Spring 5-1-2015

Effective Antisense Design Using An Ensemble of Energetically Sub-Optimal Secondary mRNA Structures

Andrea M. DiVenere
University of Connecticut - Storrs, andrea.divenere@uconn.edu

A. Abstract

There is a current deficit of effective therapies against bacterial infection. Many strategies seek using small molecules to target the infectious pathogen. One approach involves direct manipulation of the pathogen at the RNA level. Messenger RNA (mRNA) is a genetic transcript that encodes the fundamental instruction for protein production. Inhibiting mRNA translation effectively prevents protein synthesis.

The therapeutic agent must physically access mRNA to effectively block its message from being read. A technique has arisen where a complementary nucleic acid binding strand, called antisense, is generated to impede protein synthesis. An issue in creating effective antisense is finding mRNA target sites for inhibition. This problem is largely due to fluctuating secondary structure blocking target sites. In fact, the kinetics of physical accession are suggested to be the rate-limiting factor and thus the inefficiency of antisense.

However, these secondary structure fluctuations are energetically inherent to the nucleic acid sequence and are predictable. A program, GenAVERT, has been developed to thermodynamically determine effective mRNA target sites. It identifies the ensemble of the most probable energetically suboptimal states and determines which regions are therefore most accessible. With proper identification, a successful, effective antisense can be synthesized.

This work has proven GenAVERT's capacity by constructing antisense to downregulate green fluorescent protein expression in *Escherichia coli*. The fluorescence reduction affirms this model's efficacy.

Supported by fundamental energetic principles, this method of producing an antisense sequence complement for the most accessible mRNA target region has the potential to greatly reduce the strife of pathogenic infection.

B. Introduction

There is a blatant deficit in the current research against bacterial infection. It is common practice to search for small molecules to target infectious pathogens. A successful candidate must meet many criteria. The molecule must be effective, inhibitory, specific, nontoxic, and consistent. Many potentially targeted pathogenic proteins have multiple, spatially and environmentally dependent, conformations. An effective molecule must be versatile enough to inhibit the various forms, but specific enough to not affect other pathways. Industrialized high thoroughput screening allows scanning of thousands of existing synthetic chemicals against an inhibitory criterion. Successful candidates are found 43% of the time [1]. Fewer than 20% of these initial candidates make it to market [2] due to reasons such as nonexclusive selectivity of the desired target or toxic PKPD profiles. Even when success is achieved, bacteria can evolve their action to overcome this inhibition. This strategy is wastefully inefficient in both time and money.

A more efficient method is to target at the level of messenger RNA (mRNA). mRNA encodes the fundamental instructions to protein production. Its properties are thermodynamically inherent and can therefore be readily predicted for effective targeting. Designing a successful therapeutic agent to impede its function will ensure no protein can be made. Antisense, a complementary nucleic acid binding strand, is capable of this inhibition.

The use of antisense RNA as a therapeutic agent is an emerging and promising field. Due to the specificity of nucleotide base pairing, the antisense RNA will only bind to its complimentary sequence, ridding concerns of nonexclusive selectivity [3]. The resulting duplex of RNA impedes ribosomal binding, prevents ribosomal migration, and signals for the internal degradation of the strands [4,5]. There has been large scale success in the clinical trials for several cancer antisense drugs [3].

The biggest hurdle in this field of research is of accessing the mRNA. mRNA molecules are known to exhibit secondary structure whereby internal hydrogen bonds cause folding of the strands. Though strong, these bonds are transient, and the strands are continually in a state of flux [6]. The driving internal energy gradient pushes for acquisition of the lowest free energy form. A thermodynamic analysis of the formed structures can find those with minimal free energy and those with slightly higher energies. The regions that change between these structures are the most volatile – and thereby most accessible – regions of the mRNA. Designing an antisense molecule to target this region will provide a therapeutic agent with maximal effectiveness and can provide the answer to bacterial inhibition.

C. Theory

The GenAVERT computational program has proven the validity of this antisense design hypothesis. When given genes with strong natural antisense, it was able to output sequence predictions greater than 77% similar to the natural sequences. 100% similarity was seen in the case of the *hokA* gene [7]. Only the sequence of the target gene is required. This simplicity is especially pragmatic with the growing antibiotic resistance crisis [8]. When current small molecules fail to work, another full round of screening is necessary to find a drug to inhibit the evolved infectious function. This antisense design proposal simplifies the work. All that would be required is a re-sequence which will then be thermodynamically evaluated and a new applicable output will be given. There is no wide-spread, potentially nonexclusive guessing to seek competitive functions.

Other existing antisense prediction packages require user inputs, influencing the output. SFold is one such program. The Soligo module predicts target accessibility to output predicted antisense oligonucleotides [9]. However, it requires a user to determine the antisense length. This

could produce a constrained output, the most accessible applicable to the designed system than overall. Without prior knowledge, recommendation is to use 50 base pairs. Optional filters are also recommended, allowing more user influence.

Downregulation of green fluorescent protein (GFP) provides an easily quantifiable way to confirm antisense's capability. The long half-life of this protein, however, poses a time burden on this measurement [10]. The unstable mutant GFP-LVA has a half-life of just 40 minutes [11]. Therefore, continued fluorescence beyond antisense induction can be attributed to ineffective antisense than pre-existing GFP.

It is hypothesized that upon induction, the most thermodynamically accessible antisense sequence, that is still long enough to maintain specificity, will produce the most dramatic reduction in fluorescence. GenAVERT will analyze the thermodynamics of the entire system and provide the most volatile sequence for antisense accessibility. Utilization of all recommendations for the software package SFold should provide an antisense sequence less accessible and therefore be capable of a lesser fluorescence reduction.

D. Materials and Methods

The model organism *Escherichia coli* was used for its easy manipulation. The DH5α strain was chosen for its high transformation efficiency and supply in lab.

Two unique plasmids were constructed and dually transformed into DH5α cells for protein and antisense expression. For compatibility in the same cell, these plasmids were constrained to have unique origins of replication, antibiotic selection, and inducers. Additionally, the lack of a native T7 RNA polymerase in DH5α cells limited plasmid selection to those without a T7 promoter.

The construction and optimization of the GFP-LVA plasmid took up the majority of the time and efforts of this project. Several considerations also had to be made for the protein expression plasmid pBAD18/Kan. The steps, both successes and failures, are listed in the **Supplemental Materials** along with recommendations for future endeavors. The simplified steps that led to the final antisense quantification results are listed here.

The GFP-LVA sequence was extracted from an iGEM supplied plasmid. Primers were constructed (**Appendix II.A.iii**) through Invitrogen's Primer Design and amplification was performed with a Vent^{R} driven polymerase chain reaction (PCR) (**Appendix I.B.i**). A pACYC184 plasmid was acquired from New England BioLabs ® inside *E. coli* K12 ER2420 cells. It was purified from these cells with a plasmid prep (**Appendix I.A.ii).** The pACYC184 and GFP-LVA insert were then double digested by HindIII and EagI-HF (**Appendix I.C.i**). The double digests were separated with gel electrophoresis and extracted (**Appendix I.A.iii**). The pieces were then ligated over-night (**Appendix I.E**) and chemically transformed (**Appendix I.D.i**) into blank DH5α cells. Sequencing confirmed identity of GFP-LVA inserted into the plasmid (**Appendix II.B.ii**).

The GFP-LVA sequence was analyzed both with GenAVERT and S-Fold for antisense prediction.

GenAVERT required only an input of the mRNA sequence. It outputted the top 14 most volatile regions with their coding and antisense sequences and location on the gene. (**Appendix III.A.**) The results were screened for those above 35 base pairs to ensure sequence specificity. This criteria eliminate the most volatile region. The second and third antisense predicted sequences of 54 and 168 base pairs in length respectively were then chosen for construction.

S-Fold was accessed online and use was made of the Soligo module. The sequence was inputted. Assuming no prior knowledge, the preferred length of antisense oligos was specified at 50 base pairs in accordance with common predictions. The fields of folding temperature and ionic conditions were provided by the program without means to change the values. The supplied inputs were 37C and 1M NaCl, no divalent ions respectively. The results were e-mailed in a zip-folder with many subfolders for different analyses. The oligoout.txt file contained information of target position, target sequencing, antisense oligo, GC content, oligo binding energy, and GGGG indicator. The 843 outputs had to be manually evaluated for the optimal sequence by which the oligo binding energy was lowest. This was an oligo spanning bases 621-674 with a binding energy of -30kcal/mol. Another submission was then made with the specification of 54 base pairs to compare to GenAVERT, but was not synthesized.

The candidate antisense sequences were then synthesized for construction of the antisense plasmid. The pBAD18/Kan plasmid was acquired from ATCC ® 83797. Primers were designed for each of the sequences and the pBAD18/Kan plasmid (**Appendix II.A.v-viii)**. An NcoI cut-site was inserted at the start of the gene and a ScaI cut-site downstream to allow proper insertion. PCR was performed with pACYC184 template for the antisense sequences and pBAD18/Kan template for the linearized plasmid (**Appendix I.B.ii-iv**). Individually, the inserts and plasmid were double digested with NcoI and ScaI (**Appendix I.C.iii**), ran on a gel and extracted, ligated overnight, and chemically transformed into DH5α cells containing the pACYC184-GFP-LVA plasmid (**Appendix I.D-F)**. The transformation samples were plated with double antibiotic selection of kanamycin and chloramphenicol. The plates were then incubated overnight at 37°C. Colonies indicated successful transformation. Freezer stocks were immediately made.

GFP expression and downregulation were quantified with a Qubit® 2.0 Fluorometer. A Raw Mode program was written for fluorometric measurement at the emission wavelength of the GFP-LVA protein. (**Appendix I.F.i**). Plasmid amplification with 300ug/mL of spectinomycin elevated the GFP fluorescence levels. The induction of pBAD18/Kan for antisense expression was optimized with arabinose. Data collection was then performed on separate days for each antisense system tested. The detailed experimental protocol, which can be found in **Appendix I.F.ii,** is summarized here. Blank DH5α cells, DH5α cells with only the pACYC184-GFP-LVA plasmid, and DH5α cells with both the pACYC184-GFP-LVA plasmid and pBAD18/Kan-αs were inoculated in the morning of Day 1. Ten hours later, the cells were subcultured in 1:50 volume ratio into fresh LB media with appropriate antibiotic selection in 250 mL flasks. Two and a half hours after subculture, spectinomycin was added to a final concentration of 300ug/mL. These cultures were grown overnight. Initial culture samples were extracted at twelve hours after subculture. 2% w/v dry arabinose was then added to induce antisense expression of the experimental cultures. Arabinose was not added to three antisense-containing flasks to serve as a control. The fluorescence was then measured and normalized by optical density over a time span of 4.5 hours.

E. Results

The predicted antisense sequences were unique for the two different prediction software systems. **Figure 1** shows the location on the GFP-LVA sequence that each prediction matches. Even when supplied with the bias of antisense length as predicted by GenAVERT, SFold still predicted a 54 base pair region apart from GenAVERT's. That SFold's 50 base pair region was within the 54 region, the unbiased assumption of 50 base pairs was used for antisense design.

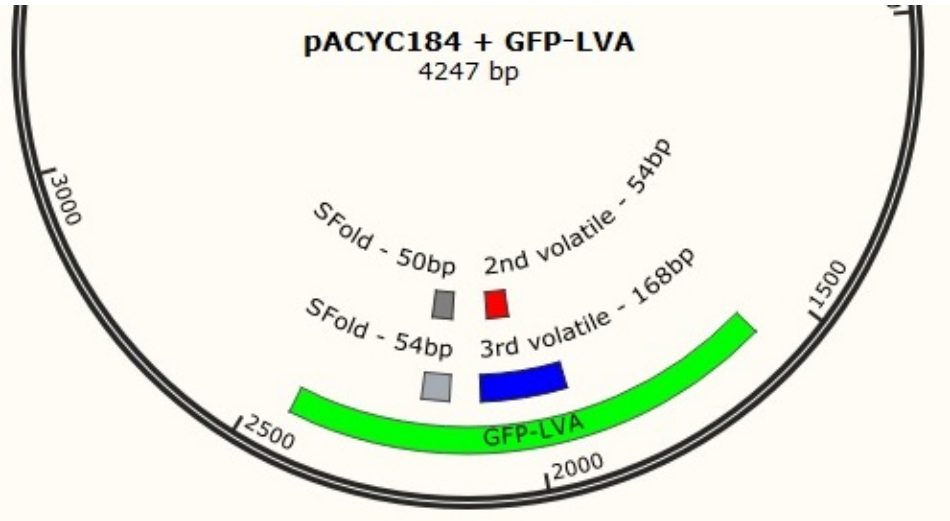

Figure 1. The Target Locations of the Predicted Antisense Sequences. The regions in red and blue are the regions predicted by GenAVERT as second and third most volatile, respectively. The gray are both predictions by SFold. The 50 base pair region was used for antisense construction as its length would be used without prior knowledge of bondage region.

GenAVERT's first (not shown here) is within the second most volatile region which too is a subset of the third most volatile region. This suggests strong volatility in this region and affirms the consistency of GenAVERT's computational solution.

Spectinomycin was necessary for amplification of the pACYC184 plasmid as chloramphenicol amplification was made impossible by the plasmid's resistance gene. The results from the two most report protocols (50ug/mL and 300ug/mL) are shown in **Figure 2.** Both cultures that received 300ug/mL (shown as circles) had the greatest increase in normalized fluorescence above the blank cell's autofluorescence. This concentration was therefore used for subsequent fluorescent quantifications.

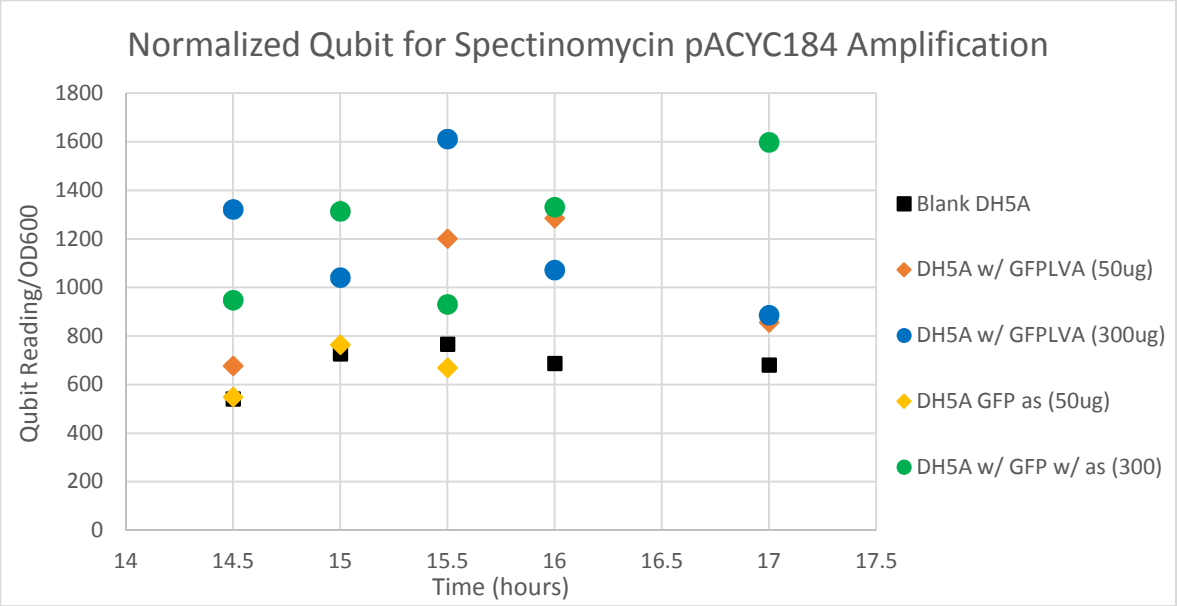

Figure 2. Results from Spectinomycin Amplification Protocol Development. As shown here, 300ug/mL was the most effective concentration for fluorescence amplification and was used in subsequent experiments.

Each antisense sequence was tested in triplicate on an individual day. Conditions of subculture time, duration of overnight growth, concentration of spectinomycin for plasmid amplification, temperature, beaker size, agitation speed, and concentration of arabinose for antisense induction were kept constant for each day. The cells were subcultured at time zero. Two and a half hours later, spectinomycin was added. Samples were extracted at 12 hours after sub-culture. Arabinose was added after extraction. Optical density and fluorescence measurement continued for hours 12-16.5. The normalized fluorescence was averaged over the triplicates of induced and uninduced antisense. The results can be seen in **Figures 3, 4, and 5.** The statistical variation of each of the samples is shown with error bars.

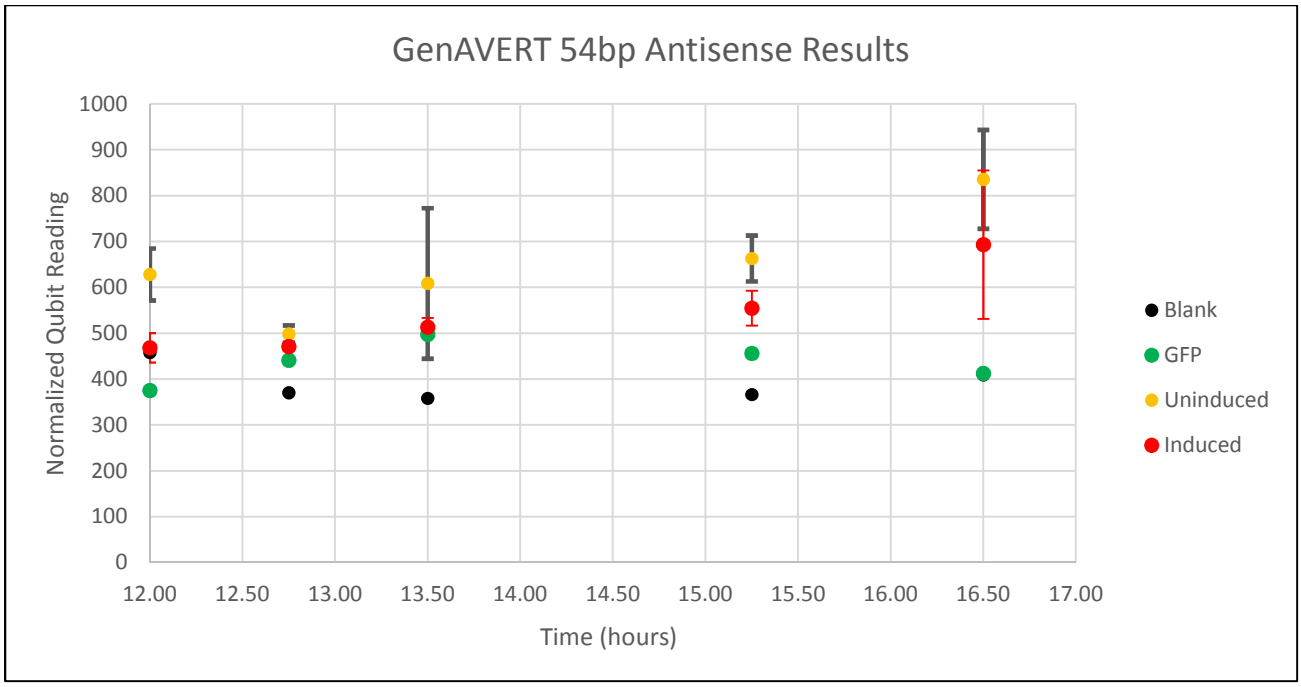

Figure 3. The Normalized Fluorescence Results of the GenAVERT 54 Base Pair Antisense Sequence. The uninduced and induced values here are average values of the triplicate samples with error bars depicting the standard deviation. Antisense was induced immediately following the samples being taken at Time = 12 hours.

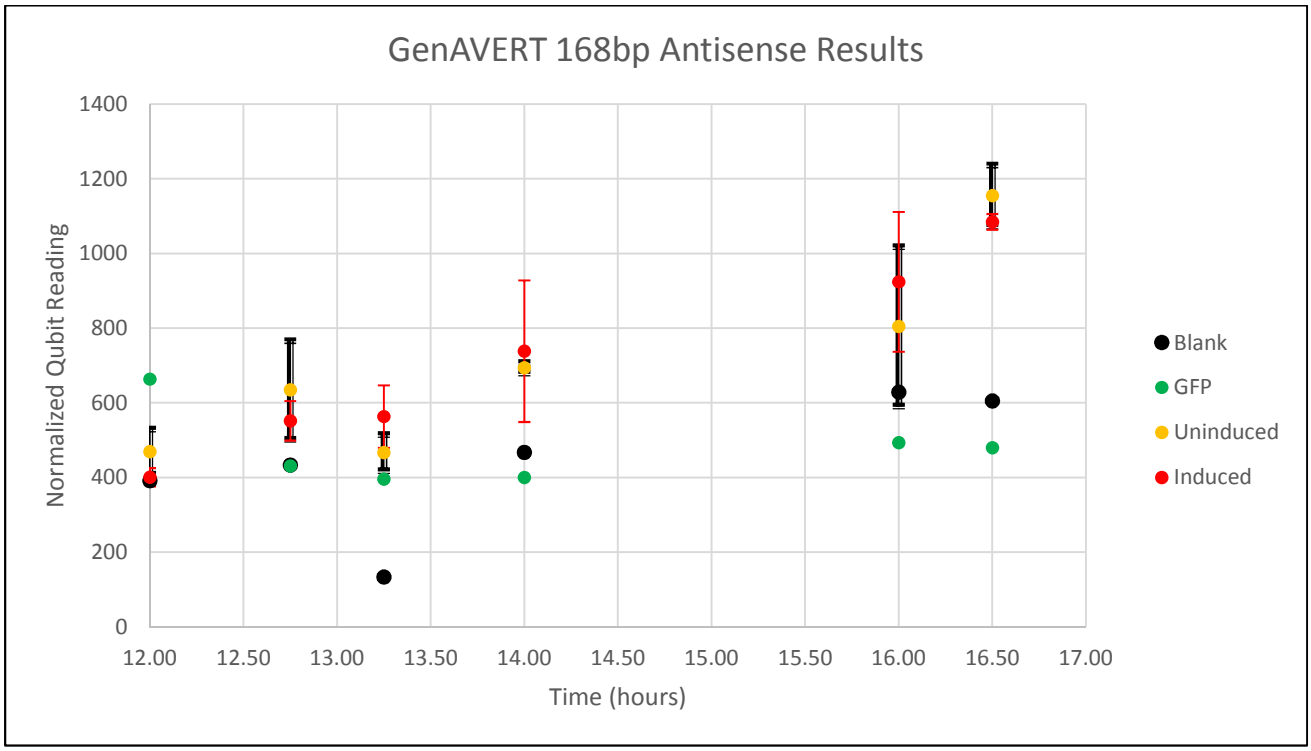

Figure 4. The Normalized Fluorescence Results of the GenAVERT 168 Base Pair Antisense Sequence. The uninduced and induced values here are average values of the triplicate samples with error bars depicting the standard deviation. Antisense was induced immediately following the samples being taken at Time = 12 hours.

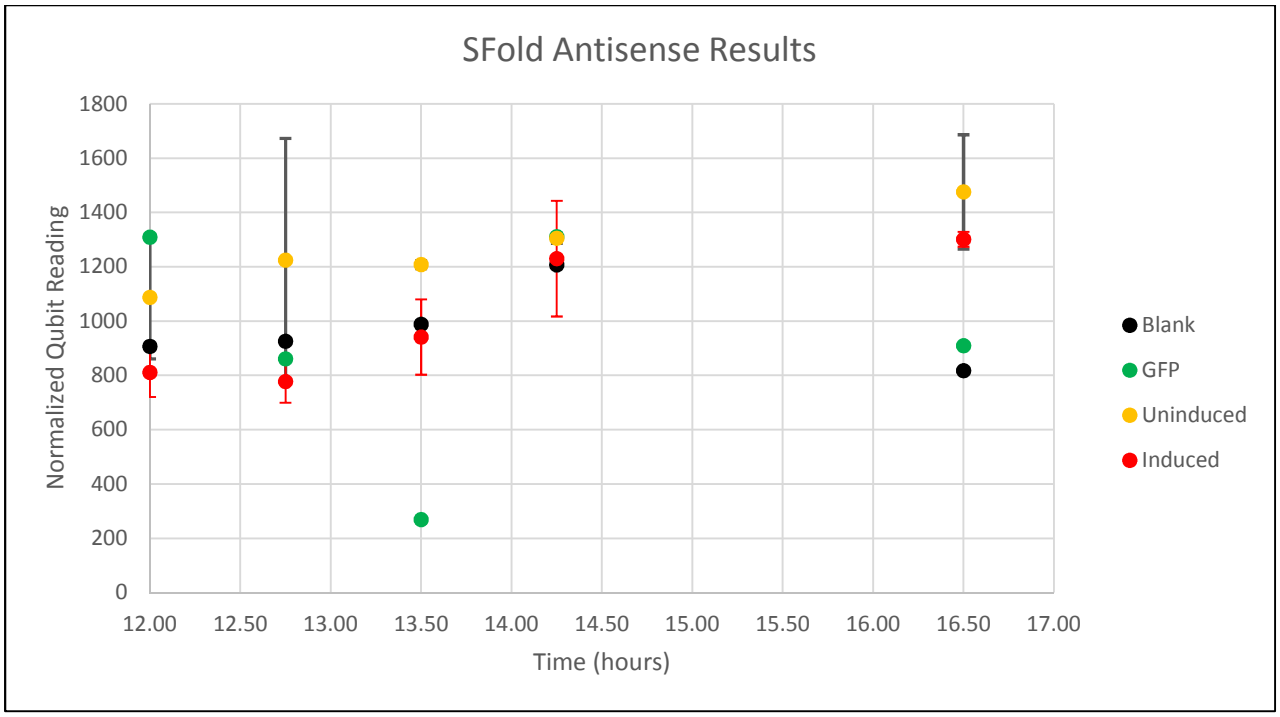

Figure 5. The Normalized Fluorescence Results of the S-Fold Generated Antisense Sequence. The uninduced and induced values here are average values of the triplicate samples with error bars depicting the standard deviation. Antisense was induced immediately following the samples being taken at Time = 12 hours.

General observations can be made from the graphical depictions here. The antisense containing cells exhibit increased fluorescence over that of the blank DH5α and DH5α cells with just pACYC184-GFP-LVA. The data of GenAVERT's 54 base pair antisense sequence and SFold's prediction show the induced pBAD18/Kan-as containing cells have decreased normalized fluorescence from that of the uninduced antisense-containing control. This trend is also seen in GenAVERT's 168 base pair antisense system, but only at the latest time point. The decreased fluorescence suggests the antisense may be the cause of reduced protein expression.

A comparative summary of the three systems can be seen in **Figure 6.** Here, the normalized triplicate measurements of uninduced controls were divided by the values of the induced triplicate samples to provide a ratio indicative of fluorescence reduction. The time points of two and four and a half hours were chosen for assurance of complete antisense induction and

degradation of the existing GFP-LVA proteins in the cells.

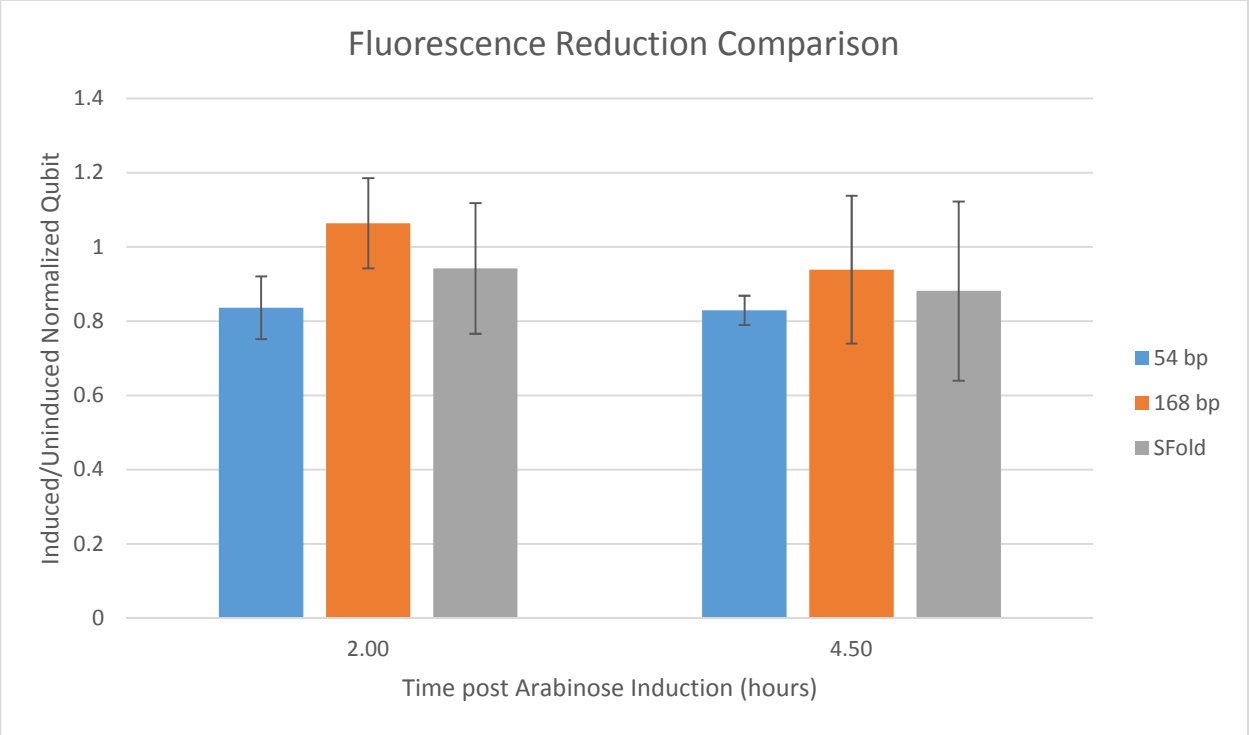

Figure 6. A Comparison of the Fluorescence Reduction Achieved for the Three Antisense Sequence Systems

The later time point exhibits a greater reduction for all systems than the two hour time point. This could indicate antisense induction takes longer than two hours for complete induction or that the GFP-LVA mutant has a longer stability than the 74 minute half-life claimed by Andersen et. al. and the iGEM supplier. In both time points evaluated, the 54 base pair sequence provided by GenAVERT was able to achieve a greater degree of fluorescence reduction than the other two systems. SFold's prediction consistently achieves reduction to a lesser degree. The 168 base pair sequence of GenAVERT provides less definitive trends. The error bars should be considered in observation of the trends.

F. Discussion and Conclusions

The uninduced antisense-containing cells theoretically should have had the same normalized fluorescence as that of the DH5α cells with only pACYC184-GFP-LVA. **Figures 3, 4, and 5** show a consistent trend where the fluorescence of the cells with the pBAD18/Kan

plasmid fluoresce to a greater degree than those without. This could indicate that the existence of two plasmids maximizes spectinomycin amplification or else there is another intracellular connection. Numerical quantification of plasmid content could further reason this observation.

The general increase in fluorescence of all the cultures over time could also be a factor of the spectinomycin action. The spectinomycin was added at time 2.5 hours from the time 0 of subculture. By the later time points here, the cells have been under protein synthesis arrest for over 12 hours. As the plasmid synthesis will continue right up until cell death, the cells sampled later will have a greater plasmid content. The normalization of fluorescence takes into account cell number. Comparison across the time points could only be made with the assumption that cell number is directly related to plasmid number. This may not be the case. Therefore, the data must be interpreted at individual time points. Comparison can only be made with the other cultures at the same time point, not across the collection time.

The cells without antisense fluoresced very close to the baseline cells that only had the pACYC184-GFP-LVA plasmid. This could still be a factor of the low pACYC184 copy number or could be from a low expression promoter. Future studies should work with a greater discernment between the blank and GFP cells. Many low copy number plasmids have been amplified with growth for 18-30 hours [12]. Growing the cultures longer than the 16.5 hours here could present two problems. It is only recommended to grow cells under spectinomycin amplification for 12-17 hours. Longer durations will cause cell death from inability to make their essential protein machinery. Furthermore, the selective antibiotics in the samples will begin to degrade. The addition of more antibiotics to goose the system could be detrimental to the weakened and slow growing cells. There are many factors involved in this system that make direct manipulation infeasible. Rather, two options are proposed. The promoter strength could be

engineered or a plasmid with both higher expression and a greater copy number could be used. These options will be discussed in detail in **Section G. Future Work**.

With the discrepancy between uninduced and GFP-only containing cells, the induced antisense results will be further discussed in comparison to the uninduced as a baseline.

The comparison shown in **Figure 6** indicates that the GenAVERT 54 base pair antisense sequence is the most effective for fluorescence reduction. This is likely due to the targeted region being more volatile and therefore accessible to the other regions. That the shorter antisense sequence predicted by SFold achieved a lesser reduction indicates that size is less of a factor than volatility in antisense effectiveness. Even though a smaller strand may have less of a spatial chance of impedance, accessibility is essential. That the third most volatile region GenAVERT predicted was less capable of down-regulating fluorescence indicates that there may be an upper threshold of effective antisense size. At 168 base pairs, the strand has more chance to be impeded by other components in the cell or even by itself. Secondary structures and antisense folding should also be considered in the design criteria.

The conclusions deduced here from **Figure 6** should be made carefully. The samples were taken at 12 hours immediately prior to the addition of arabinose. The three uninduced and three induced samples were at this point identical in composition and treatment. The results should have been the same. The addition of the arabinose to the induced cells should have been the cause for affected fluorescence. That they were different before its addition implies another factor of variability is present in this system.

The variability is also seen from the large error bars. The trends depicted by the average data point may not be truly representative of the factors investigated but of other variables

present. This further emphasizes the need for a new system before hard conclusions are made of the antisense effectiveness.

G. Future Work

Working to improve the existing system provides, at best, only a chance for greater discernment. Quantification of the plasmid concentration could provide insight to the mechanism of spectinomycin action. It would also provide a normalized fluorescence by plasmid count which would lift the incorrect assumption imposed by normalization by cell count that all cells have the same number of plasmids. However, the low copy number of this plasmid introduces many challenges. The amplification methods then add variables to the system making it difficult to infer the direct effects of antisense.

It would instead be recommended to use a different plasmid entirely. The compatibility constraints that necessitated using pACYC184 could be lifted with a different *E. coli* strain. The capacity of a T7 RNA polymerase would allow expression of many more plasmids. Many well-known strains have this gene, denoted with DE3. The notable two for recommendation would be BL21(DE3) or JM109(DE3). This option would bring another strain into the lab, but it would provide a chance to acquire more sound results. Extracting the GFP-LVA gene from the pACYC184 into a higher expression vector would allow use of the same antisense prediction results. The pBAD18/Kan plasmids already constructed could then be used.

Ideally, the GFP-LVA gene would be placed under a comparable promoter as the antisense. Should one promoter be significantly stronger, the mRNA of that gene will be in a much greater quantity than the other. This could over or underestimate the effectiveness. A very effective, but low copy antisense would appear the same as a minimally effective, but high copy antisense sequence. This should be taken into consideration with construction of a new GFP-LVA expression plasmid.

Once the possibility for greater discernment is made, the different antisense sequences can be more accurately compared. After the initial comparison of uninduced to induced as performed here, the kinetics of antisense migration could be investigated. The change in fluorescence after induction is dependent upon antisense production, physical migration to the target mRNA, access and inhibition of target mRNA translation, and degradation of existing GFP in the cell. The many steps will cause a time lag and add in chances for submaximal repression. A maximally effective antisense sequence would effectively stop all GFP protein synthesis. This can be artificially imposed by addition of rifampcin, the protein synthesis inhibitor. Addition of this drug will be comparable to immediate mRNA accession and inhibition of new GFP production. The degradation rate of GFP can then be determined from the kinetic profile of reduced fluorescence. Comparison of this complete inhibition to the antisense imposed reduction will describe the limitations inherent in antisense.

This phenomenon becomes especially relevant in the downstream application of antisense therapeutics. A maximally effective antisense sequence by design may appear ineffective if the mRNAs are spatially separated from each other. A means of efficiently delivering the antisense to the target is necessary. This work surpasses the scope of this project, but is something that should be considered in the study of antisense design.

H. Acknowledgements

This work was made possible by the help of several people. Carol Norris was influential in protocol development and FACS analysis. Susan Soucy and Leah Winterberger aided in placing and receiving all of the ordered supplies for this project. The temperature sensitivities of many of the reagents, and the storm delays, made their help indispensable. Stephen Johnson and Ian Rogers helped with the computation and graphical representations of GenAVERT. Their help permitted both the antisense design of this project and the delivery of a presentation at the

National AIChE conference on this work. The guidance, protocols, and mentorship from Eddy Bautista were incredibly influential in the development of my passion for research. I would not be as confident or have as many laboratory skills without her. Erik Johnson was always responsive to all of the questions that arose with this project. He helped work through many of the difficulties he had experienced with antisense development, but was even more willing and gracious to provide feedback and suggestions to my proposed new systems for this work. The experience he was able to provide saved enormous amounts of time and gave me a deeper understanding of many biological components. Christine Endicott was just a text or a knock of the door away. She provided me with the reference to protocols and computational resources that made carrying out this project possible. She would provide feedback and truly helped define the directions of this project. My graduate school experience will be enhanced by what she's helped me learn. Finally, this entire project could not have been performed without Dr. Ranjan Srivastava. He introduced me to the world of research by accepting my request to join his lab. Without this opportunity my future would be drastically different. His encouragement to take on my own project has truly shaped my learning experience at the University of Connecticut. I have gained laboratory skills, but more importantly the critical thinking of true scientific research. There were many hiccups in this project that experience could have prevented, yet he still gave me the opportunity to tackle this feat. In doing so, I can confidently say my experimental insights have significantly matured. I will take this experience with me into graduate school at the Univeristy of Texas at Austin as I pursue my Ph.D. in Chemical Engineering and beyond to all my future scientific research endeavors.

I. Supplemental Materials

i. Construction of GFP-LVA-Containing Plasmid

The GFP-LVA construct was obtained from iGEM Part BBa_K145280 from the iGEM08_KULeuven group on a pSB1A2 backbone. This backbone was ampicillin resistant and the GFP-LVA gene was under a tetracycline promoter that could be turned on with addition of anhydrotetracycline. The plasmid was shipped in New England Biolabs® (NEB) 10β *E. coli* cells. Once received, the cells were grown under ampicillin selection and the amplified plasmid was prepped from cells using a QIAprep Spin Miniprep Kit (**Appendix I.A.).** Sequencing primers were designed (**1a** and **1b** as listed in **Appendix II**). Submission to sequencing confirmed the GFP-LVA sequence both by the Anderson reference and by the iGEM annotation (**Appendix III**). It was determined the sequence was the short-lived mutant desired. This plasmid stock was then transformed into blank DH5α cells. The original attempt of polyethylene glycol induced competency proved unsuccessful, so a calcium chloride method was used as listed in **Appendix I.E.i.**

Baseline fluorescence experiments were then performed. Tetracycline analogs provide a more tolerable means of induction as they do not have the toxic side effects as tetracycline itself. The derivative anhydrotetracycline actually has been found to have a much greater binding constant to this iGEM Part than tetracycline while also offering a much lower antibiotic activity [13]. To ensure analog comparability and optimize effectiveness, two induction protocols were tested. As recommended by the KULeuven group 140 ng/ul anhydrotetracycline [14] and an attempt of a tolerable dose of 50ng/ul tetracycline. Allowing time for protein production after induction, the cells and controls were taken to fluorescence activated cell sorting (FACS) 2.5 hours after addition of the respective inducers. There was no noted fluorescence increase in the GFP-LVA containing cells than the blank cells (**Appendix V.A.**). After reviewing the literature,

it was found that the Part's Ribosomal Binding Site was weak, 0.3 instead of 1.0 (Wang et. al). Though Wang et. al had produced a part with a stronger RBS, a more reputable plasmid source was sought.

The incapability of having a T7 promoter was again an issue as the majority of bacterial expression vectors use one [15]. The new plasmid also had to be compatible with pBAD18/Kan, which would later house the antisense sequence. The vector pACYC184 was chosen as it had dual chloramphenicol and tetracycline resistance, a p15A origin of replication, and was constitutively expressed, eliminating the need to optimize induction protocol. This plasmid was purchased from New England BioLabs® Inc. as product carried by *E. coli* K12 ER2420 cells. The plasmid was prepped in accordance with the protocol described above.

ii. Primer Construction

The first attempt of sequence insertion was to build EcoRI and NcoI cut sites into primers surrounding the GFP-LVA gene for insertion into the chloramphenicol gene on pACYC184. The A Plasmid Editor (ApE) software was used to analyze both the plasmid and insert. Assurance was made that these enzyme recognition sites were unique and thermodynamic evaluation of the insert ends found possible annealing regions. Primers were designed with the criteria that they have 40-60% G/C content, greater than 53C T_m, and an annealing region of 18-22 base pairs. With amplification by polymerase chain reaction (PCR), double digestion, and gel extraction this piece could then be inserted into the doubly digested pACYC184 backbone (See map – **Appendix IV**). This would disrupt the chloramphenicol resistance while still allowing selection with tetracycline. Many iterations of PCR amplification, ligation, and transformation were attempted. Several sources on NEB were consulted to develop a PCR protocol based on the DNA polymerase, length of the template, and primers to be used. All of the DNA polymerases in stock were tried and several chemical transformations all failed. A third optimization of

electroporation achieved successful transformation. However, the sequence had leader amino acids from the chloramphenicol reading frame and thus would not make a functional GFP-LVA protein. With lack of plasmid annotation, the exact start of the reading frame was uncertain. To ensure proper insertion, decision was made to insert directly onto the tetracycline gene's ATG start site. This necessitated copying all nucleic acids upstream to ensure encompassing the promoter's -10 and -35 sites.

The nearest upstream cut site was HindIII, found within the -10 region. Inclusion of extra nucleotides, the cut site, the nucleic acids between the -10 region and the gene start site, and enough to anneal to the sequence resulted in a primer of 92 base pairs (**Appendix II.A.iii**). This imposed the risk of secondary structures and primer dimers in the operation of PCR. The proposed primer sequence was submitted to the Integrated DNA Technologies OligoAnalyzer 3.1. The most energetically favorable structures were provided with their ΔG and T_m values as seen in **Figure 7**.

Structures

Structure Name	Image	ΔG(kcal.mole^{-1})	T_m (°C)	ΔH(kcal.mole^{-1})	ΔS(cal.K^{-1}mole^{-1})	Output	
1		-5.09	39.9	-107.1	-342.13	Ct	Det
2		-4.89	40.2	-100.5	-320.68	Ct	Det
3		-4.86	39.9	-101.9	-325.49	Ct	Det
4		-4.53	39.5	-97.6	-312.17	Ct	Det
5		-4.33	47.8	-61.1	-190.39	Ct	Det
6		-4.3	42	-79.9	-253.56	Ct	Det
7		-4.27	41.7	-80.5	-255.66	Ct	Det
8		-3.96	35.5	-116.2	-376.45	Ct	Det
9		-3.82	34.3	-125.8	-409.12	Ct	Det
10		-3.67	38.7	-83.6	-268.08	Ct	Det
11		-3.67	38.6	-83.9	-269.11	Ct	Det

Figure 7. The Integrated DNA Technologies OligoAnalyzer 3.1 Output of the Long Forward Primer Sequence

The highest melting temperature was found to be 47.8C with the 5th most energetically favorable structure. The annealing region of this primer had a T_m of 58C. Therefore, it was determined that operation of the melting of the PCR at 60C should be high enough to inhibit the secondary structure hairpin from forming.

A compatible and unique downstream enzyme was then sought with ApE and restriction enzyme evaluation. EagI was found to fulfill both criteria. The high fidelity version was purchased as this enzyme was not already in stock in the lab. A primer with a similar melting temperature to the forward HindIII was then designed (**Appendix II.A.iii**) and both were purchased from Invitrogen.

iii. pACYC184-GFP-LVA Plasmid Assembly

The GFP-LVA insert was amplified with the DNA polymerase Vent^R acquired from New England Biolabs. The NEB T_m calculator was used to develop a PCR protocol with the primers to be used (**Appendix II.B.i**.). The PCR products were run on a gel and extracted (**Appendix I.A.iii.**). Double digestion and subsequent heat inactivation of the enzymes followed (**Appendix I.C.ii**). The digested pieces were then ran on a gel for extraction. Overnight ligations were run with varying insert-to-backbone ratios. The previously optimized chemical transformation successfully yielded 1 colony on chloramphenicol plates. This cell line was then grown overnight for freezer stock. A plasmid prep was submitted to the Sequencing Center for confirmation, but failed to be read. A consultation with a technician led development of a hypothesis that this was due to very low plasmid yields of 2.1ng/uL and the lengthy forward primer. It was also warned that the first 50bp at the end of the primer will not be confident and the maximum read length is 1100 base pairs. A new forward sequencing primer was designed over 60 base pairs upstream of the GFP-LVA start site and to be compatible with the reverse primer taken from cloning.

(**Appendix.II.A.iv**). The sequence was compared on ApE to the known GFP-LVA sequence and insert identity was confirmed with 993 matches and 2 mismatches. (**Appendix II.B.ii**)

iv. Antisense Determination and Synthesis

The GFP-LVA sequence provided by iGEM rather than the one acquired from the Sequencing Center was used for input to both software programs. This was done to begin the development and ordering of primers quickly. The Sequencing Center was closed for storms and so its result was delayed. It should be noted that the two mismatched nucleotides could skew thermodynamic analysis results, however the very close comparison made this concern negligible. GenAVERT and SFold both required inputs of the nucleotide sequence for analysis.

GenAVERT very specifically outputted the antisense candidates in order of decreasing volatility. The most volatile region was a 13 base pair region. With concerns of non-specificity a threshold of 35 base pairs was imposed. The second and third most volatile regions fulfilled this criteria at 54 and 168 base pairs respectively. It was also noted that the regions were all subsets of each other, indicating that this is a very volatile region.

In addition to the sequence, SFold required other user inputs. Of most concern was the need to specify sequence length. Most antisense prediction methods search for 50 base pairs and this is a recommendation by the program itself. Therefore, 50 was chosen to assume no prior knowledge. An analysis of 54 base pairs was performed for direct theoretical comparison to GenAVERT, but this output was not synthesized.

v. Antisense Plasmid Construction

The pBAD18/Kan plasmid has a multiple cloning site designed for protein production. Insertion into this region would produce leading mRNA transcript for ribosomal binding and initiation. The accessibility of the antisense would be impaired by these excess nucleotides. Therefore, the plasmid had to be manipulated to eliminate this region. It was decided to linearize

the rest of the plasmid backbone and create sticky ends for attachment. This required finding restriction enzymes that did not have recognition sites on the rest of the plasmid or in any of the antisense sequences as they too would need the same sticky ends. NcoI and ScaI were chosen and the primers listed in **Appendix II.A.viii** were designed. The amplification required extreme fidelity and an increased extension time. The VentR DNA polymerase's high proof-reading capability met the first need. Optimizations of the PCR protocol were necessary as the VentR extension time of 1min/kb template was not sufficient. An even longer extension time led to success (**Appendix I.B.iv**.).

The inserts were extracted off the pACYC184-GFP-LVA plasmid with the primers listed in **Appendix II.A.** by the PCR protocols in **Appendix I.B.** Double digestion of the plasmid and the inserts with NcoI and ScaI created sticky ends (**Appendix I.C.**) for insertion. The different antisense sequences were individually ligated into the pBAD18/Kan plasmid (**Appendix I.D.**) and transformed into DH5α cells that contained the pACYC184-GFP-LVA plasmid. Successful transformants were those that could grow on both chloramphenicol and kanamycin. Colonies were grown up overnight and freezer stocks were made.

vi. Fluorescence Measurement

The purchase of Qubit ® 2.0 Fluorometer allowed in-house fluorescence quantification. FACS was still desired as an end quantification tool, but the Qubit was used for its accessibility, no cost, and ability to collect a time series fluorescence profile. Its Raw mode feature measures raw fluorescence values. A program was written as a .txt file using the template and directions provided on the Life Technologies website [16]. The Raw mode options of excitation and emission were set ranges rather than being able to specify a wavelength. The optimized values found for the GFP-LVA part of 475nm excitation and 515nm emission had to be matched to the choices. The program was then written for excitation of "Blue" light and emission of "Green"

light, 470nm with a filter of 430-495nm and 510-580nm respectively. No calibration was chosen to collect raw fluorescence data which eliminated need to fill out many of the other fields. The finished .txt (**Appendix V.B.**) was then changed to a .qbt file and imported onto the Qubit USB.

Blank DH5α cells, cells with pACYC184-GFP-LVA, and cells with pACYC184-GFP-LVA and a pBAD18/Kan-αs plasmid were then grown for fluorescence measurement. The blank cells were used to provide a fluorescence control and the antisense-containing cells were used without induction of the antisense gene to ensure no leaky antisense expression. Cells were grown overnight and subcultured in the morning. Samples were taken, measured for OD600, pelleted, doubly washed with PBS, and resuspended in PBS for Qubit analysis. The raw fluorescence reading was normalized by the OD600. **Figure 8** shows the fluorescence of the GFP containing cells was insignificant over the autofluorescence of the blank cells. Optimization was then necessary.

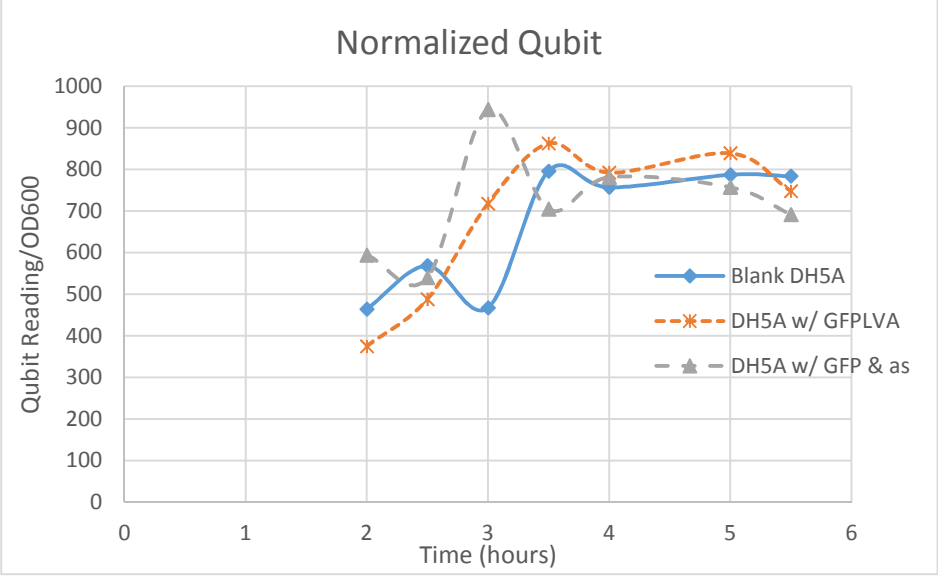

Figure 8. Normalized Initial Fluorescence Data. These results showed that the GFP was not being expressed to a high enough degree to be significant above that of autofluoresence. This was thought to be due to both the low copy number of the plasmid and the constitutive nature of the promoter. Further optimizations were subsequently performed.

vii. GFP Expression Optimization

The low plasmid count of pACYC184 was already known to be a limitation when performing plasmid preps for Sequencing and PCR DNA templates. It was hypothesized that this was the cause of the low GFP-expression. Chloramphenicol amplification, a common protocol, was not possible with this system as the pACYC184 plasmid has a chloramphenicol resistance gene (**Appendix**). Rather, spectinomycin was recommended for this system. Spectinomycin acts by binding to the 30S ribosomal subunit which inhibits protein synthesis and therefore slows the growth of cells [17]. The pACYC184 plasmid however, which uses RNA for the initiation of plasmid replication, can still be replicated. A greater plasmid yield per cell is achieved than under physiological conditions. Sources stated that this amplification could be achieved with concentrations of spectinomycin in the range of 50-300ug/mL [18]. The cells need to be growing robustly before addition of the spectinomycin. Upon addition, the cells must grow long enough for pACYC184 to be amplified at the reduced growth conditions. After extended growth, the cells will die from inability to synthesize essential protein machinery. Three optimizations were then necessary: the time before spectinomycin addition sufficient for robustness, the concentration for sufficient, yet nonlethal amplification, and the time after addition to achieve plasmid amplification without extending into death phase. The times were found by trial and failure. The successful condition then permitted investigation of spectinomycin concentration.

Spectinomycin was received as a free sample of product S0692-1ML from the Sigma Aldrich sales representative Jeremy Lehmann. This product was a ready-made solution of 100mg/mL spectinomycin in a 1:1 DMSO/H_2O solution and was kept in the -20C freezer at all times when not in immediate use.

Freezer stabs were grown for ten hours at 37°C, 250rpm in 10mL of media with appropriate antibiotic selection. At hour ten, the cells were subcultured with 500uL inoculation

into 24.5mL of fresh LB media with antibiotics in 250mL flasks. The agitation was increased to 300rpm for proper aeration and the temperature was left as 37°C. The cultures were allowed to grow for 2.5 hours before addition of either 50ug/mL or 300ug/mL spectinomycin. The cells were then grown overnight before measurement of optical density and fluorescence. **Figure 9** shows the normalized fluorescence results of these samples. The 300ug/mL concentration was able to achieve the increase in fluorescence necessary for these experiments and was used for all subsequent experiments. It should be noted the increased fluorescence was not seen when a larger volume of the spectinomycin stock Catalog Number NC9844390 was received from Fisher Scientific. This stock differed in concentration and was dissolved in pure water. Therefore, it would freeze in the -20C. The shipment was delayed by a snow storm and therefore the solution had a prolonged thaw. Still cold to the touch, it was kept. The ineffectiveness of this solution may have come from any of the discrepancies addressed here. Rather than risk incorporating another variable, this product was discarded and more of the Sigma Aldrich product were purchased.

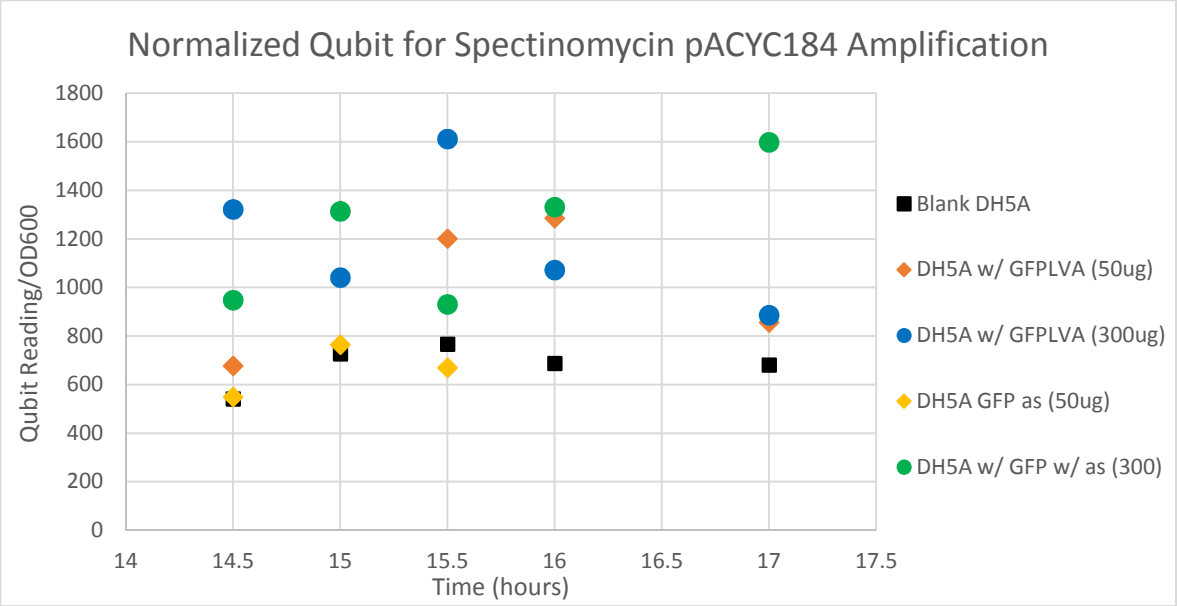

Figure 9. The Normalized Qubit Reading of GFP Fluorescence with Spectinomycin Plasmid Amplification. The spectinomycin in the specified concentrations was added at time = 2.5 hours post subculture. The higher concentration (300ug/mL) was capable of achieving higher GFP expression and was chosen for subsequent experimental protocol.

vii. Antisense Induction

The reputability and wide-spread use of pBAD/18-Kan made optimization of this system simple. Maximum induction is achieved two hours after addition of 2% w/v arabinose to a culture [19]. The small culture volume made dissolution of a concentration stock volume challenging. Instead, dry 99% (L)-arabinose, product A11921 from Alfa Aesar, was added to the non-control flasks 12 hours after subculture. Optical density and Qubit fluorescence measurements were then carried out for collection of the experimental results as shown in the **Results Section**.

viii. Concluding Remarks

This Thesis was an in-depth learning experience. The largest amount of time was spent in trouble shooting the plasmid construction for more optimal protein expression. To achieve the desired results, this project should be redeveloped in another expression system. Use should be

made of an *E. coli* strain that has a T7 polymerase so that a higher expression plasmid can be used.

Though the results here were suboptimal, this work had truly provided a deeper understanding of cloning, genetic manipulations, and most importantly research strategies. This Thesis was a success in two ways: it found that this system would not work for future studies and it provided an undergraduate student the perseverance and knowledge to carry into graduate school. This University Scholar Thesis and the lessons learned will have lasting effects in the Ph.D. Thesis to come.

J. References

[1] Zaman, G. "Targeting RNA: new opportunities to address drugless targets." *Drug Discovery Today* 8, no. 7 (April 2003): 297-306.

[2] DiMasi JA. Risks in new drug development: approval success rates for investigational drugs. *Clinical Pharmacology & Therapeutics* 2001, 69: 297-307.

[3] Bennett, C. Frank, and Eric E. Swayze. "RNA Targeting Therapeutics: Molecular Mechanisms of Antisense Oligonucleotides as a Therapeutic Platform." *Annual Review of Pharmacology and Toxicology* 50, no. 1 (February 2010): 259-93.

[4] Wanger EG. Kill the messenger: Bacterial antisense RNA promoters mRNA decay. *Nat Struct Mol Biol*, 2009, 16(8): 804-806.

[5] Darfeuille,F., Unoson,C., Vogel,J. and Wagner,E.G.H. (2007) An antisense RNA inhibits translation by competing with standby ribosomes. *Mol Cell*, 26, 381-392.

[6] Mahen EM, Watson PY, Cottrell JW, Fedor MJ (2010) mRNA Secondary Structures Fold Sequentially But Exchange Rapidly In Vivo. *PLoS Biol* 8(2): e1000307.

[7] Johnson, E., and R. Srivastava. "Volatility in mRNA secondary structure as a design

principle for antisense."*Nucleic Acids Research* 41, no. 3 (February 2013): e43-e43.

[8] Silver LL, Bostian KA (1993) Discovery and development of new antibiotics: the problem of antibiotic resistance. *Antimicrob Agents Chemother* 37: 377–383.

[9] Ding, Y., Chan, C.Y. and Lawrence, C.E. (2004) Sfold web server for statistical folding and rational design of nucleic acids. Nucleic Acids Res. 32 Web Server issue, W135-W141.

[10] Tombolini R, Unge A, Davey M E, de Bruijn F J, Jansson J K. Flow cytometric and microscopic analysis of GFP-tagged Pseudomonas fluorescens bacteria. *FEMS Microbiol Ecol.* 1997;22:17–28.

[11] Andersen, Jens Bo, Claus Sternberg, Lars Kongsbak Poulsen, Sara Petersen Bjorn, Michael Givskov, and Soren Molin. "New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria." *Applied and Environmental Microbiology* 64, no. 6 (June 1998): 2240-46.

[12] Addgene.org,. Protocol - Inoculating a Liquid Bacterial Culture https://www.addgene.org/plasmid-protocols/inoculate-bacterial-culture/.

[13] Antiquity, G. (2003). *Part:BBa_R0040:Experience*. *iGEM*.

[14] Wang, S., Yen, Y., Fan, R., & Yu, M. (2012). *Part:BBa K750111*. *Registry of Standard Biological Parts*. Retrieved from http://parts.igem.org/Part:BBa_K750111.

[15] HelmholtzZentrum Munchen,. (2012). *Protein Expression and Purification Facility (PEPF): Bacterial expression vectors*. Retrieved from http://www.helmholtz-muenchen.de/en/pepf/materials/vector-database/bacterial-expression-vectors/index.html.

[16] Lifetechnologies.com. 2015. MyQubit Custom Assays on the Qubit® 2.0 Fluorometer.

[17] Benedik, M. Amplification of Plasmids with Chloramphenicol Resistance. *Research Gate*, 2014.

[18] *Plasmid Amplification*; 1st ed.; Applied Biological Materials Inc.: Richmond, BC, CANADA, 2015.

[19] Guzman L. M., Belin D., Carson M. J., Beckwith J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* 177, 4121–4130.