8-2007

Characterization of the Group II Intron Gs. Int1 from the Thermophilic Bacterium *Geobacillus stearothermophilus*.

Huijing Sun
East Tennessee State University

Characterization of the Group II Intron Gs. Int1 from the Thermophilic Bacterium *Geobacillus stearothermophilus*

A thesis

presented to

the faculty of the Department of Health Sciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Biology

---

by

Huijing Sun

August 2007

Dr. Bert C. Lampson, Chair

Dr. Eric L. Mustain

Dr. Dhirendra Kumar

Keywords: Group II Intron, *Geobacillus stearothermophilus*, Thermophilic bacterium

ABSTRACT

Characterization of the Group II Intron Gs.Int1 from the Thermophilic Bacterium *Geobacillus stearothermophilus*

by

Huijing Sun

Group II Introns are small segments of DNA that reside in the chromosome of bacteria or the organelles of primitive eukaryotes. These elements have some very interesting properties. First, they are retrotransposons that can move from one location to a new location in DNA via a reverse transcription mechanism. Second, they form a large ribozyme that mediates self-splicing of the intron from pre-mRNA. A Group II Intron type protein with similarity to reverse transcriptase was discovered in the thermophilic bacterium *Geobacillus stearothermophilus* strain 10 (Vellore *et al*., 2004, Appl. Environ. Microbiol. *70*: 7140-7147). Numerous copies of the intron, designated Gs. Int1, are present in the chromosome of strain 10 but absent from a related strain ATCC 12980. Experiments to detect the *in vivo* splicing of intron Gs.Int1 from *G. stearothermophilus* cells did not work. Plasmids to that will over-express the Gs. Int1 intron to detest splicing *in vivo* in *Escherichia coli* have been constructed.

DEDICATION

I dedicate this thesis to my parents, Rongmin Sun and Aixia Nan, and my sister, Jingru Sun, whose invaluable support made me finish my master degree and further my education for Ph.D. possible. I would have never made it this far without your support.

ACKNOWLEDGEMENTS

I would like to thank Dr. Bert Lampson, my advisor and committee chair, for being a wonderful mentor during this research project. I appreciate all the guidance, patience, and support I received from Dr. Lampson. I would also like to thank my committee members, Dr. Eric Mustain and Dr. Dhirendra Kumar, for their help and support.

Many thanks go to Mr. Ralph Coffman and Mrs. Robin Grindstaff for their help in the laboratory. I also would like to thank Mrs. Nancy Coffman for her help in departmental official procedure and kindness. I would specially like to thank Dr. Michael Gallagher and Dr. Allan Forsman for their encouragement. In addition, I would like to thank the entire Department of Health Sciences for giving me this opportunity to study in this program that would be an invaluable experience in my life.

Also, I would especially express my appreciation to my classmates and my friends. Their help made my study and my life easier in this country over the last two years. Words can not express how blessed I feel to have all of your help and friendship.

I am very grateful to my parents and my sister's love and invaluable support. I would have never made it this far without you. Thank you!

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This thesis describes the characterization of a group II intron from the thermophilic bacterium *Geobacillus stearothermophilus*. Group II introns are small segments of DNA that code for a catalytic RNA enzyme (ribozyme) and also code for a protein with reverse transcriptase (RT) activity. These products provide group II introns with some interesting properties. First, group II introns can self-splice; that is, they can excise themselves out of a pre-mRNA transcript without requiring a protein product (Lambowitz and Zimmerly 2004; Toro et al. 2007). Self-splicing group II introns of bacteria are considered as the evolutionary progenitors of eukaryotic spliceosomal introns because they use a similar mechanism for splicing out the intron from pre-mRNA. Second, group II introns are also mobile DNAs (called retro-transposons) that can move from one location to a new location in a DNA molecule. This mobility requires the intron encoded RT (Lambowitz and Zimmerly 2004). Bacterial group II introns are also of considerable interest because they have potential practical uses. Some characteristics of group II introns can be exploited for use in genetic engineering in biotechnology and for possible gene therapy (Toro et al. 2007). For example, derivatives of the group II intron LI.LtrB from *Lactococcal lactis* have been used to produce targeted gene disruption in some gram-positive and gram-negative bacteria. This results in a desired mutant strain (Toro et al. 2007). Group II introns can be engineered to insert efficiently into virtually any target DNA and retain activity in human cells (Guo et al. 2000). The group II intron from the thermophile *G. stearothermophilus* is a new class of introns and its adaptation to

function at hot temperatures may provide new insights into how these interesting genetic elements work.

The Genus *Bacillus*

The *Bacillus* genus includes bacteria that are aerobic and facultatively anaerobic, rod-shaped, Gram positive to Gram variable, and form endospores (Nazina et al. 2001). Different species can grow in different environments. They are divided into thermophilic, psychrophilic, acidophilic, alkalophilic, halophilic, and freshwater bacteria (Nazina et al. 2001). They can use a wide range of carbon sources to grow.

According to 16S rRNA gene sequence analysis, the genus *Bacillus* contains 5 genetic subgroups (group 1-5). A phylogenetic tree, based on 16S rRNA, of the genus *Bacillus* and related genera is shown in Figure 1 (Zeigler 2001).

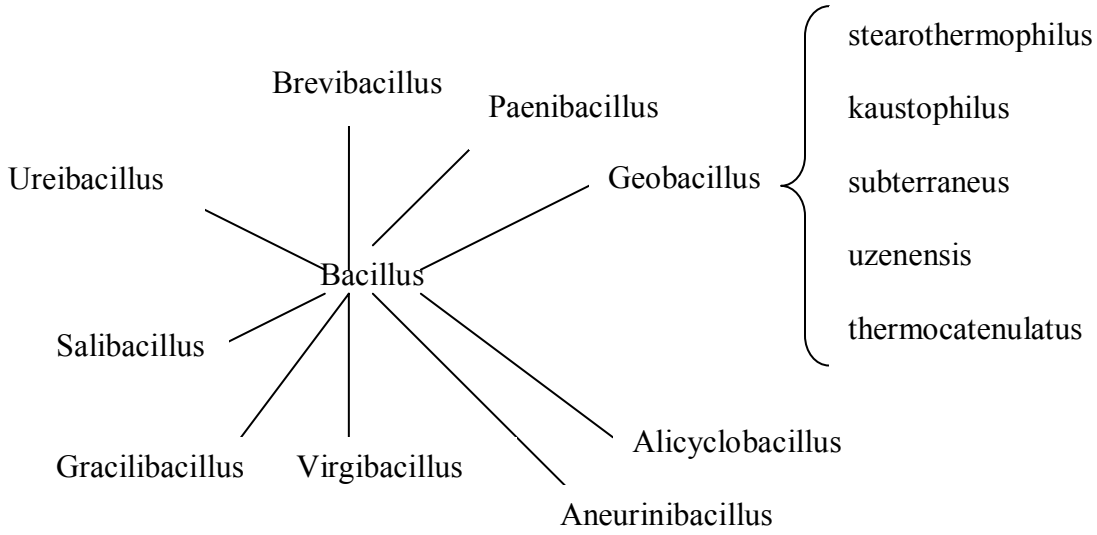


Figure 1 Phylogenetic Tree of the Genus *Bacillus* and Related Genera

(Adapted from Zeigler 2001)

For the genus *Bacillus*, based on their 16S rRNA sequence, the majority of the thermophilic species are divided into group 1 and group 5 (Nazina et al. 2001). Group 5 is a thermophilic bacilli group that has very high similarity among their 16S rRNA sequences (98.5-99.2%) (Nazina et al. 2001). According to physiological characteristics, the results of fatty acid analysis, DNA-DNA hybridization studies and 16S rRNA gene sequences analysis, a new genus was proposed: *Geobacillus* gen. nov., (Nazina et al. 2001). Nazina T.N. and other researchers proposed the transfer of the following species of group 5, *B. stearothermophilus, B. thermoleovorans*, *B. thermocatenulatus, B. kaustophilus, B. thermoglucosidasius,* and *B. thermodenitrificans* into this new genus *Geobacillus* (Nazina et al. 2001). Thus, *Geobacillus stearothermophilus* was reclassified from *Bacillus stearothermophilus.*

The Genus *Geobacillus*

The genus *Geobacillus* includes a diverse group of bacteria that are found in soil. *Geo-* means soil or earth and *bacillus* means small rod-shape. So, the *Geobacillu*s means earth or soil bacillus. There are currently 9 established species within this genus (Zeigler 2001).

For morphology, they are rod-shaped cells, occurring either singly or in short chains and motile by means of peritrichous flagella. They are Gram-positive bacteria, but the Gram stain reaction may vary between positive and negative. One ellipsoidal or cylindrical endospore occurs per cell and is located terminally or subterminally in slightly swollen or non-swollen sporangia. They have variable shape and size and pigments may be produced on certain media (Nazina et al. 2001; Zeigler 2001). They are obligately

thermophilic aerobic or facultatively anaerobic bacteria. The growth temperature range is 37°C -75°C with an optimum at 55°C-65°C. Growth occurs in a pH range of 6.0 -8.5 with an optimum at pH 6.2-7.5 (Nazina et al. 2001; Zeigler 2001).

Thermophilic bacilli, including *Geobacillus*, are widely distributed and have been isolated from all continents where geothermal areas are found (McMullan et al. 2004). Geobacilli are also isolated from shallow marine hot springs, deep-sea hydrothermal vents, and artificial hot environments such as hot water pipelines, heat exchangers, waste treatment plants, burning coal refuse piles, and bioremediation biopiles (McMullan et al. 2004). Most work has concentrated on the isolation from natural and artificial high-temperature "biotopes".

The genome sequences of some numbers of *Bacillus* species have been completed, such as *B. subtilis*, *B. halodurans,* and *B. anthracis* (McMullan et al. 2004). Recently the complete genomic sequence of *Geobacillus kaustophilus* HTA426 has been determined. It was the first complete genome sequence for a thermophilic *Bacillus* related species that is composed of one circular 3.54 million base pairs chromosome and one 47.9 kilo base pairs plasmid. About 839 genes found in the genome may contribute to thermophilic ability and 91 genes were found to encode putative transposases. These enzymes, encoded by DNA transposons, allow transposons to be cut from genomic DNA and be inserted at another location (Takami et al. 2004).

Geobacillus species have a lot of potential applications in biotechnological processes, for example, as sources of various thermostable enzymes, such as proteases from *G. stearothermophilus* strain TLS33 (Sookkheo et al. 2000), amylases from *G. thermoleovorans* (Uma et al. 2003), and lipases and pullanases (McMullan et al. 2004).

Also, *Geobacillus* species can generate particular products for industrial uses such as exo-polysaccharides. Some species, such as *G. caldoxylosilyticus* T20, can be used in agricultural biotechnology for metabolizing herbicides because they can metabolize a range of organophosphonates including the herbicide glyphosate. In addition, two strains of *G. thermoleovorans* can produce large bacteriocins that exhibit lytic activity on other strains of *G. thermoleovorans* and a range of bacteria including *Salmonella typhimurium* (McMullan et al. 2004).

Geobacillus stearothermophilus

Species name: *stearo* (fat) + *thermo* (heat) + *philus* (loving) =fat- and heat- loving bacterium (Zeigler 2001). For *G. stearothermophilus*, there are several strains: *G. stearothermophilus* strain 10, *G. stearothermophilus* strain ATCC 12980, *G. stearothermophilus* strain B-4419, *G. stearothermophilus* strain XL-65-6, and *G. stearothermophilus* strain NUB36 (Zeigler 2001).

G. stearothermophilus strain 10 and *G. stearothermophilus* strain ATCC 12980 were used in this project. *G. stearothermophilus* strain 10 was isolated from a hot spring in Yellowstone National Park (Zeigler 2001). This species was the subject of some early studies to characterize the differences between mesophilic and thermophilic *Bacillus* species in DNA polymerase properties and DNA base composition (Stenesh and Roe 1972). Currently, the *G. stearothermophilus* strain 10 genomic sequence is being determined at the University of Oklahoma, Norman Campus's Advanced Center for Genome Technology.

G. stearothermophilus strain 26 is from the collection of the National Canning Association. This organism is also called G. *stearothermophilus* ATCC 12980. Cells are motile, 0.6-1.0 μm long, and 2.0-3.5 μm wide with "ellipsoidal endospores" (Zeigler 2001). The growth temperature is from 37°C to 65°C. This strain has several functions. First, it is the source of DNA restriction endonuclease BstP; second, it is used in paper strips as a biological indicator for steam sterilization and it is also a source for thermostable enzymes for industrial applications, for example, aspartate transaminase that can be used to produce herbicide (Bartsch et al. 1996; Zeigler 2001).

Group II Introns

Group II introns have three characteristics: first, they are introns; second, they are ribozymes; and third, they are transposons.

What is an Intron?

An intron is a non-coding segment of DNA that is initially copied into RNA but is cut out from the final RNA transcript. Introns are common in eukaryotic RNAs of all types, but they can be also found in some prokaryotic organisms. The regions of a gene that remain in spliced mRNA are called exons. The number and length of introns varies widely among species and among genes within the same species. The discovery of introns led to the Nobel Prize in Physiology or Medicine in 1993 for Phillip A. Sharp and Richard J. Roberts. The term *intron* was introduced by American biochemist Walter Gilbert in 1978 (Gilbert 1978).

Four classes of intron are known to exist: Group I Introns, Group II Introns, Group III Introns, and Nuclear Introns. Group I, Group II, and Group III introns are self-splicing introns. They can catalyze their own splicing out of the primary RNA transcript. Some Group III introns are also identified as Group II introns because of their similarity in structure and function. Group III introns are smaller and more streamlined but the splice site consensus sequences are not as well conserved. Group I introns are the only class of introns whose splicing require a guanine nucleoside. They possess a secondary structure different from that of group II and group III introns. Group I introns are often found in bacteria and protozoa. Nuclear or spliceosomal introns are spliced by the spliceosome, which is a series of small nuclear RNAs (snRNAs) plus proteins. There are certain splice signals (or consensus sequences) that "abet" the splicing (or identification) of these introns by the spliceosome (Roy and Gilbert 2006).

A group II intron is a small genetic element found in the chromosome of bacteria and some eukaryotic organelles, such as mitochondria and chloroplasts. Because any intron is a non-coding part of a gene, it has to be removed from mRNA before the RNA can be translated.

What is a Ribozyme?

A ribozyme is an enzyme that is composed totally of RNA rather than protein. It is also called an RNA enzyme or catalytic RNA because it can catalyze a chemical reaction.

The first ribozyme was discovered in the 1980s by Thomas R. Cech, who was studying RNA splicing in the ciliated protozoan *Tetrahymena thermophila* and also

Sidney Altman, who was working on the bacterial RNase P complex .They won the Nobel Prize in chemistry for the discovery of catalytic RNAs (Lehmann and Schmidt 2003). Seven types of ribozymes have been identified based on their different structures, but this may not be the final number (Lehmann and Schmidt 2003). Due to their difference in size and reaction mechanisms, the seven types of ribozymes are divided into two different groups, the small ribozymes and the large ribozymes (Lehmann and Schmidt 2003).

Group II introns belong to the large ribozyme group. They perform splicing similar to the spliceosome without requiring any protein (see splicing reaction below). This similarity suggests that group II introns may be evolutionarily related to the spliceosome.

What is a Transposon?

Transposons are also called jumping genes or mobile genetic elements. They can move from one location to another location in a DNA molecule. There are several reasons to account for why transposons are very important. a) They can carry antibiotic resistant genes; b) they can carry virulence properties; c) they can act as vehicles for gene exchange; d) they can generate insertional mutations; and e) they can promote evolution. There are two types of transposons: DNA transposons and Retro-transposons. The difference between them is the mechanism they use to move from a donor DNA to a recipient DNA. DNA transposons include insertion sequences, class I composite transposons, class II non-composite transposons, class III transposing bacteriophages, integrons, mobilizable transposons, and conjugative transposons. Among them, insertion

sequences (IS elements) are the simplest type of DNA transposons and those other DNA transposons are more complex in structure and often carry antibiotic resistance genes, catabolic genes, vitamin synthesis genes, nitrogen fixation genes, and heavy metal resistance genes, so they are very important in bacteria (Whittle and Salyers 2002). Retro-transposons include group II introns. They need an RNA intermediate to transpose (see description of retro-transposition below).

The Structure of a Group II Intron

The structure of a group II intron is shown in Figure 2.

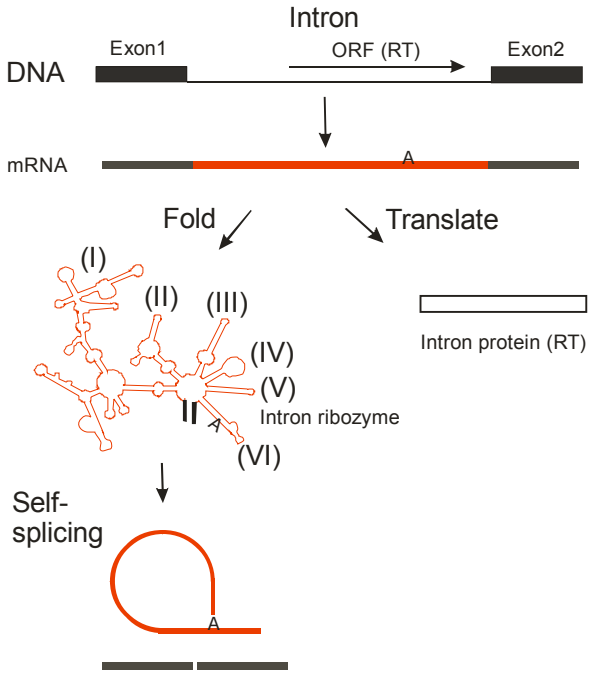


Figure 2 Structure of a Group II Intron

The RNA Molecule

The DNA segment containing the intron (thin black line, Figure 2) is transcribed into mRNA, and then the intron RNA sequence (red line, Figure 2) folds into a conserved secondary structure consisting of six domains (labeled I-VI, Figure 2). This folded RNA structure forms the intron ribozyme. The whole secondary structure can excise (self-splice) out from mRNA and form a typical lariat form (red loop structure, Figure 2). The two exons can join together to form a functional gene (black bar segments, Figure 2). Also, the intron RNA sequence can be translated into an intron protein which has reverse transcriptase activity (open black bar, Figure 2).

Therefore, a group II intron codes for two functional molecules: The first functional molecule is a ribozyme that comes from mRNA. Like nuclear spliceosomal introns, group II introns are spliced by two sequential transesterification reactions that produce ligated exons and an excised intron lariat with a 2'-5' phosphodiester bond (Lambowitz and Zimmerly 2004). The splicing reaction of group II introns is typically characterized by two transesterification steps (Figure 3). The first step is initiated by the attack of an unpaired intron-internal adenosine (A, Figure 3), located close to the 3' end of the intron, on the 5' splicing site, resulting in a free 5' exon and a branched intron-3' exon intermediate. In the second step, the 5' exon attacks the 3' splicing site, leading to the ligation of the exons and the release of the intron to a lariat form, this process is also called branch-point splicing (Lehmann and Schmidt 2003; Toro et al. 2007). In the case of a group II intron, the splicing reactions are catalyzed by the intron itself.

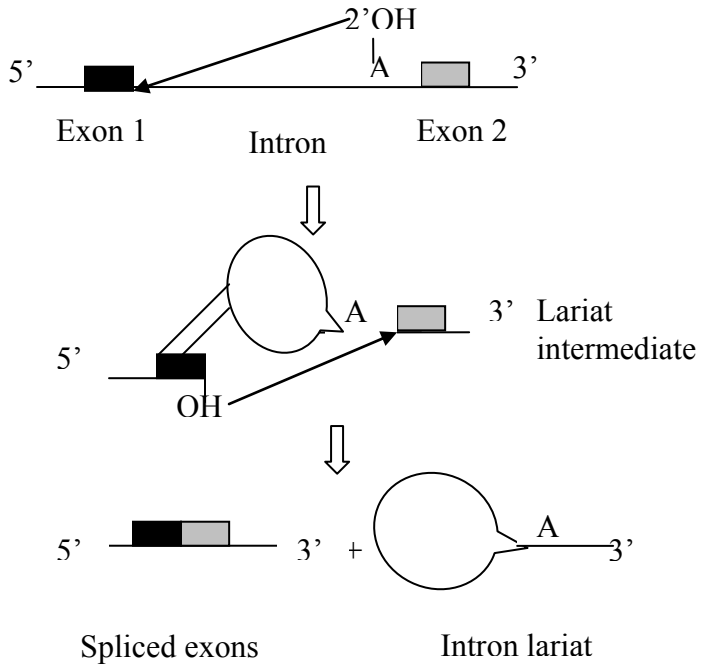


Figure 3 Group II Intron RNA Splicing Reaction Mechanism

(Adapted from Toro et al. 2007)

In human cells when an intron is spliced out by the spliceosomal complex, it also forms the lariat structure suggesting some evolutionary conservation. The spliceosomal complex is an assembly of RNA and protein molecules within the nucleus that performs splicing of mRNA in eukaryotes.

The Intron Protein

The second functional molecule is the intron-encoded protein (IEP). The intron RNA sequence is translated into a protein that has reverse transcriptase activity. The protein usually contains four domains: Reverse transcriptase (RT), maturase (X), DNA binding (D), and DNA endonuclease (En) (Lehmann and Schmidt 2003; Lambowitz and

Zimmerly 2004). The typical open reading frame (ORF) of a group II intron is located in intron domain IV. The N-terminal part of ORF contains the RT domain which contains conserved amino acid sequence blocks RT-1 to 7. The RT domain contributes to retro-transposition of group II introns. Domain X, downstream of the RT domain, contains a poorly conserved sequence region. The so called maturase activity is found in this region of the protein and can help the intron splice. So far, the maturase domain is present in all known IEPs of group II introns. This provides a very important function of domain X in RNA splicing. The RT and X domains bind the intron RNA together to initiate RNA splicing and as a template for reverse transcription (Lambowitz and Zimmerly 2004). At the C-terminal end of the IEP are DNA binding (D) and endonuclease (En) domains. They are not necessary for RNA splicing but important in intron mobility. The D domain is poorly conserved in sequence and the same region also includes a conserved DNA En domain that helps the intron insert into a specific location in a DNA molecule (Lambowitz and Zimmerly 2004). But more than half of the bacterial group II intron IEPs lack the En domain and the D domain (Toro et al. 2007).

Retrotransposons

Group II introns are also transposons that can move from one location to a new location by two different mechanisms. First, they are retrotransposons that can "reverse splice" and insert back into a specific DNA site within an allele that lacks the corresponding intron in a process called homing (Figure 4). Second, on some occasions, group II introns can also insert into nonhomologous sites in DNA in a process called

retrotranspositon; this occurs at low frequency (typically 10^{-4} or 10^{-5}) (Lambowitz and Zimmerly 2004).

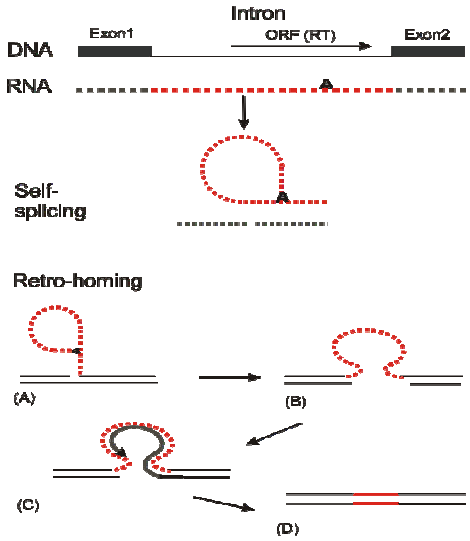


Figure 4 Introns are Mobile DNAs

Intron RNA sequence (red line, Figure 4) can self-splice out from mRNA to form a typical lariat shape (red loop, Figure 4) and use a branching splicing mechanism as described before. Then the spliced intron uses an intriguing mobility mechanism, known as target-primed reverse transcription (TPRT) for retrohoming (Dai and Zimmerly 2002b; Lambowitz and Zimmerly 2004). A specific target sequence is required for retro-homing and occurs at ~30 bp from target site. Also, flanking exon sequences that are located downstream of the intron for a short distance and upstream of the intron for a longer variable distance are needed for retro-homing (Dai and Zimmerly 2002b; Lambowitz and Zimmerly 2004). The IEP binds to the intron to form an active ribonucleoprotein particle (RNP) that recognizes specific sequences in the DNA target site and the intron lariat recognizes some nucleotides in the target DNA near the intron insertion site and other

nucleotides in the homing site are recognized by the IEP (Dai and Zimmerly 2002b). The sense strand cleavage is catalyzed by the intron RNA reverse splicing reaction and the antisense strand of DNA was cleaved by the endonuclease domain of the IEP (Dai and Zimmerly 2002b; Lambowitz and Zimmerly 2004). The intron RNA is inserted into the cleaved sense strand and acts as a template for reverse transcription and the 3'-OH of the cleaved antisense strand of the DNA can serve as a primer for reverse transcription of the inserted intron by TPRT mechanism (Dai and Zimmerly 2002b; Lambowitz and Zimmerly 2004). As mentioned previously, many bacterial group II intron IEPs lack the En domain. For En^- introns, the retrohoming process requires an alternate mechanism to prime reverse transcription. The mechanism involved in retrohoming is not understood. To some extent, nascent lagging strands will be used as primers (Zhang and Lambowitz 2003; Lambowitz and Zimmerly 2004).

Types of Group II Introns

Group II introns are classified as IIA, IIB, and IIC based on their RNA secondary structures; also IIA and IIB are further divided into A1, A2, B1, and B2 (Toor et al. 2001; Toor et al. 2006). These three classes have the same general secondary structure elements but there are significant differences in structure and function (Toor et al. 2006). The secondary structure of a group IIC intron is shown in Figure 5.

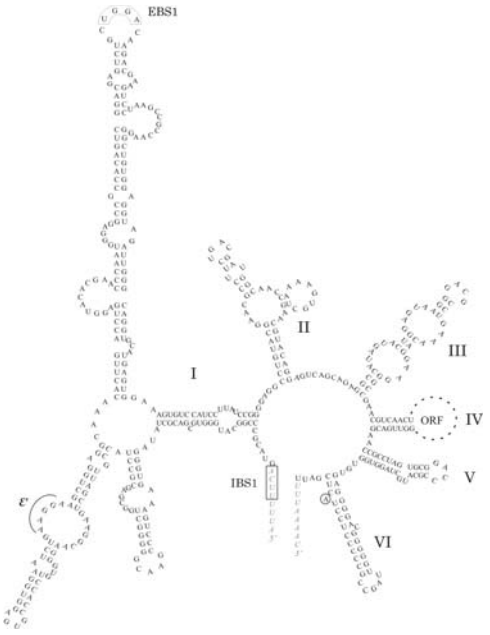


Figure 5 Postulated Secondary Structure of Group IIC Intron Ribozyme of Intron Gs. Int1

Group IIA and IIB introns share a common mechanism for 5' exon definition but differ in 3' exon definition. Group IIC introns, which are a newly recognized subclass, differ dramatically from IIA and IIB secondary structures. Group IIC introns are abbreviated in size and its distal stem domain IV (Figure 5) is 2 base pair shorter than for IIA and IIB introns, and its basal stem domain IV has a CGC in place of the highly conserved AGC triad (Toor et al. 2006). Group IIC introns have highly distinct target specificity, and they are only found in bacteria. Also, their IEPs lack an En (endonuclease) domain (Lambowitz and Zimmerly 2004; Toor et al. 2006). Group IIA and IIB introns include IBS1-EBS1 (intron binding site1 and exon binding site 2) and IBS2-EBS2 (intron binding site 2 and exon binding site2); however, group IIC intron only includes IBS1-EBS1. Based on the secondary structure of its ribozyme, *G.*

stearothermophilus strain 10 group II intron Gs. Int1 belongs to IIC type introns (Figure 5).

Intron Insertion Sites

DNA target sites can be recognized by Group II intron ribonucleoprotein particle (RNP) that is composed by intron RNA lariat and the IEP. Group II intron RNPs recognize specific sequences in the DNA target site by using both the IEP and base pairing of the intron RNA, and the base pairing of the intron RNA contributes most of the recognition of the DNA target site (Lambowitz and Zimmerly 2004). Group IIC introns typically insert downstream of palindromic rho-independent transcription terminators; also, group IIC introns are often found inserted at multiple target sites having little sequence similarity but share the palindromic terminator motif (Lambowitz and Zimmerly 2004). So far, group II introns are found widely distributed in bacteria and discovered in increasing numbers in bacterial genomes (Dai and Zimmerly 2002a). Because group II introns have splicing ability at the RNA level and they often insert after transcription terminators, they generally avoid interrupting genes that can lead to host damage (Dai and Zimmerly 2002a). Many group II introns are located outside of genes, so a selection mechanism may exist against insertion into genes.

Because group II introns have this ability to recognize DNA target sites mainly by base pairing of the intron RNA, they can serve as gene-targeting vectors for gene therapy. For example, group II introns were designed and inserted into the HIV 1 provirus and the human gene encoding CCR5, an important target site in anti-HIV therapy (Lambowitz and Zimmerly 2004).

Prior Investigations

Group II Intron Gs. Int1

A group II intron (Gs. Int1) was previously discovered in the thermophilic bacterium *G. stearothermophilus* strain 10 (Vellore et al. 2004). Based on a BLAST search of bacterial genomes, the *G. stearothermophilus* intron protein was compared with three other intron proteins from *Bacillus halodurans*, *Clostridium acetobutylicum,* and *Pseudomonas alcaligenes* (Vellore et al. 2004). Seven highly conserved amino acid domains were found. These conserved blocks of amino acids are found in all reverse transcriptases including the group II intron proteins. Also, a block of amino acids called domain X was found that is associated with the maturase function. This domain X is only found in group II intron-encoded proteins. The intron sequence was cloned into a plasmid vector for expression in *Escherichia coli.* The purified protein retains its RT activity even after exposure to 75°C (Vellore et al. 2004).

Group II Intron Gk. Int1

Another group II intron (Gk. Int1) was discovered in a related bacterium *G. kaustophilus* (Chee and Takami 2005). A housekeeping gene *recA* was interrupted by this intron. Based on the BLAST search, the *G. kaustophilus* intron protein was compared with other different proteins from bacteria and found partially similar to other bacterial group II introns.

RT-PCR was used to amplify the cDNA copy of the group II intron from total RNA extracted from *G. kaustophilus*. The size of the amplified cDNA was shorter than

that of the unspliced mRNA. So, it demonstrates that the intron Gk. Int1 can splice *in vivo*. Splicing *in vitro* was confirmed from a plasmid that contains the group II intron. After RT-PCR amplification, the size of the product was found to be equal to that of the *in vivo* splicing with total RNA from *G. kaustophilus* (Chee and Takami 2005).

Hypotheses

Group II introns are very interesting elements. First, they can code for a catalytic RNA enzyme (ribozyme) and also code for a protein product that is a type of reverse transcriptase. These properties provide group II introns to excise themselves out of an mRNA transcript without requiring a protein product. Second, group II introns are mobile DNAs. They can move from one location to a new location in a DNA molecule. This mobility requires the intron encoded RT. The group II intron from the thermophilic bacterium *G. stearothermophilus* is a new class of group II introns and its adaptation to function at hot temperatures may provide new insights into how these interesting genetic elements work.

Based on the prior investigations, there are some scientific problems I want to resolve. First, study the characteristics of the group II intron Gs.Int1 from the thermophilic bacterium *G. stearothermophilus* strain 10. Second, determine if this intron can self-splice. Third, determine if the intron can splice at hot temperatures. Finally, for the importance of this investigation, I want to find out more about this new type of group II intron, learn how proteins and especially how a ribozyme can function at hot temperatures, and learn more potential practical uses of thermostable enzymes.

I proposed the following hypotheses:

1. There maybe more than one copy of this intron in the genome of *G. stearothermophilus* strain 10.
2. This intron can be found in other related strains of *G. stearothermophilus*.
3. The intron is functional and will splice *in vivo*.

CHAPTER 2

MATERIALS AND METHODS

Bacterial Strains, Plasmids and Growth Conditions

Escherichia coli strains (DH5α, JM109) were grown by standard methods (Sambrook and Russell 2001). They were inoculated in Luria-Bertani (LB) broth (10 g tryptone, 5g yeast extract, and 5g sodium chloride per liter of distilled water) and grown overnight in a 37°C incubator with shaking. Solid culture media were made by addition of 7.5g agar to the LB broth for per liter and *E. coli* strains were grown on agar in a 37°C incubator overnight.

G. stearothermophilus strains were inoculated in LB broth or LB solid media as above. *G. stearothermophilus* were cultivated at 60 °C overnight, with shaking in the case of liquid cultures.

Antibiotics and other supplements were added into media at the following concentrations: Ampicillin 50 or 100 micrograms/milliliter (µg/ml); Kanamycin 50 microliter/milliliter (µg/ml); Chloramphenicol 30 micrograms/milliliter (µg/ml); and isopropyl-β-D-thiogalactopyranoside (IPTG) final concentration 1mM.

Isolation of Plasmid DNA and Genomic DNA

Many methods have been used to isolate plasmid DNA. All of them involve three basic steps: growth of bacteria and amplification of the plasmid; harvesting and lysis of the bacteria; and purification of the plasmid DNA.

In this study, the isolation of plasmid DNA was using the alkaline lysis method of DNA isolation as described in Sambrook and Russell (2001). *Escherichia coli* strains containing the plasmid wanted were grown overnight in selective medium containing appropriate antibiotics. Preparations of *G. stearothermophilus* chromosome from strain 10 and ATCC 12980 were carried out using a protocol described in this section.

Isolation of Plasmid DNA from *E. coli* cells

The method used was the same as that described by Moretz (2003).

Isolation of Genomic DNA from *G. stearothermophilus* Cells

Genomic DNA was isolated by first growing *G. stearothermophilus* on a LB agar plate at 55 °C. Individual colonies were scraped from the LB plate with a loop and transferred to a sterile microcentrifuge tube containing 500 µl 50 mM Tris-HCL, pH 8.0 until the buffer was very turbid with cells. The cell suspension was centrifuged at 14,000 rpm for 45 seconds, and the supernatant was discarded. The cell pellet was re-suspended in 350 µl TES buffer (50 mM Tris-HCL, 1 mM EDTA and 25 mM sucrose, pH 7.4). 50 µl of freshly made lysozyme solution (100 mg/ml in TES) was added to partially degrade the cell wall. The mixture solution was incubated at 37 °C for 30 minutes in a water bath with vortexing every 10 minutes. After the initial incubation, 20 µl of 20% SDS and 7.5µl of proteinase K (20 mg/ml in H_2O) were added to completely lyse the cell. The mixture was incubated at 37 °C for 30 minutes in a water bath with gentle mixing every 10 minutes. 400 µl TE saturated phenol (pH 8.0) was added to the above mixture that was extracted once (mix them thoroughly with hands but not too rough). It was centrifuged at

14,000 rpm for 5 minutes and the aqueous layer was removed to a new sterile microcentrifuge tube. Chloroform: isoamyl alcohol (24:1) 400 µl was added to extract twice. The chromosome was then precipitated from the remaining aqueous phase by adding 1 ml cold 100% ethanol. The genomic DNA was then spooled out using a pipette tip to a new sterile microcentrifuge tube with 400 µl of 70% ethanol. It was rinsed by inverting gently and spun at 14,000 rpm for 25 seconds to remove all liquid. The genomic DNA was air dried briefly and then re-dissolved in 100µl of 0.1X TE buffer containing RNase (20mg/ml).

Preparation of Competent Cells and Transformation Conditions

Calcium Chloride ($CaCl_2$) Competent Cells

Preparation of competent cells for *E. coli* strain JM109 or DH5α was started by a seed culture. A single colony was inoculated to 2 ml of LB broth then grown at 37°C with shaking overnight. Next day, 1 ml of the seed culture was transfered to 40 ml of LB broth in a side-armed flask to grow to early log phase. The Klett is about 20 and the OD 600nm to 0.4. The culture was chilled on ice for 5 minutes. Then the culture was transfered to a sterile oakridge tube and centrifuged at 5,000 rpm for 5 minutes at 4°C. The supernatant was discarded and the cell pellet was re-suspended in 20 ml of cold 10 mM $MgSO_4$. The mixture was placed on ice for 15 minutes. The cells were then centrifuged at 5, 000 rpm for 5 minutes at 4°C. The supernatant was discarded and the cell pellet was re-suspended in 20 ml of cold 50 mM $CaCl_2$- 10 mM Tris, pH 8.0. The cell mixture was placed on ice for 15 minutes. Next, the cells were centrifuged and the

cell pellet was re-suspended in 2 ml of sterile cold 50 mM $CaCl_2$- 10 mM Tris (pH 8.0) + 20% glycerol. At last, about 110µl of solution in small tubes (0.5 ml) were frozen at -80°C.

Transformation Conditions

$CaCl_2$ Competent cells were removed from -80°C to ice to thaw. After the competent cells thawed, 50 µl of competent cells were mixed with a 2 µl of plasmid DNA and 10 µl of sterile distilled water in a sterile 5 ml Falcon ® plastic tube and allowed to place on ice for 15-30 minutes. The mixture was heat shocked at 42 °C for 45 seconds and then placed at room temperature for 5 minutes. 0.5 ml of LB broth was added to the heat shocked mixture in the tube and then incubated at 37 °C for 1 hour with shaking. Then undiluted, 10^{-1}, 10^{-2}, and 10^{-3} dilutions of transformed cells were spread in 50 µl aliquots on LB agar containing the appropriate antibiotic and incubated at 37°C overnight.

Total RNA Extraction from *G. stearothermophilus*

RNA Extraction Method 1

Total RNAs were extracted from *G. stearothermophilus* strain 10 and strain ATCC 12980 by the methods Chee and Takami (2005) and Igo and Losick (1986).

RNA extraction was started by inoculating a single colony into a 2 ml LB broth and then the culture was grown in a 60 °C water bath with shaking at 200 rpm overnight. The next day, 0.5 ml of the seed culture was transferred to a 30 ml LB broth in a sterile

side-armed flask and then grown in a 60 °C water bath with shaking at 200 rpm until the OD 600nm reached 1.0 and the klett to 55.The culture was quickly cooled in an ice bath and then the cells were harvested by centrifuging the culture at 6,000 rpm at 4°C for 5 minutes. The cells were washed in 10 mM Tris buffer (pH 7.4). After the cells were washed, several components were added to lyse the cells: 1 ml of baked 0.45-0.5 mm (in diameter) glass beads (225°C, overnight), 2 ml of LETS buffer (100 mM LiCl, 10 mM EDTA, 10 mM Tris-HCl (pH 7.4), 1% w/v SDS, and 2 ml of phenol: chloroform: isoamyl alcohol (25: 24: 1). The mixture was then vortexed completely for 4 minutes and then centrifuged at 6,000 rpm at 4 °C for 5 minutes. Three layers appeared. The bottom layer was glass beads and the middle layer was some protein stuff and a top aqueous layer. The top liquid phase was removed carefully by pipette tip to a new tube and then 1.8 ml phenol: chloroform: isoamyl alcohol (25: 24: 1) was added to re-extract the RNA. The mixture was vortexed completely and centrifuged at 6,000 rpm at 4 °C for 5 minutes. The top liquid layer was removed to a new microcentrifuge tube. The total RNA was precipitated by adding 1/10 volume 2 M of LiCl solution plus 2.5 volume of ethanol. The microcentrifuge tube was then put at -20°C overnight.

RNA Extraction Method 2

Cells from a 50 ml culture (in LB broth at OD 600nm 0.6~0.7) were washed in 50 mM Tris pH 8.0, then aliquoted into 1.5 ml microcentrifuge tubes (about 2.5 ml of culture per microcentrufuge tube), and then the cell pellet was frozen at -80°C. 100 µl lysozyme soln (20% sucrose, 150mM NaCl, 1mM EDTA, 0.4mg/ml fresh lysozyme) was added to the frozen cell pellet. The mixture was vortexed and incubated at 37°C for 15

minutes. The following components were pre-warmed in a 65°C water bath and added one by one. After pre-warmed, 200 µl 4M Guanidine Soln (4M guanidine throcyanate, 50 mM Tris, pH 7.5, 4% sarkosyl Soln, 1% Beta-Mercaptoethanol); 100 µl Buffer I (100 mM NaOAc, 10 mM Tris, pH 7.5, 1 mM EDTA); 200 µl Acid Phenol (phenol equilibrated with 50 mM NaOAc pH 4.0); 400 µl Chloroform: isoamyl alcohol 24:1 were added. The microcentrifuge tube was vortexed and put on ice for 10 minutes. It was then centrifuged at 14,000 rpm for 5 minutes and the aqueous phase was collected to a new sterile microcentrifuge tube. 200 µl Phenol and 200 µl Chloroform were added to extract the aqueous twice followed by 400 µl Chloroform: isoamyl alcohol 24:1 to extract twice. Then 2 volumes EtOH was added and stored at -80°C overnight. It was centrifuged at 14,000 rpm for 20 minutes and the pellet was re-dissolved in 20 ~30 µl RNase-free H_2O (DEPC treated).

RNA Gel Electrophoresis

The RNA sample was precipitated overnight and centrifuged at 14,000 rpm for 20 minutes at 4°C. The supernatant was discarded and the pellet was re-suspended in 25 µl RNase-free water (DEPC treated) plus 0.25 µl of RNasin (RNasin 40 U/ µl). The RNA marker was removed from ultra freezer (-80°C) to ice to thaw. Following re-suspension, 5 µl RNA sample, 10 µl RNA sample buffer, and 2 µl RNA loading buffer were mixed to one new sterile microcentrifuge tube. Also, 1 µl RNA maker, 10 µl RNA sample buffer, and 2 µl RNA loading buffer were mixed to another new sterile microcentrifuge tube. The two tubes were placed into 65°C water bath to heat for 7 minutes, and then they were put on ice for 2 minutes followed by centrifuge for 10 seconds. The RNA samples

were separated by electrophoresis in a 1.3% agarose gel in autoclaved 1 X TAE buffer (40 mM Tris, 20 mM Acetic acid, 1 mM EDTA, pH 7.6). The gel was run at 100-125 volts for 60 minutes in 1 X TAE buffer (autoclaved), and then RNA bands were checked directly on UV trans-illuminator.

RNA Sample Buffer:

10 mls Formamide

3.5 mls Formaldehyde (37%)

2.0 mls 5x MOPS buffer (0.2 M MOPS, pH 7.0, 50 mM NaOAc, 5 mM EDTA)

RNA Loading Buffer:

50% glycerol

1mM EDTA

1mg/ml EtBr

Preparation of cDNA from total RNA of G. *stearothermophilus*

Total purified RNA 10 µg from G. *stearothermophilus* was treated with RNase-free DNase I. At first the microcentrifuge tubes with RNA were centrifuged at 14,000 rpm for 5 minutes and the liquid was aspirated completely. The pellet was dissolved in 22 µl H_2O (DEPC treated), and then 3 µl 10 X DNase I buffer and 5 µl RNase-free DNAse I (1 unit/ µl) were added to the microcentrifuge tube. The microcentrifuge tube was vortexed to mix the mixture completely, and then the mixture was incubated in a 37 °C water bath for 15 minutes. Following the incubation, 1 µl stop solution was added to the mixture, and then the mixture was incubated at 65 °C for 10 minutes, followed by the mixture was put on ice. The treated RNA was mixed with RT primer 12B43D2233

(5'-TGCCCGAGCTCGACGATTACGCTCAA-3') or 8A6D2350 (5'-TTGGATCCAGCGACATCCAAACCGAT-3'). For each microcentrifuge tube, treated RNA 15 µl, H_2O (DEPC) 15 µl, Primer 1.5 µl (20µM Soln), 3 M NaOAc (DEPC treated) 3 µl and Ethanol (for RNA) 86 µl were added. The mixture was mixed completely and then these microcentrifuge tubes were put at -20°C for 1 hour. The Microcentrifuge tubes were centrifuged at 14,000 rpm for 20 minutes and then the pellets were rinsed in 200 µl 70%-75% Ethanol. These microcentrifuge tubes were spun again at 14,000 rpm for 5 minutes and the solution was aspirated out. Then the pellet (RNA + Primer) was re-dissolved in 15 µl H_2O (DEPC). Following the re-dissolution, 15 µl RNA+Primer, 10 µl extension mix and 2 µl MMLV-RT (400 units) were mixed and then incubated at 42 °C for 90 minutes followed by addition of 1 µl EDTA (0.25M). Then 0.5 µl RNase A (10mg/ml) was then added and the mixture was incubated at 37 °C for 15 minutes to remove any remaining RNA template. Following the incubation, 100 µl 7.5 M ammonium acetate was added and the mixture was vortexed completely. And then 125 µl of phenol-chloroform (1:1) was added to the mixture to extract. Following the extraction, 100 µl of aqueous phase cDNA was recovered and precipitated by adding 300 µl Ethanol. It was stored at -20°C for 1 hour or overnight. The tube was centrifuged at 14,000 rpm for 20 minutes and the pellet was rinsed in 500 µl 75% Ethanol followed by a spin for 5 minutes at 14,000 rpm and air dry the pellet briefly. The pellet was re-dissolved in 30 µl of H_2O (DEPC) and then it was frozen at -20 °C.

PCR of cDNA Preparation

Specific amplification of cDNA copy of intron RNA was prepared by PCR technique. In preparation for PCR, 38.5 µl PCR H_2O, 5 µl 10X Taq buffer (w/$MgCl_2$), 1µl PCR nucleotide mix (10 mM dNTPs), 1.25 µl upstream primer (20 µM), 1.25 µl downstream primer (20 µM), 0.5 µl Taq DNA Polymerase (5U/µl) and 2.5 µl cDNA (10^{-4}) were added to PCR microcentrifuge tubes and the intron was amplified in PCR thermocycler. Upstream primer 12B43U92 (5'-TACACTGCAGACGGTTTTCCTTGCAC-3') and downstream primer 12B43D2233 (5'-TGCCCGAGCTCGACGATTACGCTCAA-3') were used. Another upstream primer 8A6U45 (5'- AACTGCAGATGGAAAACATCTCGCAA-3') and downstream primer 8A6D2350 (5'-TTGGATCCAGCGACATCCAAACCGAT-3') were also be used in this experiment.

At the same time, a negative control for PCR of cDNA was set up without upstream primer.

The PCR program Bert STD1 was used:

1. 95°C 2 minutes
2. 95°C 1 minutes
3. 56°C 2 minutes
4. 72°C 3 minutes
5. Go to 2, repeat 29 cycles
6. hold 4°C

After the program was finished, the PCR microcentrifuge tubes were stored at - 20°C.

DNA Electrophoresis

The DNA samples were separated by electrophoresis in 0.75% agrose gels in this study. After the DNA samples with loading dye were loaded into the gel wells, the gel was run in 1X TBE buffer (50 mM Tris, 1 mM EDTA, 45 mM boric acid, pH 8.3) at 125 volts for 1 hour. For Southern blots, chromosomal DNA digested with restriction endonucleases was loaded into the gel wells as above and was run in 1X TBE buffer at 13 volts for a minimum of 16 hours to ensure proper separation of DNA fragments. Gels were stained in 15 μg/ml ethidium bromide for 15 minutes then UV light was used to check the DNA bands. The gels were washed with water with shaking for 20 minutes then the gels pictures were taken.

For smaller size of DNA molecules, 5% polyacrylamide gels were used in 1X TBE buffer. In preparation for this, 8.6 ml H_2O, 1.2 ml 10 X TBE, 2.0 ml 29:1 Acrylamide: Bis and 100 μl 10% ammonium persulfate were added to a smaller vacuum flask to degas, then 8 μl TEMED was added to start polymerization. The solution was poured between two plates to solidify for 1 hour, and then 100 bp DNA maker and 10 μl of appropriate samples with loading dye were loaded to gel wells. The gel was run in 1X TBE buffer at 120 volts for 1 hour. Gels were stained and washed as above.

Electro-Elution of DNA

The method used was the same as that described by Moretz (2003).

Southern Blotting

The method used was the same as that described by Moretz (2003).

Random-Primed Labeling of Probe DNA with Digoxygenin-II dUTP (DIG)

The method used was similar to that described by Moretz (2003).

Determination of Yield of Probe DTG Labeled DNA

The Protocol used was the same as that described by the manufacturer (Rhoche).

Hybridization

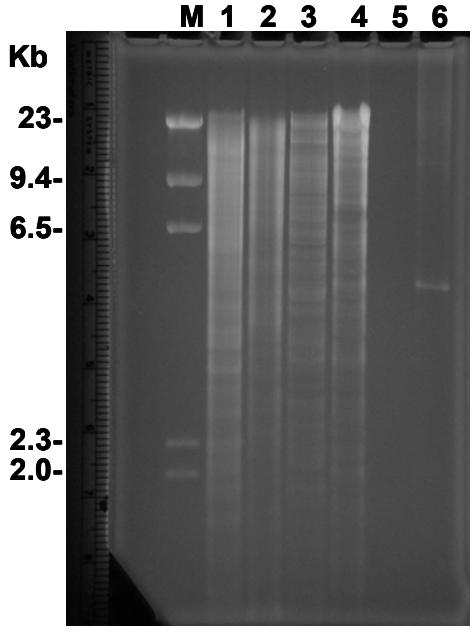
The method used was the same as that described by Moretz (2003).

CHAPTER 3

RESULTS

How many copies of the Intron are found in the chromosome?

A Southern hybridization experiment was done to detect the approximate number of copies of the intron Gs. Int1 in the chromosome of *G. stearothermophilus* strain 10 and a related ATCC 12980 strain. Genomic DNA from each strain was digested to completion with restriction endonuclease *Eco*RI or *Hin*dIII. The resulting DNA fragments were separated by electrophoresis in a 0.75% agarose gel (Figure 6). The DNA fragments from the gel were transferred to a nitrocellulose membrane via Southern blotting.



M. λ/*Hin*dIII Molecular Weight Standard
1. Strain 10 Chromosome cut with *Hin*dIII
2. Strain 10 Chromosome cut with *Eco*RI
3. ATCC 12980 Chromosome cut with *Hin*dIII
4. ATCC 12980 Chromosome cut with *Eco*RI
5. Blank
6. pBluescript-12B43

Figure 6 Agarose Gel Electrophoresis of Digested Genomic DNA from *G. stearothermophilus*

In Figure 6, lane 6 is a pBluescript plasmid containing the intron Gs. Int1 as positive control.

To detect the presence of the Gs. Int1 intron in these DNA fragments, a probe DNA containing part of the intron DNA was hybridized to the Southern blot. The probe DNA was produced by cutting out an internal *Eco*RI fragment from the intron (Figure 7) and then labeling this probe DNA by incorporation of a DIG labeled nucleotide (methods).

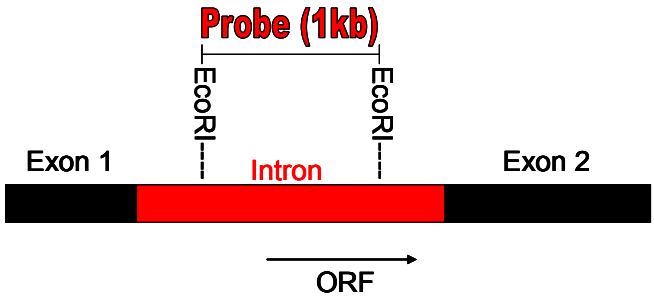
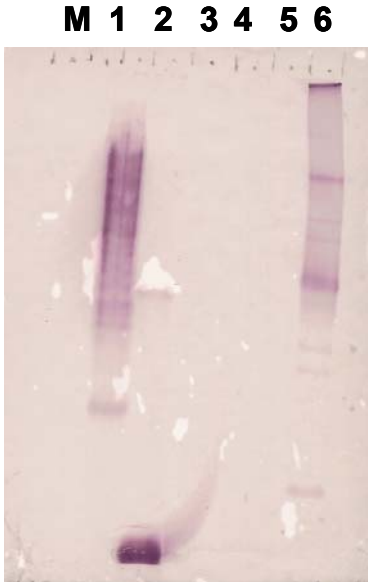


Figure 7 Design of a Probe

After the labeled DNA probe hybridized with the fragments of genomic DNA on the Southern nitrocellulose membrane, the hybridization result is shown in Figure 8. Based on this Southern hybridization result, the probe DNA hybridizes with many different restriction fragments from the genomic DNA of strain 10 (lane 1, Figure 8), but does not hybridize at all to the genomic DNA of strain ATCC 12980 (lanes 3 and 4, Figure 8). For the lane 2 (Figure 8), there is one strong band. That is because every intron in the genomic DNA has the same two *Eco*RI enzymatic sites and there are many copies of the intron in the genomic DNA. Therefore, after digestion, there are a lot of intron

DNA fragments which have the same size. After hybridized with the DIG labeled probe, one strong band appeared.



M. λ/*Hin*dIII Molecular Weight Standard
1. Strain 10 Chromosome cut with *Hin*dIII
2. Strain 10 Chromosome cut with *Eco*RI
3. ATCC 12980 Chromosome cut with *Hin*dIII
4. ATCC 12980 Chromosome cut with *Eco*RI
5. Blank
6. pBluescript-12B43

Figure 8 Southern Hybridization of an Intron Probe DNA Against Genomic DNA from *G. stearothermophilus*

Splicing *in vivo* in *G. stearothermophilus*

To detect splicing *in vivo*, total RNA was extracted from *G. stearothermophilus* strain 10 and ATCC 12980 bacterial cells grown in LB broth medium at 60°C in a water bath with shaking. RNA was isolated when the optical density of bacterial cells reached 1.0 at 600 nm (Figure 9).

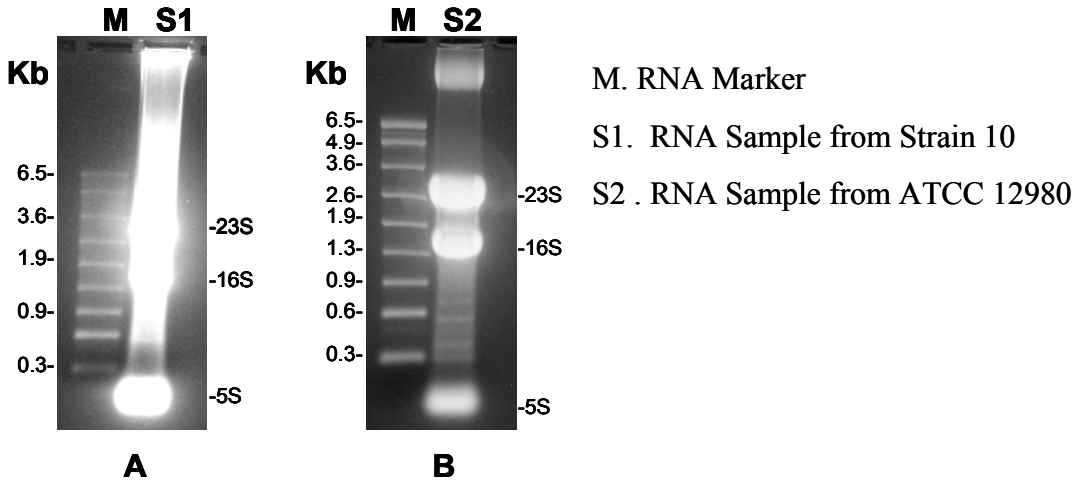


Figure 9 Agarose Gel Electrophoreses of RNAs Extracted from Strain 10 (A) and ATCC 12980 (B)

From this total RNA, a specific primer (Green arrow in Figure 10) was used to copy the intron containing mRNA into a cDNA copy by reverse transcription (Figure 10).

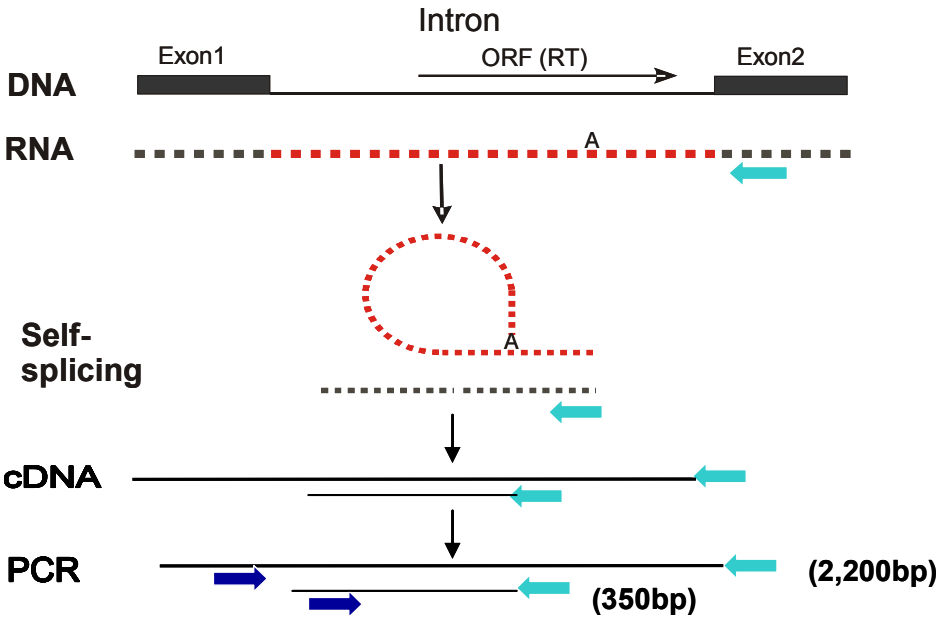


Figure 10 Method to Detect *in vivo* Splicing in *G. stearothermophilus*

PCR is then used to amplify the cDNA copy that is more stable than RNA. PCR primers are located such that they will amplify the intron sequence plus flanking exons from the cDNA copy. The PCR primers were designed based on the DNA sequence recovered from a plasmid clone (p12B43 or p8A6) of the intron from the chromosome of *G. stearothermophilus*.

Upstream primer: 5'-TACACTGCAGACGGTTTTCCTTGCAC-3' (12B43U92) and Downstream primer: 5'-TGCCCGAGCTCGACGATTACGCTCAA-3' (12B43D2233); or

Upstream primer: 5'-AACTGCAGATGGAAAACATCTCGCAA-3' (8A6U45) and downstream primer: 5'-TTGGATCCAGCGACATCCAAACCGAT-3' (8A6D2350).

There are two possible kinds of amplified DNA that could appear. One still contains the intron RNA sequence that didn't splice out (2,200bp) and the other just contains the two exons joined together (350 bp). During PCR, a sample that has no primer or no template is included as PCR negative controls. Gel electrophoresis was run to check the cDNA amplified by PCR. Two different kinds of gels were used, agarose gel and acrylamide gel. Acrylamide gel has higher resolution ability for smaller sized DNAs than agarose gel. The results after gel electrophoresis of the DNA are shown in Figure 11.

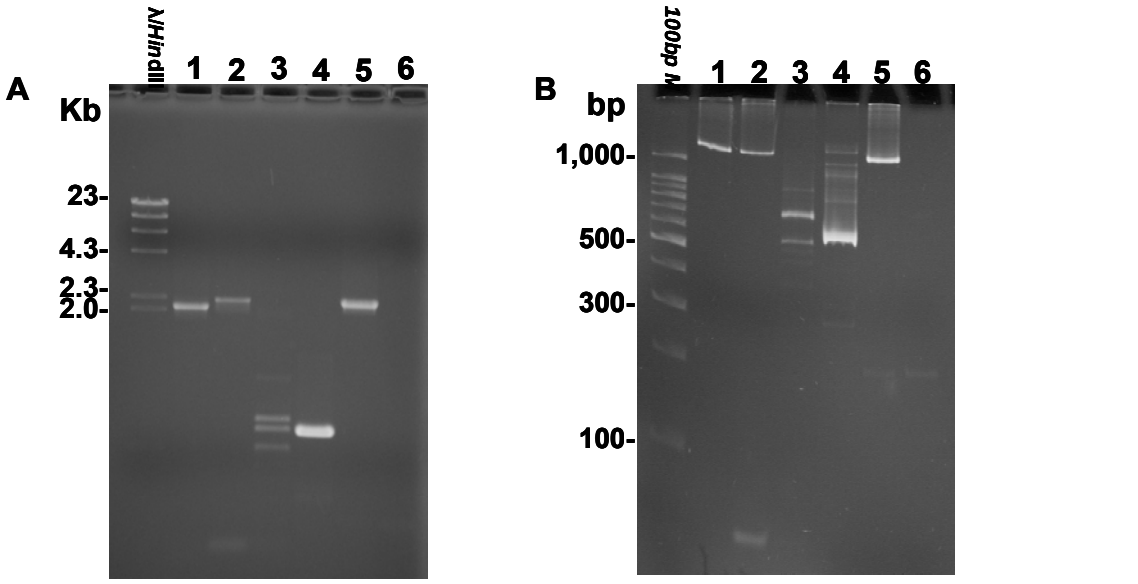


Figure 11 Detection of *in vivo* Splicing Reactions

The *G. stearothermophilus* strain 10 samples produced amplified cDNA of about 2,200 base pairs (lanes 1 and 2, Figure 11) that was bigger than the expected 350 base pairs for a spliced intron cDNA, so the intron did not splice out. Splicing *in vivo* in *G. stearothermophilus* did not appear to work. The amplified cDNA that appears from strain ATCC 12980 RNA (lanes 3 and 4, Figure 11) is about 550 base pairs in size and is also too large to be a spliced product. Also, based on the Southern hybridization experiment described above, no intron related to Gs. Int1 appears to be present in the genome of strain ATCC 12980. Thus, this amplified DNA appears to be unrelated to any intron DNA.

Constructed Plasmids for Expression of Gs. Int1 in *Escherichia coli*

The intron Gs. Int1 plus flanking exons was cloned from *G. stearothermophilus* strain 10, and then the cloned DNA was inserted into an expression plasmid pET21 (+) (Figure 12). This plasmid has a T7 promoter, so isopropyl- β-D-thiogalactopyranoside (IPTG) can be used to induce the expression of the T7 RNA polymerase. Transcription of all genes started with the binding of the T7 RNA polymerase to the T7 promoter in the expression plasmid. IPTG acts as an inducer and binds to the repressor to induce the expression of the cloned gene. The constructed plasmids allow the intron to be over-expressed in *E. coli* cells to detect splicing *in vivo* in *E. coli*.

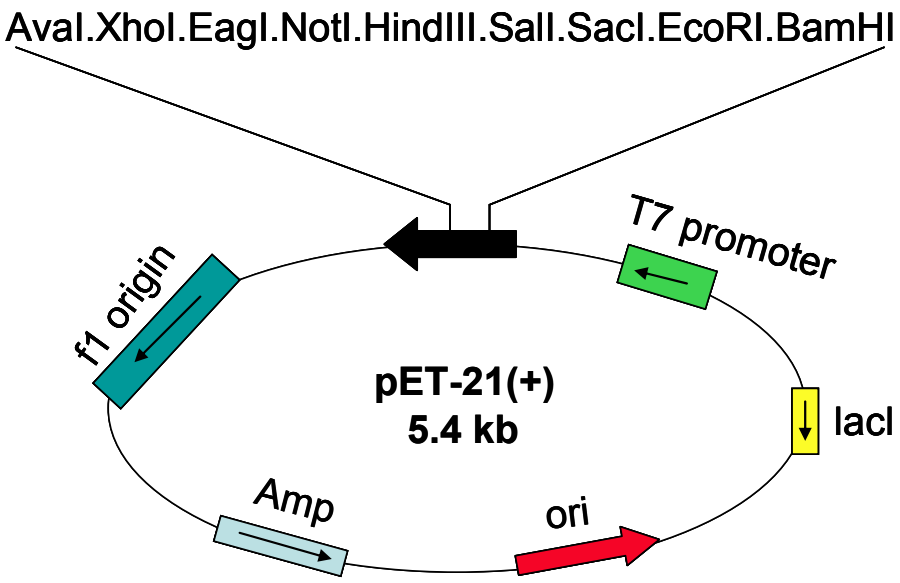


Figure 12 Restriction Map of Expression Vector pET-21(+)

Constructed plasmids for intron Gs. Int1 *in vivo* splicing in *E. coli* were set up as shown in Figure 13.

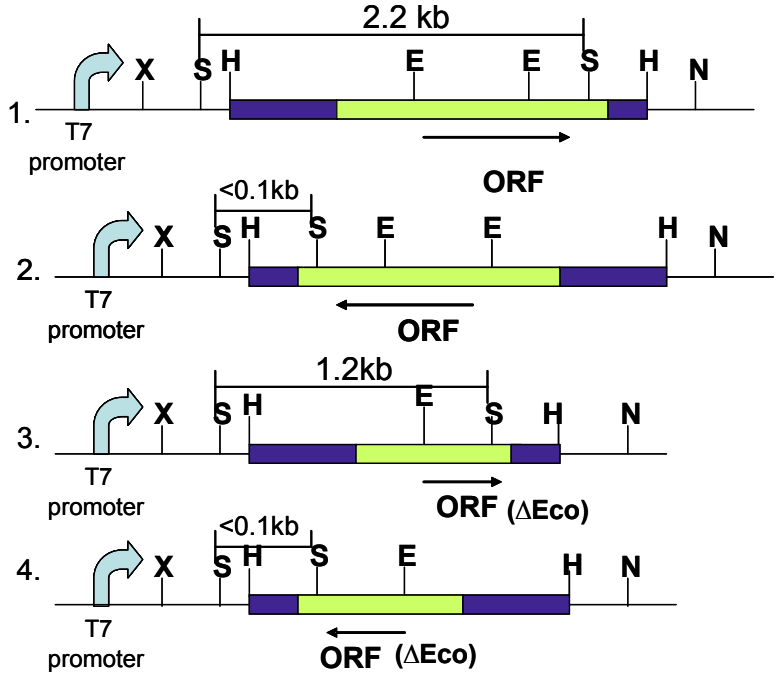
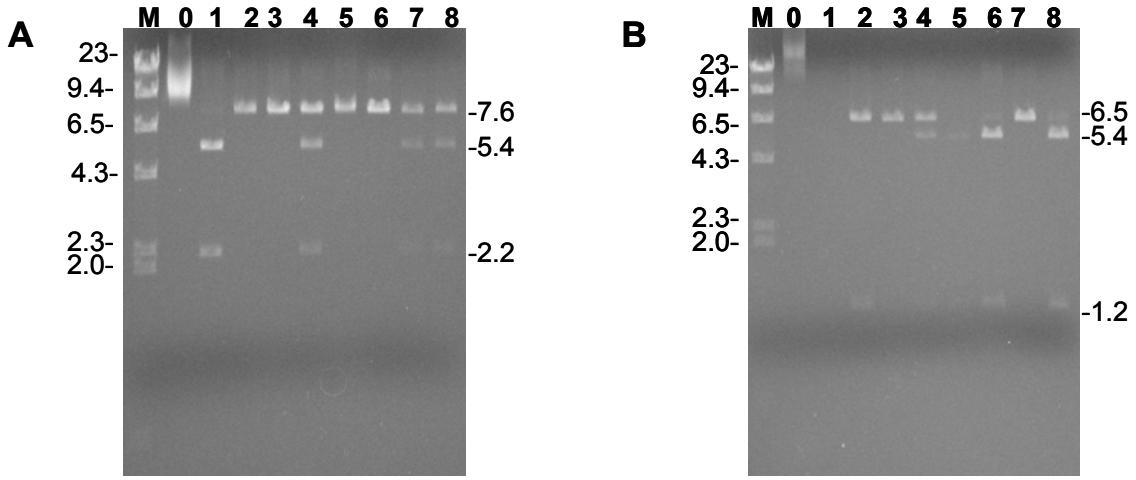


Figure 13 Constructed Plasmids for Intron Gs. Int1 Splicing *in vivo* in *E. coli.* Plasmid pET21 DNA is the thin line in drawing. Inserted intron (green) plus exons (blue) DNA is colored bars. Restriction enzyme sites are as follows: E is *Eco*RI, H is *Hin*dIII, N is *Not*I, S is *Sac*I, and X is *Xho*I.

The DNA fragment containing the intron (green bar, Figure 13) and exons (blue bars, Figure 13) was cloned from *G. stearothermophilus* strain 10. After digestion by *Hin*dIII, the DNA fragment was inserted into the expression plasmid pET21, which was also digested by *Hin*dIII. This DNA fragment was named 12B43 (about 2,200 base pairs). If the direction of the intron ORF (open reading frame) and the T7 promoter is the same, then this constructed plasmid can be expressed in *E. coli* (number 1, Figure 13).

Otherwise, if the direction is opposite to the T7 promoter, then the constructed plasmid can not be expressed in *E. coli* cells (number 2, Figure 13). So, this kind of plasmid can be used as a negative control in this experiment. These plasmid constructions were digested with the endonuclease SacI to confirm the orientation of the intron next to the T7 promoter (S sites, Figure 13). The distance between the two SacI sites is different depending on the orientation of the intron DNA in the expression plasmid. For example, if the intron ORF is in the same direction (orientation) as the T7 promoter then digestion of this plasmid with SacI will produce two restriction fragments of 5.4 kb (the pET21 vector) and 2.2 kb (the intron DNA). This is shown in Figure 14A, lanes 1, 4, 7, and 8.



M. λ/Hind III Molecular Weight Standard

A. 0 lane: pET-21+12B43 undigested

1~8 lanes: pET-21+12B43 cut with Sac I

B. 0 lane: pET-21+12B43ΔEco undigested

1~8 lanes: pET-21+12B43ΔEco cut with Sac I

Figure 14 Agarose Gel Electrophoresis of pET-21+12B43 cut with Sac I (A) and pET-21+12B43ΔEco cut with Sac I (B)

An inserted region of the intron DNA, which codes for most of the ORF, was removed by digestion of the plasmid with EcoRI followed by re-ligation of the plasmid. This plasmid was named $_{P}$12B43ΔEco (plasmid number 3, Figure 13) and the intron plus exon is about 1.2 kb. This deleted intron DNA was also ligated into the pET21 expression vector using the HindIII site. Again, the correct orientation of the inserted intron DNA was confirmed by digestion with SacI. For example, if the deleted ORF DNA 12B43ΔEco is in the same orientation of the T7 promoter then digested with SacI will give fragments of 5.4 kb (vector) and 1.2 kb (intron DNA). This is shown in Figure 14B, lanes 4, 6, and 8.

In this experiment, I planed to use *recA*[int1^{+}] group II intron as a positive control because this intron has been previously shown to have splicing ability *in vivo* in *G. kaustophilus* which also is a thermophilic bacterium (Chee and Takami 2005). pUC*recA*[int1^{+}] was donated by Dr. Chee Gab-Joo. According to the sequences of intron and flanking exons, primers were designed.

Upstream primer:

5'-CGGGATCCTTGGCAATGGCCGCAAAAC-3' (ICUP-Bam)

And downstream primers:

5'-CCCAAGCTTAGGCTTATCGCTCTCCTTGGCA-3' (ICDOWN-Hind)

5'-CGGGATCCAGGCTTATCGCTCTCCTTGGCA-3' (ICDOWN-Bam)

After PCR amplification and DNA electrophoresis, the intron *recA*[int^{+}] is shown the size about 4,000 base pairs. The PCR program Bert STD1 was used. Electro-eluted intron *recA*[int^{+}] was digested by *Bam*H1, then was ligated to pET21 that was digested by *Bam*H1 first then treated with CTP. Also, *recA*[int+] was

digested by *Hin*dIII and *Bam*H1 double enzymes first, then was ligated to pET21 which was digested by HindIII and BamH1 first then treated with CIP.

CHAPTER 4

DISCUSSION

Group II introns are very interesting elements. First, group II introns code for a catalytic RNA enzyme (ribozyme), which is a unique non-protein enzyme. It can enable the group II introns to excise themselves out of a pre-mRNA transcript without requiring protein products; this process is also called self-splicing. Second, group II introns can code for a multifunctional protein that has RT activity. RT is a very important enzyme in biotechnology and medical research. For example, it can be used in making cDNA libraries and RT-PCR (reverse transcription- polymerase chain reaction). RT-PCR is a very sensitive technique for mRNA detection and quantification. Finally, group II introns are retro-transposons that can move from one location to new locations in a DNA molecule which contributes to gene mutations in bacteria.

Intron Gs. Int1 from *G. stearothermophilus* strain 10 is a new type of group II intron that has not been studied well. Based on the intron encoded protein and conserved intron RNA secondary structures, group II introns have been divided into three main phylogenetic subclasses IIA, IIB, and IIC (Toro et al. 2007). Gs.Int1 belongs to group IIC introns that may have new properties that are different from other group IIA and group IIB introns. And all these new properties may have important applications in biotechnology. Bacterial group IIC introns appear to be inserted downstream of the stem-loop structure of rho-independent transcription terminators or other inverted repeats in bacterial genomes (Dai and Zimmerly 2002b; Toor et al. 2006).

Gs. Int1 intron was found in *G. stearothermophilus* and this intron was the first discovered in a thermophilic bacterium. There are some questions I want to know, such as how do bacteria adapt to live at hot temperatures? How do enzymes and proteins work at hot temperatures? And how does intron Gs. Int1 function at hot temperatures?

Based on the Southern hybridization experiment the intron probe DNA hybridized with multiple fragments of the *G. stearothermophilus* stain 10 chromosome. So, there are numerous copies of this intron located at different sites in the strain 10 chromosome. But there is no copy of this intron in the related ATCC 12980 strain genome. Other bacteria have also been shown to contain numerous copies of the same intron in their genome. For example, five full-length and three fragmented copies of intron B.h.I1 were found in the genome of *Bacillus halodurans* (Dai and Zimmerly 2002b); five group II introns were found in *E. coli* populations (Dai and Zimmerly 2002a).

Splicing of the intron Gs. Int1 from pre-mRNA of *G. stearothermophilus* was tested *in vivo*. Based on the splicing experiment, the intron Gs. Int1 did not appear to work. There are some possible reasons that can explain why the *in vivo* splicing did not work. For example, maybe RNA was lost during the experiment because RNA molecules are so vulnerable that they can be degraded very easily. In addition, the intron Gs. Int1 splicing could be so rare that it can not be detected. Also, I only looked at one copy of this intron and maybe this copy was defective, so the splicing reaction can not be detected.

For *in vivo* splicing in *E. coli*, the cloned intron with flanking exons was inserted into the expression plasmid pET21. The reason this expression system was chosen is the presence in this plasmid of a T7 promoter. Thus, the intron Gs. Int1 in the constructed

plasmid can be induced by IPTG to over express in *E. coli*. This will result in a lot of copies of mRNA of this intron. After cDNA is amplified by PCR, it will be easier to detect the splicing reaction. In this experiment, various controls will be used. For example, the constructed plasmid (number 2, Figure 13) in which the T7 promoter is in the opposite direction to the inserted intron ORF that can not be expressed in *E. coli,* so it can act as a negative control. The second control will be pET21-*recA*[int^{+}] which has not been completely constructed yet. The reason to choose pET21-*recA*[int^{+}] as a positive control in this experiment is because this plasmid contains the intron Gk. Int1 from *G. kaustophilus*. Both *in vivo* and *in vitro* experiments have shown this intron will splice (Chee and Takami 2005). To detect if the intron Gs. Int1 has self splice ability, the constructed plasmid (number 3, Figure 13) was set up. The region of the intron DNA that codes for most of the ORF was removed followed by re-ligation of the plasmid.

However, an over-expression system of intron Gs. Int1 for detection of *in vivo* splicing in *E. coli* still has some problems. For example, maybe the copy of this intron is defective so that the splicing reaction can not be detected in *E. coli* either. Also, the temperatures for bacterial growth are different. *G. stearothermophilus* is a themophilic bacterium with optimum growth at 60 °C. For *E. coli*, the optimum temperature for growth is 37 °C. The intron Gs. Int1 was cloned from *G. stearothermophilus*, so it should be adapted to the thermophilic bacterium environment to retain its splicing ability. It will be interesting to determine if this intron can splice at hot temperatures but this will require an *in vitro* experimental system.

For *in vivo* splicing, to my knowledge only a few of bacterial group II introns have been shown to have the autocatalytic activity *in vivo*, such as LI. LtrB from

Lactococcal lactis (Mills et al. 1996; Shearman et al. 1996); A intron from *Clostridium difficile* (Roberts et al. 2001); RmInt1 from *Sinorhizobium meliloti* (Martinez-Abarca et al. 1998); B.a.I2 from *Bacillus anthracis* (Robart et al. 2004); RIR-3 intron from *Trichodesmium erythraeum* (Meng et al. 2005); and Gk.Int1 from *Geobacillus kaustophilus* (Chee and Takami 2005). All these introns have shown *in vivo* splicing ability. It is very challenging to prove that group II introns have splicing capability *in vivo*.

REFERENCES

Bartsch, K., Schneider, R., and Schulz, A.1996. Stereospecific production of the herbicide phosphinotricin (glufosinate): purification of aspartate transaminase from *Bacillus stearothermophilus*, cloning of the corresponding gene, aspC, and application in a coupled transaminase process. ***Appl. Environ. Microbial***. 62: 3794-3799.

Chee, G-J., and Takami, H. 2005. Housekeeping *recA* gene interrupted by group II intron in the thermophilic *Geobacillus kaustophilus*. ***Gene***. 363: 211-20.

Dai, L., and Zimmerly, S. 2002a. The dispersal of five group II introns among natural populations of *Escherichia coli*. ***RNA*** 8:1294-1307.

Dai, L., and Zimmerly, S. 2002b. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. ***Nucleic Acids Rev.*** 30: 1091-1102.

Gilbert, W. 1978. Why genes in pieces. ***Nature*** 271: 5645-501.

Guo, H., karberg, M., Long, M., Jones, J.P., Sullenger, B., and Lambowitz, A.M. 2000. Group II introns designed to insert in to therapeutically relevant DNA target sites in human cells. ***Science***. 289: 452-457.

Igo, M.M., and Losick, R. 1986. Regulation of a promoter that is utilized by minor forms of RNA polymerase holoenzyme in *Bacillus subtilis*. ***J. Mol. Biol.*** 191: 615-624.

Lambowitz, A.M., and Zimmerly, S. 2004. Mobile group II introns. ***Annu. Rev. Genet.*** 38: 1-35.

Lehmann, K., and Schmidt, U. 2003. Group II Introns: Structure and Catalytic Versatility of Large Natural Ribozymes. ***Critical Rev. in Biochem. and Molecular Biol.*** 38: 249-303.

Martinez-Abarca, F., Zekri, S., and Toro, N. 1998. Characterization and splicing *in vivo* of a *Sinorhizobium meliloti* group II intron associated with particular insertion sequences of the IS630-Tc1/IS3 retroposon superfamily. ***Mol. Microbiol***. 28: 1295-1306.

McMullan, G., Chrisite, J.M., Rahman, T.J., Banat, I.M., Ternan, N.G., and Marchant, R. 2004. Habit, applications and genomics of the aerobic, thermophilic genus *Geobacillus*. ***Biochem. Soc Trans***. 32: 214-217.

Meng, Q., Wang, Y., and Liu, X.Q. 2005. An intron-encoded protein assists RNA splicing of multiple similar introns of different bacterial genes. ***J Biol Chem***. 280: 35085-35088.

Mills, D.A., McKay, L., and Dunny, G.M. 1996. Splicing of a group II intron involved in the conjugative transfer of pRS01 in *Lactococci*. *J Bacteriol*. 178: 3531-3538.

Moretz, S.E. 2003. Iron acquisition in *Rhodococcus erythropolis* strain IGTS8: characterization of a mutant that does not produce a siderophore. Thesis, East Tennessee State University, Johnson City, TN.

Nazina, T. N., Tourova, T. P., Poltaraus, A.B., Novikova, E.V., Grigoryan, A.A., Ivanova, A. E., Lysenko, A. M., Petunyaka, V.V. , Osipov, G.A., . Belyaev, S.S., and Ivanov, M.V. 2001. Taxonomic study of aerobic thermophilic bacilli: description of *Geobacillus subterraneus* gen. nov. , sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kanstophilus, Bacillus thermoglucosidasins* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G .thermocatenulatus, G. thermoleovorans, G. kanstophilus, G. thermoglucosidasins,* and *G. thermodenitrificans. Int J Syst Evol Microbiol.* 51: 433-446.

Robart, A.R., Montgomery, N.K., Smith, K.L., and Zimmerly, S. 2004. Principles of 3' splice site selection and alternative splicing for an unusual group II intron from *Bacillus anthracis. RNA*. 10: 854-862.

Roberts, A.P., Braun, V., Elchel-Streiber, C., and Mullany, P. 2001. Demonstration that the group II intron from the Clostridial conjugative transposon Tn5397 undergoes splicing *in vivo*. *J Bacteriol*. 183: 1296-1299.

Roy, S.W., and Gilbert, W. 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature reviews genetics*. 7: 211-221.

Sambrook, J., and Russell, D.W. 2001. Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor laboratory Press, Cold Spring Harbor, NY.

Shearman, C., Godon, J.J., and Gasson, M. 1996. Splicing of a group II intron in a functiional transfer gene of *Lactococcus lactis. Mol. Microbiol*. 21: 45-53.

Sookkheo, B., Sinchaikul, S., Phutrakul, S., and Chen, S.T. 2000. Purification and characterization of the highly thermostable proteases from *Bacillus stearothermophilus* TLS33. *Protein Expr. Purify*. 20: 142-151.

Stenesh, J., and Roe, B.A. 1972. DNA polymerase from mesophilic and thermophilic bacteria I-purification and properties of the enzyme. Biochem. Biophys. Acta. 272: 156-166.

Takami, H., Takaki, Y., Chee, G.J., Nishi, S., Shimamura, S., Suzuki, H., Matsui, S., and Uchiyama, I. 2004. Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. *Nucleic Acids Res*.32: 6292-6303.

Toor, N., Hausner, G., and Zimmerly, S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* 7: 1142-1152.

Toor, N., Robart, A.R., Christianson, J., and Zimmerly, S. 2006. Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res* 34:6461-6471.

Toro, N., Jimenez-Zurdo, J.I., Garcia-Rodriguez, F.M. 2007. Bacterial group II introns: not just splicing. *FEMS Microbiol Rev*. 31: 342-58.

Uma, Maheswar. Rao. JL., and Satyanarayana, T. 2003. Enhanced secretion and low temperature stabilization of a hyperthermostable and Ca^{2+}-independent alpha-amylase of *Geobacillus thermoleovorans* by surfactants. ***Lett. Appl. Microbiol.*** 36: 191-196.

Vellore, J.M., Moretz, S.E., and Lampson, B.C. 2004. A Group II Intron-Type Open Reading Frame from the Thermophile *Bacillus (Geobacillus) stearothermophilus* encodes a Heat-Stable Reverse Transcriptase. ***Appl. Environ. Microbiol***. 7: 7140-7147.

Whittle, G., and Salyers, A. A. 2002. Bacterial Transposons- An Increasingly Diverse Group of Elements. *In* "Modern Microbial Genetics". 2nd Edition. Streips UN and Yasbin RE (eds). Wiley-Liss, Inc.

Zeigler, D.R. 2001. Bacillus Genetic Stock Center, Catalog of Strains. Seventh Edition, Volume 3: The Genus *Geobacillus*.

Zhong, J., and Lambowitz, A.M. 2003. Group II intron mobity using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J* 22: 2555-4565.

VITA

HUIJING SUN

Personal data: Place of Birth: Handan, Hebei
Country: China

Education: East Tennessee State University, Johnson City, Tennessee;
Biology, M.S., 2007

Chengde Medical College, Hebei, China; Clinical Medicine,
M.D., 1999.

Professional Experiences: Graduate Research Assistant, East Tennessee State University,
Department of Health Sciences. 2005-2007

Research Assistant, Institute of Microbiology, Chinese
Academy of Sciences (IMCAS),
Department of Microbiology. 1999-2002

Researcher, Beijing Gold Chain Biotech Center,
Department of Biotechnology, 2002-2005

Publications: Sun, H., Moretz, S.E., Vellore, J.M., and Lampson, B.C. (2007)
Characterization of the Group II Intron Gs. Int1 from the
Thermophile *Geobacillus stearothermophilus*.
Abstract, H-040. 107th General Meeting of the American
Society for Microbiology, Toronto, Canada.

Sun, H., Moretz. S.E., and Lampson, B.C. (2007)
Characterization of G.S.INT1, a group IIC Intron from
the Thermophile Geobacillus stearothermophilus.
Appalachian Student Research Forum, East Tennessee
State University, Johnson City, Tennessee.