Engineering-Based Design Strategies in Biological Systems

Introduction to Engineering Biological Parts / Modularity in Biological Systems

We refer to *parts* as basic biological functions that can be encoded as genetic material (DNA). Drawing parallels to engineering design, if we have set of *standardized biological parts*, we can assemble biological *devices* as combinations of 1 or more parts that encode human-defined functions. And finally, we can assemble *systems* (or larger networks) as combinations of 1 or more devices that encode human-defined functions. One goal of synthetic biology is to create a framework through which standardized biological parts can be reliably assembled into functional devices and systems. It is an open question as to whether this can be done in biological substrates, but research is underway in several laboratories to attempt to build such a framework.
In many of the parts we will use there is a level of *modularity* that allows us to create synthetic circuits.
This idea of modularity also carries through to molecular design as many biomolecules can be comprised of modular domains, which can be used to piece together different functional domains to create biomolecules with different activities.
Before going over that these parts look like, let's review basics of gene expression. (This will provide hints as to what parts we will use to construct circuits.)
Recall: The Central Dogma

DNA → RNA → protein (which performs catalytic reaction on small molecules)

When we think about constructing a biological system we have to consider design at two different levels. Biological systems are programmed at two different levels:

- the *functional level*, which specifies the components that comprise the activities that make up a system
- the *control level*, which specifies the components that govern the operation of the functional level components

The functional level components will be specified by the sequences that encode the activities of a system of interest. Much of biological engineering design goes into how to program or specify the control level, which will regulate when and how these activities are present in the cell to interact. We will go over different mechanisms through which to program the control level of a biological system. These control levels fall under different *regulation levels* along the gene expression pathway:

- transcriptional regulation
- post-transcriptional regulation
- post-translational regulation

Transcriptional Regulation in Bacteria

Transcription is a process by which complementary RNA is made from a DNA strand.

RNA polymerases are enzymes that perform transcription. RNA polymerase is a multi-subunit complex (σ factors are detachable subunits that are responsible for reading start sites of genes).
Different σ factors recognize different types of promoters. After initial few nucleotides are added on the σ factor falls off and elongation occurs.
The rate of transcription in bacteria is approximately 50 nt/sec.

The *promoter* is a sequence of DNA indicating the starting point for the RNA polymerase. In bacteria there are well-defined promoter sequences.

Transcription termination occurs through terminator sequences that direct the polymerase to stop and release the DNA and RNA templates. Terminator sequences are typically A-T pairs proceeded by symmetric sequence forming RNA hairpins (rho-independent).
In rho-dependent terminators they lack A-T pairs and depend on the binding of the rho protein.

By making slight variations to sequences of promoter and terminator elements can generate various 'strengths' in these elements.

Constitutive versus Inducible Promoter Systems
Constitutive promoters are promoters in which gene expression is 'on' at all times at constant levels.
Inducible promoters are promoters in which gene expression can be turned 'on' or 'off' with the addition of an inducer molecule.
Activators are molecules, typically proteins, that turn 'on' gene expression.
Repressors are molecules, typically proteins, that turn 'off' gene expression.
It should be noted that practical definitions of 'on' and 'off' are relative and inducible promoters can have different characteristics. For instance, regulatable promoters can provide *tight control* or be relatively *leaky*, they can be *titratable* over a significant inducer range or provide a *sharp*, non-titratable response, and they can have large output swings (*strong*) or be relatively *weak*.

(Go over examples of regulated promoter systems.)

Post-transcriptional Regulation in Bacteria
Cells regulate gene expression levels by varying the half-lives or turnover rates of transcripts. Higher half-lives correspond to higher steady-state levels in the cell.

RNases are cellular enzymes that act to process or degrade RNA through a number of different mechanisms. Endoribonucleases are RNases that process RNA by cleaving it internally. Exoribonucleases are RNases that degrade RNA by removing nucleotides one at a time, starting at one end of a transcript and working processively to the other end.

One can engineer *cis*-acting RNA sequences that act to enhance or block the activities of these enzymes when located within a particular target transcript.
One can also engineer *trans*-acting molecules (both RNA and proteins) that act on a target transcript to regulate gene expression.

Translational Regulation in Bacteria
Translation can be broken down to three stages in bacteria:

- Initiation: steps encompassing the assembly of the ribosome onto the transcript
- Elongation: steps encompassing synthesis of the protein from the coding region
- Termination: steps encompassing the disassembly of the ribosome from the transcript

Initiation
Initiation includes the steps for ribosome assembly onto the transcript. These steps included binding of the 30S subunit (through sequences in the 16S *rRNA* (ribosomal RNA)) to the *Ribosome Binding Site* (RBS) or *Shine Delgarno Sequence* (SDS). The larger subunit of the ribosome (50S) binds at the AUG (*start codon*), this event sets the reading frame of the translated protein.

The consensus RBS (or SDS) in *E. coli* is 5' AGGAGGU 3'. This sequence is complementary to part of the 16S rRNA. In addition, there is typically a space of 8-9 nt between the SDS and the start codon.

Ribosomes are large multi-subunit machines (comprised of RNA and protein) that decode RNA (or read off the triplet code and synthesize the corresponding protein product).
Initiation factors (IFs) are proteins that orchestrate interactions and assembly.

Elongation
The sequence of RNA is read consecutively in groups of three nucleotides. Each group of three nucleotides is called a *codon* and makes up the *genetic code*.
Therefore, there are three different *reading frames,* and the initiation signals set the reading frame.

The genetic code is *degenerate,* meaning that there are 61 codons and 3 stop codons, but only 20 amino acid monomers.
There is a different bias, or *codon bias*, in different organisms. This is because not all tRNAs are present at the same level.

tRNAs (transfer RNAs) are RNA molecules that play a role in elongation by 'reading' the codon sequence and providing a link between the nucleic acid monomers and the amino acid monomers that transition between the different biopolymers.

Elongation occurs through a step-wise reading and assembly process.
Elongation factors (EFs) are proteins that play a role in the elongation process.

Termination
Termination occurs when the ribosome reaches a *stop codon*. There are three stop codons (UAA, UAG, UGA).
Release factors (RFs) are proteins that play a role in the termination process.

<u>Pathway Coupling</u>
Coupling can occur between the different gene expression pathways. This occurs in bacteria often times because translation occurs co-transcriptionally.

Transcriptional attenuation occurs when there is switching between terminator and anti-terminator structures within a transcript. In many examples, the structure that forms depends on the rate of transcription.

Translational coupling occurs when translation of a downstream gene in a multi-gene transcript is actively linked to translation of an upstream gene.

Dynamic RNA Regulators that Act Through Post-transcriptional Mechanisms

There are numerous types of RNA regulators that act to control gene expression through various post-transcriptional mechanisms, including processing (cleavage), transcriptional (termination), and translational (initiation).

Recently, dynamic RNA regulators, or *riboswitches*, have been described (both natural and synthetic) that exhibit ligand-regulated control over gene expression or allostery.

These dynamic RNA regulators are comprised of two distinct types of functional domains – a ligand-binding or sensor domain (comprised of an *aptamer sequence*) and a regulatory or gene expression platform domain (comprised of a RNA regulatory element). Allostery is achieved as binding of the ligand to the sensor domain alters the activity of the regulatory element.

It has been demonstrated that these dynamic RNA regulators can exhibit compositional modularity, both between the regulator elements and other parts in the system and between the domains of the regulator itself.

Engineering Biological Parts for Regulating Gene Expression

Flexible, transferable biological parts have the property that the function of a part is not linked to the specific gene(s) (or other parts) in the system.

Important properties of such parts are modularity and standardization.

We would like to be able to combine parts and devices in order to construct increasingly complex systems.

Summary of control methods that are available for use (and their corresponding parts):

- *Transcriptional regulation*: engineered promoter systems (constitutive and inducible)
  Promoters with different 'strengths', transfer functions (input and output swings), and regulatory schemes (induction and repression) can be employed. In addition, new chimeric (or hybrid) promoters can be constructed from existing promoter systems.
  Parts: promoter (including operator and regulatory elements), transcription terminator

- *Post-transcriptional regulation:* engineered *cis*- and *trans*-acting elements that regulate RNA processing, transcription, or translation.
  Parts: RNA control elements such as shRNAs, antisense RNAs, ribozymes, and riboswitches

- *Translational regulation:* there are several control points along the translation process.
  i. Translation initiation – Rates of translation initiation can be modified by engineering RBS, including the RBS sequence and the spacer regions.
     Parts: RBS, spacer sequences
  ii. Translation elongation – Rates of translation elongation can be modified by optimizing codon bias of the encoded protein. This makes use of the degeneracy of the genetic code in order to keep amino acid sequence the same, while changing the nucleic acid sequence. This is a less modular tool, but is generalizable.
     Parts: degenerate codons

iii. Engineering functional diversity into the protein building blocks – Protein function can be altered by generating tRNAs that carry non-natural amino acids for site-specific incorporation of new chemistries into proteins.
Parts: engineered tRNAs and associated amino acyl carrier (tRNA loading) proteins

Modularity in Biological Parts

Many of the parts we use in constructing biological systems exhibit some form of compositional modularity (i.e., RBS, TT, promoters, RNA regulatory elements), either with other parts or with domains within the part itself. This means that such parts can be treated as discrete units and couple to other units independent of the particular sequences of the parts.
One example of this, is seen from a given promoter (P1) that can be coupled to two different gene coding regions, independent of the sequences of the genes as:
P1-gene1 and P1-gene2

Construction of genetic devices
Basic device that encodes the expression of a gene of interest:
Promoter-RBS-gene-TT
Note, that such a device can also include other regulatory parts such as RNA control elements, which are not shown here.

Device that encodes the inducible, monitorable expression of a gene of interest:
Operator-Promoter-RBS-gene1_reporter gene-TT
Note, that this device encodes a gene fusion, but can be encoding in a slightly different manner as:
Operator-Promoter-RBS1-gene1-RBS2-reporter gene-TT
This device encodes a polycistronic transcript.
Question: Why use one of these devices over the other?

Device that encodes the expression of a gene of interest such that the protein can be purified and its half-life controlled (or implementation of functional tags):
Promoter-RBS-His tag_gene_Lite tag-TT
This device encodes a protein that is tagged on both its N- and C-termini.

The *Lite tag* targets the protein to rapid decay in the cell. It is modular in that it can be placed at the C-terminal end of any protein. It is typically used for transcriptional regulators so that they do not accumulate to high levels in *E. coli*.

The *His tag* (and other purification tags) is used to purify the tagged protein from cell lysates. They can be placed at the C- or N-terminal ends of a protein and are also modular. Note, that we will not be using purification tags in this class.

On each part denoted above, you can alter its sequence and therefore its cellular activity. For instance, the strength of a promoter can be tuned by altering its sequence (similarly for a RBS or a TT).

Also, the translational efficiency of a gene can be altered to tune its rate of synthesis without changing its activity in the cell by making use of codon bias as described previously.

Modularity of inducible promoter systems
Operator and promoter sequences can be considered to be modules within a part that can be swapped and combined.
For example if given two inducible promoters:
OP1-P1 and OP2.1-OP2.2-P2 these can be combined to
OP2.1-OP2.2-P1, OP1-P2, or OP2.1-OP1-P1, etc.
This allows one to start building in logic with transcriptional networks.

Transcriptional networks
In the laboratory you will be building transcriptional networks to explore response properties of different network motifs.

Example of a transcriptional network:
$A \rightarrow B$
P_{C1}^{*}-RBS1-λCi_lite(A)-RBS2-CFP-TT1
$\lambda P_{RM}OR_3^{-}$-RBS3-gene2(B)-RBS4-YFP-TT2
*Note, that by changing this promoter to λP_R will get A -| B

By building genetic devices out of transcriptional regulators (i.e., where B is a transcriptional regulator) and adding other genetic parts, different types of regulated control loops can be constructed (such as negative and positive feedback, and feedforward). These control loops will exhibit differing dynamic response properties.

Example of a negative feedback loop:

A

$P_{LlacO-1}$-RBS1-lacI_lite-RBS2-CFP-TT1

Systems can also be constructed that exhibit interesting dynamic behaviors by operating them in an unstable regime to observe responses like oscillatory behavior.

Example of a toggle switch:

A B

λP_R-RBS1-lacI_lite-RBS2-CFP-TT1
$P_{LlacO-1}$-RBS3-λCi_lite-RBS4-YFP-TT2