Optimization-Based Approach for Transcriptional Regulatory Network Refinement

Dipak Barua[1], Joonhoon Kim[1,2], and Jennifer L Reed[1,2]

***Short Abstract* — We have developed a mixed integer linear programming (MILP) algorithm to identify and refine incorrect transcriptional regulatory interactions in a regulatory and metabolic network model. The algorithm can also predict genetic perturbations that allows for cell growth in specific lethal environments. The algorithm was applied to a genome-scale metabolic and regulatory network model of *E. coli,* and refinements were made based on the algorithm's predictions to improve the model's prediction accuracy. Additionally, few genes were identified by the algorithm that can rescue *E. coli* growth in lethal conditions. Experimental data exists supporting some of these predicted rescue lethal genes.**

A challenging task in the construction of regulatory network models is identifying accurate interactions between transcription factors and their corresponding gene targets [1,2]. We have developed a mixed integer linear programming (MILP) algorithm that identifies which TF-target interactions are inconsistent with experimental data, and can aide in the refinement of integrated metabolic and transcriptional regulatory models [3-5]. In addition, the algorithm can be used to identify genes that if allowed to violate their regulatory constraints (such as by over-expression) can rescue lethal phenotypes. The optimization algorithm allows an integrated metabolic and regulatory model to choose and violate specific regulatory rules in order to achieve non-zero growth phenotypes under different environmental conditions. We applied this algorithm to an existing regulatory network model for *E. coli* (iMC1010 [6]) to identify Boolean regulatory rules which caused incorrect growth predictions. Out of 13,750 knockout mutant/environmental cases evaluated, 1,143 (8.3%) were incorrectly predicted by the existing model to be unable to grow. The algorithm was applied to these 1,143 cases to identify the the incorrect no-growth phenotype prediction. Nearly a third of these cases correspond to growth on N-acetyl-glucosamine, N-acetyl-mannosamine, and N-acetylneuraminate and many of these can be corrected by relaxing the regulatory rule for *glmU* (whose gene product is required for peptidoglycan and lipopolysaccharide biosynthesis). Allowing for constitutive expression in the Boolean rule for *glmU* expression would result in substantially higher agreement between the regulatory model predictions and growth phenotyping data. In an additional 926 cases (6.7%) the existing transcriptional regulatory model correctly predicted knockout mutants would be unable to grow even though the metabolic network has other enzymes which would allow for growth. These cases correspond to lethal strains that could be rescued by the over expression of other gene products. On average, most lethal mutant strains would require 2.75 genes to be over-expressed so that they could grow on a condition. The developed approach allows for the automated evaluation of experimental data for both wildtype and knockout mutant strains in order to identify what transcriptional regulatory interactions are consistent and inconsistent with experimental observations. The approach can also be used to hypothesize strategies to rescue lethal mutants which if confirmed would help validate metabolic and regulatory network models.

[1]Department, of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53706. E-mail: barua.dipak@gmail.com

[2]Great Lakes Bioenergy Research Center, Madison, WI 53706

References

[1] Herrgard MJ, Covert MW, Palsson BO (2004) Reconstruction of microbial transcriptional regulatory networks. *Curr Opin Biotechnol* **15,** 70-77.

[2] Venancio T, Aravind L (2009) Reconstructing prokaryotic transcriptional regulatory networks: lessons from actinobacteria. *J Biol* **8**, 29

[3] Covert MW, Schilling CH, Palsson B (2001) Regulation of Gene Expression in Flux Balance Models of Metabolism. *J Theor Biol* **213,** 73-88

[4] Covert MW, Palsson BO (2003) Constraints-based models: Regulation of Gene Expression Reduces the Steady-state Solution Space. *J Theor Biol* **221**, 309-325.

[5] Shlomi T, Eisenberg Y, Sharan R, Ruppin E (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol* **3.**

[6] Covert MW, et al. (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature **429***, 92-96.