

**UNDERSTANDING THE ROLE OF NON-PROTEIN-CODING
RIBONUCLEIC ACIDS IN PROKARYOTES.**

by

Manuel Rafael Jiménez Díaz

A thesis submitted to the Faculty of the University of Delaware in
partial fulfillment of the requirements for the degree of Honors Bachelor of
Chemical Engineering with Distinction.

Spring 2010

© 2010 Manuel Rafael Jiménez Díaz
All Rights Reserved

**UNDERSTANDING THE ROLE OF NON-PROTEIN-CODING
RIBONUCLEIC ACIDS IN PROKARYOTES.**

by

Manuel Rafael Jiménez Díaz

Approved:

Eleftherios T. Papoutsakis, Ph.D.
Professor in charge of thesis on behalf of the Advisory
Committee

Approved:

Anne Skaja Robinson, Ph.D.
Committee member from the Department of Chemical
Engineering

Approved:

Susan E. Groh, Ph.D.
Committee member from the Board of Senior Thesis Readers

Approved:

Alan Fox, Ph.D.
Director, University Honors Program

ACKNOWLEDGMENTS

Many people helped along this project to ensure its proper development and outcome. I would like to thank Shawn Jones for his continued guidance and supervision in my thesis and previous research. Shawn was always available and always willing to listen, teach me, and guide me in the correct direction, and for this I will always be very grateful. I would also like to thank Dr. Eleftherios PApoutsakis for his continued interest in my development as a researcher and for allowing me the great opportunity of performing research alongside his very talented group even during challenging times. Thanks to the entire Papoutsakis laboratory for helping me along the way in resolving many encountered technical difficulties. In particular I would like to thank Sergios Nicolaou, Stefan Gaida, and Dr. Bryan Tracy for their invaluable advise and training.

Great thanks to my mentor and committee member Dr. Anne Robinson for her continued interest in my personal success throughout my career but especially during the last two semesters of my career. Also, thanks to my other committee member Dr. Susan Groh for her guidance and enthusiasm during the project.

Finally I would like to thank Dr. Douglas Buttrey and Dr. Maria Palacas for their constant motivation and advice.

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF TABLES	vii
ABSTRACT.....	viii

Chapter1

1 INTRODUCTION	1
1.1 Non-Protein-Coding RNA	1
1.2 Intergenic 16S	3
1.3 Signal Recognition Particle.....	4
1.4 Hypothesis.....	12
2 VECTOR CONSTRUCTION.....	13
2.1 Insert construction via PCR	13
2.1.1 Theory and methods.....	13
2.1.2 Results.....	17
2.2 Selection of Vector.	20
2.2.1 Theory and Methods	20
2.2.2 Results.....	21
2.3 Selection of Promoter and Terminator.....	22
2.3.1 Theory and Methods	22
2.3.2 Results.....	23
2.4 Cloning.....	25
2.4.1 Theory and Methods	25
2.4.2 Results.....	27
3 CHARACTERIZATION	33
3.1 Selection of Media and Strain.....	33
3.1.1 Theory and Methods	33
3.1.2 Results.....	34

3.2	Growth curves and qRT-PCR	36
3.2.1	Theory and Methods	36
3.2.2	Results.....	39
4	PERFORMANCE	42
4.1	Acid Stresses	42
4.1.1	Theory and Methods	42
4.1.2	Results.....	43
5	CONCLUSION	48
	REFERENCES	50
	APENDIX	54
A.1	Supplemental Protocols	54
A.2	Sequencing.....	59

LIST OF FIGURES

Figure 1	mRNA regulation via ncRNAs.....	2
Figure 2	Simplified Mechanism of the SRP protein translocation	7
Figure 3	Variant1 of Ffh structure as determined by Keel, et al.....	9
Figure 4	Cartoon representation of the M-Domain.....	10
Figure 5	Structure of the <i>ffs</i>	11
Figure 6	PCR Overhangs	16
Figure 7	Different attempts for PCR amplification	18
Figure 8	PCR amplification	Error! Bookmark not defined.
Figure 9	Restriction enzyme map of pACYC184 and pUC19.....	Error! Bookmark not defined.
Figure 10	Predicted structure of the <i>trp</i> mRNA terminator	24
Figure 11	pACYC184- <i>ffs</i> construct.....	28
Figure 12	S pACYC184-int16S construct.....	29
Figure 13	pUC19-Ffh transinfection.....	30
Figure 14	pACYC184- <i>ffs</i> -Ffh transinfection	31
Figure 15	Minimal media comparison. Growth after 48hrs	35
Figure 16	TaqMan illustration	38
Figure 17	Growth curves in LB media	39
Figure 18	Typical qRT-PCR outcome from the Bio-Rad iQ5 cycler.....	41
Figure 19	12 hr Acid stress in LB 5ml tubes	44
Figure 20	24hr Acid stress in LB 5ml tubes	44
Figure 21	Acetic acid stress after 12 hours in Succinate minimal media ...	45
Figure 22	Lactic acid stress after 12 hours in Succinate minimal media....	46
Figure 23	Butyric acid stress after 12 hours in Succinate minimal media..	46

LIST OF TABLES

Table 1	Summary of primers utilized for PCR amplification.....	15
Table 2	Summary of attempted constructs.....	16
Table 3	Legend explanation of the acid stress figures	43

ABSTRACT

Previous acid stress studies in *Clostridium acetobutylicum* and *Streptococcus mutans* were applied in *Escherichia coli* to further understand the mechanisms of acid tolerance and the effect of non-protein-coding RNA (ncRNA) in prokaryotes. The strain overexpressing the 4.5S ncRNA, as well as the intergenic upstream 16S ncRNA sequences with a low copy plasmid, pACYC184, in *E. coli* Top10 F cells. These cells were grown to stationary phase under the following acid stresses: lactic acid, acetic acid, and butyric acid. The cells were grown in a minimal medium containing succinate as the main carbon source and in LB medium. Optical density was measured at various time points to assess the growth of the strains. In most cases the data indicates that the aforementioned RNA sequences did not provide acid tolerance to *E. coli*. It was evident that the Ffh protein when overexpressed in high copy numbers was lethal to the *E. coli* strains. A similar phenomenon was observed when the intergenic 16S and the *ffs* sequences with natural promoters and terminators were transformed into *E. coli*. The Ffh protein was also intended to be cloned in the lower copy number plasmid pACYC184 but similar patterns of cell toxicity were also observed. The quantitative reverse transcription PCR assays was not sensitive enough to detect the short ncRNAs thus it is uncertain whether the constructs containing the *lac* promoter overexpressed the ncRNAs.

CHAPTER 1

INTRODUCTION

1.1 Non-Protein-Coding RNA

Prokaryotic cells are very tightly regulated chemical systems which permits them to live in various environments at different times. These cells have a multitude of regulatory systems extending from protein allosteric regulation to complex signaling transcription factors. The central dogma of molecular biology states that RNA serves primarily as an informational translator between a coded section of DNA—a gene or an operon—and the constructed protein. This simple model proposes a mechanism where messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA) would interact among each other to effectively decode DNA into the correct sequence of amino acids that would ultimately conform an active protein with specific function (Lodish, 2007). So far it has been observed that this is mainly true for prokaryotes as their tightly packed genome is constituted mainly by protein coding sequences which encompass a relative larger percent of the genomic output when compared to eukaryotes (Mattick, 2006). In contrast, higher eukaryotes can have up to 95% of its transcriptional output composed of non-protein-coding RNAs (ncRNAs), as is the case in humans (Mattic, 2003). Until recently were most known ncRNAs viewed as merely assistants in the translation of mRNA or as debris from inefficient translational machinery. In recent studies it has been demonstrated that these ncRNAs are involved in the

control of chromosome architecture, mRNA splicing, and generally in roles involving highly specific recognition of nucleic acids (Mattick, 2003; Lau, 2001; Eddy, 2001). This paper focuses on one ncRNA which aids in the recognition of nucleic acids and another which could repress the catalytic function of an rRNA. Currently less than 100 ncRNAs have been confirmed in *Escherichia coli* and about 200 novel candidates have been predicted computationally (Rivas, 2001; Saetrom, 2005; Kawano, 2005).

There are different ways in which ncRNAs can repress or activate translation of mRNA as well as repress or activate the catalytic activity of other ncRNAs. It is known that ncRNAs in *E. coli* can destabilize mRNA making it more prone to endonuclease activity, they can sequester the mRNA thus preventing the ribosome from translating it; they can bind to a 5'UTR antisense mRNA strand inherently blocking the attached translated sense strand from the ribonuclease and thus preventing degradation, and that they can stabilize mRNA by base pairing to the 3' end preventing degradation from 3'-exonucleases (Stroz, 2004; Lewin, 2006). Figure 1 illustrates some of these possibilities.

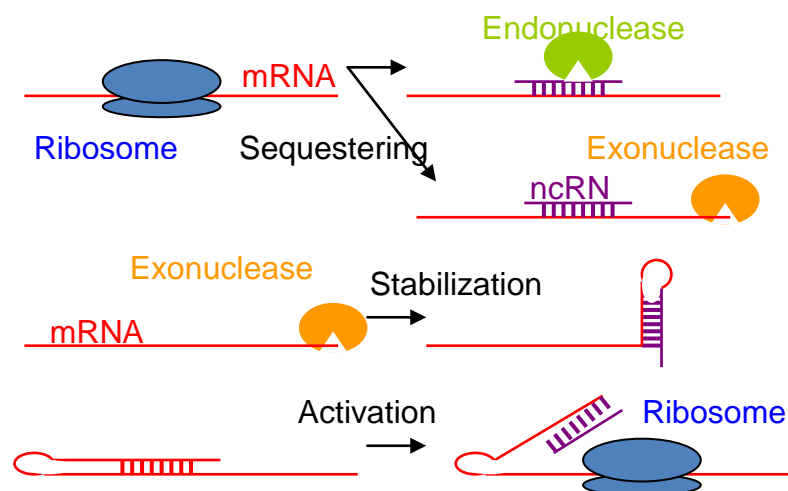


Figure 1. mRNA regulation via ncRNAs (Kawano, 2005).

1.2 Intergenic 16S

The sources of ncRNAs are many. In *E. coli*, ncRNAs have been traced back to several, intergenic regions of different genes, some of which lay at the antisense strand of the overlapping gene (Kawano, 2005). Most ncRNA systems studied in *E. coli* have been *trans*-encoded, which means that the antisense RNA strand is encoded at a different site than where the sense strand of the interacting gene is located (Kawano, 2005; Stroz, 2004). Most of the *cis*-encoding genes found in *E. coli* were found on repeat units such as the rRNA, tRNA, and the LDR regions (Kawano, 2005). In prokaryotes, genes coding for rRNA are one of the major DNA repeats; for example, *E. coli* has 7 repeats of the 16S rRNA (Lewin, 2006).

Recent discovery of non-coding RNA (ncRNA) in bacteria has shown their importance as catalytic controls for many different reactions and metabolic pathways, making their elucidation of great interest due to the high amount of enabled biotechnological tools (Urban, 2007). Previously it was found that the intergenic region upstream of the 16S rRNA sequence in *Clostridium acetobutylicum* conferred acid tolerance characteristics (Borden, 2010). A similar intergenic region upstream of the 16S rRNA sequence was found in *Escherichia coli* which could confer similar acid tolerance characteristics.

The selection of genes from a genomic library in *C. acetobutylicum* stressed with butyric acid found a variety of gene fragments (Borden, 2010). These fragments were then placed into plasmids with different permutations that were again tested for higher acid tolerances (Borden, 2010). The outcome of this second selection, after sequencing and identification found a 353 bp

sequence in the intergenic region upstream of the 16S and downstream of the 5S gave the highest tolerance (Borden, 2010). Computational analysis further showed that all of the intergenic regions upstream of the 16S rRNA sequences (11 in total) shared high homology.

This region has been computationally analyzed in other organisms such as *E. coli*. All bacteria analyzed have shown a similar pattern, that is the intergenic regions upstream of all 16S rRNA sequences show high homology within that organism's genus. However, there is little to no homology between bacteria of different genus, like *E. coli* and *C. acetobutylicum*. The intergenic region found in clostridia showed that it could potentially be transcribed into a ncRNA that regulates the expression of the 16S rRNA (Borden, 2010). The results of a quantitative reverse transcription polymerase chain reaction (Q-RT-PCR) on *E. coli* show that its intergenic region upstream of the 16S rRNA sequence, or the reverse complement of it, is transcribed into RNA. These two facts have lead to the belief that the mechanism endowing *C. acetobutylicum* with enhanced acid tolerance could possibly function in a very similar manner in many other prokaryotes such as *E. coli*. It is also known that most small interfering RNAs (siRNAs) are cis-encoding RNAs, and that trans-encoding RNAs are less common (Storz, 2004). This suggests that it is the opposite strand to the 16S that might inhibit its activity; however, a trans-encoding siRNA might be responsible for such regulation.

1.3 Signal Recognition Particle

The 4.5S or *ffs* as annotated in the Kyoto Encyclopedia of Genes and Genomics (KEGG) is another important ncRNA which is also related to acid stress tolerance. This ncRNA is part of the Signal Recognition Particle (SRP).

From a study done in *Streptococcus mutans*, acid tolerance is correlated to this particle's protein counterpart, Ffh. It was found that the Ffh protein's mRNA increases eightfold when *S. mutans* are exposed to low pH (Gutierrez, 1999). Kremer reports that the 4.5S RNA is an essential part of the signal recognition particle (SRP_{bact}) in *S. mutans* and that the SRP plays an important role in this organism's acid tolerance (Kremer, 2001). The effect of the SRP in acid tolerance in *E. coli* will also be analyzed in a similar fashion as the 16S upstream intergenic region.

The Ffh protein in *E. coli* is 48k-Da and homologous to the SRP54 subunit of the eukaryotic signal recognition particle and thus much knowledge can be gained from studying the prokaryotic system about the eukaryotic system (Park, 2002). In a study performed on temperature-sensitive *ffh* mutants it was shown that overexpression of the 4.5S or *ffs* ncRNA suppressed the deleterious effects of the dwindling protein (Park, 2002).

The signal recognition particle plays an important role in membrane protein synthesis in all organisms. In cellular biology, proteins, macromolecules responsible for the majority of cell activity, are synthesized from the information encoded in messenger RNA transcripts. The ribosome is the major piece of biological machinery responsible for assisting in 'translation', resulting in newly synthesized polypeptides. Proteins have functions within every compartment of the cell, and many need to be transported to their functional destination.

Membrane proteins are transported to the plasma membrane in prokaryotes and to the endoplasmic reticulum in eukaryotes by the SRP mechanism. Each membrane protein has a signal sequence on the N-terminus

end of the polypeptide chain that is rich in nonpolar amino acids, such as alanine or leucine (Abell, 2004). The SRP is a complex of protein and RNA. A hydrophobic pocket on the protein portion of the SRP binds to the signal sequence on the membrane protein. This hydrophobic region is comprised of mainly methionine residues. Because of the flexible nature of the M domain, the SRP is able to bind to membrane proteins of many different shapes and sizes. The protein portion of the SRP contains three distinctive domains: the N, G, and M domains. The N domain contains the N terminus and is closely linked to the G domain, also known as a GTPase domain, which chemically interacts with GTP (guanosine triphosphate), which energetically switches the protein's function to the "on position". The M domain contains the methionine rich sequence required for binding to membrane proteins.

The way that the SRP mechanism works is that it binds to the membrane protein signal sequence while it is still being synthesized in the ribosome complex. This interrupts the translation process in Eukaryotes; however, it has not been observed in prokaryotes (Keenan, 2001). The SRP then transports both the nascent polypeptide and the ribosome to its target membrane and finds the SRP receptor on the surface of the membrane. The SRP and its receptor bind, and the ribosome and nascent protein are attached to the surface of the membrane. In eukaryotes, the signaling sequence of the nascent protein is released at the same time that the ribosome and nascent protein complex (RNC) is transferred to a large protein conducting channel via GTP hydrolysis by the SRP-SRP receptor complex (Skatch, 2009). In *E. coli* GTP binding is required for Ffh to attach to the RNC, the hydrolysis of GTP

then releases Ffh from its receptor but hydrolysis is not necessary for the release of the SRP to the RNC after being bound to its Receptor (Keenan, 2001; Valent, 1998). Translation resumes, the SRP dissociates to find other membrane proteins to transport, and the newly synthesized protein is released into the membrane. Figure 2 shows a simplified model of the SRP mechanism.

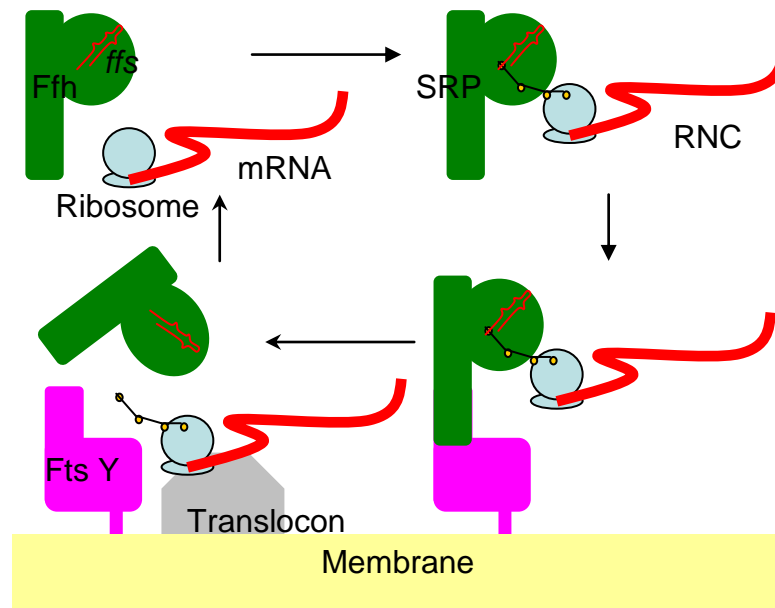


Figure 2. Simplified Mechanism of the SRP protein translocation

SRP structure differs between prokaryotes and eukaryotes. The prokaryotic SRP is made up of one protein, called Ffh, and one 4.5 S RNA, called Ffs. The eukaryotic SRP is made up of 6 proteins and one larger RNA. The SRP54 protein in eukaryotes is homologous to Ffh, and one of the domains of the eukaryotic SRP RNA is homologous to the 4.5 S RNA in prokaryotes. Therefore Ffh and *ffs* are evolutionarily conserved because they are present within every organism's cells. The fact that the SRP mechanism has been conserved since the first signs of life shows just how important it is to cell function.

In the area of biofuels and toxic waste biological remediation, as well as in any other field where organism are utilized for purposeful motives in harsh chemical environments, understanding this mechanism and how to make it more efficient is very important. The more efficient this mechanism is the better the membrane will be able to handle chemical attacks and maintain homeostasis.

The structure of Ffh was determined by X-ray crystallography by Keenan, et al and its findings are summarized in *The Signal Recognition Particle*. Figure 3 shows the structure of the Ffh as a stick cartoon and a ribbon cartoon. As stated earlier Ffh is comprised of three main domains: the M, the G, and the N domain (Keenan, 2001).



Figure 3. Variant1 of Ffh structure as determined by Keel, et al.

Signal sequences have two main requirements: hydrophobicity and α -helical content; without any of these two, the proteins have been shown not to enter the mechanism. These sequences bind to the M-domain was named as such due to its high methionine content which confers it a high hydrophobicity. The M-domain residues lie in bristles giving the domain a high level of

plasticity which allows the protein to bind to several non-specific signaling sequences (Keenan, 2001). This domain creates its hydrophobic pocket (M-domain groove) by placing three amphipathic α -helices in its surroundings. The “Finger Loop” connects the first two helices and has been determined to be very flexible and thus probably rearranges to adapt to different signaling sequences (Keenan, 2001). The M-Domain is also where the 4.5S ncRNA binds. It binds between α -M2 and α -M4 and a highly conserved motive including arginine, serine, and glycine residues is essential for high affinity (Keenan, 2001).

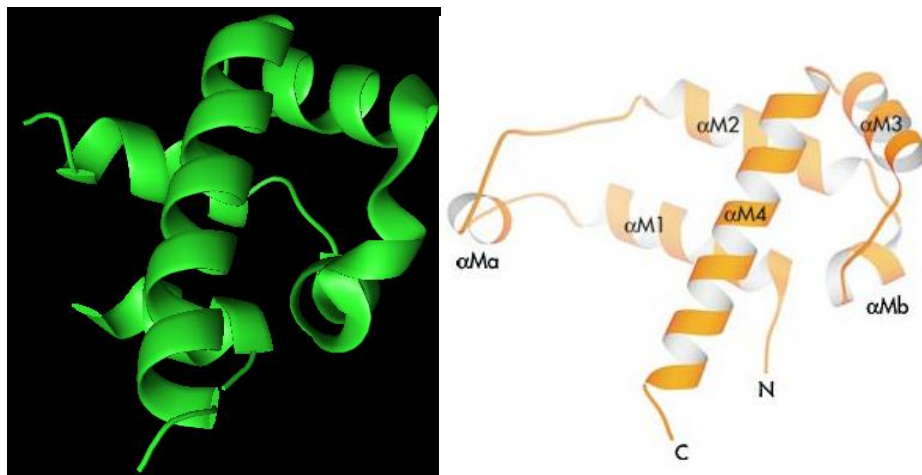


Figure 4. Left: cartoon representation of the M-domain from Keel Variant1.

Right: cartoon representation of the M-Domain from Keenan.

The binding site (called domain IV) of the ncRNA is about 50 nucleotides long. This domain contains an asymmetric and a symmetric loop (Keenan, 2001). The main function of the ncRNA has been suggested to interact with the Ribosome (Keenan, 2001). Figure 5 shows the structure and the two loops of the ffs. Another study showed no contact between the NG

domain and the ncRNA (Buskeiewicz, 2005). In its bound state the ncRNA's domain IV symmetric and asymmetric loops are no longer separated by the three base pairs and form a contiguous surface (Keenan, 2001). It is important to note that the stability of the SRP is directly tied to the concentration of *ffs* relative to the concentration of Ffh. The overproduction of Ffh makes the protein unstable while overproducing *ffs* at the same time stabilizes the protein (Jensen, 1994). The usual ratio of protein to ncRNA is 1 to 4 (Jensen, 1994).

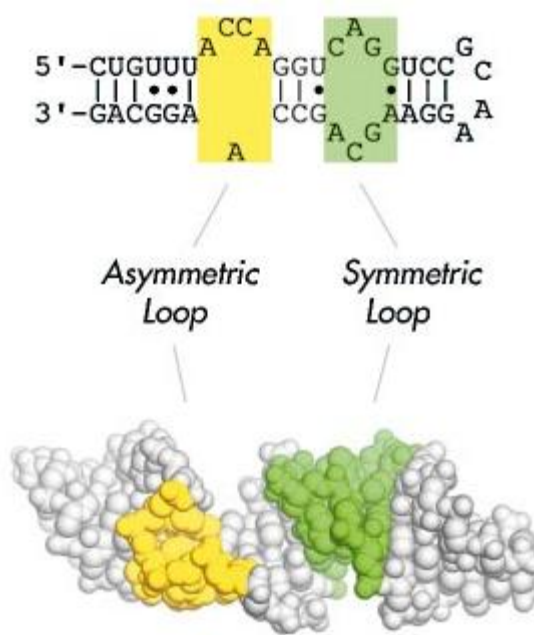


Figure 5. Structure of the *ffs*. (Keenan, 2001).

The N and G domains are closely bound together and are sometimes referred as a single subunit called the NG domain. The G domain confers the GTP binding of the molecule. Crystal structures of the Ffh show that the M domain and the NG domain are linked by a disordered chain suggesting the high mobility between these two main domains (Buskeiewicz, 2005).

1.4 Hypothesis

It is expected that by overexpressing the ncRNA *ffs*, the upstream intergenic 16S aforementioned sequence, and the Ffh protein by themselves or in unison in any combination, *E. coli* will have a higher tolerance to high carboxylic acid concentrations. This will be reflected by the faster growth over time while challenged with different carboxylic acid concentrations.

Chapter 2

VECTOR CONSTRUCTION

2.1 Insert construction via PCR

2.1.1 Theory and methods

To construct the DNA sequences that would later be inserted into a plasmid for transinfection of *E. coli*, Ampitaq Gold® Polymerase from Applied Biosystems was utilized in the amplification of the desired sequences. Every desired DNA sequence was present in the genome of *E. coli*; thus, the PCR reaction was able to be performed from genomic DNA. The genomic DNA was isolated utilizing the magnetic bead DNA isolation kit from InviMag®. The following is the recommended standard preparation of a 50 µl rxn

Amplitaq Polymerase PCR reaction mixture (for 50 µl rxn)

1. 5 µl of Buffer II
2. 5 µl of 25mM MgCl₂
3. 3.5 µl of 10mM each dNTP
4. Add appropriate amount of template (typically ~1 µl)
5. 0.5 µl of primer mix (1ug/ µl concentration)
6. 34.5 µl ddH₂O
7. 1 µl of Amplitaq Gold polymerase

The specific DNA sequence was amplified by nested PCR to acquire a more pure precursor DNA sequence. The final PCR amplification involved the

addition of overhangs which contained the sequence of a restriction enzyme for facilitated cloning. Figure 6 illustrates this procedure.

Some sequences needed to have an added promoter and terminator flanking the amplified sequence of interest. For these cases where the promoter and the terminator were not the naturally occurring ones, the overhangs included not only the necessary restriction enzyme overhang but also the necessary sequence for the correct construction in the correct orientation of the desired sequence. Table 1 lists the primers utilized in this process and Table 2 summarizes the attempted sequences through PCR amplification. Figure 6 serves as an illustration of the PCR amplification.

Table 1. Summary of primers utilized for PCR amplification.

ID	Sequence	PCR round	Enz.
ffs PCRI for	GCG GCG ATT ACC GTG AAT GAC AAA	1	-
ffs PCRI rev	CAA TAC GAA GCG CGC CAA ACT CTT	1	-
ffs NatII for	NNN NGT CGA CGA GTG AAG TCG CAT TGC GCA AGA A	2	Sall
ffs NatII rev	NNN NGT CGA CTT CAT CAG ACA TAC CTC CGA AGC G	2	Sall
ffs prom Sall	NNN NGT CGA CCC AGG CTT TAC ACT TTA TGC TTC CGG CTC GTA TAA TGT GTG GNG GGG GCT CTG TTG GTT CTC CCG CA	2	Sall
ffs term com Sall	NNN NGT CGA CAA AAT GCC GCC AGC GGA ACT GGC GTG TGG GAT TAG GGT GGG GGC CCT GCC GC TAC AT	2	Sall
Prom Sall rec	NNN NGT CGA CCC AGG CTT TAC ACT TTA TGC TTC C	recuperate	Sall
Term comp Sall rec	NNN NGT CGA CAA AAT GCC GCC AGC GGA ACT GGC G	recuperate	Sall
int16S NatI for	TGG TGC CGG GTT CAT ATT CAC CTT	1	-
int16S NatI rev	TGG TCT TGC CAG GTT ATG CGG TAT	1	-
int16S NatII for	NNG GAT CCT TCT GTG CAG CTA ACT GTT GTG CG	2	BamHI
int16S NatII rev	NNG GAT CCC AGC AAG CTG TTT CCT GTT ACC GT	2	BamHI
int16S desI for	AAC TCC CTA TAA TGC GCC ACC ACT	1	-
int16S desI rev	TGG TCT TGC GAC GTT ATG CGG TAT	1	-
int16S term com BamHI	NNG GAT CCA AAA TGC CGC CAG CGG AAC TGG CGG CCT GTG GGA TTA ACT GCT CTT TAA CAA TTT ATC AGA	2	BamHI
int16S prom BamHI	NNG GAT CCC CAG GCT TTA CAC TTT ATG CTT CCG GCT CGT ATA ATG TGTGGA AA AG TTT GAT GCT CAA AGA ATTA	2	BamHI
Ffh PCRI for	ATG TAG AAC AGG CCA GTG CAA AGC	1	-
Ffh PCRI rev	TTC ACA ATT CGT CAG CAG GTT CGC	1	-
Ffh NatII for	NNG GAT CCC GGG CAT AGT GAT GTC GAC AGT TT	2	BamHI
Ffh NatII rev	NNG GAT CCA ACA ACC TGG TAG AAC GGA CGC TT	2	BamHI
Ffh prom BamHI	NNG GAT CCC CAG GCT TTA CAG TTT ATG CCT CCG GCT CGT ATA ATG TGT GGA ATC AGC GCA AGA CTG ACG GAG TA	2	BamHI
Ffh term comp BamHI	NNG GAT CCA AA TGC CGC CAG CGG AAC TGG CGG CCT GTG GGA TTA ACA ACC TGG TAG ACC GGA CGC TTT	2	BamHI
Prom BamHI rec	NNG GAT CCC CAG GCT TTA CAG TTT ATG CCT CC	recuperate	BamHI
Term comp BamHI rec	NNG GAT CCA AA TGC CGC CAG CGG AAC TGG CG	recuperate	BamHI

Every primer was obtained from Integrated DNA Technologies and most primers were modeled through their website's software (IDT). It was imperative that some primers were designed entirely by hand as they had to end at the beginning of the ncRNA sequences. The melting temperature of these primers was assessed by Integrated DNA Technologies as well.

Table 2. Summary of attempted constructs

DNA sequence	promoter	terminator	restriction enzyme
Ffh	lac	<i>trp</i>	BamHI
Ffh	natural	natural	BamHI
<i>ffs</i>	lac	<i>trp</i>	Sall
<i>ffs</i>	natural	natural	Sall
int 16S	lac	<i>trp</i>	BamHI
int 16S	natural	natural	BamHI

The DNA sequences were extracted from Kyoto Encyclopedia of Genes and Genomes from the organisms data base of *E. coli* K12 MG1655 (KEGG).

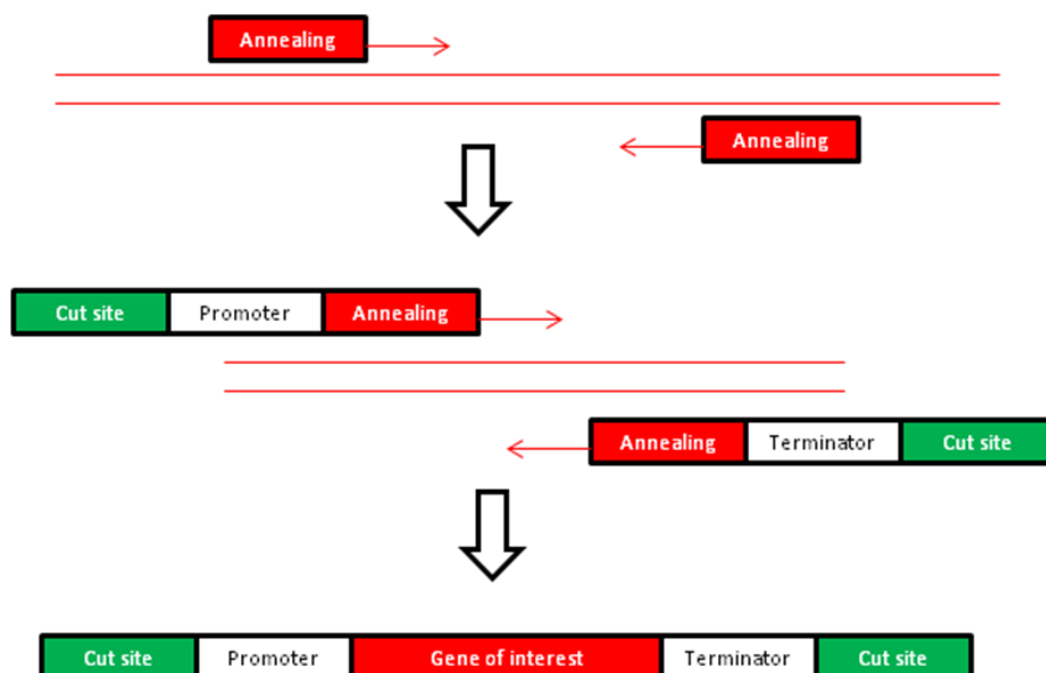


Figure 6. PCR Overhangs

2.1.2 Results

In PCR amplification several components determine the annealing of the primers to the DNA sequence at a set temperature: the concentration of the primers, the concentration of the salts in the buffer, the sequence itself, and the concentration of the DNA template are all very important variables that need to be kept fine tuned for an effective amplification. Unfortunately, most programs available as well as the programs utilized lack the capacity to account for overhangs and their interaction with the environment as well as for possible primer-primer interaction as the one described in (Ahsen, 2001). Noting that some of these overhangs are almost 100bp long one can expect the predicted temperature not to be accurate.

To overcome this problem, several trials were performed where the different variables were modified until a clean product was achieved. As the product is also dependant on the capabilities of the enzyme one is also bound to the extent at which one could vary the parameters without tampering with the polymerase capability to elongate DNA. The main tested variable was temperature and even though salt concentration was also modified, the temperature alterations proved to be sufficient to produce every product listed in Table 2 except for the very first one.

The construction of *ffh* with the *lac* promoter and the *trp* terminator was not achieved. As seen in Figure 7 the product achieved was consistently much larger than its expected value. Several PCR runs were implemented at various temperatures but no fruitful results were seen.



Figure 7. Different attempts for PCR amplification

Finally, two separate preceding PCR attempts were performed where the primer with the promoter overhang was coupled with the reverse primer with no overhangs in one reaction and the forward primer with no overhangs was coupled with the reverse primer containing the overhang of the complementary sequence of the *trp* terminator. The products of these two preceding PCRs were of the expected size and were produced stringently. These products were then utilized in a further second PCR reaction but still the product was yet again too large. This led to the belief that the primers with overhangs were interacting between each other in an undesired manner.

Figure 8 shows one of the several runs to obtain the correct clean inserts. It can be seen that the *ffh* on the right has a faint product that is almost twice the size of the expected size of 1.5kb as it lies above the 1,900 bp mark. The ladder tagged as lambda is the lambda DNA cut with restriction enzyme BstEII.

Figure 8 shows a much clearer image when most of the PCR of the other sequences was finally mastered.

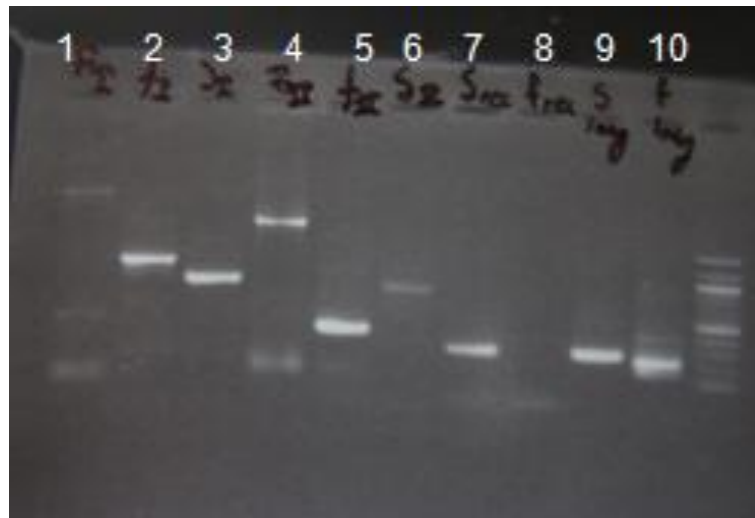


Figure 8. PCR amplification. from lanes 4 to 10: Ffh protein natural promoter and terminator, *ffs* natural promoter and terminator, *int16S* natural promoter and terminator, the recovery of *int16S* engineered promoter and terminator, an unsuccessful trial, *int16S* engineered promoter and terminator, *ffs* engineered promoter and terminator, and the 100bp ladder. Lanes 1-3 are previous nested PCRs to obtain the sequence of the pure DNA sequences without any promoter or terminator of *ffh*, *ffs*, and the intergenic 16S, respectively.

2.2 Selection of Vector.

2.2.1 Theory and Methods

As developed by Bailey in Biochemical Engineering Fundamentals, the concentration of a protein and mRNA can be modeled by the following equation:

$$\frac{d(P)}{dt} = k_q \xi [mRNA] - k_e [P] - \mu [P] \quad \text{Eq. 1}$$

$$\frac{d[mRNA]}{dt} = k_p \eta [G] - k_d [mRNA] - \mu [mRNA] \quad \text{Eq. 2}$$

In simple terms it relates the concentration of a protein or mRNA directly to the growth rate, the concentration of that element and the concentration of the element from which it originated. Here G is the concentration of the gene, K_c and K_d are the rate constant of degradation, K_p and K_q are the rate constants of generation, μ is the specific growth rate, P is the concentration of the protein of interest, mRNA is the concentration of the messenger RNA, and η and ξ account for efficiencies of production (Lee, 1984).

Translating this model to the rate at which the concentration of ncRNAs change would be simple and would have a similar form where the rate of generation is dependant on the concentration of the genome.

That means that the higher the number of copies of a single gene in the cell the faster the change in concentration of the mRNA, protein, and ncRNA assuming that the efficiency of production is maintained equal, which might not always be the same especially at very high concentrations of the producing element.

As a matter of fact, at very high concentrations of a single gene the cell will become more and more inefficient. The copy number of a plasmid is directly related to the number of copies which a cell is capable of holding, and it is related to the origin of replication of the plasmid (Lewin, 2006). The growth rate monotonically decreases as the copy number of a vector expressing foreign protein increases (Bentley, 1990).

Understanding that the two ncRNAs and the protein dealt will have a potential for unbearable metabolic burdens, it was decided to try a high copy number as well as a low copy number and to investigate the effects of inducible promoters.

2.2.2 Results

Figure 9 presents the two vectors chosen. pACYC184 has a low copy number and a resistance to chloramphenicol and tetracycline which made it very attractive for the possible coexpression of *Ffh* and any of the ncRNAs but specially *ffs*. pUC19 has a high copy number but is probably one of the most versatile vectors available due to its capacity to rapidly detect insertion and due to its multi-cloning loci.

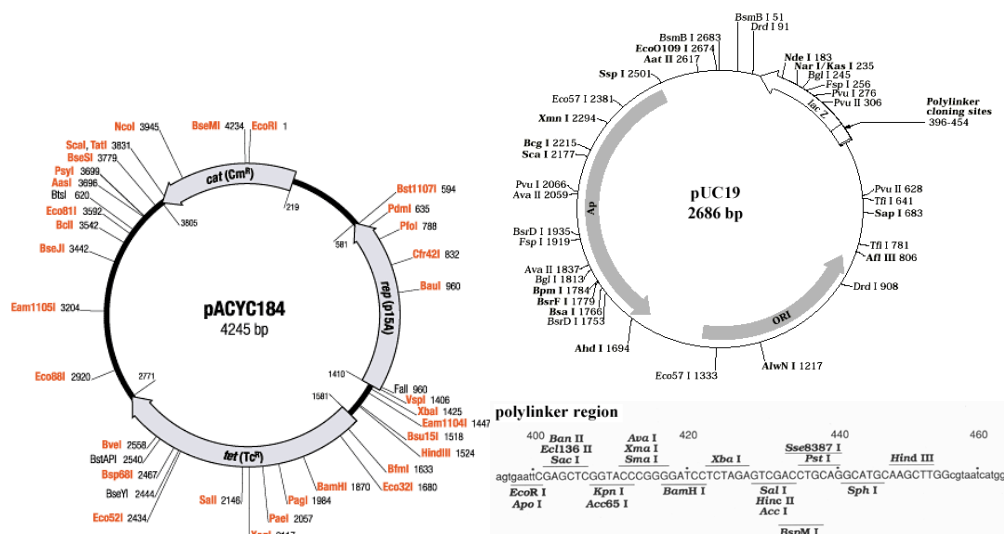


Figure 9. Restriction enzyme map of pACYC184 and pUC19

Several cloning sites were identified in these two plasmids that allow for the interexchange of inserts between them. In particular, the BamH1 and Sal1 restriction enzyme sites are suitable for the insertion of the different constructs in both of these plasmids. The sites in pACYC184 were selected to deactivate the capability of the enzymes to resist tetracycline. This was done with the aim that selection of different related strains would be able to be done by excluding the produced strains via addition of tetracycline. In this report such possibility did not arise as the strains created were limited in number; however, the potential for discrimination is still there.

2.3 Selection of Promoter and Terminator

2.3.1 Theory and Methods

It has been observed that several detected intergenic putative natural ncRNAs have potential rho-independent natural terminators as well as (Kawano, 2005) Potential σ^{70} -dependent promoters (Kawano, 2005). However,

most tRNAs have been observed to have rho-dependent terminators. Either terminator for the construction of ncRNA should suffice, but it is recommended to utilize a natural rho-independent terminator.

In the construction of the addition of the terminator and promoter it is important that both can function by immediately transcribing and immediately terminating the ncRNA. Any extra extension can result in the inhibition of the ncRNA.

The promoter region should have the necessary upstream sequences that have been proven to induce transcription at the correct places such as the -35 sequence and the TATA box about 20 base pairs upstream. The consensus sequence and nucleotide frequency can be found in (Lodish, 2007), which describes the following sequence as the most probable to function correctly: TATA(AorT)A(AorT)(AorG) .

For the constructs involving the over expression of the Ffh, it was important that the Shine-Dalgarno sequences are included in the construct for its accurate translation.

2.3.2 Results

Figure 10 shows the predicted structure of the *trp* terminator.

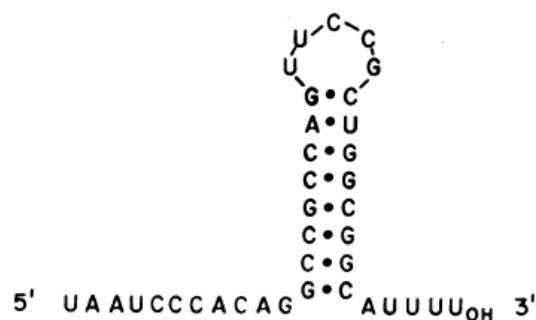


Figure 10. Predicted structure of the *trp* mRNA terminator (Wu, 1978).

The *trp* terminator has been observed to be rho-dependent; however, up to 40% of the usual terminator efficiency is expected to occur in the absence of a functional rho terminaton factor (Wu, 1978).

The promoter utilized was the *lac*UV5 promoter which has been reported to adequately transcribe the attached sequence accurately and at high levels after being induced with 1.0M isopropyl β -D-thiogalactoside (IPTG) (Boer, 1983). The following is the reported sequence of the *lac*UV5 promoter.

CCAGGCTTTACACTTTATGCTTCCGGCTCGT**TATAATGTGTGGA**

The TATA box is in bold and the -35 sequence is underlined.

Transcription is reported to start right after this sequence. Other promoters hybrids of the *lac* and *trp* promoters are listed in (Boer, 1983). The *lac* promoter was chosen because it is well studied and it has the capacity of acting as a constitutive promoter or as an inducible promoter with IPTG.

2.4 Cloning

2.4.1 Theory and Methods

Finally to insert the DNA into the vectors a multitude of enzymes were utilized for the splicing, dephosphorylation and ligation of the DNA. The following procedures were implemented at different times for different purposes:

Klenow Treatment

1. 50 µl of Qiagen pure PCR product/restriction digest gel band excision
2. Add 5.5 µl of NEB buffer 2
3. Add dNTPs to final concentration of 33 µM
4. Add 1 µl of DNA Polymerase I, Large (Klenow) Fragment (M0210S, low concentration 5,000 units/mL)
5. Incubate at room temperature for 15 minutes.
6. Immediately Qiagen purify to remove the Klenow

Kinase Treatment

1. Take 50 µl of Qiagen pure PCR product/restriction digest gel band excision
2. Add 5.5 µl of 10X T4 DNA ligase buffer (already has ATP)
3. Add 1µl of T4 PNK
4. Incubate 37 °C for 30 minutes
5. Immediately Qiagen purify to remove the T4 PNK

Dephosphorylation

1. Take 50 µl of Qiagen pure PCR DNA
2. Add 5.5 µl of 10X Antarctic phosphatase buffer

3. Add 1 μ l of Antarctic phosphatase
4. Incubate at 37 °C for 15 minutes
5. Heat inactivate at 70 °C for 10 minutes

Ligation

1. Add 50 ng of vector into microfuge tube
2. Add 3X molar amount of insert or 6X if blunt end ligation
3. Bring volume up to 10 μ l with ddH₂O
4. Add 10 μ l of 2X Quick Ligase Buffer
5. Add 1 μ l of Quick Ligase, mix with pipette tip swirling
6. Incubate at room temperature for 5 minutes and then proceed to transformation

TOP10 cells Transformation (Invitrogen)

1. Thaw competent cells on ice for 5 min, prepare a 42°C water bath
2. Add 2 μ l of ligation mixture to competent cells and swirl with pipette tip
3. Allow to incubate on ice for 20-30 min, no more than 30 min.
4. Incubate 42°C water bath for exactly 30 seconds, then place back onto ice.
5. Add 200 μ l of SOC within 1 min of returning cells to ice.
6. Allow to outgrow for exactly 1 hr at 37°C with shaking (225 rpm)
7. Plate 100 μ l onto appropriate plate.

For the most part the purification of DNA was done with a Qiagen PCR purification kit or by a Qiagen miniprep or maxiprep kit.

In most cases the DNA inserts constructed via PCR were purified then treated with the specific restriction enzyme. At the same time the vector was

obtained from a maxi or mini prep and was also treated with the same restriction enzyme thus leaving sticky ends in both DNA pieces for a facilitated ligation. To ensure a lower percentage of self ligating vectors, the plasmid was dephosphorylated after being treated with the restriction enzyme.

For other more challenging cases blunt ends were created by treating both the insert and the plasmid with Klenow after the digestion with the specified restriction enzyme. Dephosphorylation of the linearized plasmid was also conducted to achieve a lower percentage of recircularized plasmids with no insert.

Immediately after the ligation TOP10 cells were transfected with the newly formed constructs.

2.4.2 Results

This part of the study proved to be more challenging than previously foreseen. Several problems were encountered and different trouble shooting strategies were undertaken.

One of the main problems faced was the lack of stability of the vector pACYC184. This vector tends to degrade very rapidly even at its most stable form, which is double stranded and circularized. Whenever the vector was digested with any of the implemented restriction enzymes, the DNA would rapidly degrade and disassemble into the solution. To counter this problem, the enzyme concentration was lowered to up to 10 times less than the recommended amount. The incubation period was also shortened and constant monitoring of the DNA reaction via gel electrophoresis was necessary. A much larger amount of vector was also utilized to account for the heavy losses due to its continued degradation. Maxi preps were executed to achieve the higher

concentration levels as this is a low copy number plasmid and most minipreps resulted in too low of a concentration.

The vector pACYC184 also had very low yields when purified with the Qiagen PCR purification kit. Blunt end ligation proved to be most difficult for this plasmid as several purification steps were needed and much of the DNA was lost after every purification. Figures 11 and 12 are the two constructs that were achieved after many trials.

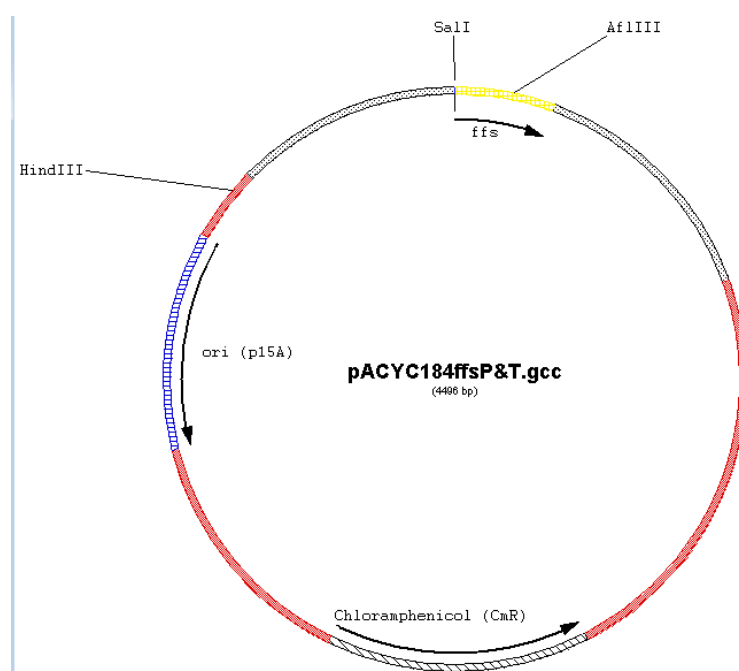


Figure 11. pACYC184-*ffs* construct.

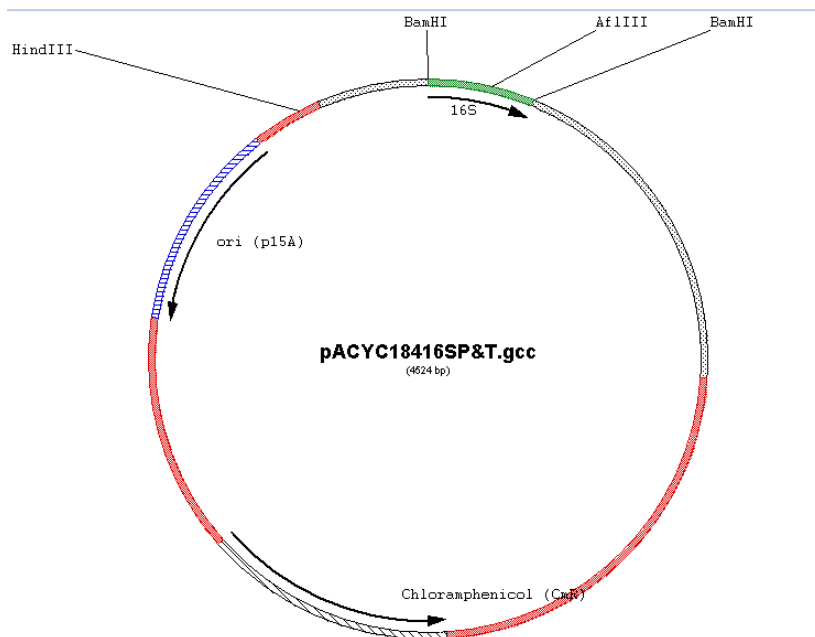


Figure 12. pACYC184-int16S construct.

As seen in Figures 11 and 12 both of the plasmids can be single digested with AflIII to confirm that the surviving cultures do in fact have the DNA insert and that the plasmid has not simply recircularized. The sequencing of these constructs in the appendix confirms the single or double restriction enzyme ligations. When an insert was successfully inserted into pACYC184 its stability dramatically improved. This has been observed in previous studies done with the same plasmid.

Vector pUC19 was much more manageable as it was easy to acquire though simple minipreps due to its high copy number, its purification yield was normal and it had no stability problems. The main problems arose when the ligated constructs were transfected into chemically competent cells. It is difficult to assess the problem at this stage of cloning as there is no way to assess for the correct ligation of the construct. Since this step involved a multitude of preceding steps the probability that this last step (transinfection)

fails due to previous steps is higher. Gel electrophoresis and DNA mass spectrometry are the best methods for assessing the viability of the vector and the insert. There is no practical direct method to assess for the accuracy of dephosphorylation as well as other steps. To assess the quality of the entire procedure, a control is utilized which consists of transfecting the cells with the dephosphorylated linearized vector and no insert. If the vector has effectively linearized, and has been dephosphorylated then there should be no surviving cells as the vector will not be able to ligate due to the lack of phosphate groups attached to the ends of the DNA. However, there is always a low percentage of vectors that recircularize or that were never digested. The control serves as a background to qualitatively assess the viability of the cloning procedure. If the surviving cultures of the desired construct are as many as the surviving cultures of the control then the insertion of the DNA has been unsuccessful. It is only if the number of surviving cultures of the desired construct is higher than the control that there exists a high probability that the DNA was inserted effectively into the linearized vector.

Figures 13 and 14 show the outcome of a couple of the multiple trials.

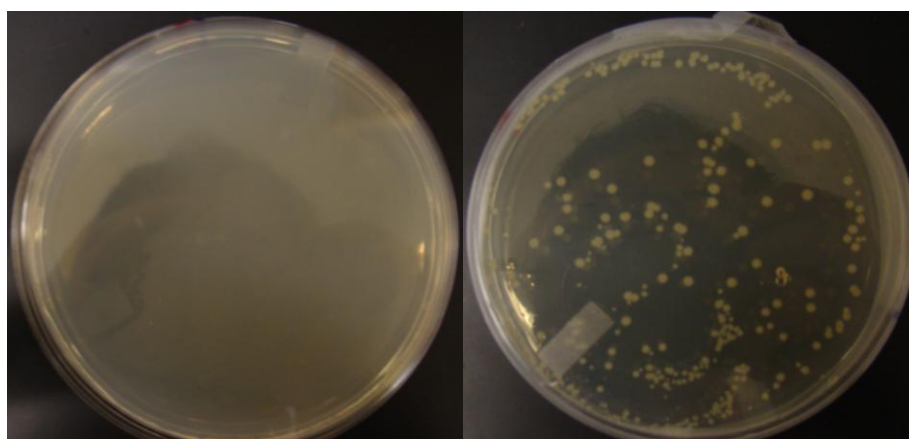


Figure 13. Right: pUC19-Ffh transfection, Left: pUC19 BamH1 digest control

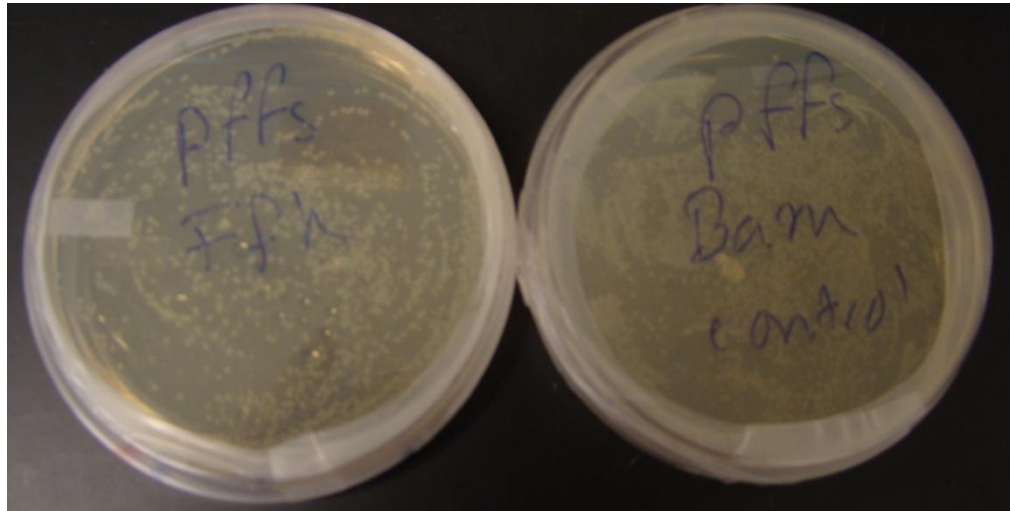


Figure 14. Right: pACYC184-*ffs*-Ffh transinfection, Left: pACYC184-Ffh BamH1 digest control

It has been determined that the most probable reason why none of the pUC19 constructs worked was because the high copy number of the pUC19 represented a lethal overexpression of the inserted genes. If the control has more surviving colonies than the intended construct and given that there is no difference between the transinfection procedures then it is highly probable that the successful insertion of the DNA into the vector was attained but that carrying this construct has negative repercussions for the cell. The observed cultures in the control are most likely due to recircularization of linearized vectors that still had phosphate groups attached to them while this self-ligation was impeded in the non-control due to the high concentration of inserts which will capture any linearized vector that has not been dephosphorylated.

As clearly depicted in Figure 13, the intended transinfection of the Ffh gene into pUC19 resulted in no surviving colonies. The usual outcome of this procedure was a much lower amount of surviving colonies than those seen in the control Figure 13 is an exemplar case where nor a single colony was

produced in the regular procedure. All other inserts placed into pUC19 had a resembling pattern except for the attempted constructs that had the designed *lac* promoter instead of the natural promoter and only when transinfected into cells with an active *lacI* gene.

Ffh resulted in a high toxicity even when inserted to a low copy number plasmid as seen in Figure 14, where Ffh was inserted to the pACYC-*ffs*, a low copy number carrying the *ffs* gene.

Every successfully constructed insert from Table 2 was attempted to be integrated into these plasmids in a various places and in different ways. Blunt end ligation was implemented for the insertion of the inserts into different places of the vector but no other construct was able to be attained in due time.

Chapter 3

CHARACTERIZATION

After the construction of the strains their characterization was performed to assess their growth rate and different RNA levels.

3.1 Selection of Media and Strain

3.1.1 Theory and Methods

The *lac* promoter belongs to the *lac* operon which is induced by the addition of lactose to the system as well as the removal of the glucose as the main source of nutrition (Lodish, 2007; Lewin, 2006). Its mechanism is based primarily on its repressor which inhibits the transcription of the genes downstream from it by inhibiting the binding of the RNA polymerase to the DNA. When the repressor protein encoded by *lacI* is bound by lactose or IPTG the protein can no longer bind to the DNA and thus prevent the *lac* promoter from transcribing the genes downstream from it (Lewin, 2007; Lodish, 2006).

It has also been determined that the *lac* operon also has an activator region which enhances its functionality (Lee, 1984). In (Sabrook, 2001) it is noted that the levels of cAMP are a key factor in enhancing the promoter's ability to trigger transcription of the downstream genes from the *lac* promoter. Glucose is said to reduce the levels of cAMP and that higher levels of cAMP occur when the cell is under a lactose diet (Sabrook, 2001). To increase the activity of the *lac* operon and thus ensure transcription media containing

glucose should be avoided. Succinate, maltose, and lactose have been reported to increase the levels of cAMP in *E. coli*.

As mentioned before, the gene *lacI* is responsible for expressing the repressor protein that will inactivate the *lac* promoter. There are several commercial strains which have had this gene knocked out. Cells without this gene would have a constitutive expression of the genes upstream of the *lac* promoter. Cells with the *lacI* gene would necessitate the induction of this gene with the aid of IPTG.

3.1.2 Results

An analogous minimal media to the minimal media quoted in Molecular Cloning was developed by adding succinate as a carbon source (Sabrook, 2001). The succinate solution added to the minimal media was prepared by directly dissolving succinate into water at a concentration three times as dilute as the solubility limit and then NaOH was added until it reached a pH of 7.

The main concern regarding the new minimal media was that carboxylic salt would inhibit the normal growth of *E. coli*. To test whether this new minimal media was viable, two different strains containing an unaltered vector puc19 were grown in 500ml baffled flasks and 5ml conical tubes. Figure 15 shows that there exists little difference between the glucose media and the succinate media.

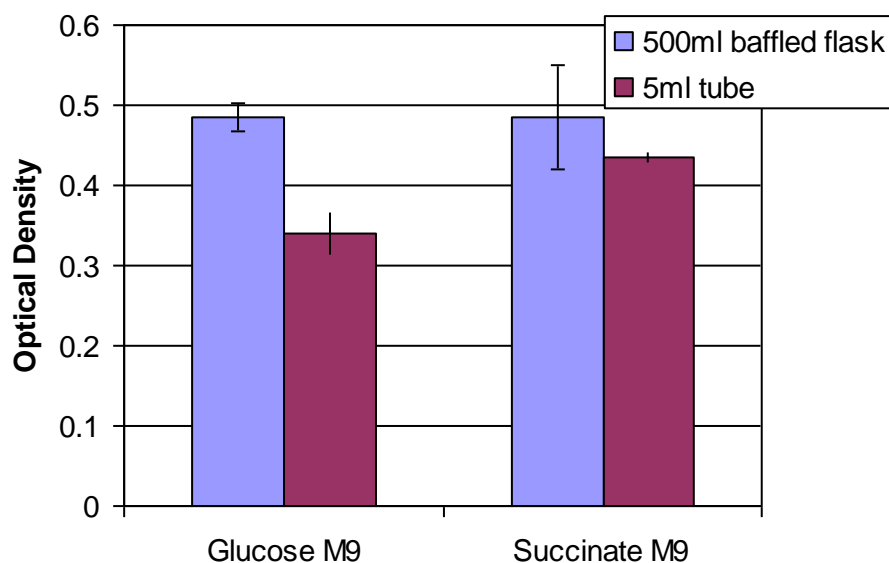


Figure 15. Minimal media comparison. Growth after 48hrs. The optical density of the samples was taken from a pair of technical replicates taken each from two biological replicates. The four measurements were averaged and their standard deviation reported as error bars.

As seen in Figure 15, the new implemented minimal media does not show any growth advantages or disadvantages when grown in 500ml baffled flasks. Succinate does demonstrate an improvement over glucose minimal media when the cells are grown in 5ml conical tubes. This probes that the new minimal media is a viable solution and was utilized in further tests.

LB was also utilized as it has no glucose and is a standard utilized media. The only problem with this media is that it is not specific and every batch will differ from each other. This media is attained from yeast extracts that vary in composition; therefore, the same batch was used for every test throughout the entire course of the experiment to maintain a lower variability.

Two strains were utilized TOP10F and TOP10F'. TOP10F cells differ very little from TOP10 F'. Their main difference lies in that TOP10F cells lack the repressor gene for the *lac* promoter and TOP10F' cells do have it. This would mean that for constructs containing a *lac* promoter they will be able to be induced by the addition of IPTG and will be constitutively expressed in TOP10F cells.

Since the *lac* promoter was being utilized for most of the constructs it is recommended not to use media that contains glucose as the *lac* promoter will not be activated as strongly as without glucose due to the different levels of cyclic AMP (Sabrook, 2001).

3.2 Growth curves and qRT-PCR

3.2.1 Theory and Methods

The optical density (OD) is a good measure to count the density of the population and thus correlates with the population number. Most spectrometers measure the absorbance of light that the sample captures. The light is passed through the sample and depending on the properties of the liquid, the path's length, and the concentration of the sample more or less light will be allowed through the liquid. The Beer-Lambert law simplifies this relationship: $A = \epsilon bc$. Here ϵ is the molar absorptivity of the liquid, b is the path's length, and c is the concentration of the compound in the solution. The denser the population is, the higher the concentration will be, and the higher the absorbance will be for the specific sample.

This simple linear relationship of concentration and absorbance only holds when the absorbance is low. A previous study was executed where the

spectrometer measured the absorbance at 600nm and counted the number of cultures that appeared in a plate. It was found that the measurements should be done when the absorbance is less than 0.3. Therefore, any sample should be diluted tenfold to achieve an accurate and precise relationship. Most samples were diluted 10-12 fold in the course of these experiments.

Measuring the levels of RNA is an important part of this study as it will prove whether the constructed plasmids have any effect in overexpressing the inserted genes. To do so one must isolate the RNA, and be able to quantify it. qRT –PCR is a powerful tool that can detect varying levels of RNA in the cell. Either TaqMan® or SYBR green PCR light cycling are applicable techniques that can detect relative levels of cDNA.

The protocols for the isolation of RNA and reverse transcription can be found in the appendix under supplemental protocols. It should be noted that RNA is very sensitive to degradation and can present several problems if not done carefully.

TaqMan® has the capability of discriminating between the sense and antisense strands of the DNA, but SYBR green cannot. Detecting different strands of DNA would be beneficial for the detection of the intergenic 16S DNA insert as this sequence lies in the antisense portion of the 16S rRNA promoter region and SYBR green might detect levels of this promoter region and not only the targeted overexpressed sequence.

Figure 16 shows an illustration of the principle behind TaqMan®.

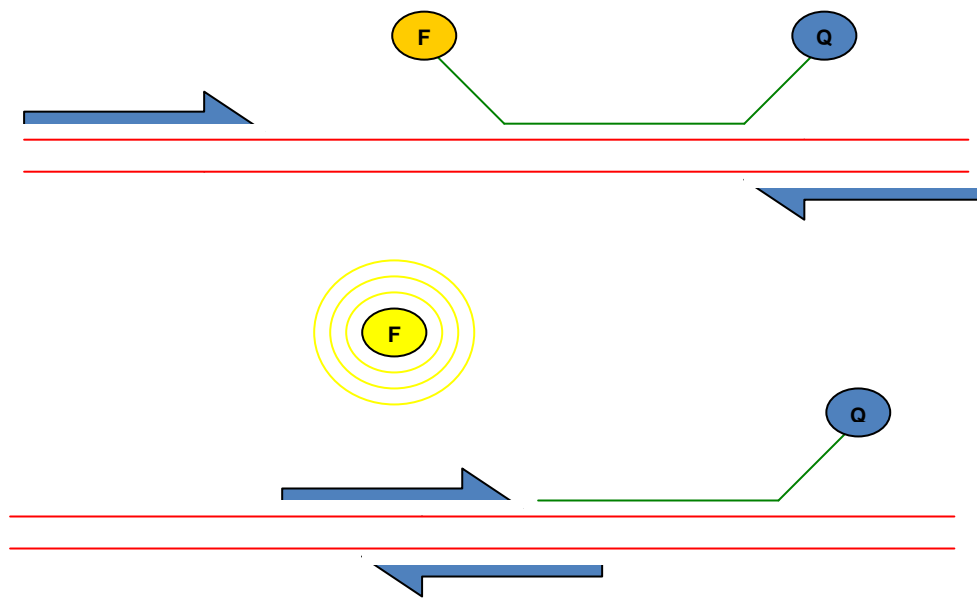


Figure 16. TaqMan illustration.

TaqMan® utilizes a probe that contains a fluorophore and a quencher. The probe base pairs to the targeted DNA strand. As the DNA polymerase replicates the DNA strand the TaqMan®. As the DNA replicates the DNA polymerase will break the TaqMan® probe, releasing the fluorophore. Once the fluorophore is separated from the probe its fluorescence can be detected as it is no longer absorbed by the quencher. The more cDNA is present the faster the fluorescence levels will reach a determined sensitivity threshold. Thus the faster the threshold is overcome the more cDNA there is in the mixture and the more RNA was present in the sample.

For SYBR green the threshold principle is identical. SYBR green intercalates within the double stranded DNA. The acquired cDNA does not have a high enough concentration to overcome the threshold. It is only by adding extra double stranded DNA by PCR DNA amplification of specific

genes that the fluorescence levels can reach and surpass the sensitivity threshold.

3.2.2 Results

Figure 17 shows the growth curve measured via optical density at a wavelength of 600nm in 100ml of LB in a 500ml baffled flask at 37 °C.

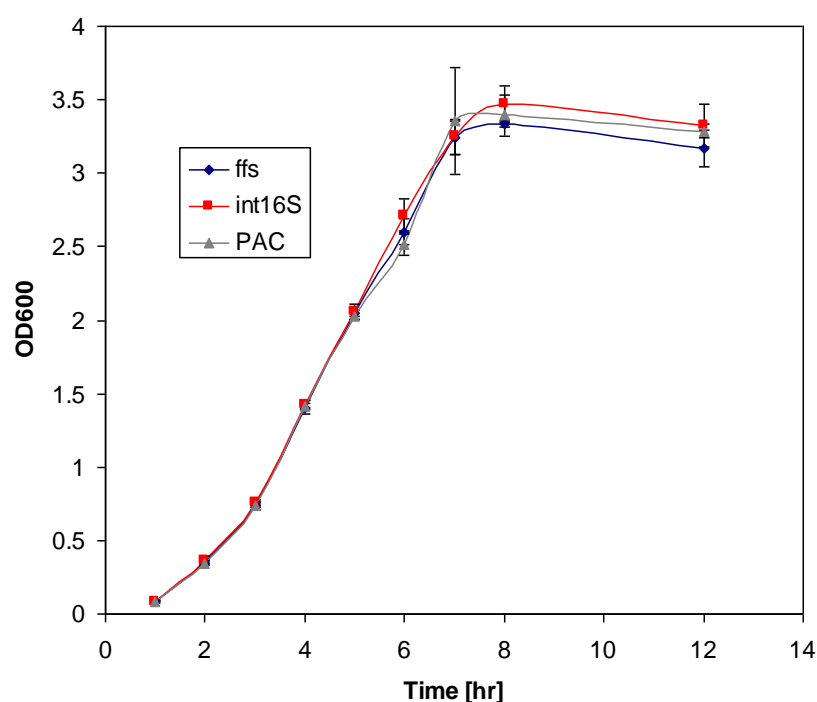


Figure 17. Growth curves in LB media. The strain utilized was TOP10 F, therefore the inserted genes were always transcribed. The control was a strain containing the pACYC184 plasmid with no insert (PAC). There were two technical replicates for every biological replicate, and two biological replicate for every tested strain with different inserted genes giving a total of four measurements of which standard deviation is reported as the error bars.

For the growth curve in Figure 12 two biological replicates were grown separately and samples were taken every hour for 8 hours and then a final sample was taken at 12 hours. It can be seen that there exists no difference between the three strains' growth rate. In this media the stationary phase is reached after 7 hours of growth.

RNA was isolated at 6, 7, 8, and 9 hrs of growth as well as at 12hr and 24hr. This was done with the purpose of measuring the RNA levels at the stationary phase as well as at the exponential phase. cDNA was created soon after the isolation process and qRT-PCR was executed on these samples. SYBR green was the fluorophore of choice for the first rounds. The qRT-PCR was done in a Bio-Rad cycler. The qRT-PCR was not able to detect any significant amount of RNA in any of the samples taken. The problem of this procedure could lie during the isolation of RNA or during the creation of cDNA. Both genes *ffs* and the intergenic 16S are present in *E. coli* naturally and one should be able to detect a certain level of RNA present in any of these samples even if the vectors are not overexpressing these genes. This process was tried several times; however the procedure was not resolved in due time. The main concern during the cDNA creation is that these ncRNAs are quite short and may not be reverse transcribed adequately. The primers are random and the annealing temperatures for most regions within these ncRNAs are somewhat high. This can result in short cDNA fragments or in no creation of cDNA at all. The purification of RNA was done with Qiagen columns that are capable of capturing the small size RNA; therefore, the only possible problem regarding this step is the degradation of RNA. However, careful isolation was done in several trials avoiding exposure to high temperatures or long time storage.

Figure 18 shows a typical melting curve obtained for these samples with this method.

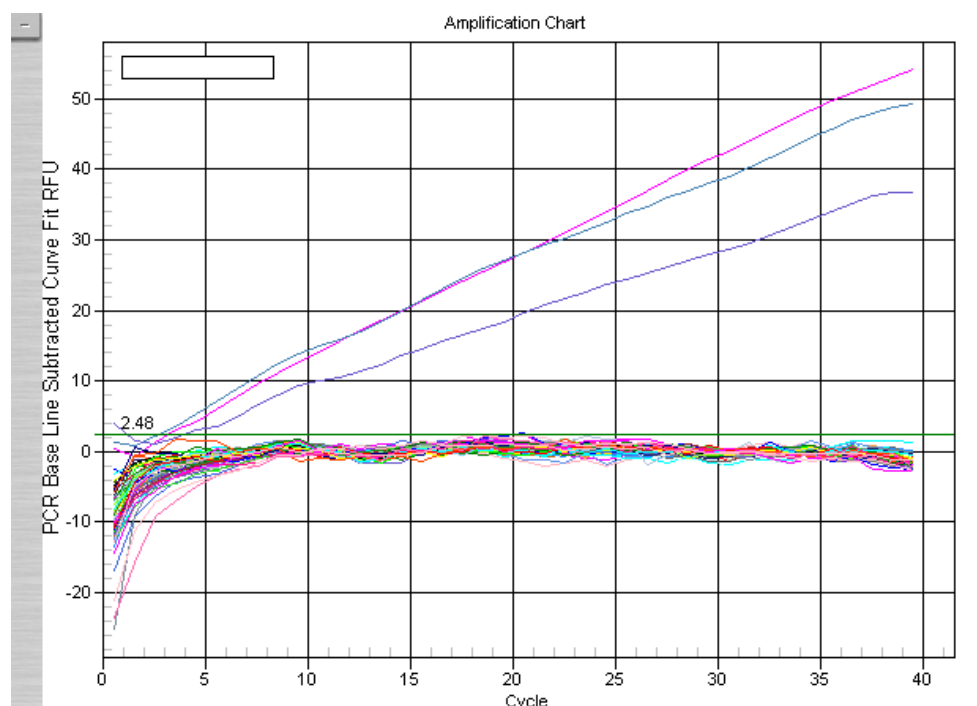


Figure 18. Typical qRT-PCR outcome from the Bio-Rad iQ5 cycler.

It can be seen from Figure 18 that the threshold cycle (CT) is never reached and the only detected CT values are of erroneous samples. A normal CT curve rises exponentially and never linearly as seen in Figure 18. Taqman® was not performed as SYBR green had no results and the necessity for strand discrimination was not able to be assessed.

The RNA probed for was the outcome of the translation of the intergenic 16S, the *ffs* ncRNA, and the 16S rRNA which was utilized as the control. Each RNA sample had a dilution series to detect for different Ct values at different RNA concentrations.

Chapter 4

PERFORMANCE

4.1 Acid Stresses

4.1.1 Theory and Methods

The strains were tested against a control containing a vector with no insert. The tested acids were lactic, acetic, and butyric acids. There is little stress expected from acetic and lactic acid as the bacteria is capable of metabolizing lactic acid as well as acetate; however the stress from butyric acid is expected to be much greater as the bacteria lack the capacity to metabolize this acid. These stresses are not pH related, as the acidity of the growing cultures is almost non-existent. The percentage of acid added is so minute that the solution's pH will not be affected in the least.

These carboxylic acid salts can present a great stress to the cultures as the cell's proton transport is heavily affected by these extracellular carboxylic salts. The build up of protons outside of the cell is necessary for the formation of ATP from electron carriers such as NADH and NADPH. The concentrated salts outside the membrane can prevent this build up to occur naturally by disrupting the regular function of the oxidative phosphorylation pathway. Another concern is that the membrane stability becomes compromised due to the presence of highly concentrated salts and acids. The longer the chain of

these acids is the higher the interaction between the membrane and the concentrated acids will be and thus the higher the stress levels will be.

The cells were grown for 12hr and 24hr under varying degrees of concentrated acids. The extent of the growth was then measured via optical density as previously described. The media utilized was either LB or the minimal succinate media M9. For all the acid stresses the TOP10 F cells were utilized; thus there was no need for IPTG induction.

4.1.2 Results

Figures 19 and 20 show the results from the acids stress done in conical tubes of 5ml for 12hr and 24hr respectively. Table 3 serves as a key for the acid stresses and the volumetric percentage concentration in figures 19 and 20.

Table 3. Legend explanation of the acid stress figures.

Acid / Concentration	0.25%	0.50%	0.75%
Lactic	L25	L50	L75
Acetic	A25	A50	A75
Butyric	B25	B50	B75

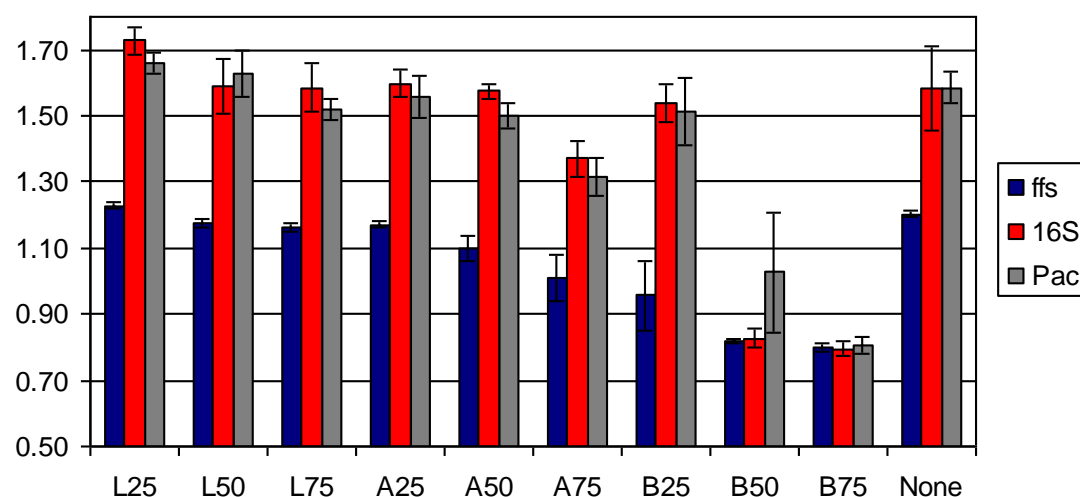


Figure 19. 12 hr Acid stress in LB 5ml tubes

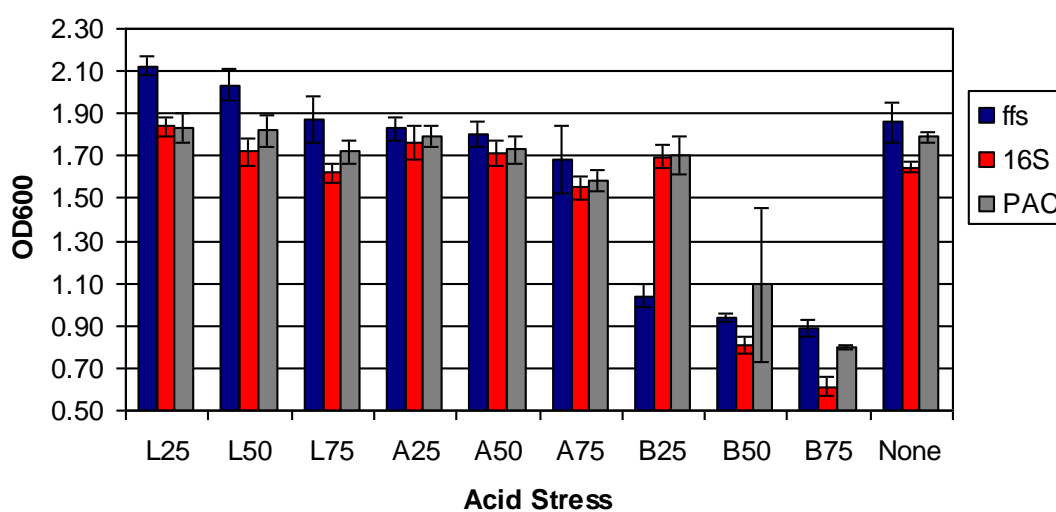


Figure 20. 24hr Acid stress in LB 5ml tubes.

As shown in Figure 19, for acid stresses other than the lactic acid, the *ffs* strain shows a worse growth rate than that of the control containing the undisrupted pACYC184. After 12 hrs it seems that both the int16S and the control grow at equal rates. The lower growing rate of the *ffs* strain is remarkable at 12. However, the OD of every *ffs* sample was measured directly

and not diluted. Thus the data regarding the 12 hr outgrow of *ffs* cannot be used for accurate comparisons. Every other sample was diluted from then on.

After 24 hrs of growth the *ffs* strain shows definite improvement over the control strain during the lactic acid stress at 0.25%. During the highest butyric acid stress the *ffs* strain has a marked advantage over the control.

Figures 21 to 23 show the outcome of the acid stresses done in succinate minimal media. The cells were grown for 12hr in 5ml conical tubes. The acid concentrations were lowered as the minimal media was expected to reduce the tolerance of the strains to the carboxylic acids.

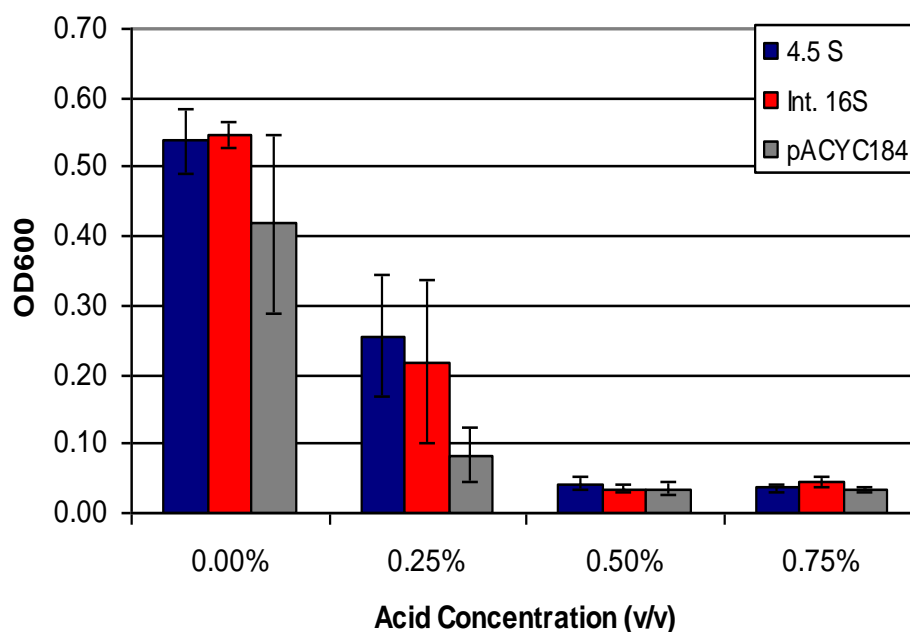


Figure 21. Acetic acid stress after 12 hours in succinate minimal media

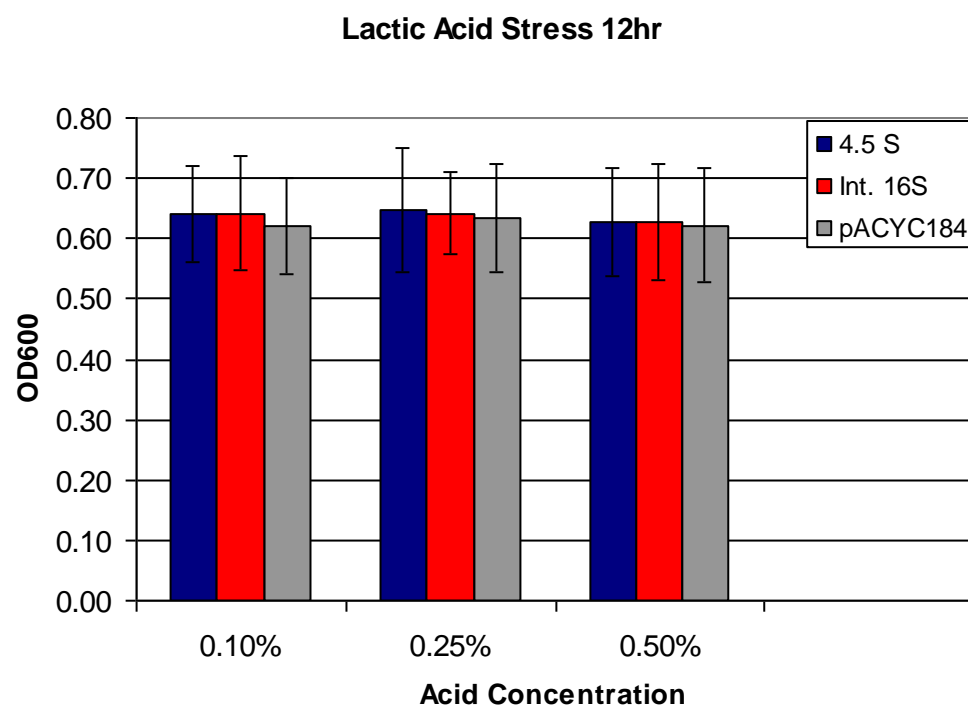


Figure 22. Lactic acid stress after 12 hours in succinate minimal media.

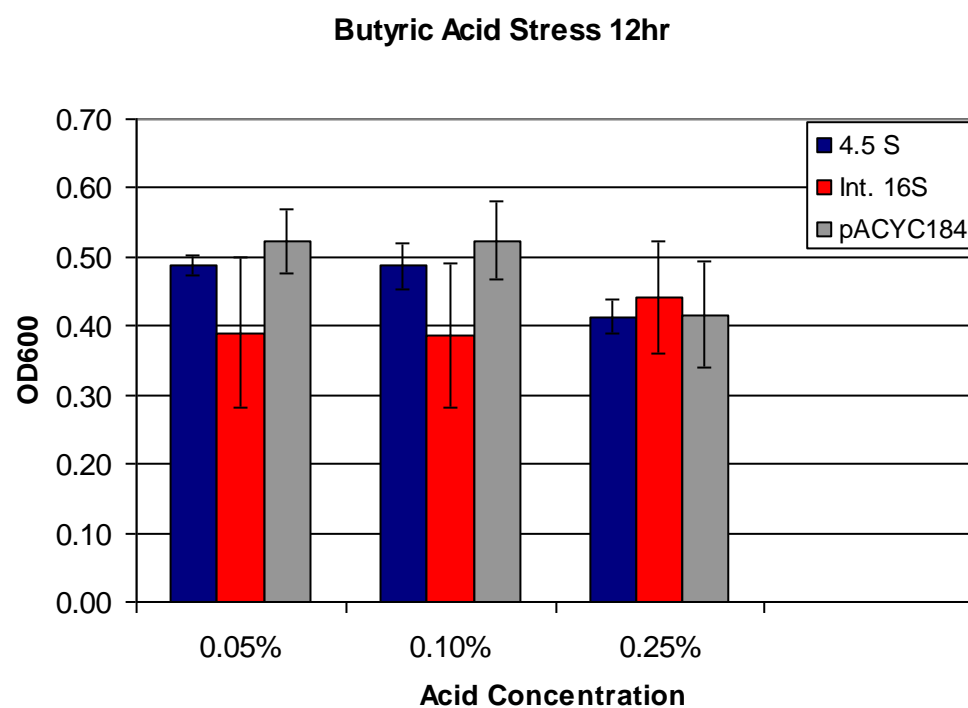


Figure 23. Butyric acid stress after 12 hours in succinate minimal media.

As seen from Figures 21 to 23, there was no statistical difference between the growth of the control and the strains with the vectors containing the *ffs* and the intergenic 16S sequences. The lactic acid stresses could be increased as there was no detected difference between the higher concentrations and the lower concentration. The cells are not challenged by these lactic acid concentrations. The butyric acid concentrations can also be raised to better investigate the tolerance of these strains. Figure 21 presents a clear trend of the challenges that these cells face. As the concentration is increased the growth decreases substantially. The cells were able to resist the acetic acid up to 0.50%. Higher acetic acid concentrations present the cells with a stress where they can no longer grow.

Most of the strains did not present a statistically significant advantage over the control, i.e., the average OD of the control strain was less than three standard deviations away from the average OD of the strain of interest. They do show a trend where they are marginally outperforming the control. However, it is so minimal that it is almost negligible.

Chapter 5

CONCLUSION

The 4.5S ncRNA, as well as the intergenic upstream 16S ncRNA sequences were overexpressed with a low copy plasmid, pACYC184, in *E. coli* Top10 F cells. These cells were grown in LB in 5ml conical tubes for 12hr and 24hr under the following acid stresses: lactic acid, acetic acid, and butyric acid. The acid concentrations were the following: 0.25% (v/v), 0.50% (v/v), and 0.75% (v/v). The cells whether grown in succinate minimal media or LB showed no statistical difference from the control strains under any acid challenge. There do exist a couple of cases where the data indicates that the growth difference is statistically significant, however, the difference is still marginal. The *ffs* strain proved to have a greater outgrow than the control after 24 hours in LB media at 0.75% butyric acid. Overall the cells had a high tolerance for lactic acid. Butyric acid proved to be the most stressful environment for *E. coli* as most cells grew marginally after 24 hrs at 0.50%(v/v).

It was evident that the Ffh protein when overexpressed in high copy numbers was lethal to the *E. coli* strains. A similar phenomena was observed when the intergenic 16S and the *ffs* sequences with natural promoters and terminators were transfected into *E. coli*. The Ffh protein was also intended to be cloned in the lower copy number plasmid pACYC184 but similar patterns of cell toxicity were also observed as the control had larger numbers of surviving

colonies than the transinfection containing the DNA inserts. The quantitative reverse transcription PCR (q_RT PCR) was not sensitive enough to detect the short ncRNAs thus it is uncertain whether the constructs containing the *lac* promoter overexpressed the ncRNAs.

REFERENCES

- Abell, et al. Signal recognition particle mediates post-translational targeting in eukaryotes. *EMBO*. 2004. Vol 23, pp. 2755–2764.
- Ahsen, Nicholas; Wittwer, Carl; Schütz, Ekkehard. Oligonucleotide Melting Temperatures under PCR Conditions: Nearest-Neighbor Corrections for Mg²⁺, Deoxynucleotide Triphosphate, and Dimethyl Sulfoxide Concentrations with Comparison to Alternative Empirical Formulas. *Molecular Diagnostics and Genetics. Clinical Chemistry* 47:11. 2001.
- Alberts, Johnson, Lewis, Raff, Roberts, and Walter. *Molecular Biology of the Cell, 4th Edition*. 2002, Garland Science, New York, NY.
- Bentley, Willaim; Mijalili, Noushin; Anderson, Dana; Davis, Robert; Kompala Dinkara. Plasmid-Encoded Protein: The Principal Factor in the “Metabolic Burden” Associated with Recombinant Bacteria. *Biotechnology and Bioengineering* Vol 35. 668-681 (1990)
- Berndt, et al. A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. 2009. *PNAS*. Vol 106 ,pp 1398-1403.
- Borden, Jacob; Eleftherios, Terry. “A genomic-library based discovery and characterization of a novel acid-tolerance mechanism in *Clostridium acetobutylicum*”. *Metabolic Engineering*. 2010. vol 12. pp. 268-281.
- Bornemann, et al. Signal sequence–independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. *Nature Structure & Molecular Biology*. 2008. Vol 15, pp. 494-499.
- Boer, Herman; Comstock, Lisa; Vasser, Mark. The tac promoter: A functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci. USA*. Vol 80. 1983.
- Buskeiewicz, Peske, Wieden, Gryczynski, Rodnina, and Wintermeyer. Conformations of the Signal Recognition Particle Protein Ffh from *Escherichia coli* as Determined by FRET. 24 June 2005. *Science Direct, Journal of Molecular Biology*, Volume 351, pp. 417-430.
- Buskeiewicz, Peske, Wieden, Gryczynski, Rodnina, and Wintermeyer. Domain rearrangement of SRP protein Ffh upon binding 4.5S RNA and the SRP receptor FtsY. *RNA*. 2005 Vol 351, pp. 417-430.

- Eddy, S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.*, 2, 919–929.
- IDT. Integrated DNA Technologies.
<<http://www.idtdna.com/Home/Home.aspx>> visited from may-2009 to may 2010.
- Jensen, Charlotte; Pedersen, Steen. Concentrations of 4.5S RNA and Ffh Protein in *Escherichia coli*: the Stability of Ffh Protein Is Dependent on the Concentration of 4.5S RNA. *Journal of Bacteriology*. Dec 1994. pp. 7148-7254.
- Kaiser, C.M. et al. Real-time observation of Trigger factor function on translating ribosomes. *Nature* 444, 455–460 (2006).
- Kawano, M., Reynolds, A.A., Miranda-Rios, J. and Storz, G. (2005) Detection of 5'- and 3'-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.*, **33**, 1040–1050.
- KEGG. Kyoto Encyclopedia of Genes and Genomes. < <http://www.kegg.com/>> visited from may-2009 to may 2010.
- Keenan, et al. The Signal Recognition Particle. *Annu. Rev. Biochem.* 2001. Vol 70, pp.755–75.
- Kremer, et al. Characterization of the *sat* Operon in *Streptococcus mutans*: Evidence for a Role of Ffh in Acid Tolerance. *Journal of Bacteriology*. 2001. Vol 183 pp. 2543-2552.
- Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 2001;294:858–862
- Lee, Sun; Bailey, James. Genetically Structured Models for lac Promoter-Operator Function in the *Escherichia coli* Chromosome and in Multicopy Plasmids: lac Operator Function. *Biotechnology and Bioengineering*. Vol 26, pp. 1372-1382 1984
- Lewin, Benjamin. *Genes IX*. Jones and Bartlet Publishers, MA. 2006.
- Lodish, et al. *Molecular Cell Biology*. W. H. Freeman and Company. 6th edition. New York 2007
- Mattick, Josh; Makunin, Igor. Non-coding RNA. *Human Molecular Genetics*, 2006, Vol. 15. ppR17-R29.
- Mattick, Josh . Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *BioEssays*. Wyley Periodicals, Inc. 2003. 25.10:930-939

- Moser, Mol, Goody, and Sinning. The signal recognition particle of *Escherichia coli* (FtsY) has a nucleotide exchange factor built into the GTPase domain. October 1997. Proceedings of the National Academy of Science, Volume 94, pp. 11339-11344.
- Ogg, Poritz, Walter. Signal Recognition Particle Receptor is Important for Cell Growth and Protein Secretion in *Saccharomyces cerevisiae*. August, 1992. Molecular Biology of the Cell, Volume 3, pp. 895-911.
- Gutierrez, Crowley, Cvitkovitch, Brady, Hamilton, Hillman, and Bleiweis. *Streptococcus mutans ffh*, a gene encoding a homologue of the 54 kDa subunit of the signal recognition particle, is involved to resistance to acid stress. 1999. Microbiology, Volume 145, pp. 357-366.
- Park, et al. Functional Analysis of the Signal Recognition Particle in *Escherichia coli* by Characterization of a Temperature-Sensitive *ffh* Mutant. Journal of Bacteriology, May 2002, p. 2642–2653
- Reinau; Otzen. Stability and structure of the membrane protein transporter Ffh is modulated by substrates and lipids. Archives of biochemistry and Biophysics. 2009. Vol 492, pp. 48-53.
- Rivas, E., Klein, R.J., Jones, T.A. and Eddy, S.R. (2001) Computational identification of non-coding RNAs in *E. coli* by comparative genomics. Curr. Biol., **11**, 1369–1373.
- Sætrom, et al. Predicting non-coding RNA genes in *Escherichia coli* with boosted genetic programming. Nucleic Acids Research, 2005, Vol. 33, No. 10. 3263–3270.
- Sambrook, Russell. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press. New York 2001.
- Skatch, William. Cellular mechanisms of membrane protein folding. 2009. Nature Structure & Molecular Biology. Volume 16, pp. 606-612.
- Storz, G., O'Dyke, J.A. and Zhang, A. (2004) Controlling mRNA stability and translation with small, noncoding RNAs. Curr. Opin. Microbiol. 7, 140–144.
- Urban, Johannes; Vogel, Jorg. Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. Nucleic Acids Research. 2007. Vol 35.
- Valent, et al. The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon. EMBO. 1998. Vol 17, pp.2504-2512.

Wu, Anna; Platt, Terry. Transcription termination: Nucleotide sequence at 3' end of tryptophan operon-in- *Eseherichia coli*. Proc. Natl. Acad. Sci. USA. Vol 75. pp. 5442-5446. 1978

APENDIX

A.1 Supplemental Protocols

E. coli RNA Isolation

Note: Wear gloves and use RNase-free tips and tubes throughout procedure.

Prepare in advance:

- Aliquots (1 ml) of 15 mg/ml lysozyme in RNase-free SET buffer and store at -20°C.
- If using the Qiagen DNase I kit, dissolve the lyophilized DNase I in 550 µl of RNase-free water. Aliquot in 50 µl aliquots and store at -20°C. They are good for 9 months at -20°C or 6 weeks at 4°C.

1. Thaw 15 mg/ml lysozyme in SET buffer.
2. Thaw cell pellets on ice.
3. Add 100 µl of Proteinase K to 1 ml aliquot of SET with lysozyme.
4. Resuspend each pellet in 200 µl of SET with lysozyme.
5. Mix by vortexing for 10 s. Incubate at room temperature for 10 min. Vortex for 10 s every 2 min.
6. Immediately put tubes on ice after lysing.
7. Add 1 ml of Trizol to mixture and pipet well.
8. Add 200 µl ice-cold RNase-free chloroform and shake tubes vigorously by hand for 15 seconds; incubate at room temperature for 3 minutes.

9. Spin samples at no more than 12,000 RPM for 15 minutes at 4°C.
10. Transfer upper (aqueous) phase (<500 µl) to a new RNase-free tube; try not to touch the interface to prevent contamination.
11. Add 500 µl of 70% EtOH and mix by vortexing.
12. Pipet up to 700 µl of sample into RNeasy Mini Spin Column. Close tube gently and centrifuge at room temperature, $\geq 8000 \times g$ ($\geq 10,000$ rpm) for 15 seconds. Discard flow-through.
13. Repeat step 15 for remainder of sample.
14. Add 700 µl Buffer RW1 to RNeasy Mini Spin Column. Incubate at room temperature for 4 minutes. Close the tube gently, and centrifuge for 15 seconds at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to wash the column. Discard the flow-through and collection tube.
15. Transfer the RNeasy Spin Column into a new 2 ml collection tube (supplied). Pipet 500 µl Buffer RPE onto the RNeasy Spin Column. Close the tube gently, and centrifuge for 15 seconds at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to wash the column. Discard the flow-through.
16. Add another 500 µl Buffer RPE to the RNeasy Spin Column. Close the tube gently, and centrifuge for 2 minutes at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to dry the RNeasy silica-gel membrane. Discard the flow-through.
17. Centrifuge in a microcentrifuge at full speed for 1 additional minute to further dry the membrane.
18. To elute, transfer the RNeasy Mini Spin Column to a new 1.5 ml collection tube (supplied). Pipet 50 µl RNase-free water directly onto the RNeasy silica-gel membrane. Close the tube gently, and centrifuge for 1 minute at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to elute.

19. To quantify samples, measure 1.5 µl on the Nano Drop spectrophotometer.

To evaluate degradation, run sample on 1.0% agarose gel.

20. Store samples at -85°C.

cDNA Generation

Updated 7-14-09, SWJ

Uses Applied Biosystems' High-Capacity cDNA Reverse Transcription Kit
(P/N# 4368814)

1. Thaw RNA samples and kit components on ice.
2. Prepare a 2X RT Master Mix (per 20 µl reaction):

Component	Volume (µl)/Reaction
10X RT Buffer	2.0
25X dNTP Mix (100 mM)	0.8
10X RT Random Primers	2.0
MultiScribe Reverse Transcriptase	1.0
Nuclease-free H ₂ O	4.2
Total per Reaction	10.0

Important! Include additional reactions in the calculations to provide excess volume for the loss that occurs during reagent transfer.

3. Keep the 2X RT Master Mix on ice.
4. Calculate the volume of RNA needed for 2 µg of total RNA.

5. For each reaction, add to the tube:

Component	Volume (μl)
2X RT Master Mix	10.0
Nuclease-free H ₂ O	Balance for a total volume of 20.0 μl (e.g. 7.7 μl of Nuclease-free H ₂ O)
RNA sample	Volume for 2 μg of total RNA (e.g. 2.3 μl of RNA)
Total per Reactions	20.0

6. Centrifuge briefly to spin down contents and eliminate any air bubbles.
7. Keep tubes on ice until thermal cycler is setup.
8. Program the thermal cycler with the following program:

	Step 1	Step 2	Step 3	Step4
Temperature	25°C	37°C	85°C	4°C
Time	10 min	120 min	5 min	∞

9. After completing the reverse transcription, remove the tubes from the thermal cycler and add 80 μl of nuclease-free H₂O to the sample to

dilute it to 20 ng/μl. If a different reaction volume was used or a different amount of RNA was added, add the appropriate amount of nuclease-free H₂O to achieve a concentration of 20 ng/μl.

10. For short-term storage, store cDNA at 4°C. For long-term storage, store at -20°C.

A.2 Sequencing

Sequencing results from the pACYC int16S 9b:

TGGTTGGATCGAGATACGGATTCTTAACGTCGCAAGACGAAAAATG
AATACCAAGTCTCAAGAGTGAACACGTAATTCATTACGAAGTTTAA
TTCTTTGAGCATCAAACCTTTTCCACACATTATACGAGCCGGAAGCAT
AAAGTGTAAGCCTGGGGATCCTCTACGCCGGACGCATCGTGGCCG
GCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGA
CATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGC
GCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCGTGGCCGGGGGAC
TGTTGGGCGCCATCTCCTTGCAATGCACCATTCCTTGCGGCGGCGGTG
CTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTC
GCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAACCCA
GTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACT
TATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAG
CGCTCTGGGTCAATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACG
ATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGACGCCCTCGC
TCAAGCCTTCGTCACCTGGTCCCGCCACCAAACGTTTCGGCGAGAAGC
AGGCCATTATCGCCGGCATGGCGGGCCGACGCGCTGGGCTACGTCTT
GCTGGCGTTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTC
TTCTCGCTTCCGGCGGCATCGGGGATGCCCCGCGTTGCAGGCCATGCT
GTCCAGGCAAGGTAGATGACGACCATCATGGAACAGCCTTCCAAGG
ATCGCTCGCGGCTCTTACCAACCTAACTTCCAATCATTGGACCGCTG
ATTCGTCCCGGCGATTTTATGCCGCCCTCGTGCAAACACATTGGAAC
GGGGTTGGGCCTGGAATTGTAGTGCGCCGCCCCCATTACCTTTGTCT
GGCCCTCCCCGCATTTGCGTCCTCGGGTCAAAGGAGACCGTGCCAC
TCTCAACCATGAAATGAGAACCCGGGCGGCCACCTCTCATTAACG
AATTCTCACCAACTCTCAAAGAATTGGTGAGAGCCAAATTCAATTTT
ATTTGTGACGGAAATAAAC

Sequencing results from the pACYC int16S 9b:

ACGATCATACGTGTTCACTCTTGAGACTTGGTATTCATTTTTTCGTCTT
GCGACGTTAAGAATCCGTATCTTCGAGTGCCACACAGATTGTCTGA
TAAATTGTTAAAGATCAGTTAATCCCACAAGCCGCCAGTTCCGCTGG
CGGCATTTTGGATCCACAGGACGGGTGTGGTCCCCATGATCGCGTAT
TCGATAGTGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGC
CAAAGCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTG
CATCAACGCATATAGCGCTAGCAGCACGCCATAGTGA CTGGCGATG
CTGTCGGAATGGACGATATCCCGCAAGAGGCCCCGGCAGTACCGGCA
TAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGAT
GACGATGAGCGCATTGTTAGATTTTCATACACGGTGCCTGACTGCGTT
AGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATA
AGCTGTCAAACATGAGAATTACA ACTTATATCGTATGGGGGCTGACTT
CAGGTGCTACATTTGAAGAGATAAATTGCACTGAAATCTACAAATA
TTTTATCTGATTAATAAGATGATCTTCTTGAGATCGTTTTTGGTCTGCG
CGTAATCTCTTGCTCTGAAAACGAAAAAACCGCCTTG CAGGGCGGT
TTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTA ACTG
GCTTGGAGGAGCGCAGTCACCAAACTTGTCTTTTCAGTTTAGCCTT
AACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTAC CAGT
GGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAA
GACGATAGTTACCGGATAAGGCGCAGCGGTTCG GACTGAACGGGGG
GTTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGA ACT
GAGTGTCAAGCGTCATGAGACATACGCGGCCATCACAGCGGATGAC
ACGGGTAAACCGAAACGGCATGAACAGGGAAGTCGCACCGAGGAA
GCCGCCAGGGGGGAACGCTTGTATCTTTATAGTCATGTCGGGTT CGC
CACC ACTGAATTGACGTCGAATTCCTGATGCCCTGTT CAGGAGACCG
TAATA

Sequencing results from pACYC int16S 10a

CGGTTGCTCGAGATACGGATTCTTAACGTCGCAAGACGAAAAATGA
ATACCAAGTCTCAAGAGTGAACACGTAATTCATTACGAAGTTTAATT
CTTTGAGCATCAAACCTTTTCCACACATTATACGAGCCGGAAGCATAA
AGTGTAAGCCTGGGGATCCTCTACGCCGGACGCATCGTGGCCGGC
ATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACA
TCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC
TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTG
TTGGGCGCCATCTCCTTG CATGCACCATTTCCTTGCGGCGGCGGTGCT
CAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGC
ATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGT
CAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTA
TGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG
CTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGAT
GATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGACGCCCCTCGCTC
AAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA
GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTG
CTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCT
TCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCACGCCATGCTGT
CCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCT
CGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCA
CGGCGATTTATGCCGCCTCGGCGAGCACATGGAAACGGGTTGGCAT
GGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGT
CGCGGTGCATGCAGCCGGGCACTCGACCTGAATGGAGCCGCCGCAC
TTCGCTACGATCACACTCAGGATTGCAGCCATCATTCTGCGGAACTG
TGATGCCAACTATCCTGGCGAACTTA