M.S. THESIS

Sequence Independent Base-excision Method using Engineered Nucleotide for Synthetic Biology

탈아미노화 염기를 이용한
유전자 서열 독립적 절단방법 연구

FEBRUARY 2017

BY

JUNGMIN KIM

DEPARTMENT OF INTERDISCIPLINARY PROGRAM FOR BIOENGINEERING
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

M.S. THESIS

Sequence Independent Base-excision Method using Engineered Nucleotide for Synthetic Biology

탈아미노화 염기를 이용한
유전자 서열 독립적 절단방법 연구

FEBRUARY 2017

BY

JUNGMIN KIM

DEPARTMENT OF INTERDISCIPLINARY PROGRAM FOR BIOENGINEERING
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Sequence Independent Base-excision Method using Engineered Nucleotide for Synthetic Biology

탈아미노화 염기를 이용한
유전자 서열 독립적 절단방법 연구

지도교수 권 성 훈

이 논문을 공학석사 학위논문으로 제출함

2017 년 2 월

서울대학교 대학원

협동과정 바이오엔지니어링 학과

바이오엔지니어링 전공

김 정 민

김정민의 공학석사 학위논문을 인준함

2017 년 2 월

위 원 장 : 백 승렬 (인)

부위원장 : 권 성훈 (인)

위 원 : 한 지숙 (인)

Abstract

Sequence Independent Base-excision Method using Engineered Nucleotide for Synthetic Biology

JUNGMIN KIM
INTERDISCIPLINARY PROGRAM
FOR BIOENGINEERING
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Restriction endonuclease(RE) is an enzyme to make cleavage site within DNA according to its recognition site. It is used to separate wanted gene fragment. However, it is difficult to cut nucleotide on demand. As RE can recognize 4 to 6 bases length of recognition site, there might be redundancy of recognition site in target DNA strand. Moreover, most of RE treatment requires additional enzyme treatment before applying to DNA ligation or molecular cloning. It is burdensome for users to be aware of enzyme recognition site and having to have to avoid repetitive recognition sites in target DNA fragment.

Here, I showed that one-base precision restriction mechanism by incorporating

unnatural base pair as a recognition site and endonuclease V which could preferentially cut unnatural base pair as a restriction enzyme. Deaminated adenosine, inosine was inserted into primer site. Amplified DNA strands, using designed primer, were targeted by specific endonuclease. It works well as a universal cleavage method. This mechanism reduces cost and time-consuming.

This novel method can be used in the field of molecular and synthetic biology. This sequence independent cleavage method requires primers containing deaminated adenosine during Polymerase chain reaction(PCR) and endonuclease V recognizing deaminated adenosine and nicking around it. This approach can facilitate removal of designated site. We can obtain DNA that consists of extra treatment of enzyme with a blunt end, which would be assembled into long gene structure or applied to diverse DNA fragments. Therefore, this method could improve oligo retrieval system at array-based synthetic technology which used Type IIs restriction endonuclease, sequence dependent cleavage system. This universal invented cleavage method can reduce labor and cost such in high-throughput scale experiment.

Keywords: Synthetic biology, Molecular methods, DNA cleavage, modified primers, Restriction enzymes, Inosine, Endonuclease V

Student Number: 2015-21211

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background of synthetic technology

.

For years, development of scalable, time and cost efficient gene assembly has been depending on quality of its starting oligonucleotides (oligos)[1]. Synthetic oligos can be produced by column-synthesized oligos and custom microarray synthesis using technology such as microfluidics, electrochemical platform and inkjet[2]. 100nt of column-synthesized oligo costs of $5~15 and 1 in 200nt error rates. Array-based oligo synthesis is cheaper than column-based oligos, 100nt costs $0.001~0.1 but error-rating is higher than column-based synthesis. Recently it is true that array

based synthesis can provide oligo pools cheaper than column based one. Therefore we expect the improvement and development in the field of synthetic biology via cheap gene synthesis.

Custom microarray synthesis requires protocols for parallel amplification and purification of complex oligo pools to improve the quality of gene assembly. Because it provides femtomolar amounts of oligos, the amplification protocol is unavoidable for assembly. Therefore they contained common flanking region when oligo pool design to secure the amount of oligonucleotide. The flanking region works as a primer site for amplification and the sequence of near flanking region can be a recognition site to be cleaved.

1.2 Retrieval target sequence from array-based synthesis

In array based synthesis, amplification step is prerequisite protocol to obtain proper amount of DNA for further application. Of course, the primer site for amplification

was designed to remove before using synthetic fragments. Generally, we use restriction enzyme to cleave between primer site and variable region. It makes total synthetic length longer than wanted. The enzyme recognition site might have overlapping regions in variable region. Such problems make researcher hard to scale up and simplify removal flanking region after amplification. To address those burdensome, I searched for an alternative restriction enzyme method. One thing is utilizing endonuclease acting like RE but sequence independently, so researchers are not required to know where restriction sites are.

Traditional restriction enzymes (RE) recognize specific sequence of 4-6 base pair in DNA double strand and make a cut in the phosphodiester backbone of DNA at a specific position with palindromic formation. In case RE mechanism is applied to retrieving synthetic sequence, such a same recognition site might exist two or more in synthetic strand. Above all, when it is supposed to assemble those fragment, palindromic formation is hard to use ligation step directly.

For efficient ligation, Type IIS restriction enzyme with unique recognition site is used to cleave common flanking region in proximity to the recognition sequence. It has advantages for users to minimize damage of variable region[3]. However, there

also might be redundancy of recognition site in target sequence so variable region can be fragmented.

For example, Kim et al. selected error-free oligonucleotides which was read by 454 sequencing using barcoded sequences[4]. The oligonucleotides were intentionally designed to have flanking sequences with two common Type IIS restriction enzyme sites, 454adapter sequence, key sequence and barcode sequence. Type IIS restriction enzyme recognition sites were targeted by specific Type IIS restriction enzyme. They used two kinds of Type IIS restriction enzyme to make a proper cleaved template for next assembly. By providing recovery blunt ended error-free oligonucleotide from analyzing data, they can assemble target gene cluster on purpose. However, such methods were so laborious and limitation of length of synthetic fragment because of required flanking region. For gene construction of 11kb DNA fragments, two kinds of Type IIS restriction enzymes were used. That means if we wanted to assemble synthesis fragments, more Type IIS restriction enzymes and its specific recognition site design are required. It would be a big waste of the expense and labor to synthetic error-free sequence as long as possible.

The other thing is that cleaved blunt ended DNA products are available to use

in assembly directly. Type IIS restriction enzyme can be a candidate material to make blunt ends. Instead of knowing complicate recognition site, Klein et al. introduced increased assembly accuracy through high throughput 'Dial-Out PCR' protocol[5]. Cleaved from array based synthetic fragment were amplified with sub-pool containing one common (outer adapters) and uracil (inner adapters) primer. It used uracil specific excision reagent (USER) and End Repair to recognize and remove uracil (inner adapters) before PCR amplification. Then sequence verified fragment were selected. They suggested a method, independent of Restriction enzyme and offered high-throughput error-free assemblies. Those methods enable to obtain verified sequence from microarray but the yield of final product is too low to use assembly step or in itself. Besides, the overall protocol is too complicate.

1.3 Sequence Independent base-excision method

Here I developed a simple and cost efficient method to remove flanking region from each synthetic fragment sequence independently after amplification which is

compensate for quite low concentration of array synthesis.

Prokaryotic endonuclease V recognizes preferentially deamination of adenosine, Inosine as a mismatch and makes a cleavage site at the second phosphodiester bond 3′ to inosine for removal of deaminated adenosine[6]. I applied this mechanism to array-based synthesis. Amplification protocol of each synthetic sequence uses designed primer containing Inosine bases. Then, those amplicons are digested by Endonuclease V to cleave flanking region without depending on sequence information.

Therefore, I designed primer containing Inosine bases and amplified synthetic template with such primers. Endonuclease V can recognize deaminated inosine and cleave the second phosphodiester bond 3’ of inosine. On the complimentary strand, left long 3’ overhanging primer sites were digested by T4 DNA polymerase which has 3’ to 5’ exonuclease ability. Finally, two enzymes and unnatural base pair can produce blunt ends DNA (figure1).

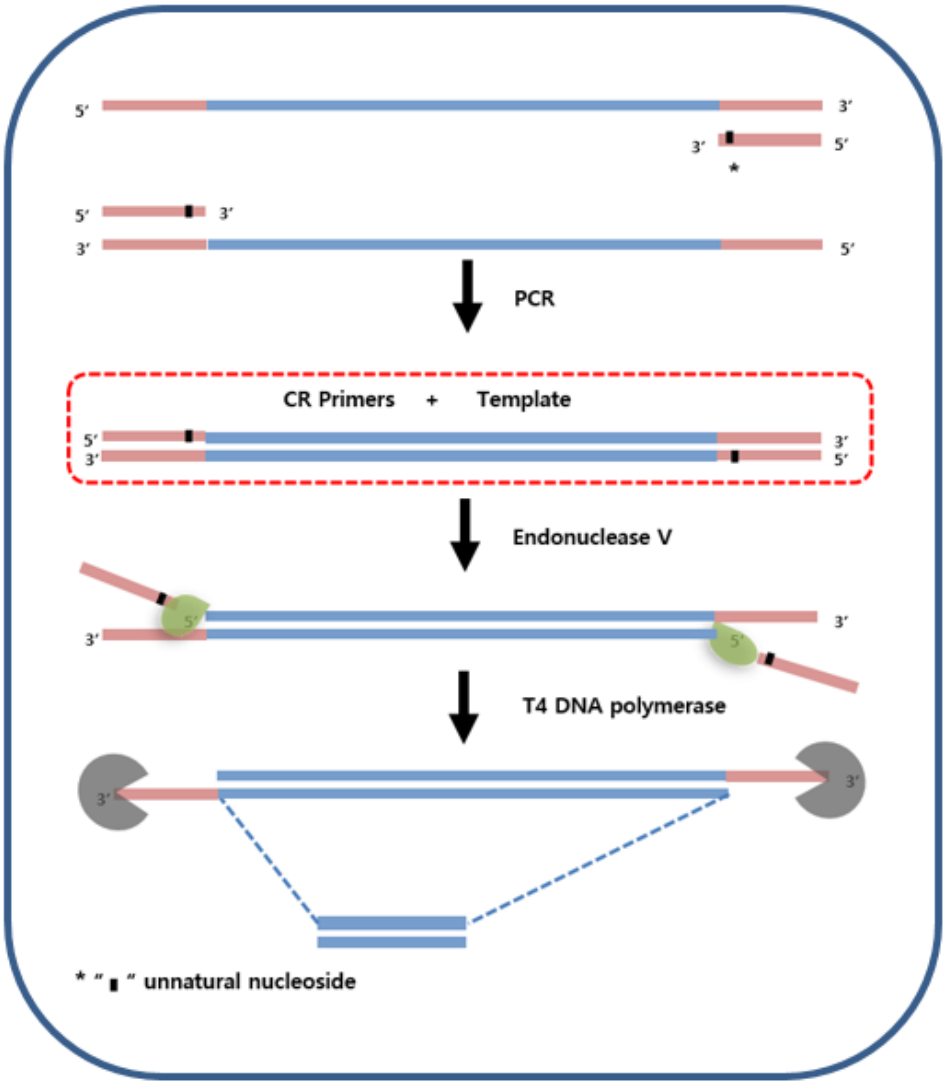

Figure 1.1 The concept of Sequence-Independent cleavage system.

1.4 Application

This mechanism can broad perspectives in the field of synthetic biology. Oligo pools derived from microarray synthesis had common primer region. Even though we use Type IIs RE, it could cleave variable region and make shorter variable region by containing Type IIs RE sequence. It would be waste of expense and time because of requirement additional enzyme treatment to make blunt ends.

More specifically, it can provide researchers more efficient starting materials for Polymerase cycling assembly (PCA), Gibson Assembly and so on. Because this base excision method can produce blunt ends DNA manually, common primer sites can serve its own ability of amplification without worrying about removing that sites. Therefore, DNA fragments cleaved by this methods allow for assembly well. If not, nonspecific fragments result in unprecedented product.

The improvement method to remove flanking region with one base pair precision could provide efficient synthetic DNA fragments. Moreover, assembly with high quality of starting oligos can be a good material by itself or resource of long genomic sequences. There are many potential applications such as the study of

genome organization, protein design and novel pharmaceuticals. The precisely designed sequences facilitate long genome assembly with low cost and error-free.

Alternatively, the precise and sequence-independent cloning method can be widely used in everywhere hard to use traditional restriction enzyme concept. Regardless of recognition site, we could cut and obtain blunt ends DNA by amplifying template with designed primer site and endonuclease function.

Chapter 2

Materials and Methods

2.1 Genomic Sequences

In this study I used genome sequence driven from Mycoplasma Genitalium (*M. genitalium*). I sheared off *M. genitalium* whole sequence every 100base pair length and 80bp overlapping region each other. 257 templates belonged to one cassette group. Total 20cassettes were prepared. Each template was designed to have common primer sites at the end of double strand DNA templates. The Forward

primer site is 5’- GTG CCT TGG CAG TCT CAG T -3' (19bp) and Reverse primer site is 5’- CGT GGA TGA GGA GCC GCA GTG T –3’ (22bp). This common primer sites on either sides are for PCR amplification and integrated other primer site having similar sequence combination but different function.

Therefore, each designed template set was synthesized with forward and reverse common primer site already. In this study, *M. genitalium* fragments (100bp) except for common primer site called target template. Briefly, I used 20 cassettes set (140bp) called polyclonal templates and one of them called monoclonal template also derived by cloning polymerase (Table 2.1). I obtained all oligonucleotides which were synthesized in our laboratory using semiconductor-based electrochemical acid production arrays (CustomArray)

Table 2.1 Mycoplasma genitalium sequence cassette set. Mycoplasma genitalium fragmented into 100bp length with 80bp overlapping region (Green color) and common primer sequence (yellow and grey) at both ends (140bp and 20cassettes set of 247)

>mg_cassette1_1

GTGCCTTGGCAGTCTCAGTTTGAACGCACTCACAGCGATTTTAAATGATTA
TAATATTTCTTTAATACTAAAAAAATACTAAGTTATTATTTAGTTAATACTTT
TAACAATATTATTAAACACTGCGGCTCCTCATCCACG

>mg_cassette1_2

GTGCCTTGGCAGTCTCAGTTTTAAATGATTATAATATTTCTTTAATACTAAA
AAAATACTAAGTTATTATTTAGTTAATACTTTTAACAATATTATTAAGGTATT
TAAAAAATACTATTACACTGCGGCTCCTCATCCACG

. . .

>mg_cassette20_255

GTGCCTTGGCAGTCTCAGTACTAAACTTTCGCAGTGACA**CTTCTTCTTTT**
ATTAGTCAATATTCCCCTAGTTTTGATGACCAATACAGTTAATTTATGA

AACTGAAGTTTTACAAACTAACACTGCGGCTCCTCATCCACG

>mg_cassette20_256

GTGCCTTGGCAGTCTCAGT**CTTCTTCTTTTATTAGTCAATATTCCCCTAG**

TTTTGATGACCAATACAGTTAATTTATGAAACTGAAGTTTTACAAACT

ACCTTTAATTAGCACGCATTGACACTGCGGCTCCTCATCCACG

>mg_cassette20_257

GTGCCTTGGCAGTCTCAGT**TTATTAGTCAATATTCCCCTAGTTTTGATG**

ACCAATACAGTTAATTTATGAAACTGAAGTTTTACAAACTACCTTTAAT

TAGCACGCATTGATTCCTTCGACACTGCGGCTCCTCATCCACG

2.2 Primer Design and Synthesis

CR primer pairs having combination of natural nucleotides and Inosine were designed to amplify each synthesized template having common primer pairs having natural nucleotides. Inosine bases incorporated in CR primer pairs used for cleaving site by unique feature of *Tma* EndoV. PCR products were cleaved at X nucleotide followed by Inosine site on double-stranded DNA. I derived 4 pairs of Forward and Reverse primers called CP(cut point) (Table 2.2). All primers are designed to contain Inosine base from none to four and they are replaced former natural nucleotides without affecting its length. Usually, Guanine base sites are exchanged into Inosine base. All primers ordered from Integrated DNA Technology (IDT) Technologies (Coralville, IA, USA).

Table 2.2 Composition of Inosine primer sets

Primer name	Primer sequence
CP 1 Forward	5′- GTG CCT TGG CAG TCT CAI T -3' (19bp)
CP 1 Reverse	5′- CGT GGA TGA GGA GCC GCA GTI T -3′ (22bp)
CP 2 Forward	5′- GTI CCT TG I CAG TCT CAI T -3' (19bp)
CP 2 Reverse	5′- CGT GI A TGA GGA ICC GCA GTI T -3′ (22bp)
CP 3 Forward	5′- GTI CCT TGI CAG TCT CA 3deoxyl -3' (18bp)
CP 3 Reverse	5′- CGT GI A TGA GGA ICC GCA GT 3deoxyl -3′ (21bp)
CP 4 Forward	5′- GTI CCT TIG CA I TCT CAI T -3' (19bp)
CP 4 Reverse	5′- T GIA TGA GIA GCC ICA GTI T -3′ (20bp)

2.3 Template preparation by PCR based amplification

In this study, *Taq* DNA polymerase was used to amplify synthetic polyclonal templates with Inosine inserted primer (CP primer). This polymerase has no proofreading function during transcription, so I can expect almost amplified templates that contain CP primer. Templates were also amplified with DNA polymerases having proofreading function. Because the use of a DNA polymerase with proofreading capability are required in larger DNA fragment for high sequence precision. If templates were not fully integrated by CP primer set, this templates have no potential to make a cleavage site between template and CP primer site.

I chose three popular proofreading DNA polymerases, especially KAPA HiFi HotStart Uracil+ Ready Mix PCR Kit by Kapa Biosystems (Woburn, MA) and Phusion U Hot Start PCR Master Mix (Thermo Scientific) are engineered to tolerate uracil, RNA or inosine bases residues. PfuTurbo Cx HotStart DNA polymerase(Agilent Technologies) is a mutant of Pfu DNA polymerase to increase proofreading activity to counter mutations introduced by uracil read-through. Those

proofreading function of DNA polymerase can extend thorough a template even containing a uracil, RNA or Inosine bases while correcting mismatch base pairs with high efficiency. Therefore, I tried to figure out which DNA polymerases with proofreading function possesses high ability of incorporating CP primer into targeted template.

2.4 Endonuclease V

Escherichia coli Endonuclease V (*E.coli* EndoV) was introduced to recognize and do deoxyadenosine 3' endonuclease in strands by recognizing deoxyinosine site on single and double strand DNA[6]. *Thermotoga maritima* endonuclease V (*Tma* EndoV) also do similar capability. Oligonucleotides were tested with *E.coli* EndoV and *Tma* Endo V. Both can recognize and nick the second phosphodiester bond 3' to deoxyinosine. I used *Tma* EndoV obtained from Thermo Fisher Scientific (St. Leon-Rot, Germany), DNA treatments were performed at 65°C for 30 min. *E.coli* EndoV were obtained from New England Biolabs(NEB) and active condition for 45min at

37°C. Inosine is derived from loss of the exocyclic amino group in adenosine and pairs with cytosine and thymine representing a less base pair. Prokaryotic Endonuclease V is a DNA repair enzyme making a cleavage site at the second phosphodiester bond 3′ to inosine for removal of deaminated Adenosine(Inosine) [7].

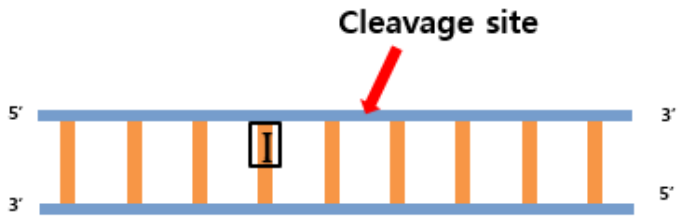

Figure 2.1 Endonuclease V cleaves at the second phosphodiester bond 3' to Inosine.

E.coli(Escherichia coli) EndoV recognizes deoxyinosine both in double and single stranded DNA. More specifically, it realizes deoxyinosine in DNA and cleaves the DNA backbone one base offset on the 3’ side of inosine with a separate catalytic active site[8]. *Tma* EndoV(Thermotoga maritima) can cleave mismatches of DNA substrates engaged in removal inosine[9]. In this study I use *Tma* Endo V as repair enzyme specialized in Inosine base.

2.5 T4 DNA polymerase

T4 DNA Polymerase catalyzes the synthesis of DNA in the 5´→ 3´ direction and has a 3´→ 5´ exonuclease activity. I used T4 DNA polymerase(Thermo Scientific) at second step enzyme to trim the 3’ overhang strand and help to make a blunt-ended DNA fragment[10].

2.6 Template amplification and cleavage procedure

All templates were amplified using diverse primer set containing inosine with multiple DNA polymerases and purified with QIAGEN MinElute PCR purification kit (QIAGEN, Valencia, CA, USA) to exclude primer dimer and non-specific product from PCR. *Taq* DNA polymerase is used for PCR based amplification, Reactions contained 0.7ng *M. genitalium* genomic DNA and 1pM each CP primer set (Table 2) was performed in 50uL of solution. The PCR condition required: (i)

95℃ for 2 min, (ii) 95℃ for 30 s, (iii) AT depends on primers℃ for 20 s, (iv) 72℃ for 30 s 10-15 cycles of PCR amplification at ii-iv and (v) 72℃ 2 min. Annealing temperature (AT) varies depending on its primers recommendation. Target template amplified with *Taq* DNA polymerase were purified with QIAGEN MinElute PCR purification kit and eluted in 15ul. The PCR product was treated with *Tma* (*T. matirima*) Endonuclease V and incubated at 65C for 30min. Reactions were then purified and eluted in 15ul. Then T4 DNA polymerase can remove 3' overhang sequence on double strand DNA by its 3' to 5' exonuclease.

In case of KAPA HiFi HotStart Uracil+ ReadyMix, PCR condition required: (i) 95℃ for 2 min, (ii) 98℃ for 20 s, (iii) AT depends on primers℃ for 15 s, (iv) 72℃ for 30 s 10-15 cycles of PCR amplification at ii-iv and (v) 72℃ 2 min. Phusion U Hot Start DNA Polymerase, amplification of template was performed following condition: (i) 95℃ for 2 min, (ii) 98℃ for 20 s, (iii) AT depends on primers℃ for 20 s, (iv) 72℃ for 30 s 10-15 cycles of PCR amplification at ii-iv and (v) 72℃ 5 min. For PfuTurbo Cx Hotstart DNA polymerase, PCR condition required: (i) 95℃ for 2 min, (ii) 95℃ for 30 s, (iii) AT depends on primers℃ for 30 s, (iv) 72℃ for 1 min 10-15 cycles of PCR amplification at ii-iv and (v) 72℃ 5

min. Agarose gel electrophoresis are used to get the quality and quantity of DNA fragment. Usually, I prepared in the condition of 2.5% agarose gels with 120V around 60-90minutes for high resolution.

2.7 Sanger sequencing and Next generation sequencing

Target templates which were amplified with *Taq* DNA polymerase using CR primer were cleaved by *Tma* EndoV and T4 DNA polymerase. Then Alkaline Phosphatase, Calf Intestinal(CIP) obtained from New England Biolabs(NEB) were treated to remove 5´- and 3´- phosphates from DNA before preparing insert TOPO vector. Those products were sent to Sanger sequencing after TOPO cloning(Figure 2.1).

For Illumina sequencing products which had blunt-ended fragment by *Tma* EndoV and T4 DNA polymerase were prepared. This product were inserted vector template and did transform plasmid DNA into *E.coli* competent cell. This cells were cultured in liquid LB and then, passed to plasmid preparation process. Kapa HiFi

Readymix(Kapa Biosystems) were obtained for next generation sequencing(NGS) library amplification. The sequence information was analyzed by Macrogen Inc. (Republic of Korea) and Illumina Mi-Seq respectively.

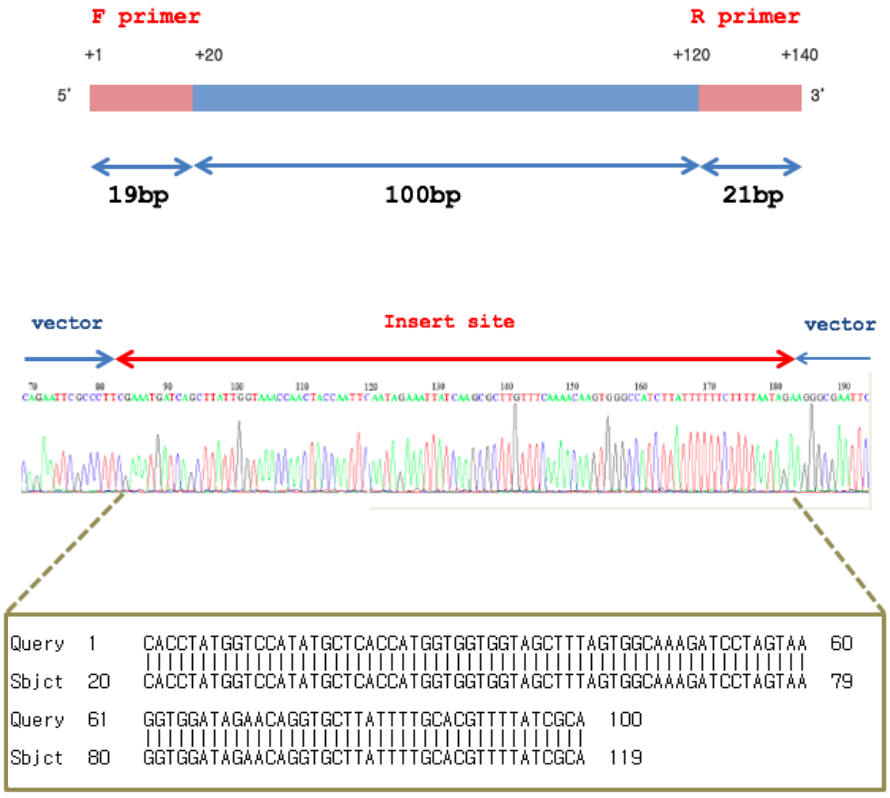

Figure 2.2 Primer sites and target variable region.

Chapter 3

Result

3.1 Sequence-Independent cleavage system

Taq DNA polymerase was used to amplify templates with common primer, CR1, 2, and 3. Common primer was already synthesized by customary, so common primer was annealed to its complementary sequence and extended. This polyclonal templates were prepared to cut its primer containing Inosine base pair. I did design

CP primer to represent the *Tma* EndoV cleavage site. CP primer has Inosine base by increasing Inosine base one by one. CP primer 1 has one Inosine base before 3' end natural nucleotide (Thymine).

CP primer 2 had three Inosine bases placed between five and seven natural nucleotides. CP primer 3 had also three Inosine bases placed between five and eight, but 3' strands were end of Inosine. It intends to digest primer region clearly. Finally, 3' end of CP primer 4 was designed not to disturb *M. genitalium* genome sequence. *Tma* EndoV could recognize one nucleotide off Inosine base and cut phosphodiester on DNA strand so, I thought CP primer 4 represents its high efficiency to detach only primer region from *M. genitalium* fragment. Figure one indicated which CP primer set could be used to precise cleavage method.

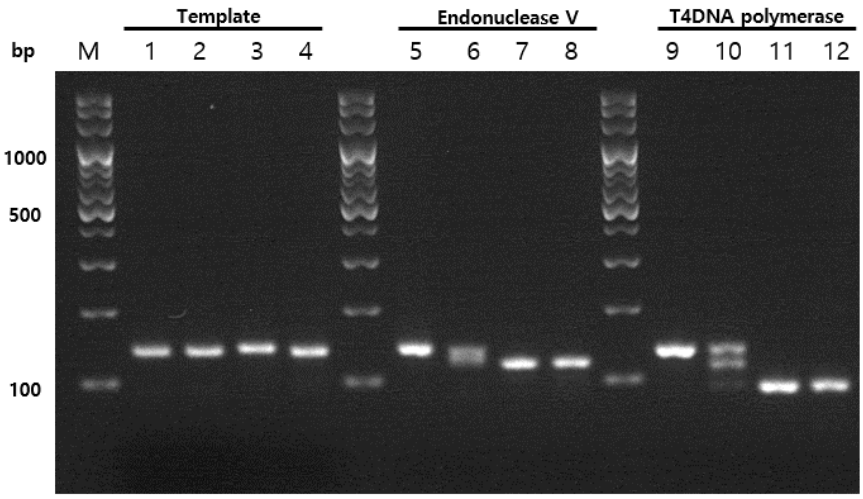

Figure 3.1 Electrophoresis gel result. Templates were cleaved by Endonuclease V which specifically recognized Inosine and made a nick at the DNA backbone and T4DNA polymerase that digests 3' overhang sequence.

Figure 3.1 shows that the (Lanes 1-4) is PCR amplicon with the same size using Common primer, CR primer1, 2 and 3 respectively. (Lanes 5-8) is cleaved at DNA backbone at one base offset of the Inosine to remove it (Figure 3.1). Therefore, lane 5 is the control sample from *Tma* EndoV and (lanes 6-8) indicates CR primers are cleaved by *Tma* EndoV. Lane 9 is the control sample from T4 DNA polymerase treatment. Lane 10 using CR primer1 produced non-cleaved fragment and cleaved fragment because there is only one Inosine base at primer. Some fragments were not digested by T4 DNA polymerase (exo) because former treatment of *Tma* EndoV was not enough to leave short length of fragmented primer which is detachable during *Tma* incubation temperature. It means that nicked strands by *Tma* EndoV can leave 3' overhanging region which is complimentary strand to CR primer. However, Lanes 11 and 12 indicate T4 DNA polymerase treat results in templates(~100bp) except for CR primer Forward and Reverse(~40bp). Then I sequenced this final product by Sanger sequencing and Mi-Seq to check the DNA length. The sequencing results from Sanger sequencing showed where they were cleaved by enzyme reaction. All templates were amplified with *Taq* DNA polymerase. Cutting ratio of flanking site was different depending on the CR primer sets. 73.7%(14 of 19) of the templates

using the CR primer 2 were cut and 93.8%(15 of 16) when the CR primer 3 was used. The number of inserted Inosine bases and the interval of Inosine base were same for both CR primer set, but the CR primer 3 was ended with deoxyinosine at the 3' end. This different could make more efficient to cleave flanking region from template. The CR primer 4 set having four Inosine bases were cut at 100%(8 of 8). From those result of Sanger sequencing, I prepared templates amplified with *Taq* DNA polymerase using CR primer 4 set. Two samples of the same protocol were tested to compare the cleavage method through Illumina Mi-Seq. Almost of all templates (98.97% of sample 1 and 2) successfully cut its flanking site having some inosine bases(Table3.1). This means that inosine base can be effectively recognized and cleaved by endonuclease V and T4 DNA polymerase with high efficiency. The unclearly cleaved results might be caused by deletion error (0.5~1%) of synthetic primer from IDT company.

Table 3.1 Illumina Mi-Seq Result of Template amplified by *Taq* DNA polymerase using CR primer 4

Sample 1		R	
		Cut	Uncut
F	Cut	**94631**	732
	Un-cut	2	0

Sample 2		R	
		Cut	Uncut
F	Cut	**88649**	361
	Un-cut	815	1

3.2 Manipulate enzyme reaction condition

3.2.1 *Tma* Endonuclease V and T4 DNA polymerase

Tma EndoV enzyme reactions were performed at 65C for 30min. T4 DNA polymerase enzyme reactions were performed at 37C 30min for incorporation of deoxyribonucleotides into a fraction or at 11C for 20min or room temperature for 5min to exert its 3' to 5' exonuclease function. I prepared PCR based template using CR3 primer set with *Taq* DNA polymerase. Then, I tried to find proper incubation time and temperature for prepared template to be cut.

First, I checked the recommended temperature for each enzyme is proper. Second, I studied that skipping the purification step between *Tma* EndoV and T4 DNA polymerse could impact on buffer condition of second enzyme reaction. If there were no difference between purification and non-purification sample, it can reduce purification cost and time-consuming. Finally, we can simplify experiment procedure, provided that I can find optimal temperature which controls two different enzymes at the same tube.

3.2.2 Matching temperature between *Tma* Endonuclease V and T4 DNA polymerase

Tma EndoV recommends its incubation temperature at 65c but we incubated at diverse temperature condition. Usually, Enzyme providers give more generous protocol to users but we need to test and define experimental protocol. Figure 3.2. A (Lanes1-4) is incubated at 25C, 35C, 50C and 65C respectively. Then, we add T4DNA polymerase at 25c 20min commonly. Electrophoresis gel image showed that cleaved fragment. Lane1 and 2 were placed more above then lane 3 and 4. That is *Tma* EndoV cannot work under 35C incubation condition. However, in the range of incubation temperature between 50c and 65c, *Tma* EndoV can recognize and cut second phosphodiester bond 3' to deoxyinosine in primer like. Then, nicked single strands of primer got a chance to be detached and removed clearly thorough PCR purification kit. Therefore, left complimentary primer region were digested by T4 DNA polymerase's exonuclease function.

T4 DNA polymerase is performed at 37C for 30min for incorporation of deoxy-ribonucleotides into a DNA substrate and 11C 20min or 25C 5min for exonuclease

function to cleave 3’ overhangs. Therefore, I tried to find the range of temperature that this enzyme could exert exonuclease function, if possible. I incubated template which is amplified with CR3 primer in *Taq* DNA polymerase. Various temperature conditions were a range with start at 25C and end at 65C. Starting temperature at 25C is tested for T4 DNA polymerase’s exonuclease activity and 65C is for *Tma* EndoV active temperature which is enzyme used in first step. Therefore, the band result demonstrated T4 DNA polymerase can stand its 3’ to 5’ exonuclease activity until 65C but there was some different efficiency of exonuclease (Figure 3.2.A). Later I would like to use this temperature condition to diminish enzyme treatment process. Figure 3.2.A (lanes 5-6) were all cut and placed at 100bp even though there might be some different enzyme activity. Progressively, I could narrow the temperature range between two enzymes. *Tma* EndoV actives at 50C - 65C and T4 DNA polymerase actives at 25C - 65C.

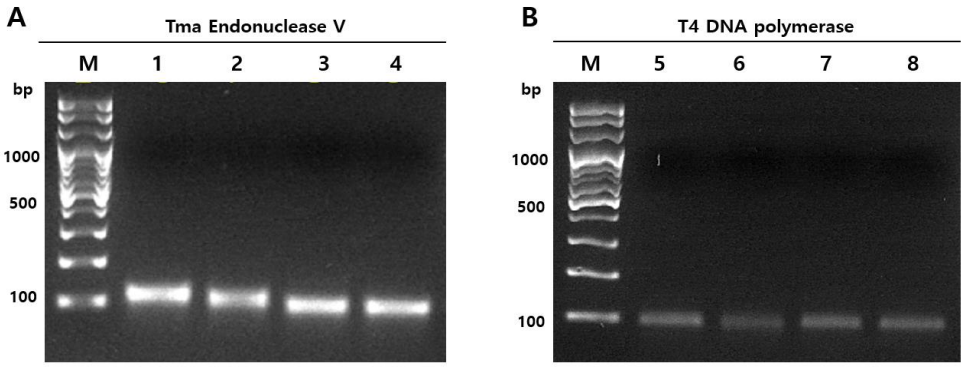

Figure 3.2 *Tma* EndoV and T4 DNA polymerase activity with temperature variation. Figure 3.2.A (Lanes1-4) is incubated at 25C, 35C, 50C and 65C respectively. Then, we add T4DNA polymerase at 25c 20min commonly. . Lane1 and 2 were placed more above then lane 3 and 4. Figure 3.2.B (lanes 5-6) T4 DNA polymerase activity at various temperature. T4 DNA polymerse recommended incubation condition is at 11C for 20min or room temperature for 5min to exert its 3' to 5' exonuclease function. Lane 5 is at 25C, lane 6 is 35C, lane7 is at 50C and lane8 is at 65C which intended to compare *Tma* EndoV condition.

3.2.3 Non-Purification and Buffer mixed condition

PCR Purification step is used to remove salt, nucleotides, enzymes and primer from PCR product. Cleaned up DNA sample is needed for next enzyme treatment, because rest of buffer pH can affect other enzyme activity. In high throughput experiment, the purification step will cost big money. QIAGEN PCR purification kit which is popular PCR purification kit costs $2.2 per sample. Moreover, purification step is difficult to be fully automated, so it is inevitable take a more effort and time. Therefore, if there is no different in between purified and non-purified sample, omitting purification step provides users to go down in price and reduce time-consuming.

The pH condition of *Tma* EndoV is pH7.4. The pH condition if T4 DNA polymerase is pH8.8. Reaction buffer pH is critical to enzyme activity. Therefore, the purification step was required before treatment of second enzyme. First, templates having CR3 primer were prepared and nicked by *Tma* EndoV at 2, 3, 4 band result (Figure 3.4). Then, 3 was just digested with T4 DNA polymerase and 4 was purified and then added T4 DNA polymerase. There was no different between

lane 3 and 4 at the cleaved length. The condition which containing both *Tma* EndoV buffer and T4 DNA polymerase buffer at 5, 6, 7 and 8 was intended to observe that the pH condition of mixed buffer could affect to enzyme activity. Lane 6 condition showed that mixed buffer compound did not inhibit *Tma* EndoV ability and rest of T4 DNA polymerase buffer tolerated at 65C. Therefore, I got cleaved band result after treating T4 DNA polymerase and dNTP. Lane 7 was treated with *Tma* EndoV in the condition of mixed two buffers. The imperfect result band appeared. Lane 8 was conducted to like lane 7 condition. Then put T4 DNA polymerase and buffer after purification. This reaction reflect two buffer's mixed condition has no influence in enzyme ability (Figure 3.3).

Achieved by the experiment, mixed buffer condition of *Tma* EndoV and T4 DNA polymerase has no different in band result. I obtained truncated sequence with same band length both purification and non-purification sample. Oligonucleotides prepared without purification have potential to enable even more convenient process. Consequently, omitting purification step between two enzyme treatments provides users fast and cost efficient product of flanking sequences.

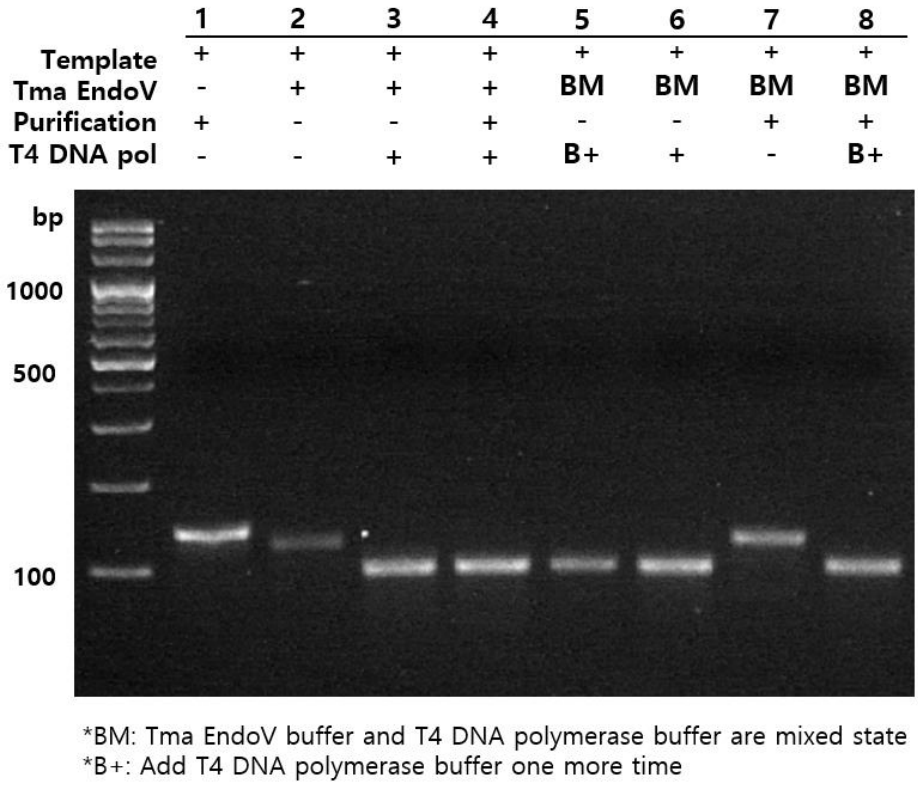

Figure 3.3 Non purification and buffer mixed experiment. Lane 1 was template using CR3 primer (140bp). The lane 2 was treated with only *Tma* EndoV. The lane 3 and 4 was sequentially treated by *Tma* EndoV and T4 DNA polymerase but different in purification. Lane 5-8 were already mixed with *Tma* EndoV and T4 DNA polymerase buffer. The 5 sample was added with T4 DNA polymerase, dNTP and buffer without purification. The 6 sample was added with T4 DNA polymerase and dNTP without purification. The 7 and 8 sample were purified. Then, 8 sample was added with T4 DNA polymerase, dNTP and buffer.

3.2.4 One shot reaction

Non purification and buffer condition were not affect for enzyme ability to recognize and digest oligonucleotides. Therefore, I expected ‘One shot reaction’ to put two enzymes and substrates together is possible. It is necessary to find optimal temperature and time for ‘One shot reaction’. I have been used 700ng of template substrate with 5unit of *Tma* EndoV which is enough to digest prepared substrate (Actual recommended amount of substrate is one ug). Then, put one micro liter of T4 DNA polymerase and dNTP each. Two buffer mixture is prepared correspond to result from (Figure 3.4). All thing that I mention is put in 100ul reactions. Then, conduct at various temperature conditions. Lane 1 at 50C for 30min and 25C 20min, Lane 2 at 40C for 30min and 25C for 20min, Lane 3 at 40C for 30min and 12C for 20min and lane 4 at 45C for 30min and 25C for 20min. All temperature condition was set based on (Figure 3.4) result. I used electrophoresis and get band image at 2.5% agarose gel at 120V for 100min. The optimal treatment for template using CR primer 3 is lane 2 condition at 40C for 30min (Figure3.5).

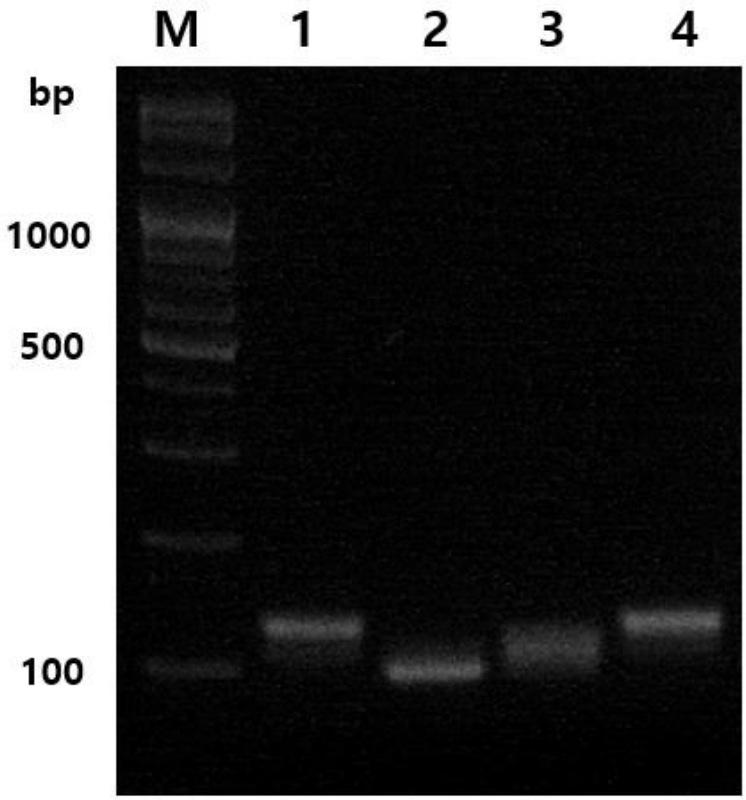

Figure 3.4 One shot reaction. *Tma* EndoV, T4 DNA polymerase, dNTP and two buffer mixture is mixed in 100ul reactions. . Lane 1 at 50C for 30min and 25C for 20min, Lane 2 at 40C for 30min and 25C for 20min, Lane 3 at 40C for 30min and 12C for 20min and lane 4 at 45C for 30min and 25C for 20min.

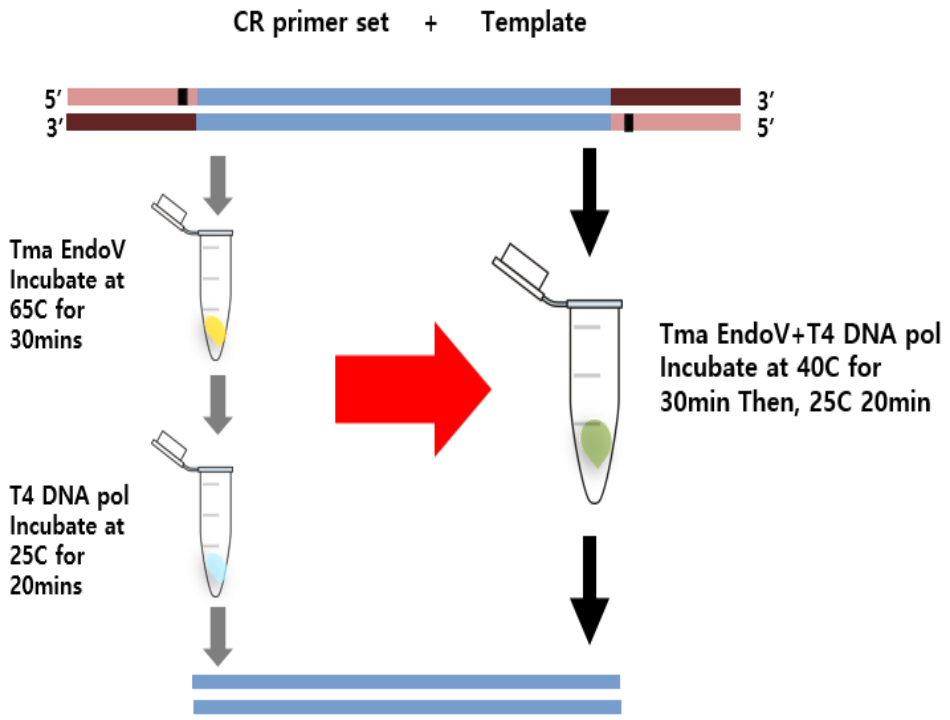

Figure 3.5 One shot reaction. Two enzyme treatment steps and purification step can be reduced to one step reaction. It shows that this cleavage method can be utilized more simply and user friendly with cost effective.

Table 3.2 Sequencing Result for One shot reaction.

One shot rxn	CR3	CR4
Cut (%) Sanger Seq	40% (4 of 11)	100% (12 of 12)

Among the variable condition, Lane 2 at 40C for 30min and 25C for 20min is the best condition. I conducted experiment with templates amplified with *Taq* DNA polymerase using CR primer 3 and CR primer 4 in this optimal treatment condition. The band results were placed at 100bp length but the sequence result had some difference (Table 3.2). In case of CR primer 3 templates were clearly detached its flanking region at 40% (4 of 11). Incompletely cleaved templates had usually one side of primer region and the length is from 108bp to 117bp (Before treatment length is 140bp). This is because such low incubation temperature was not enough to induce denaturation of primer fragment. However, in case of CR4 primer set templates were cleaved at 100%. This CR4 primer set has Inosine bases between 3-5 nucleotides, so primer fragments could be denatured at even 40C.

This one reaction step was intended to omit laborious purification step between enzyme treatments and increase experiment efficiency. Also simple method has advantages for users

3.3 Improvement of Enzyme cost efficient

When enzyme digests substrate, it is profitable for users to react to maximum amount of substrate with minimum enzyme volume. Above all, in this study I annealed CR primer containing Inosine and cut it again after template amplification. s advantage of this method is that it can generate cleavage site without Restriction enzyme (RE) and recognition sequence. Therefore, this methods can be useful to retrieve amplicon in microarray based technology such as synthetic oligo pool in microarray or PCR reactions for quantifying oligos. Traditionally, different RE was prepared so, the efficiency of each enzyme was not necessary. However, this method using *Tma* EndoV treats target templates commonly. Accordingly, to test for lager amount of substrate can reduce enzyme and other experimental material cost. In the

case of assembly of amplicon in one reaction by Polymerase cycling assembly (PCA) we put all together in a one tube and treat enzymes at the same time.

Tma EndoV costs $1.6 for 5 unit per micro-liter. 5 unit can digest DNA substrates about one μg of depurinated supercoiled plasmid DNA in product protocol (*Tma* EndoV, Thermo Scientific). However, such enzyme definition is imperfect, so I tried to figure out experimental enzyme activity based on 140bp length of double stranded DNA. Template amplified *Taq* DNA polymerase using CR primer3 were digested by *Tma* EndoV and T4 DNA polymerase. By increasing DNA amount gradually, *Tma* EndoV showed no activity at 15ug of DNA. There was not enough nicked at phosphate backbone to react with exonuclease. Lane 15ug placed at template band length (Figure3.6). 5U of *Tma* EndoV can recognize and cleave strands about 11ug of 140bp blunt-end oligonucleotides. However, 12 micro grams of template were smearing DNA band result. T4 DNA polymerase is enough to digest 3'overhangs so, 12 micrograms of template is too much to react with 5 unit of *Tma* EndoV. This helps us to eliminate enzyme cost and increase in experimental efficiency and materials like tube, pipet and so on.

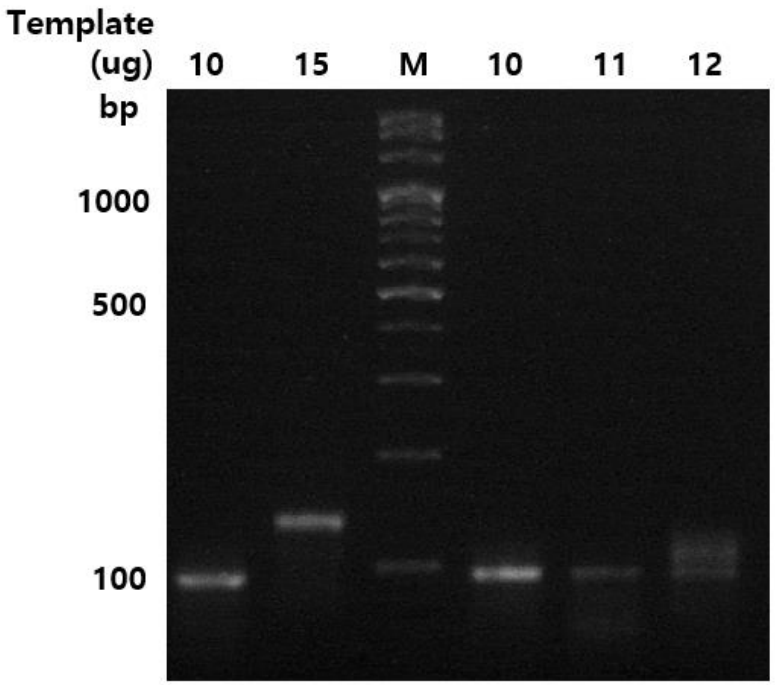

Figure 3.6 Test for *Tma* Endonuclease V activity at different amount of DNA substrates. Various amount of template using CR3 primer (140bp) were treated with *Tma* EndoV and T4 DNA polymerase. Lane of 10, 11 micro grams of template were fully digested but 12 micro grams of template were smearing DNA band result. T4 DNA polymerase is enough to digest 3'overhangs so, 12 micrograms of template is too much to react with 5 unit of *Tma* EndoV.

3.4 Multi-polymerase cleavage system

Inosine primer sets were used to synthetic template which has a pair of common primer set. *Tma* EndoV and T4 DNA polymerase helped to leave variable sequence region by cleaving primer region. However, *Taq* DNA polymerase has no proofreading function during template amplification, so there might be error in variable region. Improvement to this problem, usually we incorporate proofreading polymerase which can read mismatch sequence while performing polymerase function. *Taq* DNA polymerase could be effective in the case of using short fragment of variable sequence. In contrast, such a long sequence might occur error in oligonucleotide during PCR so, it needs proofread polymerases. Of course, error correction by using proofread polymerase can inhibit anneal Inosine base in PCR primer. Inosine base is important to be recognized as cutting point, so I tested for some provided proofread polymerase products.

KAPA HiFi HotStart Uracil+ ReadyMix and Phusion U Hot Start DNA Polymerase have been used to amplify oligonucleotides[11]. Those are engineered to tolerate uracil, RNA or inosine bases residues, so it can prevent polymerse error by correcting base mismatch. Furthermore, I selected PfuTurbo Cx Hotstart DNA

polymerase which is not yet used to tolerate Inosine base residue. However, among proofread polymerases having tolerance to uracil can stand to Inosine base so I would like to test PfuTurbo Cx Hotstart DNA polymerase.

First, three polymerases were used to PCR using CR primer set. After obtaining 140bp length of amplified oligonucleotide, I treated *Tma* EndoV to recognize Inosine base as a cutting point and T4 DNA polymerase to trim 3' overhangs. Agarose gel electrophoresis were used to show cleaved DNA fragment. Usually, I prepared in the condition of 2.5% agarose gels with 120V around 60-90minutes for high resolution. Unlike the result that we have been cutting almost primer region from amplified templates which were performed with *Taq* DNA polymerase, the band result using proofreading polymerse indicated untruncated fragments (Figure 3.7). As depicted in figure3.7 KAPA HiFi Hotstart U final band result was placed upper 100bp ladder, Phusion U DNA pol and PfuTurbo Cx Hotstart DNA polymerase were placed near 100bp ladder. I thought the unclear bands came from proofreading function. That is, Inosine bases in primer were changed to natural nucleotides during transcription. I calculated cutting ratio for three polymerases which have specially modified ability available to incorporate Uracil base(Table3.3)

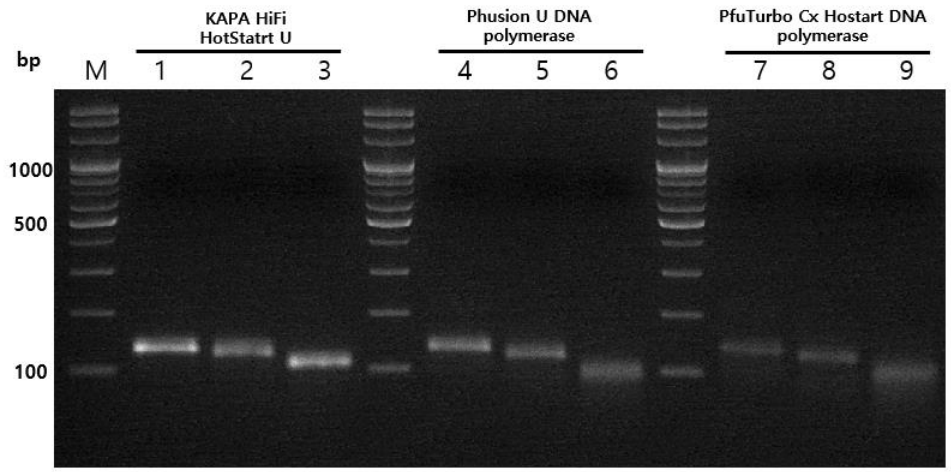

Figure 3.7 Test for Multi-polymerase cleavage system. Three proofread polymerase were tested. First lane (1, 4 and 7) denoted template(140bp). Second lane (2, 5 and 8) was treated with *Tma* EndoV. Third lane (3, 6 and 9) was final product after T4 DNA polymerase.

Table 3.3 Sequencing Result for Multi-polymerase cleavage system. Three proofread polymerase were tested.

A

CR3	KAPA U	Phusion U	Pfu Turbo Cx
Cut (%) Sanger Seq	5% (1 of 19)	29% (2 of 7)	5% (1 of 19)

B

CR4	KAPA U	Phusion U	Pfu Turbo Cx
Cut (%) Sanger Seq	48% (10 of 21)	95% (19 of 20)	93% (12 of 13)

3.4.1 longer primer set for proofreading function of DNA polymerase

The proofreading function of DNA polymerase can be used to reduce error if the wrong base is inserted. Therefore, in case of accuracy of template is important, we could take advantage of proofreading function of polymerase. I have studied sequence independent methods with templates which were amplified with various polymerase. Most of all, template amplified with Taq polymerase was cleaved at designed sites well(98.97%). However, Taq polymerase could not detect error in the sequence, so it might increase error-rate in synthetic DNA sequence fragment during PCR.

To address this problem, we used proof reading function of DNA polymerases such as KAPA HiFi HotStart Uracil+ Ready Mix, Phusion U Hot Start DNA Polymerase and PfuTurbo Cx Hotstart DNA polymerase. Though it could prevent to incorporate wrong base, such enzyme could affect inosine base in primer by considering it as a wrong base while moving along a template strands. As a result, Tma EndoV lost recognition site and could not make a nick, so the cutting ratio of amplicon can be decreased. I designed longer CR primer set (30mers) and hamper

proof reading function of DNA polymerases to exchange deaminated bases in primer site. First of all, monoclonal template (mg_cassette2_201) were prepared from *M. genitalium* polyclonal template cassette set. Then, elongated CR primer is designed to contain variable region about 10mers. The Forward sequence is 5'- GT I C CTT I GCA I TCT CA I T ACA AAG GAT GC -3'(30mers). Reverse sequence is 5'- TG I A TGA G I AG CC I C AGT I TAC TAA GCT TA-3' (30mers). I conducted experiment for PfuTurbo Cx Hotstart DNA polymerase. Among templates of PfuTurbo Cx Hotstart DNA polymerase, 93.8%(30 of 32)samples were clearly cleaved at flanking region. The cleavage ratio for PfuTurbo Cx Hotstart DNA polymerase was increased than before. As a control, templates of *Taq* DNA polymerase were also amplified using 30mers and 91.7%(11 of 12) samples were cleaved at flanking region. It means that there was no different in cutting ratio between proofreading function of DNA polymerase and lacking of proofreading function one. The Inosine bases in primer region had no change with CR primer 3 set but change with total length of primer site. It will be required further study to what would be affect the cleavage system since longer primer region would be possible to be used for proofreading function of DNA polymerase.

Conclusion

I have shown that the alternative restriction enzyme method generates desired DNA fragments. The improvement method requires PCR amplification step with designed primer containing inosine bases. The designed primer had Inosine bases at an interval of 3 to 5 nucleotides of primer region. This is intended to be nicked by Tma EndoV and then removed by melting temperature of fragmented primer. T4 DNA polymerase was used to trim 3' overhang site. Therefore, this method can be useful to cleave at any intended site of DNA sequence. Moreover, I combined all complicate steps into one shot reaction, which helps researchers save money and time. Notably, the sequence independently cleavage system provides advantages over traditional restriction enzyme.

This universal cleavage method could provide highly designed oligo products without sequence dependency. With traditional cleavage method, producing blunt

ends DNA, proper assembly material, is difficult and requires several experimental treatment. However, the new method is suitable for producing blunt ends fragments. This sequence independent base excision method will successfully facilitate the nucleotide cleavage system and de novo assembly of large DNA oligonucleotides.

I expect this novel method can be instantly exploited for retrieving synthetic oligo pools from microarray based technology. This process is effective to obtain amplified template except for primer region without any loss of template's base pair. It can save enormous cost for preparing the number of restriction enzymes and laborious work compared to traditional Type IIs restriction enzyme method. Therefore, this high-throughput method can help bring down the cost of oligomer synthesis considerably and increase experimental efficiency. Therefore, this novel methods can broaden perspectives in the field of synthetic biology like development of DNA reading technologies.

Additionally, there are many potential applications for the field of molecular biological such as the study of gene expression, protein expression or gene assembly. It could work as an improved method to remove target site with one base pair precision. Eventually, sequence independent universal cleavage method make easier manipulate DNA sequence as a cost-effective and high efficient method.

Bibliography

[1] Borovkov, A. Y., Loskutov, A. V., Robida, M. D., et al. "High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides," *Nucleic Acids Research*, 38(19), e180-e180, 2010.

[2] Tian, J., Ma, K., & Saaem, I. *Advancing high-throughput gene synthesis technology.* Molecular BioSystems, 5(7), 714-722, 2009.

[3] Pingoud A. & Jeltsch A. *Structure and function of type II restriction endonucleases.* Nucleic Acids Research 29, 3705-3727, 2001.

[4] Kim H., Han H., Ahn J., et al., "'Shotgun DNA synthesis' for the high-throughput construction of large DNA molecules," *Nucleic Acids Research*, 40:e140, 2012.

[5] Klein, J. C., Lajoie, M. J., Schwartz, J. J., et al., Multiplex pairwise

assembly of array-derived DNA oligonucleotides, *Nucleic Acids Research*, 44, e43, 2016.

[6] Yao, M., Hatahet, Z., Melamede, R. J., et al., “Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from Escherichia coli,” *Journal of Biological Chemistry*, 269(23), 16260-16268, 2016.

[7] Vik, E. S., Nawaz, M. S., Andersen, P. S, et al., “Endonuclease V cleaves at inosines in RNA,” *Nature Communications*, 4, 2271, 2013.

[8] Alseth, I., Dalhus, B. & Bjoras, M., “Inosine in DNA and RNA,” *Molecular and genetic bases of disease*, 26, 116-123, 2014.

[9] Dalhus B, Arvai AS, Rosnes I, et al., “Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair,” *Nature Structure. Mol. Biol,* 16, 138-143, 2009.

[10] Ibarra, Borja, et al. "Proofreading dynamics of a processive DNA polymerase," *The EMBO journal,* 28, 18, 2794-2802, 2009.

[11] Makarov, V., Laliberte, J., “Improved methods for processing DNA substrates,” WO 2015117040 A1, Jannuary 30, 2015.

국문 초록

탈아미노화 염기를 이용한
유전자 서열 독립적 절단방법 연구

제한효소는 전통적인 DNA 절단 방법으로 제한효소 인식 서열에 따라 원하는 위치를 절단 할 수 있어 주로 원하는 유전자 조각을 얻어내는 도구로 이용되었다. 하지만 제한효소가 절단 할 수 있는 유전자 서열이 제한효소 종류 및 특이적으로 다르기 때문에 원하는 유전자 조각을 얻기 위해서는 사용자의 번거로움과 비용이 드는 단점이 있었다. 제한효소는 보통 4-6bp 염기서열을 인식해 타깃 하는 유전자 위치를 자르기 때문에 타깃 하는 유전자 조각 내에 동일 제한효소로 인식되는 염기서열 부분이 중복되는 경우에는 원치 않은 유전자 조각을 얻게 된다. 또한 얻은 유전자 조각을 DNA ligation 또는 molecular cloning 에 이용하기 위해서는 vector molecule 과 유전자 조각이 상보적으로 결합

될 수 있도록 추가적인 효소로 조절해야 하는 번거로움이 있다. 본 연구는 현재 사용자에게 문제점으로 지적되는 제한효소 서열 인식의 한계, DNA 내 인식서열 반복성문제, 제한효소 처리 산물의 응용 시 번거로움을 해결하고자 한다.

본 연구는 염기서열 독립적으로 DNA 를 인식해 원하는 부위를 One base-pair 정확도로 절단하는 연구방법을 제시했다. 이를 위해 비자연적 염기와 이를 돌연변이로 인식하고 자르는 메커니즘을 응용했다. 비자연적 염기 중 하나인 탈아미노화 염기(이노신)과 이를 인식하고 자르는 Endonuclease V 를 통해 유전자 서열 독립적 절단방법을 고안했다. 이노신을 프라이머서열에 삽입해 타깃 하는 DNA 를 증폭하고 Endonuclease V 를 통해 증폭에 사용된 프라이머 서열만이 정확하게 잘림을 확인했다. 이는 저비용, 고효율적인 방법으로 Universal cleavage method 로 사용될 수 있음을 보여주는 실험이다.

본 연구는 분자생물학 및 합성생물학 분야에 귀중한 방법으로 사용될 가능성이 있다. 본 유전자 서열 독립적 절단방법은 탈아미노화 염기로 구성된 프라이머를 이용한 증폭과정을 거친 후 탈아미노화 염기를 인식하고 절단과정을 일으키는 효소를 통해 간단히 이뤄진다. 이와 같은 접근법은 사용자가 요구하는 DNA 절편을 PCR 및 효소 반응만을 통해

기존의 제한효소 방법보다 값싸고 정확하게 타깃 하는 유전자서열 산물을 얻을 수 있게 해준다. Gene structure 및 다양한 DNA 절편을 assembly 하는데 필요한 blunt-end 형태의 DNA 절편도 Exonuclease 효소를 추가적으로 넣어 쉽게 얻을 수 있다.

결론적으로 본 연구는 최근 유전자 합성분야에 가장 각광받는 어레이기반 합성의 단점 중 하나인 어레이기반 합성산물 추출 가격을 낮출 수 있는 방법으로 이용될 수 있다. Type IIs 제한효소의 대체제로서 서열 독립적으로 blunt end 산물을 쉽게 추출해 낼 수 있을 거라 기대된다. 또한 기존 제한효소 방법이 이용되는 분자생물학 분야에서도 high-throughput level 의 DNA 절편 획득 시에도 본 Universal cleavage method 를 통해 비용 및 노동력 절감을 기대할 수 있는 고효율 절단 방법으로 이용될 수 있다.

주요어 : 합성생물학, Molecular methods, DNA cleavage, modified primers, 제한효소, Inosine, Endonuclease V

학번 : 2015-21211