Synthetic Biology: Engineering Biological Systems

David Ouyang and Dr. Jonathan Silberg

Abstract

Recent advancements in molecular biology and biochemistry allow for a new field of bioengineering known as synthetic biology. Using biological parts discovered in the last thirty years and mathematical models grounded in physical principles, synthetic biology seeks to create biological systems with user-defined behaviors. The major focus of research in this emerging field is the characterization of genetic regulation and the abstraction of biological systems to clearly defined logic circuits. With the abstraction of individual DNA sequences to known biological functions, synthetic biologists seek to create a standard list of interchangeable biological parts as the foundation of this emerging field. Through genetic manipulation, these parts are expected to be useful for programming biological machines that process information, synthesize chemicals, and fabricate complex biomaterials that improve our quality of life.

Genomic Era and Tools of the Trade

On June 26, 2000, President Bill Clinton and Prime Minister Tony Blair, along with Francis Collins, director of the Human Genome Project at the NIH, and Craig Venter, president of Celera Genomics, announced the arrival of the genomic era with the sequencing of the first draft sequence of the human genome. With this wealth of information, scientists and policy-makers alike were eager to welcome in the genomic era of genetics. Doctors dreamed of personalized medicine, where genomic information can be used to diagnose individual predispositions to cancer and disease. Politicians pondered the implications of genetic profiling, where insurance companies can potentially use genetic information to screen policyholders. The genomic era is bright with promise and unprecedented potential but also rife with social implications and practical applications.

While a sequenced genome provides a boon of new information and the scientific community is quick to emphasize the potential of this plethora of information, there are still many challenges in its interpretation and analysis. The interpretation of genomic data requires both high throughput techniques, such as microarray analysis, and heuristic algorithms in bioinformatics to analyze large amounts of data. Microarray analysis allows researchers to understand differential expression of many different proteins between different species, ages, and diseases states. With more than four billion base-pairs in the human genome and over thirty thousand open reading frames, the sheer size of the human genome requires the use of ad hoc analytical methods. The status quo approach in analyzing individual enzymes and molecules is complemented by a recent desire to understand entire systems, regulatory networks, and gene families. Exponentially increasing information on biological organisms and increasing computational power has broadened the perspective of current biological research.

Although genomic sequences provide insight into the enzymes that make up an organism, understanding of how these parts work together to produce complex phenotypes is the focus of current research. Understanding the regulation of gene expression and multicellular development will require a deeper analysis of how transcription and stability of mRNA is regulated in response to the environmental stimuli. Despite the age old debate between nature vs. nurture, it is the interplay of the environment and gene products that determine disease states and merge to create the fascinating output of life. Greater understanding of the regulation of gene products is required in determining their effects on physiology and development. Synthetic biology seeks to understand and apply understanding of biological regulation to tackle general problems.

Recombinant DNA technology laid the foundation for manipulation of biological systems on a molecular level, but recent advances in DNA sequencing and synthesis technology have greatly expanded the potential of biological engineering projects. The decreasing cost of oligonucleotide synthesis as well as improved techniques of combining oligonucleotides allows unparalleled flexibility in synthesizing long DNA sequences. From traditional methods of subcloning using restriction endonucleases and ligases to polymerase-based techniques such as gene Splicing by Overlap Extension (gene SOEing), researchers have unprecedented power in their ability to alter and characterize DNA. We can now identify new genes or regulatory sequences in diverse systems and recombine them into novel networks that attempt to recreate our understanding of existing biological systems. The rapidly expanding molecular biologist's toolkit broadens the scope of manipulation to whole genetic systems instead of individual genes.

The current state of molecular biology has improved our understanding of the networks of biomolecular interactions that give rise to complex phenotypes and allows for unprecedented control of biological systems through clear characterization and synthetic

Biological Parts

Promoters
A promoter is a DNA sequence that tends to recruit transcriptional machinery and lead to transcription of the downstream DNA sequence.

Ribosome Binding Sites
A ribosome binding site (RBS) is an RNA sequence found in mRNA to which ribosomes can bind and initiate translation.

Protein tags and modifiers
Protein tags and modifiers are short peptide sequences cloned in frame with protein coding sequences that change the protein's behavior. Protein tags and modifiers might change the protein's location, alter its degradation rate, target the protein for cleavage, or enable it to be readily purified.

Protein coding sequences
Protein coding sequences encode the amino acid sequence of a particular protein. Note that some protein coding sequences only encode a protein domain or half a protein. Others encode a full-length protein from start codon to stop codon. Coding sequences for gene expression reporters such as LacZ and GFP are also included here.

Terminators
A terminator is an RNA sequence that usually occurs at the end of a gene or operon mRNA and causes transcription to stop.

DNA
DNA parts provide functionality to the DNA itself. DNA parts include cloning sites, scars, primer binding sites, spacers, recombination sites, conjugative tranfer elements, transposons, origami, and aptamers.

Plasmids
A plasmid is a circular, double-stranded DNA molecules typically containing a few thousand base pairs that replicate within the cell independently of the chromosomal DNA.

Plasmid backbones
A plasmid backbone is defined as the plasmid sequence beginning with the BioBrick suffix, including the replication origin and antibiotic resistance marker, and ending with the BioBrick prefix.

Figure 2: Rational design of novel networks of interrelated regulatory networks can provide a deeper understanding of biological systems. Such networks depend on a database of quantitatively characterized biological parts. (Description of Biological Parts and images courtesy of http://partsregistry.org/Catalog)

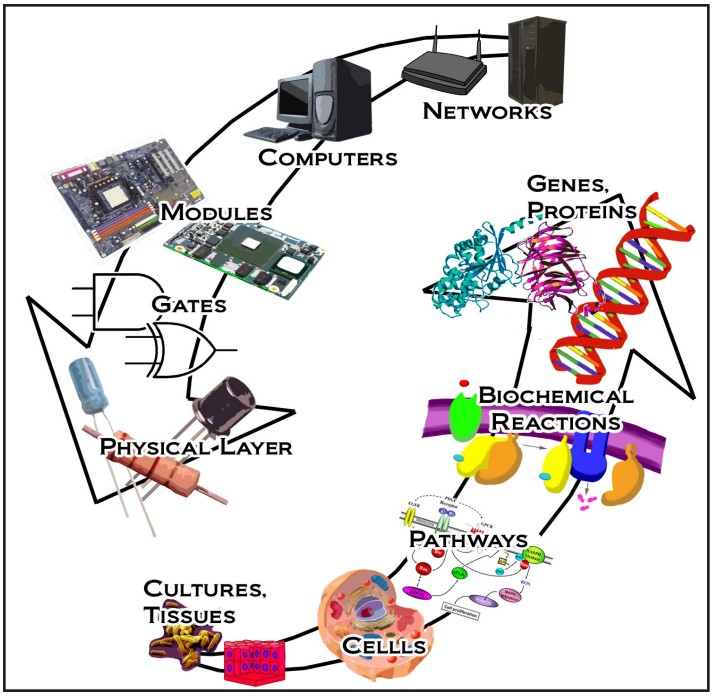

Figure 1: Understanding of the many levels of biological regulation can be applied to engineer unique cellular behaviors with desired properties. (Graphic by Katelyn Gray)

techniques. Just as electrical engineering required increased aptness in manipulating individual circuits and transistors, biology is on the cusp of synthetic potential as new technologies overcome technical difficulties challenging previous generations of scientists.

Concept

Synthetic biology can be described as a hierarchy of fundamental biological concepts. From discrete genetic parts to whole biological circuits, each level of regulation builds upon a lower level of biological function for the ultimate goal of using biological systems to perform novel tasks or improving upon natural functions. Individual genetic parts, or particular DNA sequences with known functionality, can be integrated into genetic circuits. Genetic circuits, or new combinations of regulatory and coding sequences, can be created to produce unique behavior. Ultimately, these genetic circuits can be incorporated into biological organisms or systems.

Ongoing efforts in synthetic biology are focused on the creation of reusable, modular fragments with clearly characterized behavior and functionality in biological systems. With the discovery of the lac operon, biologists recognized the possibility for digital, discrete outputs within biological systems. Detecting the presence of lactose, the LacI repressor recognizes and binds to particular DNA sequences upstream of coding regions, regulating the transcription of the gene products in an all-or-none fashion. With clearly characterized behavior, the LacI repressor is already widely used in biotechnology applications, such as PET vectors, as integral parts of simple genetic circuits. As the biological analog of electronic circuits, researchers hope to use a growing repertoire of genetic parts to mimic logic functionalities and produce com-

plex output.

The basic premise of synthetic biology is the ability to characterize and categorize a database of biological parts. A prominent example of this concept is the Registry of Standard Biological Parts (http://partsregistry.org/Main_Page) maintained by the BioBrick Foundation. Drawing upon the analogy of Lego bricks, synthetic biologists hope to use a standardized list of biological parts, 'BioBricks', to build large constructs with novel activity and unprecedented functionality. With defined activities for each component and a standardized subcloning method for combining DNA sequences, synthetic biologists hope to easily integrate 'BioBricks' to create novel biological circuits in a process analogous to the way computer scientists program computers. A database of DNA sequences, the Registry details the specific activity of individual sequences, the original sources and additional information necessary for synthetic biologists to integrate biological parts into particular biological systems.

Genetic Parts

There are three distinct levels where biological information can be regulated. In biological systems, information moves from DNA to RNA to protein. First proposed by Francis Crick in 1958, this "central dogma of molecular biology" addresses the detailed residue-by-residue transfer of sequential information. Synthetic biology utilizes regulatory elements at each level of this basic concept to create novel biological machines. On the DNA level, current understanding of genetic regulation reveals a complex system of promoters and terminators regulating transcription. The Registry contains a wide collection of parts for regulating transcription and translation, such as constitutive, inducible and repressible promoters, operator sequences, and ribosomal binding sites. Promoters, the 5' upstream DNA sequencing before coding regions, determine the amount, duration, and timing of the translation. In the Registry, there is a large catalogue of terminator sequences. The 3' region after coding regions, which form hairpin loops at the end of mRNA transcripts, cause RNA polymerase to dissociate from the template strand and end transcription. This compilation of a wide body of knowledge and literature about genetic regulation chronicles the behavior of many DNA sequences found in native systems.

Knowledge of regulation on the RNA level is applied to synthetic biology and builds upon a deep understanding of the regulation of protein production through mRNA stability and translation efficiency. Native systems display a wide range of RNA regulation that help modulate where and when particular proteins are translated. Transcribed RNA has the unique characteristic of being able to form diverse forms of secondary structure. Hairpin looping, which is intramolecular basepairing of palindromic sequences of RNA, is the basis of RNA secondary structure and can be used to create complex three dimensional structures. With the additional complexity of secondary structures, engineered RNAs function in RNA interference, as riboregulators, and catalyze key reactions. These RNA structures have been shown to mediate ligand binding and show temperature dependent activity. With temperature mediated stability, RNA sequences with designed hairpin loops can function as biological thermometers, regulating temperature sensitive expression.

The use of riboregulators is a prominent example of applying understanding of RNA behavior to regulate the expression of gene products in biological systems. Collins et al designed a system of RNA molecules that requires cooperative function of multiple RNA molecules for translation to occur. An mRNA transcript has an additional 5' sequence complementary to the ribosomal binding site, prevent binding to the ribosome from binding to and translating the gene product. This 'lock' sequence can be unlocked by the regulated production of another mRNA molecule with similar homology and tighter binding affinity; allowing translation. Riboregulators with ligand mediated activity can bind to specific mRNA transcripts, silencing translation of particular genes as the result

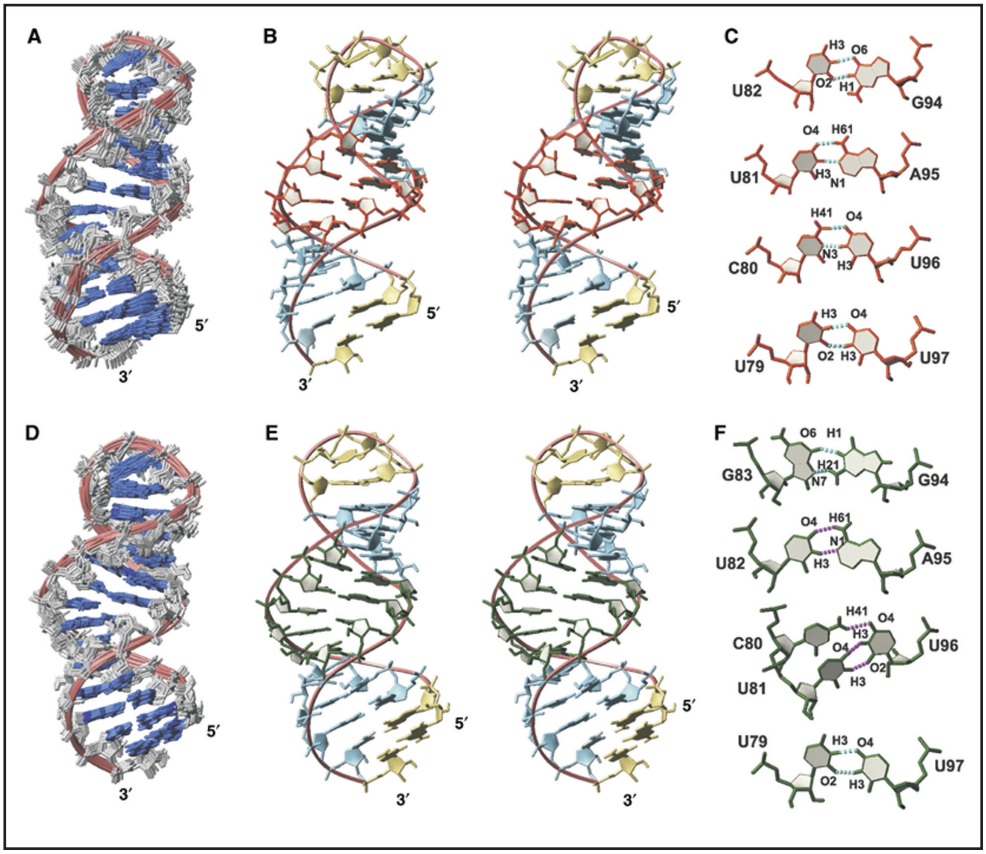

Figure 3: RNA molecules exhibit diverse three-dimensional structures dependent on temperature, nucleotide base-pairing, and steric parameters. (Chowdhury et al. EMBO Journal. 2006)

of exogenous stimuli. Both as sensors of environmental stimuli and in mediating translation, RNA has a distinct regulatory role allowing for programmable cellular behavior.

Genetic Circuits

In synthetic biology, identified regulatory components are recombined into novel networks that behave in predictable ways. An early example of a genetic circuit is the AND gate. Mimicking the functionality of digital logic of AND gates in which two unique inputs must combine to produce a positive output, Arkin and coworkers designed and modeled a genetic part to synthesize a marker protein in the presence of both salicylate and arabinose. Salicylate and arabinose are two naturally occuring, freely diffusible metabolites that bacteria normally react to; this proof of principle construct showed the ability to produce a novel reaction to simultaneous induction of both metabolites. Using two inducible promoters (NahR induced by salicylate and AraC induced by arabinose), this particular genetic part transcribed a unique T7 polymerase and the SupD amber suppressor terminator. The SupD tRNA allows translational read through at the amber stop codon, while the mutant T7 polymerase transcript includes two internal amber codons. Without the transcription of the SupD tRNA, the mutant T7 polymerase transcript would only create a nonfunctional protein product, while the SupD itself cannot induce transcription after the T7 promoter. With the combination of both gene products, a functional T7 polymerase can be expressed, which will synthesize any gene products behind the T7 promoter.

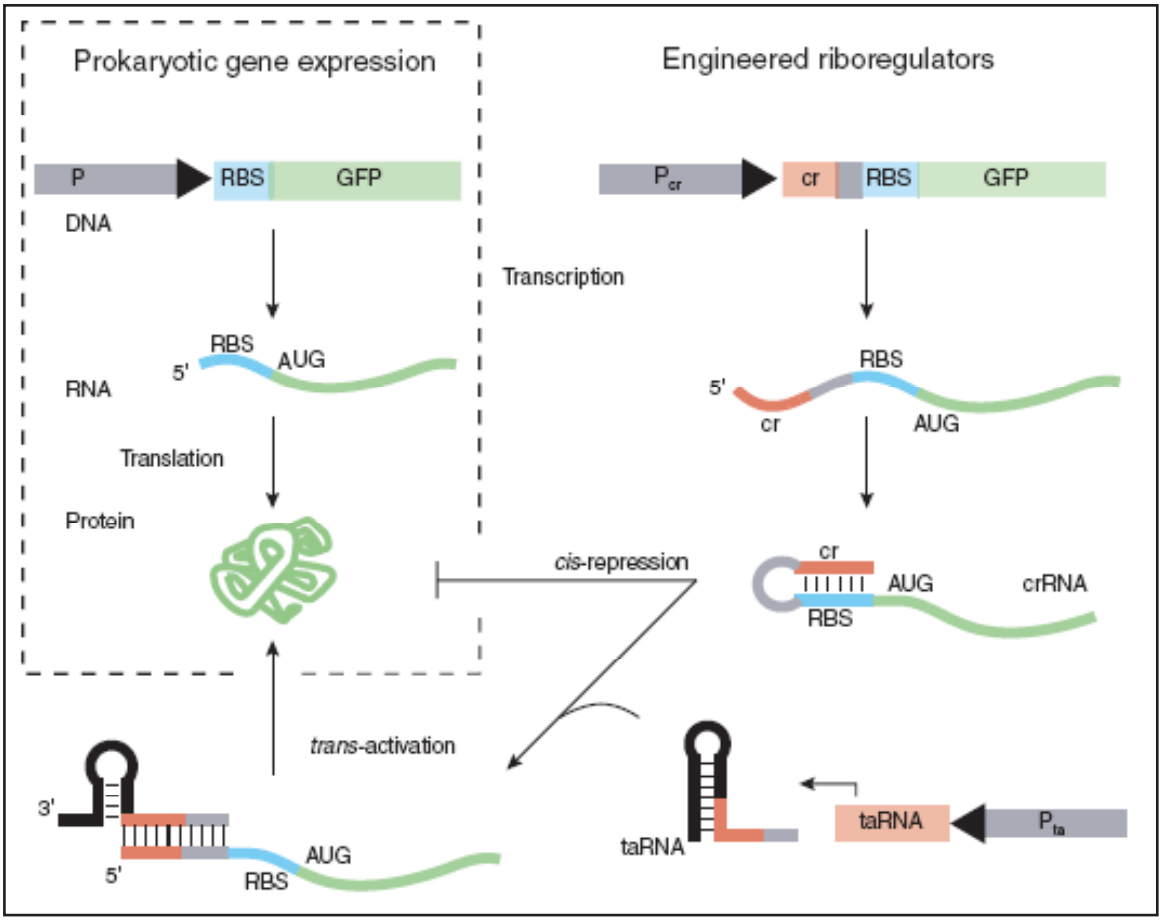

Figure 4: Engineered RNA sequences can exhibit regulation based on temperature, ligand binding, and interaction with other nucleotide sequences. Such post-transcriptional regulation have been demonstrated in E. coli, as small non-coding RNA molecules can be used to prevent or induce gene expression. (Collins et al. Biotechnology, 2004)

An ultimate goal of genetic manipulation is the creation of unique genetic devices or systems that can display unique characteristics or output not found in natural systems. An example of such a biological device is the repressilator, a biological device emulating the functionality of a digital oscillator which oscillates in its production of three different protein products. A system of inter-regulating gene products, the repressilator allows for sequential expression of three individual elements. Mimicking time dependent processes commonly found natural organisms, such as the KaiABC system and the circadian rhythm in photosynthetic organisms, this genetic circuit indicates the ability of simple DNA sequences to produce complex behaviors. Although this proof-of-concept constructare not as robust as natural systems, this biological device demonstrates the potential of deliberate genetic engineering to create novel output and emulate natural organisms.

A LacI repressible promoter regulates a tn10 transposon gene product which can repress another tn10 transposon promoter. This pTet promoter regulates the cI gene. A regulatory unit originally found in lambda phage, the cI protein regulates a lambda promoter that natively regulates switching between lytic and lysogenic phages in the lambda phage lifecycle. In a time dependent manner, the repressilator mimics the circadian clock found in most eukaryotic and many prokaryotic organisms.

Applications and Conclusions

In the last one hundred years, electrical systems have changed the face of the earth. Since the invention of the transistors, computers, phones, and other electronic systems have en-

croached upon all aspects of daily life. One can barely go through one day without use of e-mails, televisions, or cameras. Synthetic biologists dream of another world-changing revolution. Through modular parts and deliberate design, synthetic biology hopes to design biological systems to tackle challenging problems. From smart, self-regulating treatments for cancer to new solutions to the global energy crisis, the ability to engineer biological organisms has the potential to address many status quo questions. The vast natural diversity of life is a testament to the potential and opportunities available in synthetic biology.

Many different native biological organisms, such as E. coli and S. cerevesiae, are already used in many pharmaceutical and biotechnology applications. With a goal of standardization and optimization, synthetic biology allows for novel possibilities as well as improvement upon existing engineered systems. Regulating the interaction of bacteria, bacteriophage, and mammalian cells can allow for applications in medical diagnosis and treatment. The feasibility of using bacteria in biofabrication and energy generation requires designed logic functions in biological systems and biological computation. One interesting area of investigation is the removal of non-essential genes from the genome E. coli to produce an idealized minimal cell. With less chance of interfering regulatory sequences and gene products, such a minimal "cell chassis" could be the optimal shuttles for synthetic gene networks. A simplification of the cellular environment allows for greater ease in characterizing and modeling biomolecular interactions.

Utilizing the modularity of many biological systems, researchers hope to eventually produce complex behaviors through the simple combination of different biological parts. However, important considerations and research into the modularity of biological parts must still be made. In an idealized world, biological parts and coding regions could work equally well in all different cell types and organisms. Unfortunately, due to the inherent complexity of cells and intrinsically noisy nature of molecular systems, different modules might not work in different cellular environments or might not be optimized for maximum efficacy. The stochastic nature of biochemical interactions requires more work to build synthetic models and thereby understand both natural biological systems.

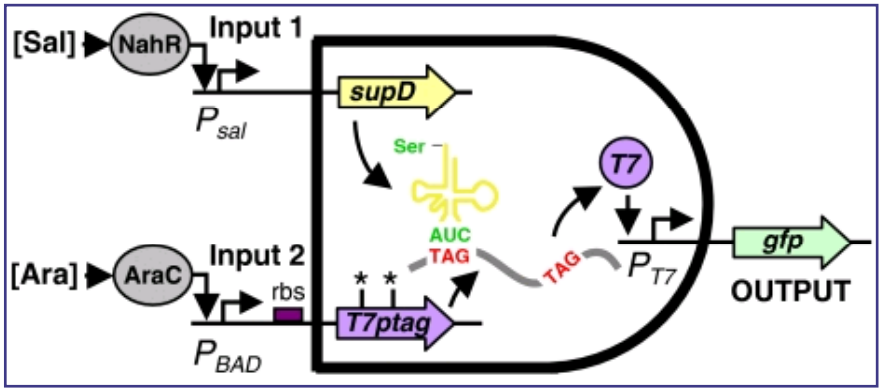

Figure 5: Engineered regulatory networks can be used to integrate environmental information to produce novel output. In a paper by Arkins et al, two inducible promoters detecting concentrations of salicylic acid and arabinose were used to cause synergistic output of green fluorescent protein. (Arkins et al. Mol Syst Biol. 2007)

As genomic sequencing costs continue to decrease, the number of characterized native biological parts and unique designed parts will increase exponentially. Ultimately, synthetic biology introduces novel biological architectures not present in nature. As synthetic biology seeks to stretch the boundary of bio-

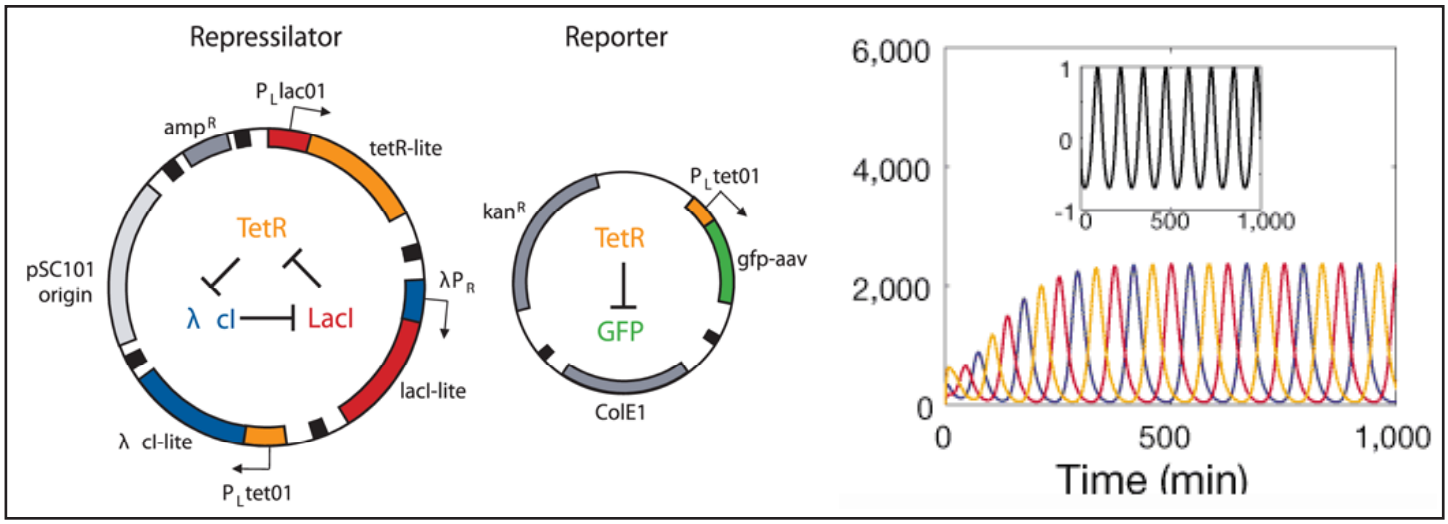

Figure 6: Biological circuits have been designed to mimic natural cyclical processes and to produce sequential output of different gene products. A series of three inducible promoters (pTet, pLambda, pLac) regulate transcription of three proteins. The three proteins (c1, LacI, tetR) repress expression of a different promoter, which causes time dependent expression of each gene product. (Elowitz et al. Nature. 2000)

logical limits and go beyond what currently exists, questions of ethics and morality need to be addressed. What should be the limitations of investigation in this powerful field? With projects like the Venter Synthetic Genome Project, will the threshold between aggregates of molecules and life be more blurred? Should there be manipulation of the human genome, both for medicinal treatments as well as non-life threatening situations? How will intellectual property be handled, as the objects in question are inherent in natural systems? This author does not have the answers to these difficult questions, but feels that one needs to balance the potential benefits with the putative risks in this potent area of research. With great power comes great responsibility; a critical and diligent eye must be maintained in this area of active research. In addition to advancing current knowledge, it is the responsibility of the scientific community to educate the public to the potential and the risks of synthetic biology. Both to prevent Luddite reactions and to address legitimate concerns, dialogue and education are required of a field that seeks to make broad impacts on society at large.

> “Ultimately, synthetic biology introduces novel biological architectures not present in nature. As synthetic biology seeks to stretch the boundary of biological limits and go beyond what currently exists, questions of ethics and morality need to be addressed.

Applying the tools and understanding of molecular biology and biochemistry, synthetic biology focuses on using current molecular tools to engineer unique biological parts and systems. Through such an engineering approach, synthetic biology also seeks to augment current approaches toward understanding regulation. Designed structures and sequences, not unique to natural systems, can be used to understand the finer details of regulation down to the very last nucleotide. As we continue to increase our knowledge of both prokaryotic and eukaryotic regulation, synthetic biologists continue to increase their repertoire of biological parts. Synthetic genome projects are currently underway and new applications such as biological computation, biological chemical fabrication, and disease treatments are being unveiled. Coupled with selection and refinement of genetic devices, deliberate genetic engineering has the potential to tackle many challenges in the near future.

David Ouyang is a sophomore Biochemistry & Cell Biology major at Baker College.

Acknowledgements

I would like to thank Dr. Jonathan Silberg for introducing me to this fascinating field of research and Dr. Daniel Wagner for his keen eye and constructive feedback. Their advice, encouragement, and support greatly aided in the writing process. I would also like to thank Erol Bakkalbasi for editing my paper.

References

1. PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes. Mol Biotechnol. 1995 Apr;3(2):93-9. http://www.ncbi.nlm.nih.gov/pubmed/7620981
2. Stephan C Schuster . Next-generation sequencing transforms today’s biology. Nature Methods - 5, 16 - 18 (2008) Published online: 19 December 2007; | doi:10.1038/nmeth1156. http://www.nature.com/nmeth/journal/v5/n1/full/nmeth1156.html
3. F. Narberhaus, T. Waldminghaus & S. Chowdhury. RNA thermometers. FEMS Microbiol Rev, 30(1):3-16, 2006. PMID:16438677
4. Farren J Isaacs, Daniel J Dwyer & James J Collins. RNA synthetic biology. Nature Biotechnology 24, 545 - 554 (2006) Published online: 5 May 2006; | doi:10.1038/nbt1208
5. Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. Nucleic Acids. Res. 32, 1610–1614 (2004).
6. James J Collins et al. Engineered riboregulators enable post-transcriptional control of gene expression. Nature Biotechnology 22, 841 - 847 (2004) Published online: 20 June 2004; | doi:10.1038/nbt986
7. Poinar, H.N. et al. Science 311, 392–394 (2006).
8. Applications of next-generation sequencing technologies in functional genomics. Morozova O, Marra MA. Genomics. 2008 Nov;92(5):255-64. Epub 2008 Aug 24.
9. Molecular basis for temperature sensing by an RNA thermometer. The EMBO Journal (2006) 25, 2487–2497, doi:10.1038/sj.emboj.7601128 Published online 18 May 2006
10. A Synthetic Oscillatory Network of Transcriptional Regulators; Michael B. Elowitz and Stanislas Leibler; Nature. 2000 Jan 20;403(6767):335-8.
11. Synthetic biology: new engineering rules for an emerging discipline. Molecular Systems Biology 2 Article number: 2006.0028 doi:10.1038/msb4100073 Published online: 16 May 2006
12. Voigt C. Genetic parts to program bacteria. Current Opinion in Biotechnology Volume 17, Issue 5, October 2006, Pages 548-557 Systems biology / Tissue and cell engineering
13. “Guttmacher, Alan E., Collins, Francis S. Welcome to the Genomic Era. N Engl J Med 2003 349: 996-998
14. J Christopher Anderson, Christopher A Voigt, and Adam P Arkin. Environmental signal integration by a modular AND gate. Mol Syst Biol. 2007; 3: 133. Published online 2007 August