

Background Materials

Synthetic Biology

Is an emerging discipline that emphasizes the application of engineering principles to the design and construction of synthetic biological systems that exhibit dynamic or logical behavior.

The first part of engineering a biological system is the *design* of the molecules, networks, circuits. What do these molecules look like? What properties do they have?

The second part is the *construction* of these synthetic systems. How do we get these molecules into cells or into the right locations?

The third part is the *characterization* of the engineered systems. What techniques are available to characterize engineered cell behavior? Fluorescence, analytical, phenotypic, etc?

Often times the molecules that you would like to be put into a cell (proteins, small molecules, RNA) can be made by the cell. This requires that these molecules be encoded on DNA, which is a genetic carrier of information, and that the synthetic DNA be placed into the cells stably or transiently. This type of design makes use of the central dogma in biology:

DNA → RNA → protein (which performs different functions in the cell, i.e. a catalytic reaction on a small molecule substrate)

DNA, RNA, and proteins are all biological polymers that are made up of building blocks.

DNA – is made up of 4 nucleotide building blocks (nucleotides: A, T, G, C)

RNA – is made up of 4 nucleotide building blocks (nucleotides: A, U, G, C)

Nucleotides are made up of three components: a heterocyclic base, a sugar, and one or more phosphate groups. The bases in DNA are adenine, thymine, guanine, and cytosine. The bases in RNA are the same except that thymine is replaced with uracil.

Proteins – are made up of 20 alpha amino acid building blocks

Nucleic acids are unique in that their secondary structure is predictive through standard folding algorithms because the bases of the building blocks interact through standard base pairing rules.

In addition, many researchers have been trying to expand the chemical diversity (and therefore function) of these biopolymers by creating non-natural building blocks to be incorporated into these molecules.

As each of these biopolymers is encoded on DNA in the cell, we need tools and technologies through which to construct and ‘write’ synthetic DNA and to bring DNA into cells.

Recombinant DNA

The most commonly used vector for carrying DNA into a cell is a *plasmid*. Plasmids are small, circular pieces of extrachromosomal DNA. Note, that bacterial chromosomes are circular, whereas eukaryotic are not.

Components of a plasmid:

replicon – A unit of DNA consisting of an origin of replication and control elements. It is the smallest piece of DNA that can replicate autonomously and defines the copy number of the plasmid.

copy number – Refers to the average number of plasmids per bacterial cell (high, medium, low); high: pUC/*pMB1 (~500-700 copies per cell); medium: pBR22/pMB1 or pACYC/p15A (~15-20 copies per cell); low: pSC101/pSC101 or F plasmid (~5-1 copies per cell).

incompatibility – Refers to the phenomenon that plasmids with the same replicon cannot be stably maintained in cells together without different selective pressures.

selection marker – Unless in low copy, cells lose plasmids if there is not selective pressure to maintain or propagate the plasmids. The burden of plasmid maintenance increases as copy number and promoter strength increase. In bacteria, selective pressure is normally associated with antibiotic resistance genes. Also, resistance markers are useful for selecting clones harboring plasmids of interest from populations of cells.

expression plasmids – Plasmids that are used for the synthesis of a particular biomolecule. They will have a promoter with a multi-cloning site (MCS) so that any gene of interest can be cloned behind the promoter.

There are many different commercial vectors that are in use. For instance, can just have a MCS to insert a piece of DNA with a promoter/gene combination in it OR it can have a MCS in front of a reporter gene to screen promoters.

As an alternative to plasmids, researchers also use special systems for chromosomal insertions or integrations. In these systems, DNA fragments are randomly inserted or targeted into the chromosome.

These systems have the following advantages:

- no requirement for selective pressure
- no issues with multiple replicons
- closer to endogenous expression levels

They have the following disadvantages:

- expression can be location-dependent
- may affect the expression of neighboring genes around insertion
- they require a longer time to generate

In bacteria, various systems have been developed with linear DNA or conditionally replicating plasmids. Most systems are based on transposons or phage-based systems. Yeast are relatively easy to integrate DNA into targeted areas of the chromosome with linear pieces of DNA with 50-60 nt of homology to the target region. DNA will integrate nonspecifically into mammalian cells and various systems have been developed for site-specific integration in mammalian cells.

De Novo Synthesis of DNA

DNA can be synthesized *de novo* on machines using a solid phase approach.

This makes use of phosphoramidite chemistry, where there is sequential addition of monomers to an immobilized nucleotide chain (typically in a column).

Unlike *in vivo* DNA synthesis, this approach synthesizes 3' → 5'.

When one writes a DNA strand there is directionality:

5' GATC 3' is different than 3' GATC 5'

There is a length limit associated with current DNA synthesis technology.

At each step the nucleotide does not add on to a certain percentage of the strands. Therefore, as the strand grows, an increased percentage of the product has deletions and mutations.

Typically, < 60-70 nt is okay without purification; 60-200 nt need purification. Sequences greater than 200 nt require special synthesis techniques – such as assemblies. Assemblies can be very expensive and time consuming.

Newer companies are working on developing foundational technologies for very long DNA synthesis (on the order of thousands of kb). These technologies use a combination of synthesis on arrays with light directed technologies, error correction schemes, and assemblies.

If the system is constructed of molecules for which the desired DNA sequences or slight modifications of these exist in nature, then the DNA sequence can be amplified from the source using *polymerase chain reaction* (PCR).

PCR requires a *template* and *primers*, which are short oligonucleotides that will bind to the template.

PCR:

- Allows exponential amplification of DNA fragments. The fragment is defined by the sequence of the primers.
- PCR thermal blocks are machines that program temperature cycles to rapidly heat or cool.
- PCR enzymes are thermostable polymerases, cloned from organisms that live at high temperatures.

PCR is a very powerful tool in constructing (and characterizing) systems. As such primer design becomes a critical part of a successful PCR.

Strategies for successful primer design:

- Primers should have melting temperatures (T_m s) between 45-65 °C, with ~55 °C being optimal.
- Primer sets (primers used in the same amplification reaction) should have similar T_m s so they are equally likely to hybridize at a given annealing temperature.
- Primer sequences should flank the region that you would like to amplify. This limits the sequence space of the primers.
- Primer sequences should also be designed to minimize (i) hairpin formation, (ii) homodimerization, and (iii) heterodimerization. There is primer design software available that will allow you to check for each of these (<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/>).
- Primers should include sequences that are specific to your DNA of interest and will not likely bind to other DNA in the sample (plasmid or genomic). You can often check for the specificity of your primer sequences with tools available in the plasmid design software (plasmid sequences) and using BLAST searches against the genome of different organisms.

Modifications to conventional PCR strategies (for each modified method similar design strategies apply):

1. One can add new sequences to the ends of the amplified region (typically small regions on the order of 20-40 nt in length), or alter sequences within the primer.

2. *De novo* synthesis of larger DNA fragments through PCR *assembly* of smaller fragments.
3. Assembly of two or more longer DNA fragments not normally linked by *overlap extension*.

As a side note, more recently people have devised PCR schemes for characterization. For instance, quantitative real-time PCR (qRT-PCR) makes use of the exponential amplification stage to monitor amplification in this range and use it to back calculate out the amount of starting material. This is a method that is used in quantifying DNA and RNA levels.

Piecing Together DNA Fragments

DNA fragments are glued together in a process called *ligation*. Ligases are enzymes that form phosphodiester bonds between 3'-hydroxyl and 5'-phosphoryl termini of nucleic acids. The process of ligating DNA fragments into plasmids (or other vectors) is referred to as cloning, and results in the formation of a circular recombinant plasmid.

Ligation occurs through a two-step reaction, where there is an intermolecular reaction in which two different fragments are connected together and an intramolecular reaction in which the ends of this fragment connect to itself.

In a ligation reaction there can be many possible products, including linear, circular, and homo- and hetero-dimers.

The concentration of DNA in a ligation reaction often can be optimized to maximize yield of circular monomeric recombinants. This is because the intermolecular reaction requires high concentration of DNA; however, the intramolecular reaction works most effectively at low DNA concentrations.

Recombinant clones must always be confirmed by restriction endonuclease mapping and sequencing.

Cloning with Protruding versus Blunt Ends

Ligation reactions with DNA fragments containing protruding ends are more efficient than blunt ends. This is because compatible, *sticky ends* (1-6 nt in length) will anneal by base-pairing between the protruding bases. This annealing brings the DNA into close alignment, which allows the ligase to catalyze the formation of phosphodiester bonds.

Example of restriction enzyme that results in 5' protruding ends: (EcoRI)

Example of restriction enzyme that results in 3' protruding ends: (PstI)

With a single restriction enzyme cut site, you have non-directional cloning.

With two restriction enzyme cut sites (with non-compatible sticky ends), you have uni-directional cloning.

Note, that MCS are sites in a plasmid with localized restriction endonuclease sites that cut once. This usually allows for flexible, localized cloning strategies.

Two fragments with *blunt ends* can also be cloned; however, this is a relatively inefficient reaction due to the lack of annealing strands to enable ligase to catalyze the linkage reaction. For blunt end ligations to occur you need very high concentrations of ligase and very high concentrations of blunt end termini.

Other useful cloning enzymes:

- Klenow fragment: A portion of a polymerase that does not exhibit exonuclease activity. Used to fill in 5' protruding ends.
- Exonuclease (I): An enzyme that removes single-stranded nucleotides in a 3' → 5' direction. Used to remove 3' protruding ends.
- Phosphatase: An enzyme that removes the 5' phosphate on a DNA strand. The most commonly used is Calf Intestinal Alkaline Phosphatase (CIP).

The first two enzymes are used to make sticky ends blunt, in order to clone DNA fragments with incompatible sticky ends. In this situation you must use blunt end ligation. The last enzyme is sometimes used in cloning strategies to increase the efficiency of ligation reactions by removing the 5' phosphate on the cleaved plasmid backbone DNA (not the insert). This reduces the efficiency with which the vector can ligate back together with itself.

Question: After you clone your recombinant DNA, how do you get it into cells?

Transformations: Getting Recombinant DNA into Cells

Several methods are available for allowing DNA to enter a bacteria. In general, the cell membrane or wall needs to be broken open, so that DNA may pass through these transient holes and the membrane must subsequently be allowed to heal. Nucleic acids do not enter bacteria on their own (they are highly negatively charged).

Several methods have been developed for *transforming* DNA into cells. Two of the most common are:

1. Chemical methods – This method is based on resuspending cells in ice-cold salt solutions (CaCl₂) and then heat shocking briefly (42 °C) to allow DNA to enter the cells. Efficiencies are ~10⁵-10⁶ colonies/μg supercoiled DNA. Disadvantages of this method are that it has lower efficiency and less reproducibility.
2. Physical methods – The most common physical method in use is electroporation. In this method, exposure to an electrical charge destabilizes the membranes of *E. coli* and induces the formation of transient membrane pores. Cells are washed free of all salts and kept ice cold. Efficiencies are ~10⁹-10¹⁰ colonies/μg supercoiled DNA (much higher than chemical methods). A disadvantage of this method is that it is expensive.

Note, that in general transformations are relatively inefficient; that is, within the population of cells, a large majority do have the recombinant plasmid introduced into them. Therefore, cells from the transformation are plated on selective media to get rid of the high background of cells that do not harbor the plasmid.

In addition, researchers sometimes use viruses as transforming agents. Viruses naturally insert their DNA into a host target. Therefore, if the virus genome can be re-engineered to harbor the DNA of interest, while retaining the ability to insert the DNA into its host, this can be used as an efficient transformation method. There are several viruses, or phages, that infect *E. coli* and have been used for such an approach.

Confirmation of Recombinant Clones

As an initial screen, recombinant clones are confirmed by restriction endonuclease (or enzyme) *mapping*. This requires that you determine the cleavage pattern for the correct plasmid and do several different combinations of cuts to confirm:

- i. insert size
- ii. insert orientation
- iii. backbone

If the banding pattern is not correct, then the recombinant clone is most likely incorrect.

An alternative method to restriction enzyme mapping is *colony PCR*. In colony PCR, you are conducting a standard PCR around your desired insert and insertion site. Such that the size of the PCR fragment or the presence of the amplified fragment indicate whether or not the correct insert is present.

If the initial screen indicates that the recombinant clone is correct, then you do a secondary screen through *sequencing*.

In the recombinant clones that you construct you will perform both screening steps to confirm the correct constructs.

This material covers general cloning strategies. Much of your work throughout the summer will utilize a particular implementation of this strategy using biobricks and the biobricks assembly strategy, which provides a standardized physical composition framework through which to piece together DNA fragments (http://parts.mit.edu/registry/index.php/Help:BioBrick_Assembly; in particular the standard and parallel assembly sections). Please read through the biobricks assembly strategy so you are familiar with it as much of our design must take this composition framework into account as part of iGEM.

Altering DNA Sequences

There are various ways through which DNA sequences may be altered.

Rational, targeted alterations through site-directed mutagenesis:

This method allows you to make relatively small scale (~1-5 nt), specific changes to a DNA sequence.

Sequence changes may be specified by either requirements for the linear sequence of the DNA/RNA, or to modify the secondary structure of the DNA/RNA, or to alter the activity of the biomolecule (DNA, RNA, or protein) through tertiary structure interactions. In the latter case, this typically requires structural data to build a model of the biomolecule and software for visualizing molecular conformation. Several programs exist for predicting secondary structures of nucleic acids from primary sequences.

Site-directed alterations in a DNA sequence are achieved through a primer-PCR type process.

- Complementary primers with the desired altered DNA sequence are designed.
- These primers are used in a PCR process where the entire plasmid (harboring the sequence of interest) is the template to be amplified.
- The parent plasmid in the PCR mixture is digested by a restriction enzyme that recognizes methylation of specific nucleotides. Note, that methylation is only present on the plasmid that has been purified from a cell host (the parent plasmid), not on the newly synthesized plasmid.

- This PCR mixture is subsequently transformed into an appropriate host (i.e., *E. coli*). Note, that this method can also be performed on a DNA fragment and subsequently cloned into a vector through standard cloning strategies.

As an aside, DNA synthesis can be performed with any combination of nucleotides at any position by simply flowing the desired nucleotide mixture through the column during the synthesis reaction. In this way, one can create libraries, doped libraries, and perform saturation mutagenesis techniques.

Directed evolution

This strategy incorporates processes which allow the ‘evolution’ of a target sequence under short time scales in a laboratory. Methods typically include a combination of recombination and random mutagenesis techniques (through PCR). Both of these processes can be random or directed/targeted to a particular region of the DNA sequence.

Random mutagenesis strategies perform PCR under conditions that increase the mutation rate of the polymerase. For example, low fidelity polymerases (Taq), altered ratios of nucleotides, high [MgCl₂] or [MnCl₂] are typically used.

By altering the combination and degree of such strategies, researchers can tune the mutation rate of the polymerase to an appropriate range for the sequence that is being evolved. For example, protein-coding regions are often evolved under lower mutation rates than non-protein-coding regions.

Recombination strategies typically start with a family of related sequences that have a certain degree of homology.

This family of sequences is often times cut up randomly with an enzyme that cleaves dsDNA (DNase). The fragments from this digestion are allowed to reanneal in a subsequent assembly reaction. As an alternative, researchers have also directed crossover events to particular areas of the sequence. These crossover sites are typically identified through software programs that can identify most likely regions of a sequence to recombine at and maintain functionality of the biomolecule. So that these programs are often time taking into account modularity and folding of modular domains.

With these strategies one can generate a relatively large library of sequences, from which only a very small percentage will have enhanced functionality (such as increased activity, altered substrate binding, etc.). Therefore, rapid, high-throughput screens or selections are necessary to select for the members with enhanced properties.

There are two strategies for searching through libraries of molecules and pulling out ones with desired properties.

A *selection* is an assay that links the desired property of the molecule to the growth of cells harboring the molecule.

A *screen* is an assay that links the desired property of the molecule to a detectable output signal from cells harboring the molecule, from which cells can be sorted.

Characterization Tools

There are basic strategies for quantification and characterization of different cellular biomolecules.

mRNA

Northern blot analysis

This is a gel-based, low throughput, low sensitivity method for quantifying mRNA levels in cells.

The basic method followed is to:

- purify RNA from cells
- run the RNA out on a gel (either agarose or polyacrylamide and denaturing or non-denaturing)
- transfer the RNA on the gel to a membrane and crosslink the RNA to the membrane
- hybridize a labeled probe specific to the transcript of interest to the membrane (Note, that the probe is typically a complementary piece of DNA or RNA that can vary from ~50-1000 nt in length, and is labeled with either a fluorescent or radioactive moiety(ies).)
- detect and quantify the amount of signal on the membrane and make comparisons to loaded controls

The signal intensity is proportional to the amount of RNA in the sample. Note, that one can conduct a very similar procedure with DNA samples (Southern blot analysis).

Real-time quantitative PCR (qRT-PCR)

This is a solution-based, higher throughput, high sensitivity method for quantifying mRNA levels in cells.

The basic method followed is to:

- purify RNA from cells
- make cDNA copies (single-stranded complementary DNA copies) of desired RNA species through reactions employing a reverse transcriptase enzyme
- amplify a small segment of the target RNA (~20-50 nt in length) with PCR

A PCR amplification curve has an initial lag period, followed by an exponential period of amplification, and an ending period where no further amplification occurs. In the exponential period, the cycle number is proportional to the amount of RNA (or starting template).

In this method, amplification is tracked by measuring the amount of fluorescence at the end of each cycle. The PCR is performed either with an intercalating dye or a specialized fluor-quencher primer (molecular beacon) such that fluorescence is correlated to the amount of double stranded PCR product generated. Differences in relative starting amounts of cDNA template result in differences in cycle numbers that it takes to get to a set amplification signal observed in the exponential period.

For qRT-PCR, cameras are mounted on top of the PCR machine that can measure up to four species simultaneously (such that instruments can resolve up to four different colors).

Microarrays

This is a very high throughput, low reproducibility method that allows for profiling the steady-stage levels of entire transcriptomes (or very large sets of mRNAs).

The basic method followed is to:

- purify RNA from cells

- make cDNA copies of all RNA species (using random hexamer primer sets or polyA primer sets)
- label the cDNAs with fluorescent labels
- hybridize the labeled cDNAs onto a chip with complementary ssDNA spotted in arrays

The ssDNA spotted in arrays on the chip can be long fragments comprised of the sequence for an entire gene or they can be oligonucleotides comprising only part of the sequence of a gene or fragment of interest.

The intensity of the spot as read off by a scanner is proportional to the amount of RNA in the sample.

Proteins

Western blot analysis

This is a gel-based, low throughput method for quantifying levels of specific proteins or protein modification states in a cell.

The basic method followed is to:

- purify protein from cells
- run the protein out on a gel (typically SDS-PAGE – slightly denaturing polyacrylamide)
- transfer the protein on the gel to a membrane
- hybridize a probe specific to the protein of interest to the membrane. Note, in working with proteins the probe is an antibody (a protein) which specifically recognizes your protein of interest.
- image and quantify the probe with a secondary antibody that is labeled and can be detected through a standard assay

Note, that the Western blot is similar to the Northern and Southern blot procedures. However, here the probe for a protein is an antibody, which is much harder to generate than a ssRNA or DNA probe that is complementary to a desired target. Therefore, the primary antibody is typically made with a conserved region to which a universal secondary antibody will specifically recognize. This secondary antibody is used as a universal mechanism to amplify, image, and quantify the signal.

Analytical techniques

The most commonly used analytical techniques include liquid chromatography-mass spectrometry (LC-MS), tandem MS-MS, and 2-dimensional gel electrophoresis (2D-GE).

Researchers use MS to identify proteins or protein fragments, and LC and GE to purify or isolate different species.

These techniques are being adapted for proteome analysis or the analysis of large subsets of proteins. Therefore, these techniques tend to be automated and very high throughput (not 2D-GE).

Activity and reporter assays

Other techniques to determine the levels of proteins in a sample include performing activity assays for different types of proteins or enzymes. These are specific to the protein of interest.

In this laboratory we will make use of reporter genes for quantifying protein levels in the cell indirectly through activity assays. *Reporter proteins* are proteins for which there are quick, standardized assays of activity.

There are commonly used reporter proteins in biology. One of the most commonly used reporter protein is the enzyme β -galactosidase (encoded by the gene *lacZ*). There are both plate-based and solution-based assays for this enzyme.

On plates, cells are grown in the presence of the substrate X-gal (5-bromo-4-chloro-3-indoyl- β -D-galactopyranoside). Cells that are expressing functional β -galactosidase will turn blue in color. This allows for blue-white screening assays.

In solution, cell extracts or cell suspensions are placed into solutions harboring the substrate ONPG (o-nitrophenyl- β -D-galactopyranoside). β -galactosidase cleaves this substrate to a compound that is yellow in color and detectable and quantifiable on a spectrophotometer. This allows for quantifiable reporter assays for levels of β -galactosidase in various solutions.

A second set of commonly used reporter proteins, that have been gaining increased favor, are *fluorescent reporter proteins*. These included the green fluorescent protein (GFP) and color-shifted variants (CFP, YFP, etc.). There are other fluorescent proteins that offer other color ranges but are not variants of the original GFP sequence such as DsRed (i.e. RFP). There are series of different commercially-available fluorescent proteins (many offered through Clontech) that can be used in single color applications or in combination.

The combinations of fluorescent markers that can be used depends on excitation and emission wavelengths, and whether the instrument you are using can resolve the different wavelengths. For example, some instruments are filter-based (such that resolution depends on the filter sets available) and some have monochromatic lasers.

GFP was cloned from a bioluminescent jellyfish, *Aequorea victoria*, other fluorescent proteins have been cloned from coral. GFP fluorescence is species independent and does not require cofactors, substrates, or additional gene products. This is in contrast to commonly used reporter enzymes such as β -galactosidase or luciferase.

Since discovery of GFP and its first use as a heterologous reporter protein in various organisms, researchers have spent significant effort in generating color-shifted variants, more rapidly folding and degrading variants, variants that form monomers versus tetramers (to be less toxic to cells and for ease in tagging assays), variants that exhibit brighter fluorescence and less photobleaching, and variants that are optimized for expression in different organisms.

Some of the commonly-used GFP variants:

EGFP – enhanced GFP, red-shifted excitation spectra; fluoresces between 4- to 35-fold more brightly than wildtype GFP

EYFP – shift emission from green to yellowish-green (4 amino acid substitutions)

ECFP – shift emission from green to cyan (6 amino acid substitutions)

GFPuv – 18 times brighter than wtGFP when in *E. coli* and excited by UV light

DsRed – a red fluorescent protein from a coral *Discosoma genus*

Often times, biological engineers use these fluorescent proteins to measure outputs from networks. For example, by coexpressing a fluorescent protein from a promoter of interest. This is referred to as a *promoter fusion*.

Researchers also use these fluorescent proteins to make tagged versions of proteins of interest or *fusion proteins*. A fusion between a protein of interest and a fluorescence reporter protein provides a fixed (1:1) ratio of the two and allows researchers to perform localization studies.

Note, that sometimes protein fusions can result in an inactive protein of interest or fluorescent reporter protein by affecting protein folding. Therefore, for these types of experiments it is often more effective to use the monomeric form of a fluorescent protein to allow for more retained activity of the protein of interest.

In addition to coexpression studies and protein fusion studies, researchers also use fluorescent reporter proteins in *Fluorescence Resonance Energy Transfer* (FRET) experiments.

This type of study requires a donor-acceptor pair, such that the donor is excited with a laser from an outside source. The emission from fluorescence of the donor is used to excite the acceptor molecule and the emission of the acceptor molecule is read off as the FRET signal.

Typically studies use CFP-YFP FRET pairs, where the donor is CFP.

FRET-based approaches have been used for determining molecular concentrations, binding interactions, and catalytic activity. The power of this method over the other applications discussed is that it allows spatial and temporal resolution due to its rapid signal response and sensitivity to distance between the donor-acceptor pair.

Limitations with this approach as currently implemented have been the dynamic range, sensitivity, and modularity.