Combinatorial analysis of transcription regulation of response to oxidative stresses in E. coli

Thiele S.[1], Klie S.[2], Jozefczuk S.[3], Selbig J.[2,3], and Blachon S.[2]

[1] Institut for Computer Science, University of Potsdam, Potsdam, Germany
[2] Max Planck Insitute for Molecular Plant Physiology, Potsdam, Germany
[3] Institute for Biochemistry and Biology, University of Potsdam, Potsdam, Germany

Transcriptional regulation is a crucial component of bacterial stress response. The existence of operons and global transcriptional regulators allows E. coli to quickly adjust to changing environment. However the great variability of environmental signals requires the ability to combine activity of different transcriptional regulators to fine tune the response. By integrating (anti)correlated gene responses with knowledge on transcription factor regulation we aim at detecting behaviors that cannot be explained with current knowledge and can result from more complex yet unknown regulations.

Materials and Methods

Using microarrays, the time response to various stresses was measured at the transcript level. Correlations between genes that vary under stress were computed. When two genes are (anti)correlated, at least one common transcription factor should be responsible for the coordinated response. Conversely, if two genes have no common transcription factor, they should not be (anti)correlated. We propose an original approach based on the Answer Set Programming (ASP) framework. ASP enables to model the problem, detect conflicts and propose sets of solutions that resolve the conflicts.

Results

Focusing on oxidative stress response, 264 genes vary significantly and have at least one known transcription factor regulating their activities. Overall, 83 transcription factors regulate these genes. Applying the ASP framework, conflicts were detected in 4472 gene pairs. Using a feature of ASP, the cardinality constraint, it is possible to compute the minimal number of repairs. Here, adding 3 unknown transcription factors is enough to repair all conflicts. These unknown transcription factors can be identified as real transcription factors for which the targets remain to be discovered.

Discussion

Exploring all possible combinations of repairs can be done only at a high computational cost. We are currently implementing new rules to refine the search space. These rules are based on the problem properties and on biological knowledge (common region in promoter sequences). So far, the approach was applied only to data produced on oxidative stress. Other stresses (heat shocks, ...) will be studied in the short run.