4as High-Throughput Time-Series Transcriptional Profiling Analysis of a Biological System Subjected to Multiple Perturbations: a Case Study in Systems Biology

Bhaskar Dutta and Maria I. Klapa

In systems biology, a holistic approach is applied for the study and analysis of biological systems. For the implementation of this approach in complex biological systems, a high resolution map of the cellular fingerprints is required. In post genomic era, analysis of biological systems has moved from measuring a small set of biological markers to the simultaneous measurement of large number of biological markers of any cellular fingerprint. This was made possible by the development of high-throughput technologies like DNA microarray, mass spectral analysis of proteins and metabolites. However, studying single cellular fingerprint is inadequate and delineates a partial picture of the cellular systems. A comprehensive analysis requires the integration of different cellular fingerprints like: genome sequence, maps of gene and protein expressions, metabolic output, and in vivo enzymatic activity. Using systems engineering approach, the system should be perturbed and the response should be studied at different cellular levels. Integration of the data of different cellular fingerprints, representing effects of the perturbation at multiple levels is required to get a comprehensive idea about the system. However, few experiments have been done to-date that actually combines information of different cellular fingerprints from systematically perturbed biological system. One of the main hurdles of carrying out an integrated analysis is that we currently lack computational techniques that allow the identification of cause-effect relationships between the cellular fingerprints.

My doctoral work consisted of designing and performing experiments that has enabled us to study integrated time-series gene expression and metabolic profiling from a systems biology perspective, followed by development of suitable data analysis technique. Specifically, *Arabidopsis thaliana* liquid cultures grown for 12 days under constant light and temperature were subjected to perturbations of (1) CO_2 level in their growth environment (2) Osmotic stress through addition of NaCl (3) Trehalose (sugar) signal and (4) their growth media by replacing sucrose with glucose, applied individually or in combination. According to the experiment design it was possible to study the dynamic effect of (a) large number of perturbations applied to the same system (b) simultaneous perturbations and (c) comparison of the combined perturbation with the corresponding individual perturbations. Full genome cDNA microarrays (printed in TIGR) were used for dynamic gene expression profiling of the system subjected to the applied perturbations. About 400 hybridizations and the corresponding total RNA and mRNA extractions carried out by me, created a vast amount of useful data, which not only was used for my doctoral work but also was provided to the scientific community. For a comprehensive analysis of data generated, I worked on modifying the existing tools as well as developing new computational methods for gene expression profiling. To the best of our knowledge, we are the first to formulate a mathematical framework for the integrated analysis of transcriptional and metabolic time series data. All the results obtained were validated in the context of the known *A. thaliana* physiology.

Through my doctoral research, I have been exposed to experimental techniques like total RNA extraction, mRNA extraction and amplification, hybridization as well as multivariate data analysis techniques like clustering, hypothesis testing, regression, constructing gene regulatory network and I would like to further contribute to the field of systems biology experimentally and computationally.