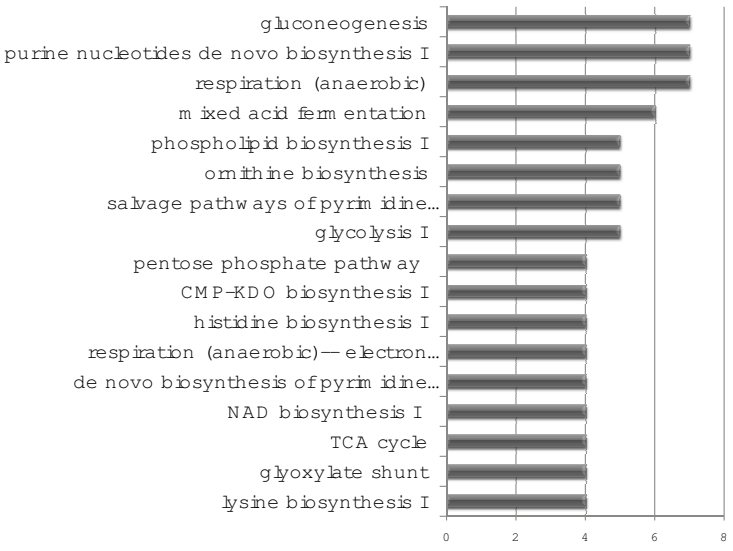

Figure 5: Top 17 selected metabolic pathways. The number shows reactions in each pathway which are regulated by significantly changed metabolites ($p < 0.05$).

Moon Soul Center for BioInformation and BioElectronics for providing research facilities.

References

[1] 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.

[2] 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.

[3] Covert, M.W. and Palsson, B.O. (2002) Transcriptional regulation in constraints-based metabolic models of Escherichia coli, *J Biol Chem*, **277**, 28058-28064.

[4] 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.

[5] 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.

[6] 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.

[7] Ihmels, J., Levy, R. and Barkai, N. (2004) Principles of transcriptional control in the metabolic network of Saccharomyces cerevisiae, *Nat Biotechnol*, **22**, 86-92.

[8] 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**, D334-337.

[9] 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.

[10] Yeang, C.H. and Vingron, M. (2006) A joint model of regulatory and metabolic networks, *BMC Bioinformatics*, **7**, 332.