

Digital Control and Communication in Living Cells

MIT9904-10

Progress Report: January 1, 2001—June 30, 2001

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 June 2001

Summer and Fall of 2000 we completed our work on careful measurements of gene expression, essentially producing the equivalent of a curve tracer for genetic inverters. The major enabling change 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.

This Spring we have started the process of intentionally modifying the properties of basic logic gates by controlled mutation of the ribosomal binding site of the target coding sequence, and by controlled mutation of the promoter. In particular, we have successfully combined these approaches to match the transfer curve behavior of the *cl* lambda promoter to that of the *lac* promoter.

Ron Weiss has completed the experimental work for his thesis in this area and plans to finish the document by August 2001.

Dylan Hirschell, an undergraduate CS major, has joined our group, and is working on extending the curve tracing work to sets of multiple gates, demonstrating that we can predict the behavior of assembled structures in the same way we can predict the behavior of individual gates. He is combining the engineered *cl* gate with a *lac* based gate and measuring the combined transfer curves.

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.

This spring we continued the sequencing effort, completing about 10% of the genomic sequence. These include long stretches, such as a 51 kilobase sequence encompassing one of the two rRNA sequences. The measured G.C ratio for this organism is 27%, varying from about 22% in coding regions to near 50% in rRNA regions.

Analysis and annotation of these sequences is under way, and will result in a sequence deposit within the next six months.

Our experiments in transforming *M. florum* with pBOT1 failed, presumably because of differences between the origin of replication of *Spiroplasma citri* and *M. florum*. Targeted sequencing was performed to locate and isolate the bacterial origin of replication of *M. florum* by a combination of degenerate PCR and direct genomic sequencing, resulting in the complete sequence of the origin. Presumably the construction of a plasmid similar to pBOT1 but with the new origin will result in a functional, replicating plasmid.

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.

During the spring we formulated a new plan using *E. coli* mutS protein as the detection method of choice, combined with HPLC as the purification protocol.

Additionally, we have expanded our collection of bacterial reporter genes with yellow and cyan derivative targeted to bacterial codon usage patterns.

Research Plan for the Next Six Months

This is the final report for this project.