# Building and Testing Scalable Fuzzy Models of Genetic Regulation in Bacteria: Analysis of Genomic Data

**Bahrad A. Sokhansanj**
*sokhansanj@llnl.gov*
University of California, Lawrence Livermore National Laboratory
L-452, 7000 East Avenue, Livermore, CA, 94550, USA
ph. (925) 422-8643
fax. (925) 422-2282

**Garry R. Rodrigue**
*ghrodrigue@ucdavis.edu*
Department of Applied Science, University of California, Davis
Hertz Hall, L-794, 7000 East Ave., Livermore, CA, 94550
ph. (925) 424-4066

**J. Patrich Fitch**
*jpfitch@llnl.gov*
University of California, Lawrence Livermore National Laboratory
7000 East Avenue, Livemore, CA, 94550, USA
ph. (925) 422-3276

**TOPIC AREA: Computational Biology (Functional/Structural Genomics)**

*This work was performed under the auspices of the U. S. Department of Energy by University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

# Building and Testing Scalable Fuzzy Models of Bacterial Regulation: Analysis of Genomic Data for the Virulence Pathway of *Yersinia pestis*

Bahrad A. Sokhansanj, Garry R. Rodrigue, J. Patrick Fitch
University of California, Lawrence Livermore National Laboratory
sokhansanj@llnl.gov, L-452, 7000 East Ave., Livermore, CA, 94550, (925)-422-8643

**ABSTRACT** Recent technological advances in high-throughput data collection (reviewed in [1]) give biologists the ability to study increasingly complex systems as a whole. One example is the DNA chip or microarray, an efficient high-throughput method for measuring temporal changes in the production of mRNA from thousands of genes. The large number of data produced in a microarray experiment pose challenges for visualization, interpretation, model building, and integration with other experimental data. Intuitive linguistic models used by biologists in the past are inadequate for complex systems studied at a genomic scale. A new methodology is needed to systematically develop and test *mathematical* biological models, both to interpret experimental observations and predict the effect of perturbations (e.g. genetic engineering, pharmaceuticals, gene therapy) on the genetic networks of organsms.

Quantitative mathematical models have been proposed for well-studied biological systems (e.g. *E. coli* phages) based on simulation of biochemical kinetics using differential equations [2] and Monte Carlo [3]. Even for well-studied systems quantitative kinetic data are inadequate to build comprehensive accurate models. High-throughput genomic and proteomic technologies, such as DNA microarray data and 2D gel electrophoresis yield poorly quantitative data for mRNA and protein abundance. Even costly test tube experiments are frequently imprecise and do not necessarily represent the reality of living organisms. Given the limitations of biochemical data, fuzzy logic modeling represent a compromise between chemical kinetics and boolean logic, which lack the resolution required to model biological processes accurately. Importantly, the mathematical framework of fuzzy logic is easily adapted to the linguistic models currently prevalent in biology, and thus fuzzy models may be aplied by biologists without the need for experts in other fields or “black-box” software.

Fuzzy modeling of microarray data [4] and biochemical kinetics [5] has been proposed, but it is limited because in its traditional formulation, the number of rules in a fuzzy model grows exponentially with the number of variables and resolution. Thus, for complex systems, too much domain-specific knowledge is required to develop a model. However, recently the union rule configuration (URC) was proposed [6], solving the problem of combinatorial explosion. URC fuzzy logic models grow linearly with problem size, and they can be used as a general scheme to model genetic regulation networks. In our lab, we are developing URC fuzzy models for the genetic regulation of virulence in *Yersinia pestis*, the bacteria that cause plague. The models are being used in conjunction with fuzzy clustering to analyze DNA microarrays (Figure 1) with assistance from visualization tools that display patterns of genome expression (Figure 2).

**ACKNOWLEDGMENTS** This work was performed under the auspices of the U. S. Department of Energy by University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

**REFERENCES**

[1] J. P. Fitch and Sokhansanj, B., *Proc.IEEE*, vol. 88, pp. 1949-1971, 2000.
[2] D. Endy, et al., *Proc. Natl. Acad. Sci. USA*, vol. 97, pp. 5375-5380, 2000.
[3] A. Arkin, Ross, J., and McAdams, H. H., *Genomics*, vol. 149, pp. 1633-1648, 1998.
[4] P. J. Woolf and Wang, Y., vol. 3, pp. 9-15, 2000.
[5] B. Lee, Yen, J., Yang, L., and Liao, J. C., *Biotechnol. Bioeng.*, vol. 62, pp. 772-779, 1999.
[6] W. E. Combs and Andrews, J. E., *IEEE Trans. Fuzzy Syst.*, vol. 6, pp. 1-11, 1998.

**TOPIC AREA:** Computational Biology (Functional/Structural Genomics)

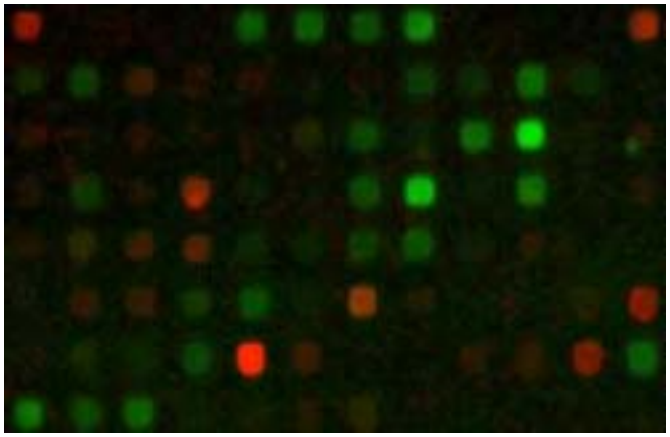

**Figure 1.** Sample microarray image, shown after image processing. The spots corresponding to controls and genes suspected of being involved with *Y. pestis* virulence. Red fluorescence results from mRNA produced under virulent conditions (37°C, low calcium); green fluorescence results from mRNA produced when the virulence mechanisms are dormant (25°C, high calcium). Thus, a spot that is mostly red corresponds to a gene that is more active when *Y. pestis* is virulence, and a spot that is mostly green corresponds to a gene that is less active during virulence. Either result is potentially significant for *Y. pestis* regulation. (Unpublished results courtesy of E. Garcia and A. Wyrobek, LLNL.)

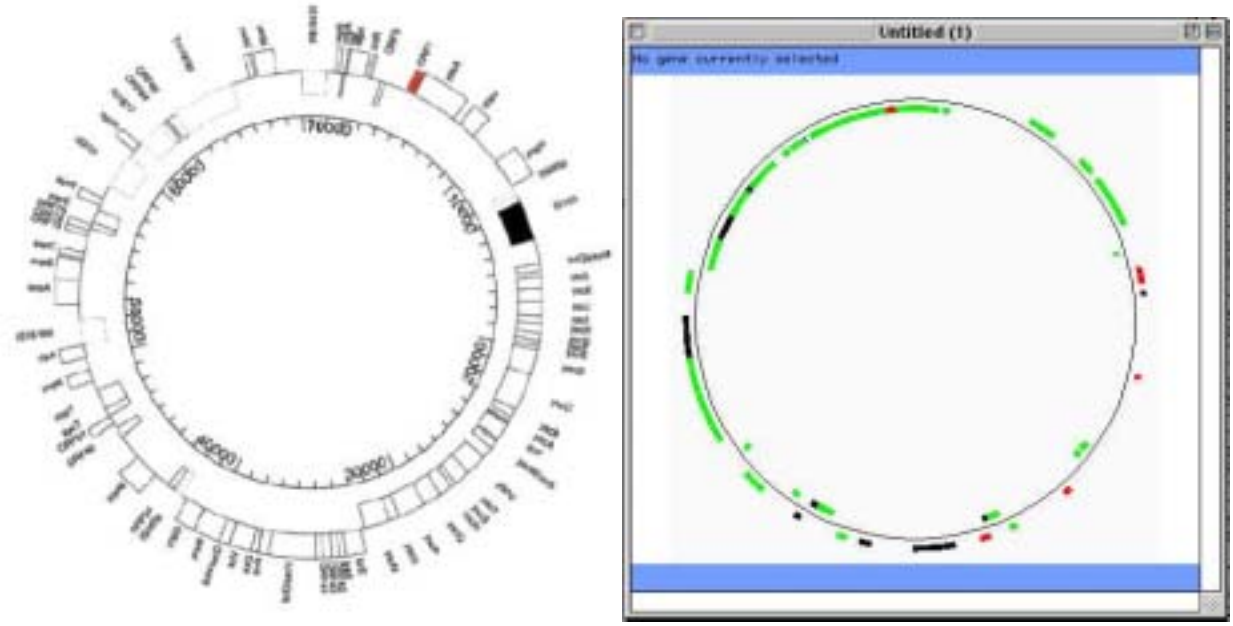


**Figure 2.** An interactive Java interface (shown on the right) illustrates expression of genes on the pCD1 virulence plasmid of *Y. pestis* superimposed on the gene order of the plasmid (shown left). This facilitates the identification of operons and regulatory sites, which contribute to the development and testing of gene network models. We are currently performing microarray experiments for the whole *Y. pestis* genome, including its chromosome.