

Biological Computing Technologies MIT9904-10

Proposal for 1999-2000 Funding

Tom Knight

We propose to create a new digital sensing, computing, and effecting technology based on biochemical mechanisms compatible with, and located within, living cells.

The key is the availability of well characterized, robust components, and the engineering methods for combining these components into genetic circuits. Genetically engineered biochemical circuits allow us to gain control, for the first time, over the digital behavior of living organisms.

Our initial efforts will be devoted to developing, characterizing, documenting, and distributing these biological components. Most (or all) of these components will be modified versions of existing biological components, because the scientific knowledge to perform *de novo* engineering of enzyme behavior is relatively primitive.

Another critical need is the development of efficient techniques for the construction of large scale genetic circuits. Current molecular biology approaches provide efficient tools for the insertion and study of individual genes through the introduction of plasmids. However, the introduction of more than a few is rare, because the standard scientific questions do not require the assembly of large engineered gene assemblies.

Software tools to assist in the design of structures, and in the design of experimental protocols for constructing those structures are also needed. Existing DNA engineering tools are relatively primitive, tuned to simple single gene insertions, and fail to provide simulation of the complex interactions of the kind of systems we anticipate building.

We believe this effort is important for several reasons. First, the development of ideas in a new domain often elucidates engineering principles which can be directly transferred to more developed technologies with important results. Understanding how biological mechanisms compute may provide a good model for solving problems in more conventional technologies. Second, we anticipate that important scientific questions which are currently inaccessible will be more

easily asked and answered with the sophisticated cellular control circuitry we envision. Essentially, our technology allows the insertion of probe points, breakpoints, and a debugger into natural biochemical pathways. The scientific returns will likely be large.

Finally, we anticipate a wide array of direct engineering applications. These application areas include chemical manufacturing technology, medical applications, nanotechnology and nanoscale electronics. We imagine creating, in effect, a process control computer, built within each cell of a bioreactor.

Our approach to developing this technology relies on the development of molecular biology expertise within an engineering context. Traditionally, the development of molecular biology has been driven by scientific and medical questions, with the engineering of systems being at best an afterthought. The result has been attention to highly complex systems, centered around eukaryotic, often mammalian, cells. These cells are inherently highly complex -- with a genome 1000 times larger than a typical bacterium, and with significant, complex, and poorly understood ultracellular structure. In contrast, our approach is to work with, and engineer the behavior of the simplest, easily grown living cells, bacteria. Working with the most commonly used *E. coli* strains has tremendous advantages in simplicity, rapid growth, and availability of infrastructure. The genome of this organism is completely known, although (very) incompletely understood.

We will be educating a small group of computer engineering graduate and undergraduate students with the required expertise in both the analytical and bench skills required to carry out this work. We also anticipate hiring a biology or chemical engineering post-doc to help us learn and develop our bench and analytic skills, and to broaden our perspectives. We will be collaborating with several other laboratories working in this area, including Roger Brent's laboratory at the Molecular Science Institute in Berkeley, and George Church's laboratory at Harvard Medical School. A significant output from our efforts will be in broadening the outlook of the computer engineering discipline to include biology as a first-class scientific contributor, along more traditional disciplines such as physics and electrical engineering.

To date, our efforts have concentrated primarily on the year-long construction and installation of a molecular biology laboratory within Technology Square. Final approval for working with recombinant molecules was granted from the Cambridge biosafety committee last month, and we are actively engaged in bringing up the required protocols for engineering DNA sequences, inserting them into cells, and verifying, through sequencing and restriction analysis, that the correct insertions were made.

Our next step is the construction of our first fully artificial gene "gate" structure. This structure consists of two parts: (1) an output reporter, constructed as combination of the transcriptional regulator region from the Lambda operator, inhibited by the *cl* gene, together with the coding sequence for a destabilized variant of the GFP (green fluorescent protein) coding sequence; and (2) an input portion, activated by an externally controllable inducer, IPTG, and producing the *cl* gene product. We anticipate fully characterizing the transfer curve of this construct before the end of 1999.

The careful characterization of these constructs is an important part of the engineering enterprise, and is often maddeningly missing in traditional biological papers. While our intention is to design and construct digital logic, which is largely robust in the presence of inaccuracy and noise, we still must understand enough about the details of the gate transfer characteristics to allow us to intelligently engineer those gates. Similarly, while the complete sequence of several organisms is now available, surprisingly, the sequence of several of the common reporter, selection, and sensing genes is difficult to obtain. We will be sequencing many of these, and adding those sequences to the existing biological databases.

A major result of our efforts will be the wide distribution of our results. Typically, this is done in the biological community through the sharing of engineered organisms, since the replication cost is near zero. The American Type Culture Collection and the Yale *E. coli* Strain Center maintain and distribute stocks of naturally occurring and engineered organisms; we anticipate being active contributors to these centers.

Our logic gate structures are only an initial, although very important, step in gaining control over cellular mechanisms. Next, we plan on expanding our range of available gates. Most of molecular biology is based on only a few, widely used sensors and reporters. After naming half a dozen, most biologists start having trouble. We need hundreds. This means either discovering naturally occurring proteins of the kind we need, or constructing them ourselves. We do not really know just how difficult this will be, but we do have a plan of attack -- the random mutagenesis of carefully chosen regions of existing DNA binding proteins, together with the massive screening capability offered by phage display and column elution. We estimate that we can easily screen 10^{10} possible proteins for activity, using the massive parallel search available through biological mechanisms.

Similarly, we must expand the range of sensors and actuators in our component inventory. This will involve the careful examination of existing sensor and actuators, such as the magnetic field detects of bacteria, to see which can be added to our component catalog. This will be a challenging and potentially

lengthy process, but is a task which, in time, can be shared between many groups.

Challenges also exist in constructing complex DNA molecules consisting of hundreds or thousands of genes in a controlled way. We are already examining how far we can go with the application of type II restriction enzymes, together with PCR techniques, to allow us to ligate tens or hundreds of distinct DNA fragments together in single reactions.

The design and analysis problem, of course, is always with us in the development of any complex technology. Again, the biological community has focussed on simple systems, rather than the hundreds of interacting genes needed for complex behavior. Computer engineering technology similar to existing VLSI tools for silicon design will have important results, we believe, not only in this effort, but in terms of understanding existing whole-genome structures whose sequence we now know.