

Digital Control and Communication in Living Cells

MIT9904-10

Progress Report: July 1, 2000 to December 31, 2000

Thomas F. Knight, Jr.

Project Overview

Genetic regulatory networks form one of the basic computational infrastructures of life. The science of such systems has been well established over the past twenty years through pioneering work of Monod and Ptashne, among others.

This project undertakes to transfer that scientific knowledge into engineering practice, by starting the serious work of characterizing components, engineering interfaces, simplifying the technology, and educating a set of students who can easily cross the boundaries between biological science and computational engineering. Stated more directly, the project is to learn how to engineer life.

Our approach has been to start our efforts with very simple structures, using well characterized organisms and genetic regulatory elements. In particular, we are working with standard laboratory strains of *E. coli* and standard promoter and reporter gene constructs.

An important initial goal was the construction and outfitting of a microbiology laboratory within the computer science building at MIT. This effort was largely complete by September, 1999. Another important initial goal was attracting and educating a core of motivated and educated graduate students and staff to perform the research. Early on we decided that it was easier to educate researchers with a CS background about biology than the other way around. Today, we have two full time graduate students, a post doctoral student, and two staff, all of whom are trained as computer scientists, but, equally, are trained in molecular biology theory and practice.

Our initial experiments primarily centered around the development of our skills in gene transfer, plasmid construction, and bacterial transformation. In these experiments, we focussed on simple gene promoter and reporter systems, such as the LacZ promoter and Green Fluorescent Protein (GFP).

Our previous work included the cloning, sequencing (genbank AF170104), and transfer of the bioluminescence and quorum sensing system from the marine bacterium *Vibrio fischeri* into laboratory strains of *E. coli*. This work was presented in June, 2000 at the DNA VI meeting in Leiden, Netherlands.

Progress Through December 2000

This summer and fall we completed our work on careful measurements of gene expression, essentially producing the equivalent of a curve tracer for genetic inverters. The major change which enabled this was the move from measurements on entire cultures of organisms (either in liquid broth or on agar plates) to the measurement of individual cells, using a dual reporter technique. Individual cells are measured using flow cytometry, which analyzes the fluorescence of cells one at a time, as they flow through a capillary. We measure two different fluorescent reporter genes, typically a cyan and yellow fluorescent protein. One gene reports the level of the DNA binding protein input, while the other reports the level of gene expression from the inverter. By varying the induction of the DNA binding protein, we can analyze, on an individual cell basis, the response of a genetic inverter. We have also found that the behavior is highly dependent on details such as the culture conditions, growth temperature, and many other variables, which we now believe we have under control.

Ron Weiss is completing a Ph.D. thesis on this work.

Additionally, we have studied the properties of a simple, non pathogenic bacterial species, *Mesoplasma florum*, which we are analyzing for possible use as a "chassis and power supply" for genetic regulatory networks. We have learned how to culture this organism, extract its genomic DNA, measured its chromosome size (870 kilobases), and begun the task of sequencing. Currently we have sequenced about 3% of the genome, using techniques of shotgun cloning of the genomic DNA into cosmid vectors, and primer walking through these clones. Regions we have sequenced look similar in most respects to the existing sequences of the pathogenic mycoplasma species *M. genitalium* and *M. pneumoniae*. Along the way, we sequenced the 16S ribosomal sequence from *M. florum* and from two other candidate species, *M. entomophilum* and *M. lactucae*. We discovered, to our surprise, that *M. florum* and *M. entomophilum* are very likely to be the same species, given that they have 100% identical 16S ribosomal DNA regions. At best, these two supposedly distinct species appear to be subspecies, perhaps strains, of the same organism.

We have characterized the antibiotic resistance of the wild type organism, and have begun the process of creating plasmid and transformation techniques which can be used in this organism. We acquired the plasmid pBOT1 from Prof. Renaudin in France, which replicates in a closely related species, and will be attempting to transform our organism with it. Similarly, we have acquired the plasmids pISM2062 from Prof. Minion (Iowa State) with which to replicate the single gene knockout strategy used by Hutchison in studying the minimal genome of *M. genitalium*.

In preparation for this work and some future protein work, we acquired new equipment consisting of an Amersham pulse field electrophoresis system, which allows us to measure the length of whole chromosome DNA strands, and an Amersham two dimensional protein gel system, which allows us to separate and quantify the number and amount of each of the estimated 700 proteins expressed in these simple organisms.

We also tried out one technique for rapid whole gene automated assembly, using mutation-cutting T4 Endonuclease V, as a means of enzymatically eliminating incorrectly manufactured strands of synthetic DNA. This technique did not work, and we are now reevaluating other approaches to this problem.

Research Plan for the Next Six Months

The transfer function measurement effort will be written up and submitted for publication.

We are engaged in discussions with the sequencing center at the Whitehead Institute in an effort to sequence the genome of *M. florum*.

We will begin the characterization of all proteins in *M. florum* by two dimensional gel electrophoresis, followed by MALDI-TOF mass spectroscopy of individual protein spots to identify coding sequences for the genes.

We plan to develop high throughput techniques which will allow us to locate the transcription start and stop locations of *M. florum* genes, which will in turn allow us to locate and characterize promoter sequences, and perhaps locate regulatory DNA binding sites.

We will begin the process of metabolic modelling of the organism, attempting to define as many enzymatic pathways as possible. We will begin the process of removing unnecessary genes and control logic from the organism by homologous recombination of plasmids.

We will be continuing our technology development efforts leading to long chain DNA direct synthesis, using some novel ideas involving binding of mismatched DNA strands to a gel matrix.

We will begin the process of standardizing and documenting a variety of genetic components, and the mechanisms for combining these components recursively into large scale genetically engineered networks.