Davidson College
Department of Biology

Honors Thesis

Title:

Solving the satisfiability ("SAT") problem using Cre-mediated recombination

Submitted by:

Mary Gearing
Class of 2011

Date of Submission: April 11, 2011

First Reader: Dr. A. Malcolm Campbell
Second Reader: Dr. Laurie Heyer

Abstract

The Boolean Satisfiability (SAT) problem asks whether there is an assignment of values to true-false variables such that a given logical expression is true. The SAT problem was the first problem to be proved NP-complete (Cook, 1971). Roughly speaking, an NP-complete problem is one in which the complexity greatly increases with small increases in size. NP-complete problems are difficult for traditional computers to solve due to their combinatorial nature.

Over the last 16 years, both bacterial and DNA-based computers have been used to solve NP-complete problems (Adleman, 1994; Lipton, 1995; Sakamoto *et al.*, 2000; Braich *et al.* 2002; Lin *et al.*, 2007; Baumgardner *et al.* 2009). In 2009, the Missouri Western/Davidson iGEM team designed a system that uses frameshift suppressor tRNAs to suppress codon frameshift mutations caused by inserting two additional bases. Suppression of the insert equals satisfaction of a logical variable; when the entire logical expression is satisfied, a reporter gene is correctly translated.

Previous work using a Lux-based system served as a proof-of-concept that frameshift suppressor tRNAs can be used to evaluate SAT problems. I constructed a Cre-based system, which requires a very low level of input for system activation and conversion to the on state. Cre containing frameshifts still mediated excision, suggesting that our Cre frameshift construct must be redesigned to disrupt the functional C-terminus of Cre. If this redesign works, the low input of the system should aid in solving SAT problems. All constructed parts are publicly available in the Registry of Standard Biological Parts.

Introduction

NP-complete and the satisfiability problem

Computational complexity theory classifies problems according to the difficulty of solving them. No efficient algorithms exist for the class of problems known as NP-complete; current algorithms must make guesses at each decision point. The growth function of the run time of an algorithm designed to solve an NP-complete problem is exponential, but once a solution has been found, it can be quickly verified (Cook, 1971). NP-complete problems are difficult for traditional computers to solve because of the combinatorial nature of these problems. The number of possible solutions that must be evaluated is very large, but the number of true solutions may be very small or zero.

The first problem proved to be NP-complete was the Boolean satisfiability (SAT) problem (Cook, 1971). In a SAT problem, two literals comprise a Boolean, or true-false, variable; for example, G and g are two literals for one variable. You can think of the literal g as the logical expression "not G," so g will be true whenever G is false, and vice versa. SAT problems can be specified in different ways; I will consider only the disjunctive normal form, in which a clause is the combination of literals in the form (G or B), and the number of variables in each clause may be restricted (*i.e.*, a 2-SAT problem.) A problem consists of an arbitrary number of clauses. To solve such a SAT problem, you must find a solution such that at least one literal in each clause evaluates to true. This solution is said to "satisfy" all the clauses. It is very difficult for traditional computers to solve the SAT problem when the number of variables is large (Garey & Johnson, 1976), but molecular biologists have worked to engineer biological computers to find solutions (Adleman, 1994; Sakamoto *et al.*, 2000; Braich *et al.*, 2002; Lin *et al.*, 2007).

Our 2009 International Genetically Engineered Machines Competition (iGEM) team developed an analogy for the SAT problem that uses locks and keys (Figure 1). If there are four doors, each of which has a single lock that can be opened with either of two possible keys, is there any subset of three keys (one of each color) that would enable one to open all four doors? The three-key set GBR can open all the doors shown in Figure 1, and therefore satisfies the four clause, 2-SAT problem [$(G \cup B) \cap (G \cup b) \cap (G \cup r) \cap (g \cup R)$].

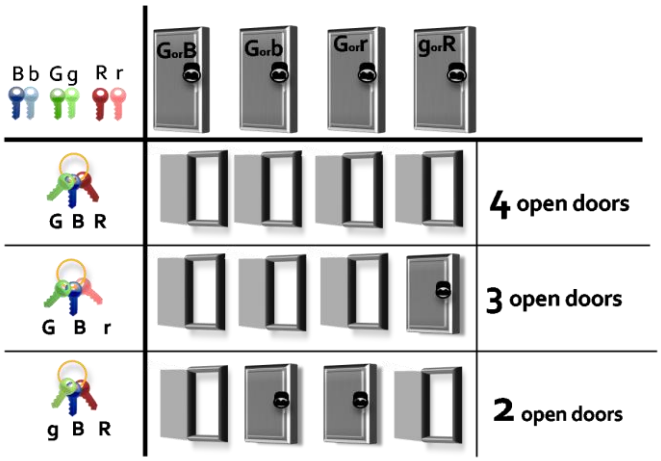


Figure 1. The SAT problem can be represented by a lock-and-key analogy. In this problem, the key combination GBR opens all four doors, whereas the other two key combinations open fewer than four doors. Since GBR opens all of the doors, the problem is "satisfied."

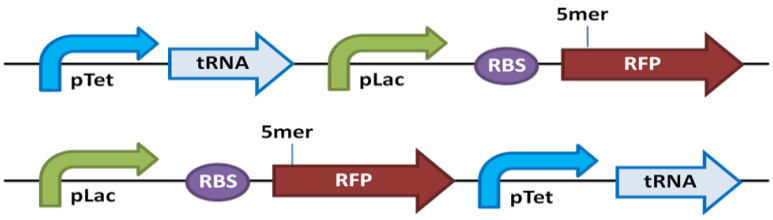
Leonard Adleman made the first effort to biologically solve the Hamiltonian path problem, a related NP-complete problem, using oligonucleotides (1994). Shortly after the Adleman study was published, Richard Lipton published a theoretical approach to the SAT problem, also using oligonucleotides (1995). Since these two groundbreaking studies, various groups have engineered ways to solve SAT problems. Sakamoto *et al.* (2000) used oligonucleotide assembly to find possible solutions and eliminate inaccurate solutions by destroying the base paired hairpin structures between the literals **a** and **not a**. Lin *et al.* (2007) used microarrays to design solutions, producing possibility trees that greatly lowered the number of possible solutions. Perhaps the most complex problem biologically solved to date is a 20-

variable SAT problem solved using a DNA computer (Braich *et al.*, 2002). None of these solutions were determined *in vivo*, and all relied heavily upon human manipulation and PCR amplification.

Bacterial computing

Past researchers in the Campbell/Heyer lab, in collaboration with students and faculty at Missouri Western State University, have focused on solving NP-complete problems *in vivo* using bacterial computing. So far, the combined Davidson and Missouri Western team has solved Hamiltonian Path (Baumgardner *et al.* 2009) and Burnt Pancake (Haynes *et al.*, 2008) problems using this approach. NP-complete problems are difficult to solve because of the massive number of possible solutions that must be tested. Bacterial computing takes advantage of the enormous parallel processing power found in just 1 mL of liquid culture, which can contain 1 billion bacterial computers. Building on previous work, the 2009 Missouri Western/Davidson College iGEM team attempted to solve the SAT problem using *in vivo* bacterial computing. The team encoded the problem into bacteria using tRNA frameshift suppression. Magliery and Anderson (2001) devised a method to isolate tRNAs that suppress four- or five-base codons using four- or five-base anticodons. For the thirteen known five-base codon suppressors, rates of suppression varied from 1.6 to 12.0% (Anderson *et al.*, 2002). The 2009 iGEM team constructed and BioBricked these suppressor tRNAs and encoded the 5 base inserts they suppress in a reporter gene, RFP (Figure 2A, methods as in Parts Registry). In terms of SAT, these are "single-literal" constructs in which a 5 base insert that must be satisfied by a respective suppressor tRNA (Figure 2B). In all cases, suppression of a 5-base codon brings a serine into the growing protein.

A)



B)

Normal translation

```
atg       gct tcc tcc gaa gac gtt atc aaa gag ttc atg cgt ttc
 M        A   S   S   E   D   V   I   K   E   F   M   R   F
```

Translation with 5mer suppression (+1 amino acid)

```
atg CCACU gct tcc tcc gaa gac gtt atc aaa gag ttc atg cgt ttc
 M    S   A   S   S   E   D   V   I   K   E   F   M   R   F
```

Frameshift translation without 5mer suppression

```
atg CCA CUg ctt cct ccg aag acg tta tca aag agt tca tgc gtt tc
 M    P   L   L   P   P   K   T   L   S   K   S   S   C   V
```

Figure 2. Frameshift suppression restores the correct reading frame. A. Schematic of single-literal constructs. There is a frameshift in the reporter gene RFP. B. Closeup of RFP translation. When a suppressor tRNA is incorporated in translation, the correct reading frame is restored and one additional amino acid is included (black S). In this case, the CCACU suppressor tRNA puts a serine into position two. When a normal tRNA is used instead of the suppressor, the 5mer insert is read as one codon plus two bases, and the reading frame is lost (red amino acids).

When we measured fluorescence for these single-literal constructs, the readings were very low (Figure 3). It was difficult to distinguish satisfied and unsatisfied phenotypes, because each suppression event produced only one copy of RFP. Since suppression events are rare, and our construct did not allow for signal amplification, our reporter output was very low.

Construct	Our % suppression
pBad-**CCAUC10**-pLac-RBS-CCAUC-RFP	0.3
pTet-**CCAUC10**-pLac-RBS-CCAUC-RFP	1.5
pBad-**CUAGU**-pLac-RBS-CUAGU-RFP	0.3
pTet-**CUAGU**-pLac-RBS-CUAGU-RFP	0.1
pLac-RBS-CUAGU-RFP-pTet-**CUAGU**	0.3
pBad-**CCACU**-pLac-RBS-CCACU-RFP	0.4
pTet-**CCACU**-pLac-RBS-CCACU-RFP	1.4
pLac-RBS-CCACU-RFP-pTet-**CCACU**	1.9
pBad-**CCCUC**-pLac-RBS-CCCUC-RFP	0.3
pTet-**CCCUC**-pLac-RBS-CCCUC-RFP	0.4
pLac-RBS-CCCUC-RFP-pTet-**CCCUC**	0.4
pBad-**CGGUC**-pLac-RBS-CGGUC-RFP	0.2
pTet-**CGGUC**-pLac-RBS-CGGUC-RFP	0.4
pLac-RBS-CGGUC-RFP-pTet-**CGGUC**	0.8

Figure 3. When a single-literal problem is encoded directly in a reporter gene, reporter output is very low. Fluorescence data from various 1-SAT problems are shown. Percent suppression represents the fluorescence of the 5mer construct divided by that of the positive control (pLac-RBS-RFP.) The low fluorescence levels make it difficult to distinguish satisfied and unsatisfied phenotypes. Constructs are color-coded by tRNA, and represent various promoters and placement of tRNAs 5' and 3', as seen in Figure 2A.

This single-literal example requires one suppression event. In true SAT problems, the satisfaction of each logical clause requires a suppression event. The probability of multiple independent suppression events is the product of the probabilities of each suppression event. For this reason, we could not use this system to solve multiple-clause SAT problems. With each suppression event needed, our already low output level would decrease even further, preventing us from obtaining clear phenotypes.

In order to increase the level of signal, we encoded the frameshift in an enzyme that is part of a positive feedback loop. By putting the SAT problem into an enzyme, we hypothesized that the catalytic nature of the encoded protein would give a more robust readout of any suppression events. The chosen enzyme, LuxI, synthesizes AHL-3OC6, which can bind to the

LuxR protein to induce the promoter pLux (Figure 4A). One copy of the LuxI enzyme produces many molecules of 3OC6, amplifying the initial suppression signal. In addition, as part of the engineered positive feedback loop, a small amount of LuxI increases production of LuxI, producing additional signal amplification. Signal amplification allows one to clearly distinguish whether or not the SAT problem has been solved, because the satisfied and unsatisfied phenotypes are clearly distinct. Figure 4B shows the full BioBrick construct.

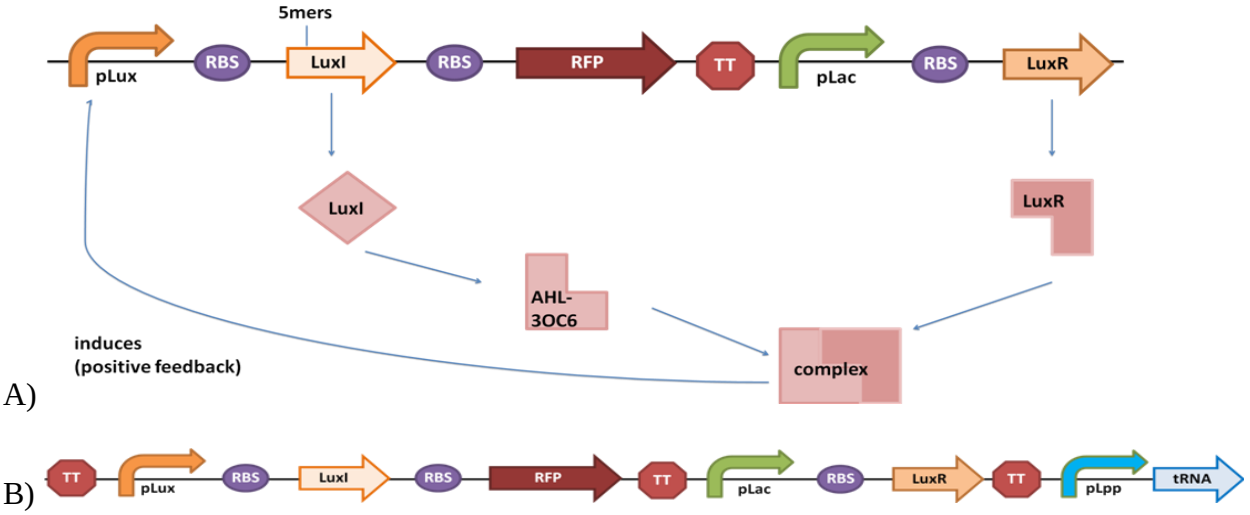


Figure 4. The Lux positive feedback loop. A. Positive feedback amplifies the signal. When the SAT problem in LuxI is satisfied, functional LuxI is produced. LuxI synthesizes AHL-3OC6, which complexes with LuxR to induce pLux and allow for RFP and additional LuxI production. When suppression does not occur, LuxI is nonfunctional, pLux is not induced, and RFP is produced at very low rates. B. The Lux positive feedback loop as a BioBrick construct. Octagons labeled TT are transcriptional terminators, used to reduce read-through transcription.

Initially, we had intended to use PCR to insert different SAT problems into LuxI. However, as illustrated in Figure 5, we engineered unique 8 bp restriction sites in LuxI to facilitate swapping out different 2-SAT and 3-SAT problems. For assembly of such small parts, oligonucleotide assembly is far more efficient than PCR amplification. We chose the enzymes AsiSI and AscI, which have 8 bp restriction sites, to maintain the BioBrick standard and minimize the probability that these sites existed anywhere else in the plasmids. We spaced the

restriction sites 5 bp apart to maintain the correct LuxI reading frame when no frameshifts are inserted.

ATG GCG ATC GCa tct gGG CGC GCC

M A I A S G R A

Figure 5. Restriction sites facilitate the swapping of SAT problems. Different SAT problems (in bold font) can be inserted between the 8 bp restriction sites. The AsiSI site is highlighted green and the AscI site is highlighted yellow. Genes with these restriction sites are referred to as "gene_name8mers."

The Lux-based design allowed us to solve a simple 1-SAT problem (Figure 6); however, there were two issues that lowered its utility. First, the Lux system was not sufficiently stable. When minipreps from colonies that had solved the problem were digested, the DNA had recombined due to the instability of the system, a result of the multiple transcriptional terminators in the construct (see Figure 5A; Sleight *et al.*, 2010). Synthetic biology depends on the stability of DNA parts, so we could not continue to use this design.

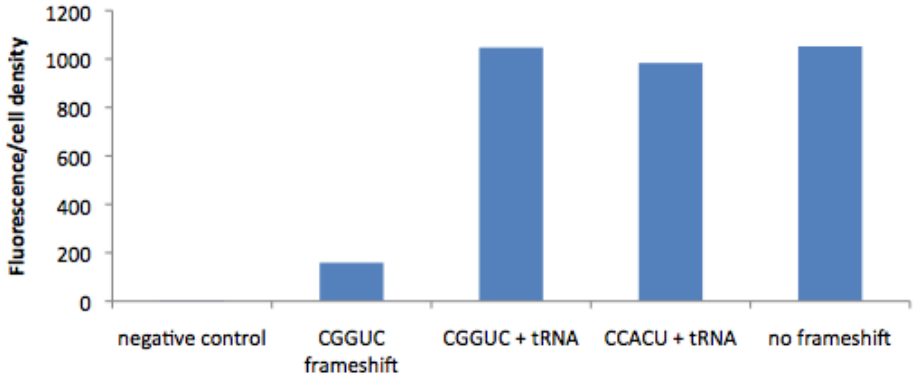


Figure 6. 1-SAT problems are solved with the Lux system. When a frameshift is present, RFP production and fluorescence are low. When the appropriate suppressor tRNAs are present, the frameshifts are suppressed and RFP is produced in very high quantities, similar to the wild-type positive feedback loop. Notice that the fluorescence level in the absence of suppressor tRNA is higher than that of the negative control.

The other problem with the system was the relatively high threshold level required to induce pLux and stimulate the positive feedback loop. We don't know how many copies of the

LuxI protein would be needed to initiate the positive feedback loop. Although reaching threshold was not a problem for the 1-SAT problem, it might be for a more complicated SAT, such as the 20-variable SAT problem solved on a DNA computer (Braich *et al.*, 2002). Since the probability of multiple independent suppression events is low, we could not be sure that we could achieve the level of LuxI input necessary to solve an extremely difficult SAT problem.

Solving the SAT problem with Cre-lox recombination

To combat the problem of extremely rare production of LuxI protein when multiple clauses are present on every LuxI mRNA, we looked for a system that would require a very low level of suppression output to produce a strong reporter signal. One such system is the Cre-loxP recombination system from P1 bacteriophage (Abremski & Hoess, 1984). In this system, Cre molecules bind as dimers to loxP DNA sequences. loxP sites are 34 base pairs long and consist of two 13 bp inverted repeats separated by an 8 bp spacer (Abremski & Hoess, 1984; Figure 7A). A stretch of DNA flanked by two loxP sites is called a "floxed" region.

A. **ATAACTTCGTATA** ATGTATGC **TATACGAAGTTAT**
13 bp 8 bp **13 bp**

B.

Figure 7. Cre/lox recombination occurs at loxP sites. A. The 34-bp loxP site consists of two 13 bp inverted repeats and an 8 bp spacer. This loxP site is in the forward orientation. B. Excision occurs when two loxP sites are in the same orientation. The floxed segment is excised as a circle removed from the original DNA molecule while one loxP site remains in the original location.

If two loxP sites are arranged in the same orientation, the DNA between them will be removed in the form of a circle of DNA (Abremski & Hoess, 1984; Figure 7B). Excision is precise and results in no addition of nucleotides, as Cre binds only to the loxP sites (Van Duyne, 2001). Our design for a Cre-based SAT problem would insert the frameshift logical clauses in

the Cre coding sequence, which is under transcriptional control of pTet (Figure 8). We can use the same restriction enzyme system illustrated in Figure 5 to insert SAT problems. Downstream of Cre is a floxed "stop sequence." When the stop sequence is translated, there are many stop codons that cause translation to end within the stop sequence rather than continuing beyond it. As a result, the reporter gene downstream of Cre would not be translated. However, when functional Cre is produced through frameshift suppression, the "stop sequence" would be excised, and translation would proceed through the reporter gene. A 2 bp spacer at the beginning of the reporter gene is required to maintain reading frame, as the loxP site remaining after excision is 34 bp.

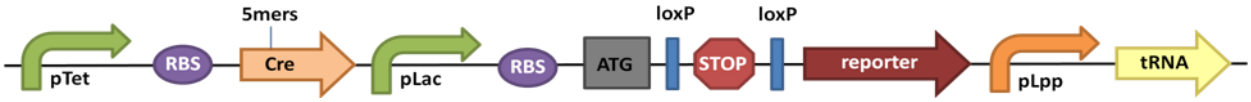


Figure 8. Solving the SAT problem with Cre. When functional Cre is produced, it excises the "stop" region, allowing for transcription and translation of a reporter. Blue rectangles represent loxP sites facing the same direction.

Scientists have used a floxed stop sequence to evaluate Cre expression in mice (Stoller *et al.*, 2008). They crossed mice homozygous for a loxP flanked STOP sequence upstream of a GFP-beta-galactosidase fusion gene with Cre expressing mice in order to evaluate where Cre was expressed. They observed GFP and beta-galactosidase activity in cells that also expressed Cre. The STOP sequence acts as an effective barrier to production of the reporter genes downstream.

We anticipated a few advantages of Cre-lox recombination over the Lux-based system. One was that the 0 state, or the off position, would be closer to a real 0, because loxP based recombination does not occur when functional Cre is absent. In contrast, in the LuxI system, pLux is slightly leaky when uninduced, allowing for detectable levels of our reporter gene (see

Figure 6). While the LuxI 1 state, or on position, is easily distinguishable from the 0 state, our 0 state in the Lux system is not as close to real 0 as we had anticipated.

Another advantage to a Cre-based system might be the low level of input needed to turn on the reporter gene. As stated above, Cre acts as a dimer, so only four molecules of Cre (two per loxP site) are needed to mediate loxP recombination (Abremski & Hoess, 1984). Cre proteins act catalytically so they should be able to excise many floxed sequences. Cre's ability to produce a detectable output is well suited for complicated SAT problems, in which the required suppression events are rare. We are uncertain of the level of input needed to induce pLux to produce detectable levels of reporter and it may be much greater than 4 molecules. An additional advantage of the Cre-based system is the ability to use reporter genes that are easily detected at low levels of expression, but lethal at high expression levels, such as tetracycline resistance (Levy, 1984). Tetracycline resistance is incompatible with a positive feedback loop, because positive feedback causes lethal expression levels of the tetracycline resistance protein while RFP is non-lethal.

The use of antibiotic resistance genes also enables us to apply selective pressure to achieve high efficiency of recombination. If we use tetracycline resistance as our reporter and culture cells in LB + tetracycline, the cells that can produce the Tet resistance protein would be at a clear advantage, and the "satisfied" phenotype would be evolutionarily favored. In contrast, in our Lux-based system, the production of RFP had no positive impact on the cell's fitness and may have been detrimental due to the large amount of resources needed for large-scale production of RFP (Sleight *et al.*, 2010). Overall, the Cre system should allow us to solve more complicated SAT problems.

Methods

Note: More detailed methods can be found in the Appendix.

Construct Assembly

Cells were cultured in LB media containing 100 µg/ml ampicillin or 25 µg/ml chloramphenicol. Plasmids used were pSB1A2 (high copy, ampicillin,) pSB4A3 (low copy, ampicillin) and pSB1C3 (high copy, chloramphenicol.) Plasmid preparation was conducted using the Promega Wizard Plus SV Minipreps DNA Purification System. Inserts and vectors were produced from existing BioBrick parts using the standard enzymes EcoRI, SpeI, XbaI and PstI (Knight *et al.*; Parts Registry). Inserts were run on agarose gels ranging from 0.4% to 2.2% agarose in TAE buffer. DNA was cleaned and concentrated according to the online protocol [www.bio.davidson.edu/courses/Molbio/Protocols/clean_short.html]. Gel purification was performed using the Macherey-Nagel NucleoSpin Extract II System. Ligations were performed using the Promega 2X Rapid Ligation System. JM109 chemically competent *E. coli* from Zymo Research were used for transformations according to the online protocol [www.bio.davidson.edu/courses/Molbio/Protocols/Zippy_Transformation.html]. Colony PCR was performed using Promega 2X Green Monster Mix according to the online protocol [www.bio.davidson.edu/courses/Molbio/Protocols/ColonyPCR_Screening.html]. Oligos were selected using an online design program [http://gcat.davidson.edu/iGem10/index.html], which added the BioBrick prefix and suffix and designed E and P sticky ends. The oligos were purchased from Eurofins MWG|Operon and were assembled according to the Building dsDNA with Oligos protocol [www.bio.davidson.edu/courses/Molbio/Protocols/anneal_oligos.html].

Construct naming:

Cre8mers: Cre with the two 8 bp restriction sites directly after the ATG as described in Figure 5, as well as the AsiSI site removed.

STOP sequence construction:

Oligonucleotide sequences:

5' AATTCGCGGCCGCTTCTAGATGAT

5' AACTTCGTATAATGTATGCTATACGAAGTTATAAGGCCTGATTAAAGAAGGC

5' GCGGTGCAGAAAACCTAAGATAACTTAACCCAGATTGGCAG

5' TTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTA

5' TATACTAGAGAGAGAATATAAAAAGCCAGATTATTAATCCGGCTTTTTTATTATTTAACTGCC

5' ATATCGTGTGAGATATTCTATGTGTCTCTGTCACGCGCTCTTCATA

5' TATTCGCTCTTACTAGTAGCGGCCGCTGCA

5' CTTCGTATAGCATACATTATACGAAGTTATCATCTAGAAGCGGCCGCG

5' TCTTAGGTTTTCTGCACCGCGCCTTCTTTAATCAGGCCTTATAA

5' CGAAGGTGAGCCAGTGTGAACTGCCAATCTGGGTTAAGTTA

5' CTGGCTTTTTATATTCTCTCTCTAGTATATAAACGCAGAAAGGCCCACC

5' GACACATAGAATATCTCACACGATATGGCAGTTAAATAATAAAAAAGCCGGATTAATAAT

5' GCGGCCGCTACTAGTAAGAGCGAATATATGAAGAGCGCGTGACAGA

Forward primer: 5' GCATGAATTCGCGGCCGCTTCTAGATGATAACTTCGTATAATGTAT GCTATACGAAGTTATAAGG 3'

Reverse primer: 5' ATGCCTGCAGCGGCCGCTACTAGTAAAGGCCTGATTAAAGAAGGC 3'

First amplification conditions: 1' 52 °C; 1' 54 °C; 1' 56 °C; 1' 58 °C. Followed by 29 rounds of standard PCR as in [www.bio.davidson.edu/courses/Molbio/Protocols/pcr.html]

Subsequent amplifications as in [www.bio.davidson.edu/courses/Molbio/Protocols/pcr.html]

Removal of AsiSI site from Cre:

Forward primer 1: 5’ TGCCAGGATATACGTAATCT 3’

Reverse primer 1: 5’ ACGGTCCGAATTTTCCATGAGTGAACGAA 3’

Forward primer 2: 5’ TCGGACCGTTGCCAGGATATACGTAATCT 3’

Reverse primer 2: 5’ ATTTTCCATGAGTGAACGAA 3’

Template: RBS-Cre8mers in pSB1A2

Reaction conditions as in [www.bio.davidson.edu/courses/Molbio/Protocols/pcr.html]

Insertion of SAT problems:

I used a restriction enzyme method to insert SAT problems into Cre. DNA was digested using AsiSI and AscI and ligated to the corresponding oligonucleotide problem.

Design of SAT problems:

2-SAT problems were designed using a single base (G) to represent OR and one codon (TCT – serine) to represent AND. This structure ensures that only one of the two tRNAs can produce a suppression event in a single clause and that one suppression event per clause must occur in order to produce a functional protein.

RBS testing

I obtained a position weight matrix (PWM) for *E. coli* RBS sequences from [http://gcat.davidson.edu/Spring2010/project2/Team2/toolsbackground.html]. I used this PWM in conjunction with an algorithm I wrote, PWM_scan. The Cre gene sequence was scanned using this algorithm. I also searched for in-frame start codons near potential RBSs.

Results

Construction of STOP sequence

I designed the STOP sequence to prevent both translation and transcription of the reporter gene downstream. The original design consisted of a 63 bp sequence, which contained stop codons in all three forward reading frames, followed by the bidirectional transcriptional

terminator Bba_B0014 (Parts Registry) and an inverted repeat of the 63 bp sequence. In order to construct the sequence, I began with a standard oligonucleotide assembly (Building dsDNA with Oligos). This method failed due to the high amount of similarity among the oligonucleotide pool, which resulted from the inverted repeat design. I redesigned the third portion of the STOP sequence to reduce the amount of similarity of oligos and prevent unwanted base pairing (Figure 9A). Although this redesign increased the amount of annealed product of appropriate size, ligations of the STOP sequence into pSB1A2 were unsuccessful.

A)

AAGGCCTGATTAAAGAAGGCGCGGTGCAGAAAACCTAAGATAACTTAACCCAGATTGGCAGTTC
ACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATATACTAGAGAGAGAATATAAAAAGCC
AGATTATTAATCCGGCTTTTTTATTATTTAACTGCCATATCGTGTGAGATATTCTATGTGTCTC
TGTCACGCGCTCTTCATATATTCGCTCT

B)

5'3' Frame 1
K A Stop L K K A R C R K P K I T Stop P R L A V H T G S P S G G P F C V Y I L E R E Y K K P D Y Stop S G F F I I Stop L P Y R V R Y S Met C L C H A L F I Y S L

5'3' Frame 2
R P D Stop R R R G A E N L R Stop L N P D W Q F T L A H L R V G L S A F I Y Stop R E N I K S Q I I N P A F L L F N C H I V Stop D I L C V S V T R S S Y I R S

5'3' Frame 3
G L I K E G A V Q K T Stop D N L T Q I G S S H W L T F G W A F L R L Y T R E R I Stop K A R L L I R L F Y Y L T A I S C E I F Y V S L S R A L H I F A

3'5' Frame 1
R A N I Stop R A R D R D T Stop N I S H D Met A V K Stop Stop K S R I N N L A F Y I L S L V Y K R R K A H P K V S Q C E L P I W V K L S Stop V F C T A P S L I R P

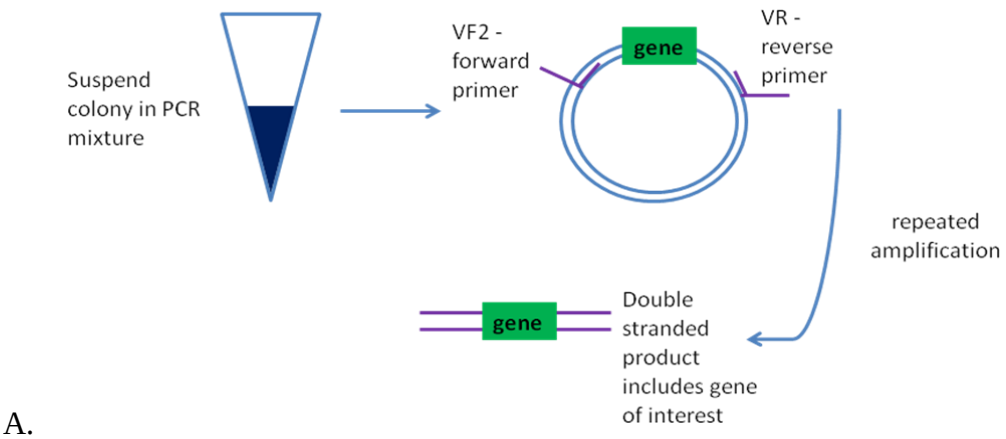
3'5' Frame 2
E R I Y E E R V T E T H R I S H T I W Q L N N K K A G L I I W L F I F S L Stop Y I N A E R P T R R Stop A S V N C Q S G L S Y L R F S A P R L L Stop S G L

3'5' Frame 3
S E Y Met K S A Stop Q R H I E Y L T R Y G S Stop I I K K P D Stop Stop S G F L Y S L S S I Stop T Q K G P P E G E P V Stop T A N L G Stop V I L G F L H R A F F N Q A

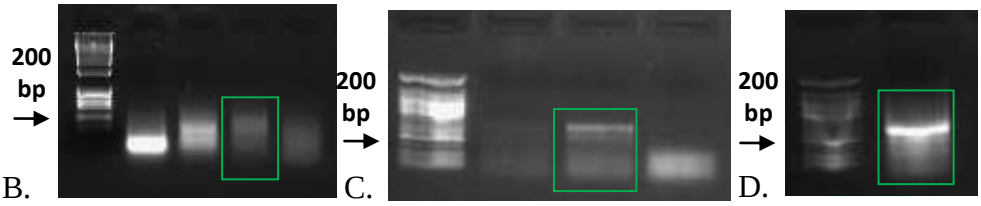
Figure 9. The STOP sequence is a barrier to both transcription and translation. A. The sequence contains a bidirectional transcriptional terminator (Parts Registry, Bba_B0014, highlighted in yellow). B. The stop sequence also contains multiple STOP codons in all six reading frames. There is one potential start codon; however, it is approximately 200 bp away from the nearest 5' RBS.

Since oligonucleotide assembly had failed to produce a high quantity of STOP sequence DNA, I decided to try a PCR approach (Figure 10A). I used the assembled oligonucleotides as a template in order to produce the STOP sequence in greater quantities. The first round of PCR produced two major bands, one of expected size and one of a smaller size (Figure 10B). I gel purified both of these bands and ran PCR reactions on each individual product and on the

combined products. The PCR reaction with the larger template produced a product of similar size in higher yield (Figure 10C). I gel purified this band and ran a final PCR reaction using it as a template (Figure 10D). After cleaning and digestion with EcoRI and PstI, the subsequent product was successfully ligated into pSB1A2.



A.



B. C. D.

Figure 10. Colony PCR is a useful construction technique for BioBrick parts. A. Colony PCR is used to construct and screen parts. B. PCR with varying quantities of oligonucleotides produced multiple products. The third PCR reaction produced products of different sizes, two of which I cut out and gel purified. C. PCR using the high MW product as a template produces more of this high MW product. The second PCR reaction produces a clear band at 300 bp, which I cut out and gel purified. D. PCR using the product in C. results in a higher yield of this product. This reaction yielded sufficient product for subsequent digestion and ligation. Green rectangles indicate wells from which products were gel purified.

Construction/Characterization of reporter genes

My reporter genes are RFP and TetA with the start codon (ATG) removed, denoted hRFP and hTetA. Both modified reporters were successfully constructed using PCR assembly (see Figure 10A). Since loxP is 34 bp long, I needed to construct my reporter genes with an extra two bases at the 5' end in order to avoid a frameshift. This two bp insertion is found in the BioBrick

prefix for each reporter. As a result, the stop codons in the BioBrick scar is not read in frame and translation continues in the appropriate frame, mimicking wild-type reporter gene translation (Figure 11).

Wild-type translation:

```
ATG GCT TCC TCC GAA GAC
 M   A   S   S   E   D
```

Translation post-excision:

```
ATG ATA ACT TCG TAT AAT GTA TGC TAT ACG AAG TTA TTA CTA GAG GCT TCC TCC GAA GAC
 M   I   T   C   Y   N   V   C   Y   T   K   L   L   L   E   A   S   S   E   D
```

Figure 11. Excision should produce in-frame translation. The top sequence represents wild-type RFP translation (Parts Registry, Bba_E1010.) The second sequence represents RFP translation after the stop sequence is excised. The two differ only in the 15 amino acids at the N terminus that result from translation of the loxP site, the BioBrick scar and the 2 bp spacer.

The usefulness of the system hinges on the nonfunctionality of reporter genes both in isolation and in the larger construct when Cre-mediated recombination has not occurred. I determined by eye that RFP was not produced by either hRFP or the full floxed construct with hRFP. To evaluate the expression of hTetA, I cultured both constructs in LB with ampicillin and tetracycline. The floxed construct with hTetA grew in liquid LB Tet. Past work in our lab has shown that Tet selection varies between liquid and solid media. I then tested the constructs on LB Tet plates, but obtained inconclusive results. Based on these results, I decided to continue with hRFP.

Construction of Cre

I added the restriction sites for AsiSI and AscI to wild-type Cre using PCR assembly (see Figure 10A). During sequence verification, I noticed that Cre contained a second AsiSI site about 400 bp from the 5’ end of the sequence. Using a standard codon usage table, I redesigned the DNA sequence of Cre while maintaining amino acid identity, without substantially altering codon usage (Figure 12).

A**GC** S (16.6)	**GAT** D (37.9)	**CGC** R (26.0)
TCG S (8.0)	GAC D (20.5)	CGT R (21.1)

Figure 12. The internal AsiSI site can be removed from Cre without altering amino acid identity and without substantial changes in codon usage. The top row of the table represents three codons of wild-type Cre and their respective amino acids, with the AsiSI site in bold. The bottom row represents the same amino acid sequence, with changes in the DNA sequence to remove the AsiSI site. Amino acid single letter codes and codon usage frequencies (per thousand) are provided (Nakamura *et al.*, 2002).

In order to construct this Cre variant, I used a variation on standard PCR assembly. I designed two pairs of PCR primers, each of which produced full length, linear Cre and plasmid pSB1A2 without the second AsiSI site (Figure 13).

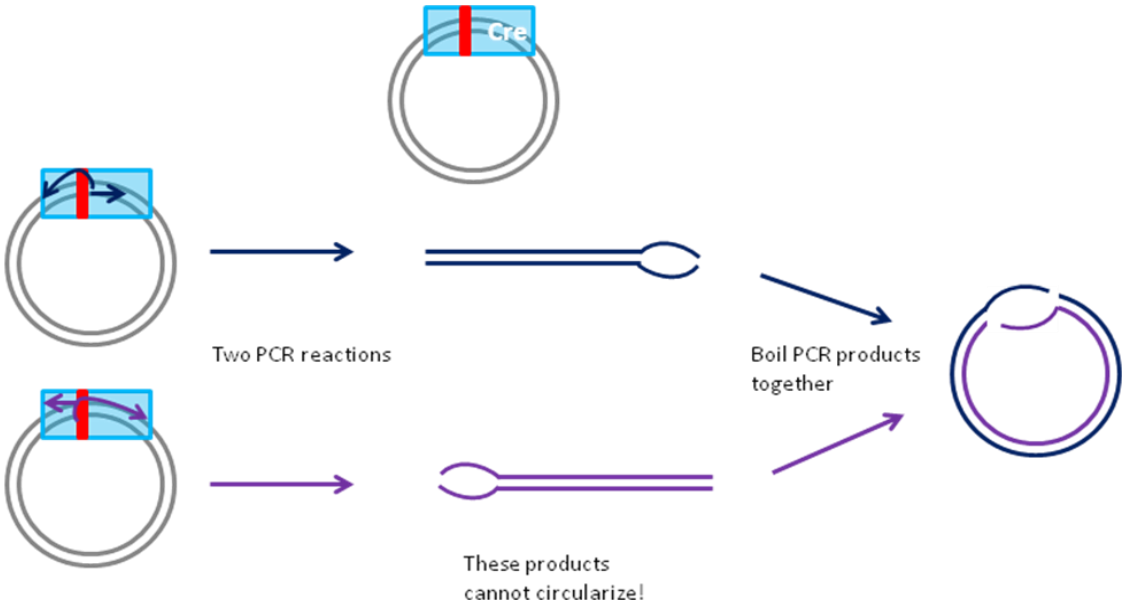


Figure 13. Constructing Cre without a native AsiSI site. A. Cre's AsiSI site is marked in red. Two rounds of PCR produce two full length, linear Cres (blue and purple) without this site in plasmid pSB1A2. These products contain the same nucleotide sequence; however, one contains

the Cre mutations (curved line) at the 5’ end and the other has these mutations at the 3’ end. B. When strands from different PCR reactions anneal, a circular product is formed. Strands from the same PCR reaction cannot form a plasmid.

One linear PCR product contained the mutations at the 5’ end and the other contained the mutations at the 3’ end. I boiled the combined products together as in standard oligonucleotide assembly and cooled to allow annealing of strands. Only pairs that contained one strand from each PCR reaction could assemble into a circle to produce a plasmid. I transformed bacteria with the cooled mixture and screened transformants using AsiSI. When DNA from colonies that retained the second AsiSI site was digested, a clear band was observed at 400 bp. Through screening, I found one colony from which the second AsiSI site had been removed (Figure 14).

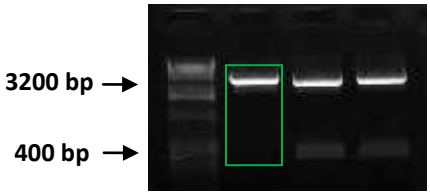


Figure 14. Digestion with AsiSI shows that Cre’s internal AsiSI site has been removed. Minipreps were digested with AsiSI only. As seen in colonies 2 and 3, there was a band at 400 bp if the internal restriction site was still present. If there was no internal restriction site, AsiSI cut once, producing a single band at 3200 bp, as in colony 1.

SAT problem design

During earlier attempts to construct a wildtype pTet-RBS-Cre that I could compare to pTet-RBS-Cre8mers, I found that pTet-RBS-Cre was not stable in *E. coli*. Sequencing results showed that recombination had taken place in pTet, rendering the construct nonfunctional. As a result, a SAT problem (frameshifts) must be inserted into Cre before adding a promoter and RBS. Since pTet-RBS-Cre was not stable in *E.coli*, I used pBad-RBS-Cre and pBad-RBS-Cre8mers for control testing. The weaker promoter pBad allows for lower production of Cre than pTet and is not lethal. Since I will be switching SAT problems in non-functional Cre genes, I need to able to differentiate them easily. During my earlier work with the Lux system, I used

oligos from the SAT problem as PCR primers to detect whether the frameshift DNA had been inserted. PCR amplification using these oligos as primers was specific to constructs that contained the newly inserted, respective SAT problems. When designing problems, I took care to design constructs for which this PCR screening approach was feasible.

The Lux system allowed us to determine proof-of-concept and solve a simple single literal problem (see Figure 6). Since a single literal problem had already been solved, I began with a 1-clause, 2-SAT problem (Figure 15). This class of problem is similar to the single literal problem in that it requires one suppression event to produce a functional protein. For our approach to be valid in solving a 2-SAT problem, suppression of either the first 5mer or the second 5mer would produce the "satisfied" phenotype as seen in the single literal problem. In addition, the "satisfied" phenotype must be present when both tRNA inputs are present. Each mRNA can be suppressed only once because the 1 base spacer prevents reading of the second 5mer as a single codon.

1-clause, 2-SAT problem

GCA**CCACT**G**CTAGT**TCTGGG

A OR B

No suppression:

GCA **CCA** **CT**G **CTA** **GT**T CTG GG

A P L L V L

Suppression of CCACT:

GCA **CCACT** G**CT** **AGT** TCT GGG

A S A S S G

Suppression of CTAGT:

GCA **CCA** **CT**G **CTAGT** TCT GGG

A P L S S G

Figure 15. The 1-clause, 2-SAT problem requires one suppression event. The problem can be represented by letters or the actual frameshifts (in red). Suppression of either frameshift results in in-frame translation of Cre (blue amino acids). If no tRNA or a tRNA not encoded in the problem is present, translation proceeds out of frame (red amino acids). Double suppression is impossible due to the 1 base spacer.

In order to test these assumptions, we must supply to the 2-SAT problem (A or B): no tRNA, tRNA A, tRNA B, tRNA C, as well as both tRNAs A and B inside cells. If every combination except no tRNA and tRNA C produced the same, satisfied phenotype, we would conclude that our bacterial computers are able to solve various SAT problems of this complexity level.

The second problem I designed was a 2-clause, 2-SAT problem of the form A OR B AND B OR C where each letter denotes a different suppressor tRNA (Figure 16). This 2-clause problem requires two suppression events in a single translation event to produce a functional product. Since multiple suppression events are necessary to solve more difficult SAT problems, it is important to determine if multiple suppression events can occur at a high enough frequency to allow for production of the reporter gene.

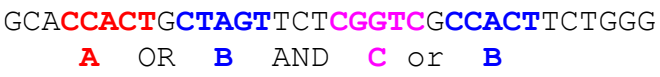


Figure 16. The 2-clause, 2-SAT problem requires two suppression events per translation event. The problem can be represented by letters or the actual frameshifts (in colors). True solutions to this problem are (A, B), (A, C), (B, C) and (B). Each tRNA in this problem has produced fluorescence as a result of a suppression event in past single literal constructs.

The 2-clause, 2-SAT design would allow us to evaluate whether multiple suppression events could result from two scenarios: 1) one suppression from each of two different tRNA alleles and 2) two suppressions from a single allele of tRNA. The structure of the problem also facilitates use of the PCR screening approach (see Figure 10A); a primer that spanned the C OR B region would be specific to that region and not bind to the 1-clause, 2-SAT, A or B described previously.

Cotransformation protocol

I chose to place the two portions of the construct on separate plasmids: pSB1C3 (chloramphenicol) for Cre and tRNAs and pSB4A3 (ampicillin) for the floxed reporter construct (Figure 17). pSB1C3 is a high copy plasmid (100-300 copies per cell) which will allow for high-level transcription of Cre (Parts Registry). The resulting high level of translation will increase the probability that frameshift suppression occurs at sufficient levels to turn the system on.

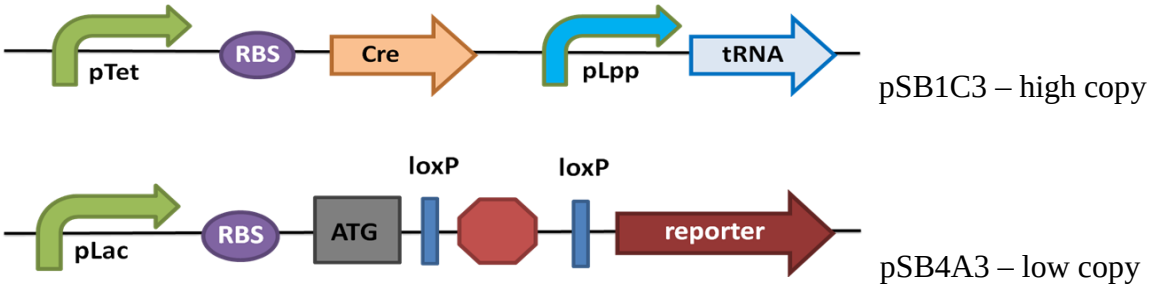


Figure 17. Cre and the floxed reporter construct are produced on different plasmids. pSB1C3, a high copy plasmid, allows for high transcription/translation of Cre and the suppressor tRNAs. This high level expression will counteract the low probability of suppression events. Since excision is common when functional Cre is produced, the floxed construct can be placed in the low-copy plasmid pSB4A3.

In contrast, pSB4A3 is a low copy plasmid, at 10-15 copies per cell (Parts Registry). Since the number of copies of the plasmid is small, there is a higher likelihood that Cre will induce recombination in every plasmid within a given cell. Although there are only a few copies of the plasmid on which excision can occur, repeated transcription and translation amplify the excision signal, so a low copy plasmid does not prevent me from obtaining clear output.

Control testing

Once I had constructed the full reporter construct for the hRFP reporter gene, I cotransformed the construct with wild-type pBad-RBS-Cre in pSB1C3. No RFP output was

visible (Figure 18).

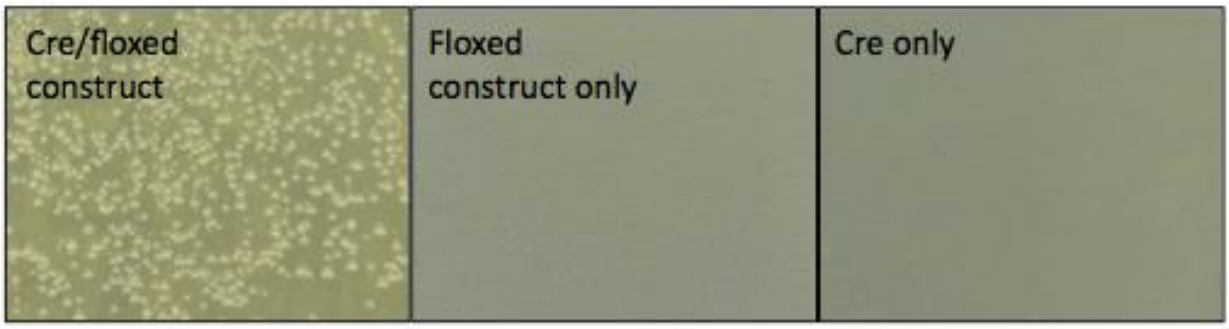


Figure 18. Cotransformation of Cre and the floxed construct does not result in reporter output. Chloramphenicol/ampicillin plates transformed with only Cre or the floxed construct produced no colonies. Cotransformation of these two plasmids is essential for colony growth. All cotransformed colonies with the RFP reporter were white.

Problems in either Cre or the floxed construct would prevent expression of the reporter gene. In order to determine if the STOP sequence had been excised, I ran a PCR reaction using a standard forward primer with a reverse primer unique to the STOP sequence. The positive control, the full RFP reporter construct, produced a band at 600 bp, indicating that the STOP sequence was present (Figure 19). Of the ten colonies I tested from the RFP reporter cotransformation plate, none produced a band, indicating that the STOP sequence had been excised. The lack of expression of the reporter gene probably was not caused by a failure to excise the STOP sequence.

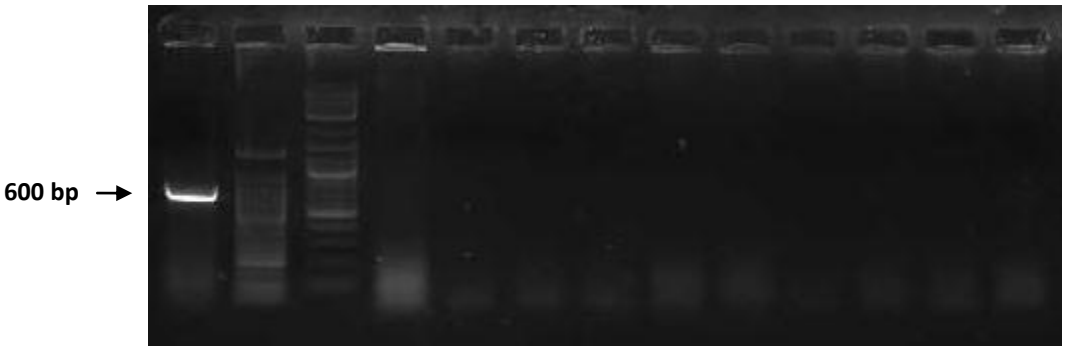


Figure 19. Lack of PCR product indicates excision. The positive control in the first lane, which was not transformed with Cre, produced a band at 600 bp, indicating that the STOP

sequence is present. The ten colonies transformed with Cre do not produce a band, suggesting that the STOP sequence was excised.

When PCR evidence suggested that the STOP sequence was excised, but RFP was not produced, we sent samples for DNA sequencing. Sequencing results confirmed that the STOP sequence and one loxP site had been excised. My analysis confirmed that the RFP reporter differs from the wild-type only in that it has 14 extra amino acids at the N terminus (see Figure 11). These amino acids result from the translation of the loxP site, the BioBrick scar and the 2 bp spacer. Unfortunately, colonies were not visibly red on plates or in liquid culture. I will address this further in the discussion.

I also tested my Cre8mers to see if it functioned as wild-type Cre. Another lab member (Bri Pearson) constructed pTet-loxP-RBS-RFP-loxP in pSB4A3 and wild-type pBad-RBS-Cre in pSB1C3. I conducted three cotransformation experiments in parallel: plate 1) floxed RFP with wild-type pBad-RBS-Cre on ampicillin only, plate 2) floxed RFP with wildtype pBad-RBS-Cre on chloramphenicol/ampicillin, plate 3) floxed RFP with pBad-RBS-Cre8mers on chloramphenicol/ampicillin (Figure 20). Colonies on the ampicillin only plate were all red, indicating that Cre was not taken up, and that the floxed construct is red before excision. Colonies on the two chloramphenicol/ampicillin plates were white, indicating successful excision of RBS-RFP. Cre8mers functions as wild-type Cre.

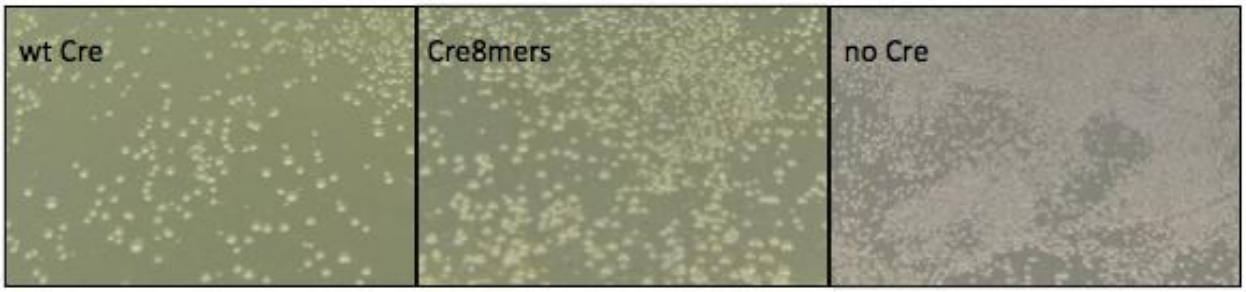


Figure 20. When Cre is present, RFP is excised. The presence of both wild-type Cre and Cre8mers resulted in excision of RFP and white colonies. Colonies on the ampicillin only plate did not take up Cre, so RFP was not excised.

To confirm excision, I ran a colony PCR reaction on colonies from plates 2 and 3, using plate 1 as a control (Figure 21). The plate 1 colony produced a band at 900 bp, indicating that RBS-RFP was still present. Plate 2 and 3 colonies produced a band at 1200 bp, representing pBad-RBS-Cre, and a band at 300 bp, representing pTet-loxP. **A 300 bp PCR fragment provides strong evidence for Cre-mediated excision** by both wild-type Cre and Cre8mers.

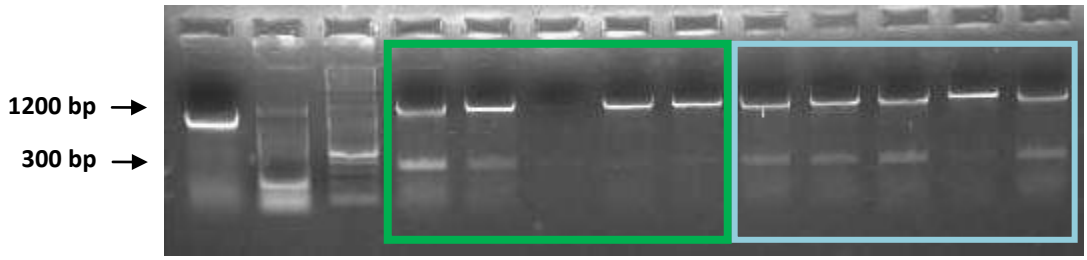


Figure 21. Colony PCR confirms excision. The control in Lane 1 shows one band at 900 bp, indicating that RFP is present. The experimental lanes show bands at 1200 bp (pBad-RBS-Cre) and 300 bp (pTet-loxP.) The first five colonies (green lanes) were transformed with wild-type Cre and the second five (blue lanes) were transformed with Cre8mers. Excision activity (300 bp band) is equal between the two Cres.

I then continued to use the pTet-loxP-RBS-RFP-loxP reporter system to test the 1-clause, 2-SAT problem. Using oligonucleotide assembly, I inserted the 2-SAT problem illustrated in Figure 15 into Cre8mers. PCR evidence and subsequent DNA sequencing confirmed the insertion of the problem into Cre8mers, now referred to as Cre2SAT (Figure 22).

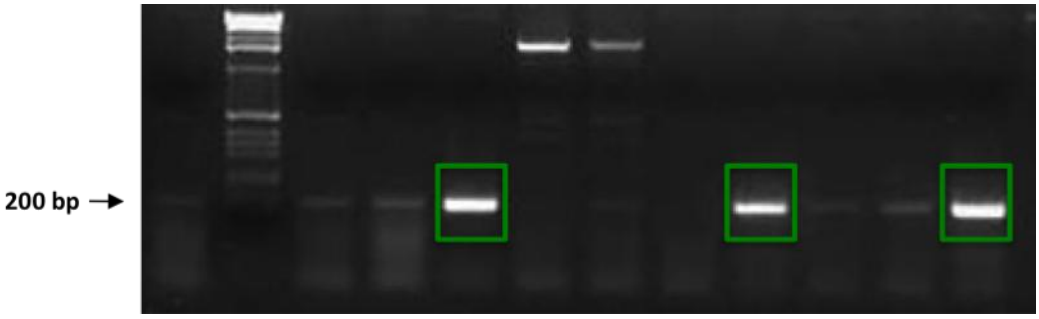


Figure 22. Colony PCR shows that the 1-clause, 2-SAT problem has been inserted into Cre8mers. An oligonucleotide from the problem was used in conjunction with a standard forward primer to screen for problem insertion. The negative control is Cre8mers, which

produces no band. Three colonies (highlighted in green) produced bright bands at 200 bp, indicating that the problem has been inserted.

I then added the pTet promoter and constructed four variants – pTet-RBS-Cre2SAT with no tRNA, CCACU tRNA, CUAGU tRNA and CGGUC tRNA. This 1-clause, 2-SAT problem requires one suppression event, which can be accomplished by the CCACU or CUAGU tRNA. Constructs with the CCACU or CUAGU tRNA should display a satisfied phenotype, whereas constructs with the CGGUC tRNA or no tRNA should not. Sequencing results showed a mutation in the anticodon of the CUAGU tRNA, so I did not include that tRNA in testing.

I cotransformed these Cre constructs with pTet-loxP-RBS-RFP-loxP and plated them on LB chlor/amp. I also plated pTet-loxP-RBS-RFP-loxP and pTet-RBS-Cre2SAT on LB ampicillin only and LB chloramphenicol only to obtain a positive and negative control, respectively. Each of the Cre cotransformation plates displayed the same phenotype. Each contained light pink colonies, which were lighter than the positive control but noticeably redder than the negative control. I could not distinguish a "satisfied" or "unsatisfied" phenotype by eye (Figure 23).

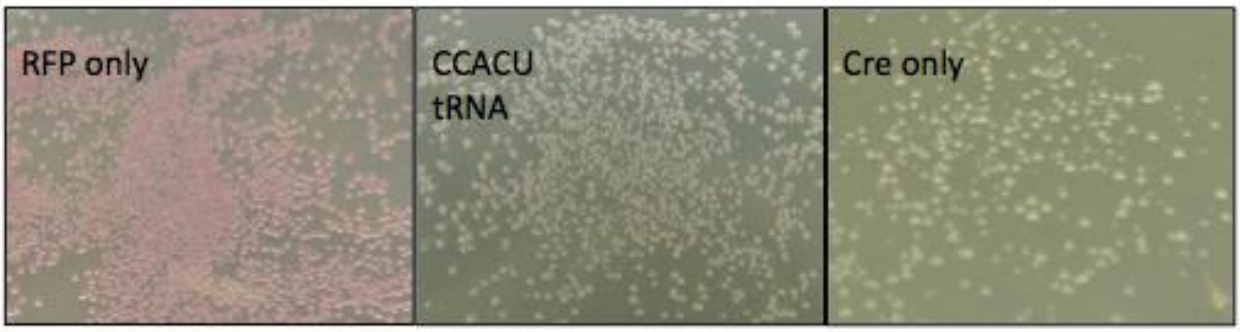


Figure 23. Cre/floxed construct cotransformation plates represent an intermediate, pinkish phenotype. Cre constructs (no tRNA, CCACU tRNA, CGGUC tRNA) were cotransformed with pTet-loxP-RBS-RFP-loxP and plated them on LB chlor/amp. Colonies from these plates are lighter than the positive control (pTet-loxP-RBS-RFP-loxP on LB amp only) but redder than the negative control (pTet-RBS-Cre2SAT on LB chlor only.) The CCACU construct shown here represents these phenotypes. Satisfied and unsatisfied phenotypes are not distinguishable by eye.

Since I could not easily distinguish unsatisfied and satisfied colonies on plates, I decided to culture five colonies of each construct to see if fluorescence varied in liquid culture. I used LB with only ampicillin to discourage the retention of the Cre portion of the construct in order to see how much excision had occurred at the time of plating. As one would expect from typical gene expression noise, there was a good deal of variance among constructs of the same phenotype (Figure 24). The no tRNA and CCACU construct measurements included outliers with very high and very low fluorescence, respectively. When I took the mean of these measurements, the fluorescence of the CCACU construct was the lowest, indicating excision and satisfaction of the problem. However, 11 of the 15 measurements among varying phenotypes fall in the same 200 point range of fluorescence. There is no clear difference between unsatisfied and satisfied phenotype.

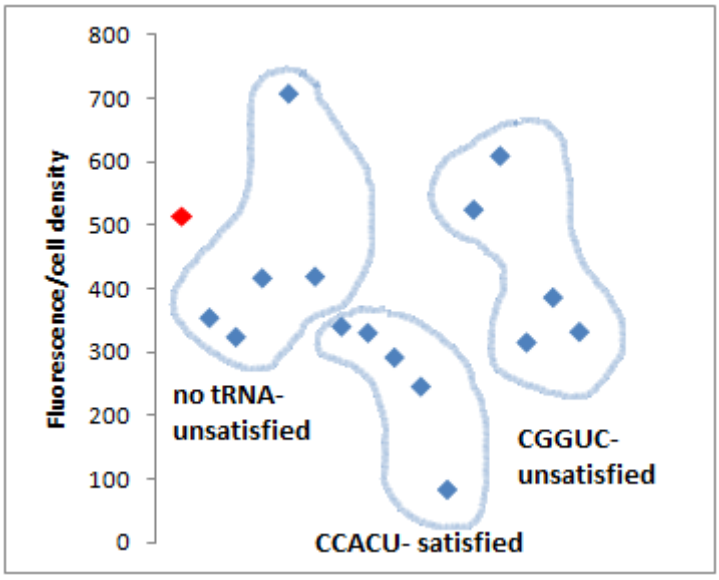


Figure 24. There is no clear difference in fluorescence between satisfied and unsatisfied phenotypes. Cultures from Cre/floxed construct cotransformation plates (no tRNA, CCACU tRNA, CGGUC tRNA) were grown overnight in LB amp. The red point represents the positive control (floxed construct only.) Constructs of the same phenotype are grouped together. 11 of the 15 measurements fall in the same 200 point range of fluorescence.

I hypothesized that differences in fluorescence would become clearer over time as excision progressed and RFP could be degraded. I began an experiment to track excision over time. I again cultured five colonies for each construct; however, I placed these colonies in LB chlor/amp to select for maintenance of the Cre construct. This selection allows for continued transcription and translation of Cre. These cultures again showed high variance among cultures of the same construct. Again, the CCACU tRNA (satisfied) construct had the overall lowest fluorescence, but some of the individual fluorescence data points were indistinguishable from those of the no tRNA (unsatisfied) construct (Figure 25).

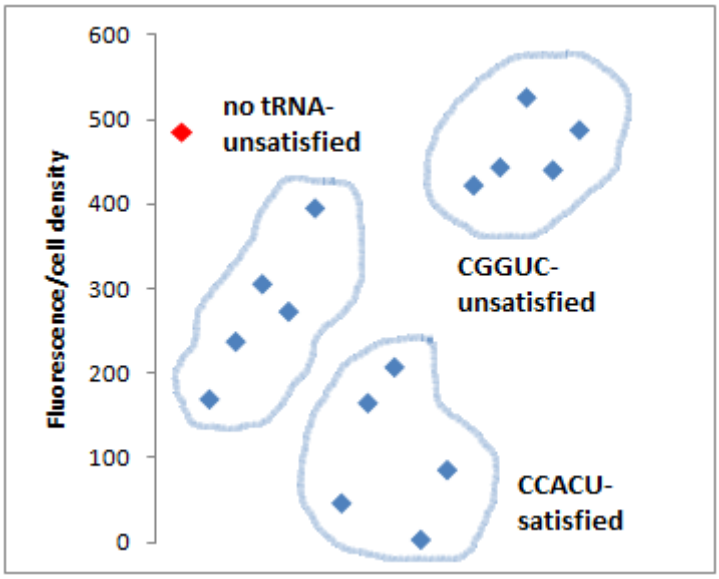


Figure 25. CCACU tRNA constructs have the lowest mean fluorescence, but there is a great deal of variance among colonies of the same construct. Cultures from Cre/floxed construct cotransformation plates (no tRNA, CCACU tRNA, CGGUC tRNA) were grown overnight in LB chlor/amp. The red point represents the positive control (floxed construct only grown in LB amp.) Constructs of the same phenotype are grouped together. There is much similarity between no tRNA and CCACU tRNA data points; CGGUC tRNA points cluster independently.

One interesting feature of this experiment was the clustering of CGGUC tRNA (unsatisfied) constructs above both the no tRNA and CCACU tRNA constructs. This clustering shows the potential for two distinct phenotypes of satisfied and unsatisfied; however, the

difference in fluorescence between the two unsatisfied constructs (no tRNA and CGGUC tRNA) is troubling.

After measuring initial fluorescence, I transferred aliquots of the cultures into fresh LB chlor/amp. I hypothesized that the fluorescence of the satisfied CCACU tRNA would continue to decrease, whereas the fluorescence of the unsatisfied no tRNA and CGGUC tRNA constructs would remain constant. I observed a different result; the fluorescence of both the unsatisfied no tRNA and satisfied CCACU tRNA decreased dramatically (Figure 26). Again, as seen in Figures 24 and 25, the fluorescence readings of individual no tRNA and CCACU tRNA constructs were indistinguishable.

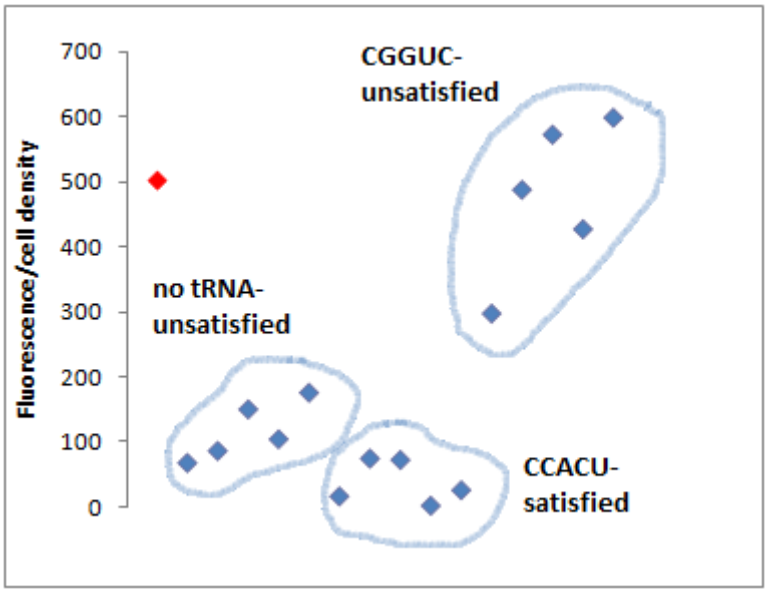


Figure 26. After a second day in culture, the fluorescence of both the no tRNA and CCACU tRNA constructs has decreased. Cultures from Figure 25 were used to seed fresh cultures in LB chlor/amp. The red point represents the positive control (floxed construct only grown in LB amp.) Constructs of the same phenotype are grouped using circles. The no tRNA and CCACU data points are virtually indistinguishable; CGGUC tRNA points cluster separately.

I again cultured these 15 colonies, this time using cells I had saved on plates the morning of the first experiment. The overnight cultures did not display any visible fluorescence as I had seen previously. Each culture was as uniformly nonfluorescent as the negative control.

I prepared samples from two colonies of the no tRNA construct for sequencing. I transformed cells and plated them onto LB amp and LB chlor plates in order to obtain data for both Cre and the floxed construct. Cre sequence data matched my prediction; the 1-clause, 2-SAT problem was maintained in the gene. The floxed construct sequence data were more surprising. The BioBrick prefix and suffix flank the pTet promoter and one loxP site (Figure 27). These data suggest that Cre-mediated recombination is occurring despite the fact that there is a frameshift at the 5' end of the Cre gene.

GAATTCGCGGCCGCTTCTAGAGTCCCTATCAGTGATAGAGATTGACATCCCTATCAGTGATAGA
GATACTGAGCACTA**CTAGAG**ATAACTTCGTATAATGTATGCTATACGAAGTTAT**TACTAGTAGC**
GGCCGCTGCAG

Figure 27. Sequence data suggest that Cre-mediated recombination has occurred in the no tRNA construct. The pTet promoter (blue), preceded by the BioBrick prefix (red), is followed by one loxP site (purple) and the BioBrick suffix (red). RFP and the other loxP site are no longer present in the construct. This precise excision suggests Cre-mediated recombination rather than random recombination.

Discussion

Construct design

The Cre-based system required *de novo* synthesis of the STOP sequence. I attempted this synthesis using an oligonucleotide assembly method. Due to the size of the sequence (>200 bp,) 13 oligonucleotides were needed to span the entire sequence. The annealing of 13 oligos is an unlikely event, and it did not occur at a high enough frequency for successful ligation into pSB1A2. In order to combat low annealing efficiency, I used PCR to amplify the STOP sequence that had assembled at low efficiency. Since I ran PCR without the initial denaturing step, the assembled oligos were able to serve as a template, and elongated PCR product could displace the oligos to form a stable double stranded product. After multiple rounds of gel purification and amplification, the STOP sequence was produced in a high enough yield to allow

for successful ligation. This PCR based amplification strategy is well suited for *de novo* synthesis. Theoretically, only one full-length set of assembled oligos is needed as a template. This solution overcomes the problems with thermodynamics associated with large-scale oligonucleotide assembly.

PCR also held the solution for another construction issue, the removal of a second AsiSI site from Cre. Since the site occurred more than 400 bp from the start of the gene, removal using standard PCR-based mutagenesis would be difficult. I designed primers to produce 2 sets of linear Cre + pSB1A2, one with the altered bases 5' and the other with the altered bases 3' (see Figure 13). I boiled the products together to allow annealing to produce a circular product and then transformed the mixture. This solution produced a colony without the AsiSI site without the normal labor associated with site-directed mutagenesis. This example also emphasizes the versatility of the BioBrick system. Codon substitution enables us to removes restriction sites that would be incompatible with various restriction enzymes without compromising gene function.

Another construction problem resulted from lethality. pTet-RBS-Cre was clonable but not stable in pSB1A2; sequencing results showed mutations in the pTet region of the construct. I then decided to use pBad-RBS-Cre as a negative control; however, the low transcription associated with pBad is not suitable for the SAT problem. I decided to continue the use of pTet to control Cre expression, but to insert the frameshift before adding the promoter. Since the frameshift disrupts wild-type Cre production, the lethality problem was easily resolved.

Construct testing

Construction problems were followed by gene expression problems. The original system was designed to produce a new phenotype when excision occurs, either RFP or tetracycline resistance. Although sequencing results suggest that functional proteins should be produced in

both cases, new phenotypes were not expressed. One potential reason for this lack of expression is the translation of the loxP site. Translation results in multiple oily residues, which may impede the folding of a barrel-shaped protein like RFP.

When this system failed to produce the desired phenotype, I began using the pTet-loxP-RBS-RFP-loxP system, which is characterized by the disappearance of the RFP phenotype. This construct worked very well when I tested the functional activity of Cre8mers; however, its utility decreased when I attempted to test the 1-clause, 2-SAT constructs. Since frameshift suppression is a rare event, it is logical that transcription and translation of RFP would occur before translation of functional Cre, rendering even "satisfied" cells red. I observed this phenomenon on culture plates; the satisfied and unsatisfied phenotypes were indistinguishable.

The loss of a phenotype rather than the gain of a phenotype also introduces problems common in synthetic biology, such as variation due to noise. If one cell translates more copies of Cre than another, more functional Cre will be produced, and we would conclude that excision would occur more quickly. My data support this idea; in each batch of testing, there was a moderate degree of variation in output among the three Cre/floxed constructs. When I took the mean of five colonies from each construct, the means were distinct, but the fluorescence readings from the individual colonies were not so easy to distinguish. I could not establish a cutoff point at which the problem was satisfied.

It is much easier to distinguish phenotypes with an antibiotic resistance reporter gene. Cell fluorescence varies widely, but antibiotic resistance can be reduced to two distinct phenotypes, survival or death in the presence of the antibiotic. For this reason, we would prefer an antibiotic resistance reporter to a fluorescent reporter.

I had hoped that, given time for excision to occur and the RFP to degrade, the two populations of cells would diverge as satisfied cells would gradually lose fluorescence. Although the fluorescence of the satisfied cells did decrease, the fluorescence of the unsatisfied cells also decreased. Sequence data show that, in the case of the no tRNA construct, this loss of fluorescence occurs as a result of RFP excision. Functional Cre is being produced despite the 5' frameshift.

Excision despite frameshifts in Cre results from the nature of Cre system. As stated in the Introduction, we selected the Cre-based system because of the low level of input required to produce recombination. Since the level of input needed is so low, rare events can have much more of an impact on phenotype. One possible rare event is a transcription error. If such an error resulted in the removal of two bases or the insertion of an additional base, reading frame would be restored, and could allow for functional protein production. Another possibility is a translation error. Again, a -2 or +1 change would restore reading frame.

An additional possibility would be that Cre is able to utilize an alternative start codon downstream of the recognized start codon. Traditional dogma states that an RBS must immediately precede a start codon, within a range of 7-8 bases, in order for translation to occur (Shine & Dalgarno, 1975), though RBSs can be located further away (Bakke, *et al.*, 2009) Multiple start codons are commonly used in prokaryotic translation (rev. in Malys & McCarthy, 2011); in only the first 200 bases of Cre, there are three such sites in the appropriate reading frame (Figure 28).

ATGTCCAATTTACTGACCGTACACCAAAATTTGCCTGCATTACCGGTCGATGCAACGAGTGATG
AGGTTCGCAAGAACCTGATGGACATGTTCAGGGATCGCCAGGCGTTTTCTGAGCATACCTGGAA
AATGCTTCTGTCCGTTTGCCGGTCGTGGGCGGCATGGTGCAAGTTGAATAACCGGAAATGGTTT
CCCGCAGA

Figure 28. Cre contains alternative start codons. The first 200 bases of Cre contain three additional ATGs (yellow) in the appropriate reading frame. Translation beginning at one of these

ATGs would bypass the SAT problem, freeing the cell from the requirement of a frameshift suppression event.

Once I had found additional start codons, I also scanned Cre for an internal RBS, which would allow translation to begin downstream of the SAT problem. To scan for highly conserved sequences like an RBS, we use a position weight matrix (PWM,) which shows the frequency of nucleotides at each position of the sequence and allows us to find and score sequences. I used a PWM for *E. coli* developed in the Bioinformatics class at Davidson and my own PWM scanning algorithm from that class to scan Cre for an internal RBS (see methods/appendix). Within the first 400 bases of Cre, I found multiple putative RBS sequences, one of which contains an ATG directly downstream. This RBS/ATG combination, as well as the BioBrick RBS/ATG, is shown in Figure 29.

AAAGAGGAGAAATACTAGATGTCCAATTTACTGACCGTACACCAAAATTTGCCTGCATTACCGG
TCGATGCAACGAGTGATGAGGTTCGCAAGAACCTGATGGACATGTTCAGGGATCGCCAGGCGTT
TTCTGAGCATACCTGGAAAATGCTTCTGTCCGTTTGCCGGTCGTGGGCGGCATGGTGCAAGTTG
AATAACCGGAAATGGTTTCCCGCAGA

Figure 29. Cre contains a putative RBS followed by an ATG. The sequence above represents the BioBrick RBS Bba_B0034, the BioBrick scar and the first 200 bases of Cre. The first 200 bases of Cre contain a putative RBS with an ATG downstream. Translation beginning at this RBS would bypass the SAT problem, freeing the cell from the requirement of a frameshift suppression event. The BioBrick RBS and ATG are highlighted in green, and the alternative RBS and ATG in pink.

Biochemical studies of Cre support the hypothesis that Cre may be functional despite a truncated N-terminus. Cre recombinase belongs to the highly diverse tyrosine recombinase family. Family members display highly diverse N-terminal ends, but maintain four residues necessary for catalysis (Argos *et al.*, 1986; Abremski & Hoess, 1992). All of these residues are located at the C-terminus: H-289, R-292, R-297 and Y-324 in Cre that is 345 amino acids long. Based on these data, it is not impossible to imagine a functional Cre with a truncated N-terminus. I will discuss potential solutions to this problem in the Future experiments section.

Evolutionary considerations

When designing and testing biological constructs, it is important to consider evolutionary implications. An ideal construct would impart a phenotype necessary for survival, such as the initial TetA design, because this phenotype would be greatly favored. A less apparent advantage results from the removal of an unnecessary phenotype, such as the floxed RFP design, because protein production beyond wild-type places a burden on the cell, as seen in the LuxI/RFP system.

We cannot be certain what advantage the loss of the RFP gives an individual cell, but it is positive and nonzero. Therefore, production of functional Cre, which allows for RFP excision, is evolutionarily favored. Cells that produced functional Cre through transcription/translation errors or alternative translation would be more likely to survive. This differential survival rate explains why the frequency of low-fluorescent no tRNA constructs increased from the first culture to the second. The additional growth time allowed for enhanced selection.

The growth conditions of my experiments also likely enhanced this evolutionary pressure. Since I placed the Cre and floxed construct on separate plasmids, each surviving cell needed to produce proteins that would confer resistance to two antibiotics, chloramphenicol and ampicillin. The cost of double antibiotic resistance is clearly detrimental to the cell. Adding in RFP production only increases this burden. When I plated cells containing pTet-loxP-RBS-RFP-loxP on ampicillin only, fluorescence was much higher than for the cells cotransformed with Cre and the floxed construct, because these cells do not have the additional burden of chloramphenicol resistance. These circumstances suggest that Cre production would be highly favored in order to mediate excision to reduce the burden of RFP production. My last round of cultures confirms this hypothesis. Although each culture was seeded from plated constructs,

some of which were originally red, each culture matched the nonfluorescent negative control. Long-term selection pressure almost completely eliminated RFP production.

Future experiments

As mentioned above, a construct that allows for the production of antibiotic resistance once excision has occurred would be best suited for the SAT problem. In order to use such a construct, hTetA must be better characterized. More thorough growth experiments in both liquid and solid media will be highly beneficial in determining the utility of this system.

The results from the no tRNA construct suggest that functional Cre is being produced and mediating excision, and we must ascertain if this is the case. One important control experiment would be transforming the floxed construct in pSB4A3 with a non-Cre construct in pSB1C3, and tracking if fluorescence decreases over time. A decrease in fluorescence would suggest epigenetic modification or excision by an alternate recombinase. If the constructs' fluorescence does not decrease, we can be more certain that the decrease in fluorescence of Cre constructs is due to Cre-mediated recombination.

It would also be prudent to test for the presence of Cre in the no tRNA constructs. One could run a Western blot with an anti-Cre antibody to determine if Cre is present. This strategy may not yield useful results because we would expect a very low level of Cre expression, but it is certainly worth doing. Other strategies may be more useful in determining if functional Cre is produced. The first is to engineer 1-clause, 2-SAT Cre without the additional start codons illustrated in Figure 28. This Cre could then be cotransformed with the floxed RFP construct to see if recombination still occurs. If recombination does not occur, this Cre could be used in future SAT constructs.

Another option is to move the placement of the frameshifts closer to the region encoding the C terminus of the protein. Since this area is a key functional domain of the protein, it follows that an alternative start codon near the 5' end could not overcome the frameshift located closer to the critical 3' end. The earliest amino acid residue necessary for catalysis is H-289, or nucleotides 865-867 (Abremski & Hoess, 1992). We could insert the SAT problems around bases 600-650, a location that would not affect catalysis. This region is also 3' to the last ATG that occurs in reading frame before the functional domain, which begins at base 576.

A one-plasmid design would decrease stress on individual cells. Such a design would place both Cre and the floxed construct on the same plasmid, allowing for selection with one fewer antibiotic. This is especially important if we can produce a functional Tet resistant construct. Currently, a satisfied phenotype would require resistance to three antibiotics, again a costly condition. The one plasmid design would decrease the resistance demand from three antibiotics to two. A one plasmid design does not work with wild-type Cre because cells produce Cre immediately after transformation, and this recombination prevents successful maintenance of ligated constructs. However, we already have solid control data from the two-plasmid design, and a frameshift Cre would not cause this problem because protein expression is so low.

If we continue to use RFP and other fluorescent reporters, the use of flow cytometry to characterize cells would be highly beneficial. My fluorescent measurements came from a standard fluorimeter, and raw fluorescence was divided by cell density to obtain a mean for each sample. Fluorescence varied among both individual cultures of the same construct and among different constructs, as well as at different time points. For example, the fluorescence of 1-clause, 2-SAT problems decreased over time. This difference could result from one of two situations, 1) the fluorescence of individual cells is relatively uniform, and decreases uniformly, and 2)

fluorescence of cells varies greatly – some cells remain highly fluorescent while others become far less fluorescent. It is important to distinguish these two phenotypes, especially in unsatisfied constructs, to determine how frequently functional Cre is being produced, and how quickly excision and RFP degradation can occur in both satisfied and unsatisfied cells.

The big question my results ask is whether or not the Cre system is appropriate for the SAT problem. The low input required to activate the system has turned out to be both an advantage and a disadvantage. Although preliminary results are not entirely encouraging, I believe that it is too early to abandon the system altogether. The changes to Cre mentioned above, particularly the movement of the frameshifts closer to the 3' end of the gene in order to disrupt the C-terminus of the protein, have the potential to fix the problem of functional Cre production without frameshift suppression. This problem is the greatest roadblock to successful implementation of the system, and if it can be surmounted, I think that the low level of input required will become much more advantageous as we move toward solving more complicated SAT problems.

Acknowledgments

I would like to thank all of the individuals I worked with in the Campbell/Heyer lab, especially Bri Pearson, Pallavi Penumetcha, Anvi Raina, Katie Richeson and Eric Sawyer. I would also like to thank the 2009 Missouri Western/Davidson College iGEM team for their foundational work on the SAT problem and the 2010 iGEM team from their work with Cre. The Davidson College Biology Department and the James G. Martin Genomics Program were important sources of funding for this project. I would also like to thank Dr. Laurie Heyer, my second reader, for patiently explaining the math behind my research. Most of all, I would like to

thank my advisor, Dr. A. Malcolm Campbell, for encouraging me to think like a scientist and to trust the data.

References

Abremski KE, Hoess RH: **Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein.** *J Biol Chem.* 1984, **259**:1509-14.

Abremski KE, Hoess RH: **Evidence for a second conserved arginine residue in the integrase family of recombination proteins.** *Protein Eng.* 1992, **5**:87-91.

Argos P, Landy A, Abremski K, Egan JB, Haggard-Ljungquist E, Hoess RH, Kahn ML, Kalionis B, Narayana SVL, Pierson LS, Sternberg N, Leong JM: **The integrase family of site-specific recombinases: regional similarities and global diversity.** *EMBO J.* 1986, **5**:433-40.

Adleman LM: **Molecular computation of solutions to combinatorial problems.** *Science* 1994, **266**:1021-1024.

Anderson CJ, Magliery TJ, Schultz PG: **Exploring the limits of codon and anticodon size.** *Chem Biol* 2002, **9**:237-244.

Baumgardner J, Acker K, Adefuye O, Crowley ST, DeLoache W, Dickson JO, Lane H, Martens AT, Morton N, Ritter M, Shoecraft A, Treece J, Unzicker M, Valencia A, Waters M, Campbell AM, Heyer LJ, Poet JL, Eckdahl TT: **Solving a Hamiltonian Path Problem with a bacterial computer.** *J Biol Eng* 2009, **3**:11.

Bakke P, Carney N, DeLoache W, Gearing M, Ingvorsen P, Lotz M, McNair J, Penumetcha P, Simpson S, Voss L, Win M, Heyer LJ, Campbell AM: **Evaluation of three automated annotations for *Halorhabdus utahensis*.** *PLos ONE* 2009, **4**:e6291.

Braich RS, Chelyapov N, Johnson C, Rothemund PW, Adleman L: **Solution of a 20-variable 3-SAT problem on a DNA computer.** *Science* 2002, **296**:499-502.

Cook S: **The complexity of theorem proving procedures.** *Proceedings of the Third Annual ACM Symposium on Theory of Computing* 1971, 151–158.

Endy D: **Part:Bba E1010.** 2004. [http://partsregistry.org/Part:BBa_E1010].

Garey M, Johnson DS. **Computers and Intractability: A Guide to the Theory of NP-Completeness.** WH Freeman & Co., 1979.

Haynes KA, Broderick ML, Brown AD, Butner TL, Dickson JO, Harden WL, Heard LH, Jessen EL, Malloy KJ, Ogden BJ, Rosemond S, Simpson S, Zwack E, Campbell AM, Eckdahl TT, Heyer LJ, Poet JL: **Engineering bacteria to solve the burnt pancake problem.** *J Biol Eng* 2008, **20**; 2:8.

Knight T, Rettberg R, Chan L, Endy D, Shetty R, Che A: **Idempotent Vector Design for Standard Assembly of Biobricks.** 2003. [http://people.csail.mit.edu/tk/sa3.pdf].

Levy, SB: **Resistance to the tetracyclines.** In *Antimicrobial drug resistance* (ed. L. E. Bryan), p. 191-240. Academic Press, Inc., New York, 1984.

Lin CH, Cheng HP, Yang CB, Yang CN: **Solving satisfiability problems using a novel microarray-based DNA computer.** *BioSystems* 2007, **90**:242-252.

Lipton RJ: **DNA solution of hard computational problems.** *Science* 1995, **268**:542-545.

Lux. [http://partsregistry.org/Lux].

Magliery TJ, Anderson CJ, Schultz PG: **Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in *Escherichia coli*.** *J Mol Biol* 2001, **307**:755-769.

Malys N, McCarthy JEG: **Translation initiation: variations in the mechanism can be anticipated.** *Cell. Mol. Life Sci.* 2011, **68**:991-1003.

Nakamura Y, Gojobori T, Ikemura, T: **Codon usage tabulated from the international DNA sequence databases: status for the year 2000.** *Nuc. Acids Res.* 2000, **28**:292.

Registry of Standard Biological Parts. [http://partsregistry.org/Main_Page].

Sakamoto K, Gouzu H, Komiya K, Kiga D, Yokoyama S, Yokomori T, Hagiya M: **Molecular computation by DNA hairpin formation.** *Science* 2000, **288**:1223-1226.

Shine J, Dalgarno L: **Determinant of cistron specificity in bacterial ribosomes.** *Nature* 1975, **254(5495)**:34-38.

Sleight SC, Bartley BA, Lieviant JA, Sauro HM: Designing and engineering evolutionary robust genetic circuits. *J Biol. Eng* 2010, **4**:12.

Stoller JZ, Degenhardt KR, Huang L, Zhou DD, Lu MM, Epstein JA: **Cre reporter mouse expressing a nuclear localized fusion of GFP and beta-galactosidase reveals new derivatives of Pax3-expressing precursors.** *Genesis* 2008, **46**:200-4.

Van Duyne GD: **A structural view of Cre-lox site-specific recombination.** *Annu Rev Biophys Biomol Struct* 2001, **30**:87-104.

Appendix

This section contains detailed lab protocols previously mentioned in the Methods section.

DNA Isolation and Purification

Mini Preps

Promega Mini Prep

1. For each miniprep, grow 2 ml culture, 37° C, overnight (O/N) in appropriate medium and anitbiotic (usually ampicillin,but not always); shake at 400 RPM and slant tubes.
   Next Day
2. Pour the contents O/N culture into one labeled tube. Replace the metal cap and save the culture tube at 4° C.
3. Spin the microfuge tube for 2 min.
4. Aspirate off the medium.
5. Resuspend pellet in 250 µl **Cell Resuspension Solution**. Resuspend cells very well by pipetting up and down.
6. Add 250 µl of **Cell Lysis Solution**. Mix by inverting the tube 4 times.
7. Add 10 µl **Alkaline Protease Solution**. Mix by inverting the tube 4 times. Incubate 3 minutes at room temperature (RT°).
8. Add 350µl **Neutralization Solution**. Mix by inverting the tube 4 times.
9. Spin full speed for 10 minutes at RT°.
10. Prepare Spin Column (with binding resin) by inserting into 2 mL collection tube.
11. Transfer supernatant to Spin Column.
12. Spin full speed for 1 minute at RT°. Discard liquid flowthrough and reinsert Spin Column to collection tube.
13. Add 750 µl **Wash Solution** (with ethonal already added).
14. Spin full speed for 1 minute at RT°. Discard liquid flowthrough and reinsert Spin Column to collection tube.
15. Add 250 µl **Wash Solution** (with ethonol already added).
16. Spin full speed for 1 minute at RT°. Discard liquid flowthrough and transfer Spin Column to a clean 1.5 mL microfuge tube.
    Be sure to label the spin column and the microfuge tube.
17. Add 100 µl **Nuclease-Free Water** to the Spin Column.
18. Spin full speed for 1 minute at RT°. **SAVE THE LIQUID with your plasmid.** Discard the spin column. Store DNA at -20° C.
19. If you want to digest some DNA, 5-10 µl of the MP DNA in a final volume of 20 µl is a nice place to start.

Zyppy Mini Prep

1. For each miniprep, grow 2 ml culture, 37° C, overnight (O/N) in appropriate medium and antibiotic (usually ampicillin,but not always); shake at 400 RPM and slant tubes.

Next Day

2. Add 600 µL of O/N culture to an appropriately labeled 1.5 mL microfuge tube. Save the rest of the O/N culture at +4 ° C and keep them sterile.
   You can increase your yield by pelleting up to 3 mL of O/N culture and resuspending pellet very well in 600 µL water or TE.
3. Add 100 µL **7X Lysis Buffer (blue color)** . Mix by inverting the tube 4-10 times. Solution should become clear instead of opaque.
   Proceed to the next step within 3 minutes.
4. Add 350 µL of **Neutralization Buffer (yellow color; RNase A already added)** and mix by inverting the tube until the entire solution and precipitate is yellow. This buffer is stored at +4 ° C.
5. Spin full speed for 2 minutes at RT°.
6. Prepare Zymo-Spin II column (with binding resin) by inserting into 2 mL collection tube. Be sure to label the spin column and the collection tube. Also label a 1.5 mL microfuge tube for use in step 14.
7. Transfer ~900 µL supernatant to Zymo-Spin II column. Do not transfer any of the solid precipitate.
8. Spin full speed for 15 seconds at RT°.
9. Discard liquid flowthrough and reinsert Zymo-Spin II column into same collection tube.
10. Add 200 µL **Endo-Wash Buffer** to the Zymo-Spin II column.
11. Spin full speed for 15 seconds at RT°. No need to empty flow through.
12. Add 400 µL **Zyppy Wash Buffer** (with ethanol already added).
13. Spin full speed for 30 seconds at RT°. Discard liquid flowthrough and and the 2 mL collection tube. The DNA is still in the spin column.
14. Transfer Zymo-Spin II column to a clean and appropriately labeled 1.5 mL microfuge tube (from step 6 above).
15. Add 30 µL **Zyppy Elution BUffer ot the center of the** to the Zymo-Spin II column. Let it stand for 1 minute to maximize yield.
16. Spin full speed for 30 seconds at RT°. **SAVE THE LIQUID with your plasmid.**
    Discard the spin column. Store DNA at -20° C.
17. If you want to digest some DNA, 5-10 µL of the MP DNA in a final volume of 20 µL is a nice place to start. You can NanoDrop the DNA if you need to know the concentration.

Ethanol Precipitation

1. If the volume of the DNA is less than 200 µl, bring the volume up to 200 µl with sterile dH_2O.
2. Add 1/10 th volume of 3M sodium acetate to the DNA solution and mix.
3. Add 2 volumes of -20° C 100% ethanol (EtOH) and vortex for 10 seconds. Put the tube in a -20° C freezer overnight or a -70° C freezer for 30 minutes.
4. Spin in a microfuge for 10 minutes. Pour out the EtOH but **save the pellet**!!
5. Wash the pellet with 500 µl of 4° C 70% EtOH, gently roll the tube, then dump the EtOH, and speedvac the pellet. **SAVE THE PELLET!**
6. Resuspend DNA in appropriate volume of TE or water.

Cloning

Ligation

X ng of insert = (2) (bp insert) (50 ng linearized plasmid-) ÷ (size of plasmid in bp)

The 2 in the numerator takes into account the fact that you want to have a 2:1 ratio of insert to vector. You want to use 50 ng of plasmid for a typical ligation.

Thaw the frozen 2X buffer at room temperature. After the first thawing, it is best if you aliquot the 2X buffer in smaller volumes to avoid repeated freeze/thaw cycles. If a precipitate is present, vortex the solid DTT until it is back in solution. This usually takes 1 - 2 minutes.

Set up the following 10 µl ligation reaction (if your DNA is too dilute, you can scale up to 20 µL ligation volume and transform all 20 µL. But try to minimize the ligation volume.

Digested Vector (50 ng)	1 µl
Insert (2:1 molar ratio insert:vector)	(< 3) **x** µl
Sterile water	3µl- x µl
2X ligation buffer	5 µl
3 units T4 DNA ligase (keep cold)	1 µl
Total Volume	**10.0 µl**

Incubate for 5 minutes at room temperature. From this point, you may either freeze the ligation or go directly to transformation.

Transformation

1. Thaw the competent cells on ice for 5 minutes.
2. Very gently, aliquot cells into smaller volumes (we have used as low as 20 µL of cells with 5 µL of ligation) using chilled microfuge tubes.
3. Add 1 - 5 µL of ligation mixture (can go as high as 10 µL).
4. Incubate on ice for 5 minutes.
5. Add SOC media with no antibiotic to a final volume of 60 - 100 µL. Spread cells onto plates containing antibiotic. You may want to let this sit for 30 minutes if the antibiotic is not ampicillin (not tested, just rumored to help)

Gel Electrophoresis (Pouring a Gel)

1. Tape the ends of a gel mold and make sure some of the tape wraps around the bottom of the mold by 1 - 2 mm. Choose the appropriate comb(s) and make sure it looks clean.
2. Make 60 mL of the appropriate gel mixture in a 250 mL flask, cover it with Saran, and microwave for 1 minute and 20 seconds on high power (a good starting time).
3. Visually check to see that all the agarose has melted. Unmelted agarose looks like tiny refractive lenses floating around. If not completely melted, nuke it a little longer. Try 20 seconds.
4. Allow the gel to cool a bit; you may hasten this by running cold water over it but do NOT let it cool too much. Don't be startled (and drop the flask) by the popping sound of the saran wrap as the flask cools.
5. Add ethidium bromide (stock EtBr; 10 mg/mL) to the 60 mL of gel so the **final concentration is 0.2 µg/mL**. Then pour the gel into the mold.
EtBr is a known mutagen so wear gloves.
6. Allow this to cool until it turns slightly white. The gel is ready to run, as soon as you pull off the tape, remove the comb, and submerge it in 0.5X TBE that has the same concentration of EtBr. Unless you are told otherwise, our gel boxes hold 450 mL of buffer.

Gel Purification

1. Run the fragment(s) on a gel and photograph the gel. Remember to use a MW marker. Also, use as low a percentage gel as you can to resolve your bands.
2. Cut out the band(s) of interest using a razor blade and the UV box with the hinged plexiglass covering. Cut as close to the band as possible to minimize the volume of the gel piece. Use as low a level UV light for as little time as possible. Protect your eyes and skin from the UV.
3. Make sure Buffer NT3 has the appropriate volume of ethanol added to it before you go any further.
4. Weigh the gel slice in a colorless tube. Add 2 volumes of buffer Buffer NT to 1 volume of gel (100 mg gel = 200 µL NT). If your gel is > 2% agarose, use 400 µL Buffer NT.
5. Incubate in 50° C waterbath for 10 minutes, or until the gel slice has COMPLETELY dissolved. You can vortex every 2-3 minutes to speed up the dissolving process. Hold the tube up to the light and look for a translucent piece of undissolved gel.
6. Place a Nucleospin Extract II spin column in one of the provided 2 mL collection tubes.
7. To bind DNA, apply the DNA solution to the Nucleospin Extract II spin column and centrifuge as full speed for one minute. The maxiumum volume you can load at a time is 800 µL. If your volume is larger, then reload the same column and spin a second time.
8. Discard flow-through from step 8 and place the column back in the same collection tube.
9. To wash, add 700 µL of Buffer NT3 to the column and centrifuge for 1 minute at full speed.
10. Discard flow-through first and then spin again for an additional two minues to dry the column.
11. Place the QIAquick column in a clean, 1.5 mL centrifuge tube.
12. To elute the DNA, add 15 -50 µL of Buffer NE (30 µL is typical) to the CENTER of the QIAquick membrane. Let this sit for 1 minute at room temperature, then spin full speed for 1 minute. Your DNA is in this small volume. If your DNA is 5-10 kb or larger, preheat buffer NE to 70° C prior to adding to spin column.

13. Use 1 or 2 µL of this DNA to quantify using the NanDrop. You are ready to do a ligation now.

Restriction Digestion

1. Decide which buffer you must use and the correct incubation temperature.
2. Decide how much (in microliters) DNA you will digest.
3. Then construct a table similar to the one below:

- 3 µl DNA (volume depends on DNA concentration, 3 µl is good if using MP DNA)
- 14 µl water (up to desired volume)
- 2 µl 10X buffer (one tenth final volume)
- 1 µl restriction enzyme (never more than 10% final volume)
- 20 µl total volume

4. As you add each ingredient to a 500 µl microfuge tube, stir it in well with the pipet tip and
5. Make sure all of the liquid is in the bottom of the microfuge tube; spin if necessary.
6. Incubate at the appropriate temperature (typically 37° C) for at least 30 minutes.

Polymerase Chain Reaction

Colony PCR

1. Determine the number of colonies to be tested. Plan to conduct PCR on control plasmids with and without the insert. Assemble the following PCR mixture:

Per Reaction (you might want to make a cocktail, rather than multiple individual reactions)
1 µL forward primer (20 pmol = 0.2 µL of 100 µM oligo stock solution)
1 µL reverse primer (20 pmol = 0.2 µL of 100 µM oligo stock solution)
10 µL dH_2O
12 µL 2X Monster Mix (Green solution from Promega)
24 µL total volume

2. Use a micropipette tip to pick a single putative colony off a plate. Insert the tip into the PCR mixture and pipette up and down.
3. Reserve bacteria from each PCR mixture by removing 1 ul and placing into 100 ul of LB + Amp in a labeled tube and put in 37° C incubator.
4. Conduct PCR according to the following thermal profile:

94° C 10 minutes
20 cycles of:
94° C 15 seconds
46° C 15 seconds
72° C 30 seconds (time varies depending on the size of insert; rule of thumb is 1 minute per kb of DNA being amplified)
Hold at RT°

5. Run reaction on appropriate percentage agarose gel. If there is no insert, then the PCR product will be 258bp in size.
6. Add 1.9 mL of media to desired clones from reserved bacteria (step 3 above) for use in plasmid preps. Do your normal MiniPrep Procedure

PCR

Common Temperature Cycle (30 seconds per kb of DNA amplified):

- Step 1: 95° C 5 minutes (denature template)
- Step 2: 95° C 30 seconds (denature dsDNA)
- Step 3: 55° C 30 seconds (T_m minus 5 degrees)
- Step 4: 72° C 30 seconds (amplify about 1 kb per 30 seconds)
- Step 5: Repeat Steps 2 through 4, 29 more times
- Step 6: Store at RT°

REAGENT*	**VOLUME (µL)**	**FINAL CONC.**
water	50 - (X + Y + Z) µL	N/A
2X Master Mix (with buffer, Taq, dNTPs, $MgCl_2$**)	50 µL	1X
template DNA (~1 ng DNA)	X ($\leq$ 1 µL)	1 ng
primer #1	Y µL	1 µM
primer #2	Z ML	1 µM
FINAL VOLUME	**100 µL**	

Gene Scanning

PWM_scan algorithm:

```
my $FILEIN;
my $PWMFile = "pwm_ecoli.txt";

open ($FILEIN, '<', $PWMFile)
   or die "Cannot open the input file: $PWMFile: $!";

my $Astring = <$FILEIN>;
chomp($Astring);
my @Avalues = split (/\t/, $Astring);

my $Cstring = <$FILEIN>;
chomp($Cstring);
```

```
my @Cvalues = split (/\t/, $Cstring);

my $Gstring = <$FILEIN>;
chomp($Gstring);
my @Gvalues = split (/\t/, $Gstring);

my $Tstring = <$FILEIN>;
chomp($Tstring);
my @Tvalues = split (/\t/, $Tstring);

my $PWMlength = scalar(@Avalues);

close $FILEIN;
my $genomefilename = "cre.txt";
open(my $fh, '<', $genomefilename)
   or die $!;
my @words = <$fh>;
my $genome = join("",@words);
$genome =~ s/\s//g;

my $sequencelength = length($genome);
my @highSequences;
my @highPositions;
my @highScores;
my @highStrands;
my $limit = 280;

for (my $i = 0; $i <= ($sequencelength - $PWMlength); $i++)
{
          my $scan = substr($genome, $i, $PWMlength);
          my $score = PWMscore($scan);
          my $revscan = reversecomp($scan);
          my $revscore = PWMscore($revscan);
   if ($score > $limit)
   {
      push(@highSequences, $scan);
      push(@highScores, $score);
      push(@highPositions, $i);
      push(@highStrands, "forward");
   }
   if ($revscore > $limit)
   {
      push(@highSequences, $revscan);
      push(@highScores, $revscore);
      push(@highPositions, $i);
      push(@highStrands, "reverse");
```

```
    }
}

print "High scoring matches to the PWM: \n";
print "Sequence", "\t", "Start", "\t", "Direction", "\t", "PWM Score", "\n";

for (my $j = 0; $j < scalar(@highScores); $j++)
{
    print "$highSequences[$j]", "\t", "$highPositions[$j]", "\t", "$highStrands[$j]", "\t",
"$highScores[$j]", "\n"
}

# Subroutines ############################
sub PWMscore
{
    my $seq = shift(@_);
    $seq = uc($seq);
    my $score = 0;
    for (my $s = 0; $s < length($seq); $s++)
    {
        if (substr($seq, $s, 1) eq "A")
        {
            $score = $score + $Avalues[$s];
        }
        elsif (substr($seq, $s, 1) eq "C")
        {
            $score = $score + $Cvalues[$s];
        }
        elsif (substr($seq, $s, 1) eq "G")
        {
            $score = $score + $Gvalues[$s];
        }
        else
        {
            $score = $score + $Tvalues[$s];
        }
    }
    return $score;
}
sub reversecomp
{
    my $seq = shift(@_);
    $seq = reverse($seq);
    $seq =~ tr/ATCG/TAGC/;
    return $seq;
}
```