**The Effectiveness of Three input RNA-based Gene Regulatory Systems**

**Main Objective**

To test the feasibility of using an RNA-based gene regulatory system utilizing three inputs and to measure the effectiveness of such systems in regulating the expression of GFP in *Saccharomyces cerevisiae.*

**Background and Significance**

The engineering of biological systems has the potential to address a variety of human needs, ranging from reducing energy consumption to ameliorating many health issues. However, biological engineering has been inhibited by the limited abilities of scientists to regulate biological systems. Because there are only a few naturally occurring promoter systems which respond to exogenous molecules, human control of gene expression is dangerous, expensive, and highly restrained. The ability to artificially associate certain genes with a regulatory system would allow for an unprecedented amount of control over biological systems at the genetic level (3).

RNA-based gene regulatory systems, specifically riboswitches, illustrate a highly effective means of controlling gene expression. Riboswitches consist of three components: an actuator, which is formed by a hammerhead ribozyme, an aptamer, which is designed to detect the presence of certain molecules, proteins, or even other RNA, and a transmitter, which relays information from the aptamer to the actuator through a shift in the equilibrium conformation of the device. The coding sequence for the RNA device is embedded in the 3’ UTR of the gene of which regulation is desired. Transcription of the DNA then leads to the creation of RNA which contains both the code for the regulated gene and the RNA device. The RNA then folds in such a way as to create the primary components of the device. In its active conformation, the hammerhead ribozyme will cleave itself, destroying the mRNA transcript and preventing expression of the gene. However, when the desired ligand binds to the aptamer region, a conformation change occurs that disrupts the active conformation of the hammerhead ribozyme. This ligand-bound conformation thus prevents the hammerhead ribozyme from self-cleavage, leaving the mRNA transcript intact and resulting in the expression of the regulated gene (3)

An important aspect of this particular RNA device is its modularity, imparted because the components can be designed independently and combined to create new functionality. For example, different ligands can be used to regulate gene expression simply by switching aptamers. Currently, there are only about 20 such aptamers with varying degrees of effectiveness. However, the introduction of more complex RNA structures which utilize multiple aptamers or sensors have exponentially increased the diversity of the RNA devices. Such diversity is realized because self-cleavage of the hammerhead ribozyme is now dependent upon the combination of two inputs rather than just one. The utilization of two input RNA switches have also allowed for the incorporation of primitive logic into these systems. Such logic systems include AND, OR, and NAND (which stands for “Not-AND”) gates, which are activated when both sensors, one sensor, and no sensors bind to a ligand. These two input systems have three methods of organization. The first method places two devices adjacent to each other so that both ribozymes would need to self-cleave in order to suppress gene expression. The second method places two “stacked” aptamers on one actuator so that once the top most aptamer is activated, it changes the conformation of the next aptamer so that it can bind another ligand, creating a cooperative binding gate. Finally, two independent apatmers can be placed on the actuator, so that the two inputs can independently affect the conformation of the actuator (Win, Science).

 While two input RNA devices have been explored in detail, the effectiveness of three input RNA-based gene regulatory systems have not been tested. Because there are only about 20 aptamers, with varying degrees of efficiency, even a combination of two inputs would only produce a maximum of less than 200 possible inputs (which is the summation of all the integers from 1 to 20) (personal communication with D. Endy). This number, however, assumes that all known aptamers can be utilized in multi-gated devices, and that two input gates would not interfere with other nearby two input gates which bind common ligands. Realistically, though, the number of different input combinations obtained from two input systems is likely much lower. Another limitation of two input systems is the limited amount of information such systems can process. Two input systems can only perform the most basic of logical operations. As a result, more robust RNA devices will be needed in order to regulate processes which depend upon the presence of many different types of ligands. A neuron, for instance, can have many thousands of dendrites which respond to neurotransmitters; thus the possibility of mimicking such complex biological structures relies upon the successful implementation of multiple input systems.

There are two means of rectifying the limited amounts of input possibilities. One would be to create more aptamers, a difficult task given the amount of effort which has gone into discovering viable aptamers and the limited amount of aptamers which have been produced. The other means of creating more input possibilities would be to utilize three input gates. Such a method would not only increase the amount of possible input combinations, but also allow for more complex logic gates. Three input logic gates would also allow for more subtle control of the RNA-based devices by utilizing properties which are unique to multiple receptor systems, such as cooperative binding, which results when receptors transit from a low-affinity state to a high-affinity state as a result of more ligands binding to available receptors. However, there are problems with increasing the number of inputs for each RNA system, perhaps the most substantial being that error rates are exponentially increased because each individual component of each gate must operate well individually in order for the system as a whole to give the desired output. Additionally, some devices seem to work less effectively when placed in proximity with other RNA devices. L2cm4 is one such device described by Win and Smolke (5).

**Methodology**

Gene synthesis: I will be building upon the basic framework used by the Smolke Lab in their creation of single and double input RNA based devices (5). The current published DNA sequences will be modified such that three actuators will be placed adjacent to each other, each with a spacer sequence separating the actuators to maintain independence of the three devices. I will be testing models where each actuator uses the same aptamer; such a model would offer me insight into the properties of cooperative bonding in these RNA devices by allowing me to observe whether the affinities of similar aptamers changes when neighboring aptamers have bound ligands. Such a model will also allow me to study other actuator to actuator interaction without needing to account for different kinds of ligands. I will also be building three actuator models which utilize three different aptamers as well and rearranging the order of the three actuators to test whether the order of the actuators affects the binding affinities of the aptamers or the levels of gene expression. I have decided to utilize three actuator systems because such systems would not require me to reorganize the DNA sequences for the components of the individual RNA devices, but rather, simply rearrange the sequences of predetermined DNA sequences. The aptamers I will be utilizing will be ones which bind to theophylline, caffeine, and tetracycline. Other aptamers not used in the Smolke Lab can also be tested. These aptamers would be found using the Ellington Lab Aptamer Database (http://aptamer.icmb.utexas.edu/). Secondary RNA structures can also be predicted using RNAstructure (http://rna.urmc.rochester.edu/rnastructure.html). I will use GeoScript, a company specializing in synthesis of oligonucleotides, to synthesize the DNA which will be transcribed into the RNA structures.

Creating the expression plasmid: I will be using plasmid pRzS with yeast-enhancing green florescence protein (yEGFP) under the control of GAL1-10 promoter as the vector to characterize the RNA devices. The DNA sequence for the device will be placed between restriction sites AvrII and XhorI, 3 nucleotides downstream of the stop codon for the green florescence protein and upstream of the ADH1 terminator sequence.

Testing for proper DNA sequencing: Once the plasmid has been produced it must be amplified and sequenced to ensure that the original DNA sequence is still intact. This is done by transforming the plasmid to a competent *Escherichia coli* strain, DH10B (Invitrogen) and then inserting the DNA through electroporation. Plasmids which have been verified by this technique can then be transformed to the *Saccharomyces cerevisiae* strain using standard lithium acetate procedures.

Test for gene expression: *S. cerevisiae* will be grown overnight with the proper nutrients and growing temperature. An input medium which would interact with the RNA switches would then be added and the *S. cerevisiae* would be allowed to grow for six hours. GFP levels can then be measured using a flow cytometer. Basal values of GFP expression, which represent the controls for this experiment, will be measured when the settings necessary for self-cleavage (or ribozyme conformation rearrangement in the inverted switches) are in place. These basal values will then be compared to the GFP expression when the conditions necessary for predicted conformation rearrangement (or self-cleavage in the inverted switches) are reached. Input switch, which is the range of input concentration over which the device output changes, can also be measured.

**Experimental Timeline**

Present to 6/22/09 – I will first need to learn the biochemistry behind the RNA structures. This will involve learning about the thermodynamic properties of the four ribonucleotides and specifically the properties of RNA folding. Understanding the thermodynamic factors which contribute to RNA conformational changes will allow me to better design RNA structures so that the binding energy will allow for a stable shift in conformation. I will also be learning any of the other basic lab techniques I might need to conduct my experiment from a graduate student at the Smolke Lab. Finally, I will make the final design plans for my RNA switches and have the oligonucleiotides synthesized by GeoScript. Upon receiving the oligonucleotides, I will design primers for my sequences and later perform PCR on my oligonucleotides so that I will have a sufficient amount of DNA for testing over the summer.

6/23/09 to 7/25/09 – This period will primarily be focused upon getting the *E. coli* to properly express GFP. I will first implant the pRzS plasmid with yeast-enhancing green florescence protein into *E. coli* using electroporation and culture the *E. coli* in order to replicate the plasmid. Next, I will remove the plasmid from the *E. coli* through minipreparation and add my sequence for my RNA device between the restriction sites described above. I will then culture the *E. coli* once more and sequence the resulting DNA to insure the desired vector is still intact. A cycle of implanting the genes, culturing the *E. coli* and testing the DNA inside the *E. coli* should take about three days, however, there will likely be numerous errors which need to be resolved before the *E. coli* will properly replicate the desired vector. Preliminary manipulation of the *E. coli* through the RNA switches will also be preformed to insure the functionality of the riboswitches.

7/26/09 to 9/16/09 – The final portion of my research will be dedicated to testing my results. I will likely need to conduct a number of trial runs so I can become familiar with what to expect. Then I will work on getting enough results so that my numbers become statistically significant. Because I will be testing nine different three actuator systems (three systems involving the same aptamers, and six systems of different aptamer arrangement), there will need to be a large number of runs and an extensive amount of data collecting. Finally, I will put aside one or two weeks at the end to trouble-shoot any potential problems. For instance, it is likely that some actuators will not respond to the presence of a ligand, in which case the transmitter region must be manipulated by single nucleotide changes in order for the actuator to become more sensitive to conformation changes in the aptamer region.

**Resources**

I will be working in Smolke Lab under the guidance of Dr. Christina Smolke. I will also have a graduate advisor, Isis Trenchard, who also works in the Smolke Lab and will be offering guidance throughout the research process. All the resources needed for this experiment will be available at the Smolke Lab. More specifically, the Smolke Lab already includes a large array of aptamers and other components of the RNA devices which will allow for more testing over the summer.

**Previous Experience**

Because this research project is as much chemical as biological, I believe that my strong background research in chemistry will be of great use to me. I have performed research with Dr. Bocarsley at Princeton University on the synthesis of a novel material known as cyanogels. During my time there, I analyzed the chemical properties of cyanogels under numerous synthesis conditions with IR spectroscopy and gas chromatography. In addition, I have studied the process of creating superplasticizers and other concrete admixtures using emulsion polymerization with Hongli Willimann of Elotex. My work ultimately led to a patent and a paper which won the William Welles Merit Scholarship.

 At Stanford, I have taken a class in synthetic biology with Drew Endy in which half the quarter was dedicated to discussing the regulation of gene expression with RNA. I have discussed my project with Professor Endy and he believes it will be one of great benefit to those looking to regulate gene expression using RNA. Additionally, I will have completed a lab course in synthetic biology by the end of spring quarter.

**Budget**

Calculated based on estimates from Stanford housing/dining services:

$2,000 Housing - Summer Research College (6/22/09-8/31/09)

$350 Housing - Interim Housing (8/31/09-9/16/09)

$1,500 19-meals a week meal plan

$1,250 Summer Stipend (10 weeks)

Total = $5,200

**References**

1. Ellington, Andy. The Ellington Lab Aptamer Database. The University of Texas. 2 Feb. 2009 <http://aptamer.icmb.utexas.edu/>.
2. Mathews, David. RNAstructure, Version 4.6. 6 May 2008. University of Rochester Medical Center. 2 Feb. 2009 <http://rna.urmc.rochester.edu/rnastructure.html>.
3. Win, Maung N., and Christina D. Smolke. "A modular and extensible RNA-based gene-regulatory platform for engineering cellular function." PNAS 104.36: 14283-88.
4. Win, Maung N., and Christina D. Smolke. "Higher-Order Cellular Information Processing with Synthetic RNA Devices." Science 322: 456-60.
5. Win, Maung N., and Christina D. Smolke. "Supporting Online Material for Higher-Order Cellular Information Processing with Synthetic RNA Devices." Science 17 Oct. 2008. 1 Feb. 2009 <www.sciencemag.org/cgi/content/full/322/5900/456/DC1>.