

Methods for Inference of Genetic Regulatory Networks

Annette M. Evangelisti
Theoretical Biology and Biophysics
Theoretical Division
Los Alamos National Laboratory
Los Alamos, NM 87545
amevang@lanl.gov

William S. Hlavacek
Theoretical Biology and Biophysics
Theoretical Division
Los Alamos National Laboratory
Los Alamos, NM 87545
wish@lanl.gov

ABSTRACT

As a result of advances in technologies we have nearly complete lists of putative genes and their products for a number of organisms, including humans. The next challenge is to discover not only the function of these genes but also the underlying network of gene regulation that controls their expression.

Because cellular regulatory networks are highly connected and complex we envision the most useful end results of studies focused on regulatory networks will be predictive mathematical and/or computational models. Because it is desirable that studies of function and regulation keep pace with our ability to collect sequence data, an important challenge is that of automatic high throughput model building.

To develop the capability to reverse engineer regulatory networks using data collected with high throughput techniques, such as cDNA microarrays, we are pursuing the following strategy: (1) formulate a plausible model for the type of regulatory network of interest, (2) simulate a high throughput experiment, generating a synthetic data set, (3) apply reverse engineering methods, and (4) evaluate the outcome of reverse engineering. With this process, we can compare the relative effectiveness of methods for reverse engineering for a given set of data. We can also determine what experimental design is most suitable when using a given method. We will be able to find the minimal amount and quality of data needed by the method to reconstruct the underlying network.

Building on the work of Santillan and Mackey[1], we have developed a model for regulation of genes that encode enzymes required for biosynthesis of tryptophan in *Escherichia coli*. Using this model and a coarse model for microarray measurement errors, we can generate simulated large scale measurements of gene expression. We plan to use these data to evaluate mutual information approaches to reverse engineering. Our goal is to determine the theoretical and experimental methods required to reconstruct the network regulating tryptophan genes in *E. coli*. We also plan to quantify the advantage obtained through measurements of protein levels and conventional low throughput but more accurate measurements of selected mRNA levels. Results are preliminary at present.

ADDITIONAL AUTHORS

The authors are members of the project team at Los Alamos National Laboratory focused on Systems Research in Genetic Regulatory Networks. Other members of the team include J. Ambrosiano (ambro@lanl.gov), M.E. Wall (mewall@lanl.gov), D.H. Sharp (dhs@lanl.gov), J. M. Ettinger (ettinger@lanl.gov), and D.N. Hill (danhill@lanl.gov).

REFERENCES

- [1] Santillan, M. and Mackey, M.C. (2001) Dynamic regulation of the tryptophan operon: a modeling study and comparison with experimental data. *Proc. Natl. Acad. Sci. USA* 98, 1364-1369.