공 학 박 사 학 위 논 문

Development and application of multiple genome editing method using CRISPR/Cas9 system and ribozyme in *Escherichia coli*

대장균에서의 CRISPR/Cas9 시스템 및 라이보자임을 이용한 다중 게놈 편집 방법 구축 및 응용

2018년 8월

서울대학교 대학원

공과대학 바이오엔지니어링전공

김 민 우

Development and application of multiple genome editing method using CRISPR/Cas9 system and ribozyme in *Escherichia coli*

A Thesis
Submitted to the Faculty of Seoul National University
by

Min Woo Kim

In Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy

Advisor: Professor Byung-Gee Kim, Ph. D.

August, 2018

Program of Bioengineering
Seoul National University

Abstract

Development and application of multiple genome editing method using CRISPR/Cas9 system and ribozyme in *Escherichia coli*

Min Woo Kim

Program of Bioengineering

Seoul National University

Engineering cellular metabolism for improved production of valuable products requires extensive modulation of bacterial genome to explore complex genetic spaces. In order to introduce genetic modifications for rebalancing the metabolic flux, target genes to be modulated should be determined, and the expression of target genes should be optimized. However, it is difficult to select the target genes precisely because it is very labor-intensive to identify all the effects of the expression changes of the all genes involved in the metabolic pathway on the production of the target product. In addition, even if several genes were selected to be modulated, it is difficult to optimize the metabolic pathway because the

combination of modulation may cause detrimental effects on the strain. Therefore, in order to overcome the drawbacks described above, it is reasonable to conduct the metabolic engineering using a combinatorial approach. A combinatorial approach is to screen strain with the best phenotype among the various mutation library expected to improve the desired phenotype. Therefore, in order to generate the mutant library, the tool to introduce mutations at multiple loci is required.

In this thesis, we established the multiple sgRNA generation strategy to enable the metabolic pathway optimization using the CRISPR/Cas system. The dCas9-ω, which fused the transcriptional activator domain to inactivated cas9 (dCas9), can activate or repress target gene expression depending on the design of the sgRNA. Therefore, if multiple sgRNAs can be generated, the expression of several genes can be modulated simultaneously without genetic modification. For this reason, the strategy to generate multiple sgRNAs was required, and we constructed a strategy to produce several sgRNAs from one primary transcript using the self-cleavage property of Rz. By combining multiple sgRNA production strategies with the dCas9-ω system, target genes that have a strong effect on phenotypic enhancement can be efficiently identified.

After identification of target genes to be modulated, genetic engineering should be performed to alter the expression of those genes. Modulating in the expression level of a single gene to fine tune the metabolic flux often results in a change in the overall flux. Therefore, when optimizing the expression levels of target genes to achieve the desired phenotype through the metabolic engineering, the combinatorial approach is more reasonable than the sequential approach. To do this, it is necessary to be able to construct a library of sufficient size that

incorporates genetic modifications to various locations within the chromosome. However, the multiplexing methods reported in the literature have limitations in producing mutant libraries for applying the combinatorial approach because the recombination efficiency is very low and negative selection is not efficient. Therefore, we have designed a system that allows a combinatorial approach by using the CRISPR/Cas system to introduce a mutation library for one site in one cycle and accumulate a mutation library for multiple sites by repeating this cycle. Therefore, we have constructed a plasmid capable of obtaining high CFU with high editing efficiency through the CRISPR/Cas system, and confirmed that a strain in which mutations were introduced into three different target genes could be generated with high efficiency.

Keywords: CRISPR/Cas system, Ribozyme, MAGE, multiplexing, metabolic engineering, combinatorial approach

Student number: 2010-23350

Contents

List of Tables

List of Figures

Chapter 1.

Introduction

1.1 Genome editing with CRISPR/Cas system

1.1.1 CRISPR/Cas system

The clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR associated proteins (Cas) systems are adaptive immune systems that are present in many archaea and bacteria (Barrangou et al. 2007). The microbes establish an array of defense mechanisms by recognizing and depositing the invasive genetic elements in their CRISPR locus. This defense mechanism consists of adaptation, expression, and interference (Makarova et al. 2011) **(Figure 1.1)**. The adaptation is a step in which some elements are integrated into the CRISPR locus as a spacer after the proto-spacer proximity motif (PAM) of the invading genetic element is recognized by the Cas protein. Expression is a stage in which CRISPR locus composed of several spacer is processed after transcription, making several crRNAs, each of which forms a complex with Cas protein. Interference is a cleavage stage after a complex of crRNA and Cas protein recognizes an external genetic element. Thus, it is being applied in the genetic engineering field due to the characteristics of the CRISPR/Cas system, which recognizes and cleaves external genetic elements in a sequence specific manner.

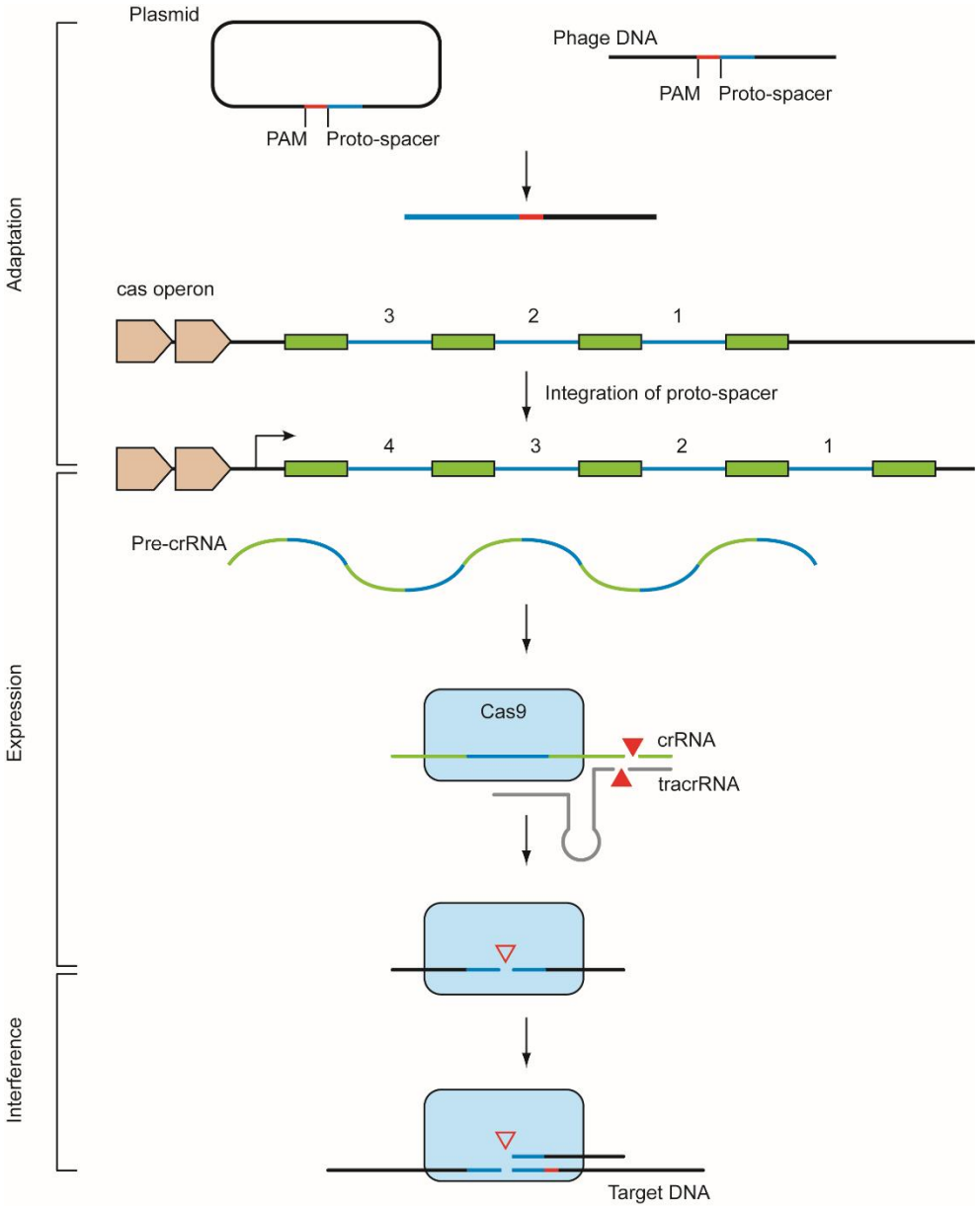


Figure 1.1 The three stages of CRISPR/Cas system

The CRISPR/Cas system is divided into several classes and types, depending on the type of Cas protein and targeting nucleotide (Makarova et al. 2011). Among the classes and types, the system commonly used for genome editing is the Class 2 type II CRISPR/Cas system using Cas9 with endonuclease activity against dsDNA. The CRISPR/Cas system is able to recognize the target in a sequence-specific manner because the pairs of crRNA and *trans*-encoded small RNA (tracrRNA) binds to a Cas9 protein having an endonuclease activity and guides to a target adjacent to the PAM sequence (Gibson and Yang 2017) **(Figure 1.2)**. The crRNA has the ability to recognize a target site, and the tracrRNA has the ability to bind Cas9 protein. Therefore, single guide RNA (sgRNA), a chimeric RNA synthesized by fusion of tracrRNA and crRNA, has been developed, and this sgRNA has been applied to guide Cas9 to a target in a sequence-specific manner **(Figure 1.2)**.

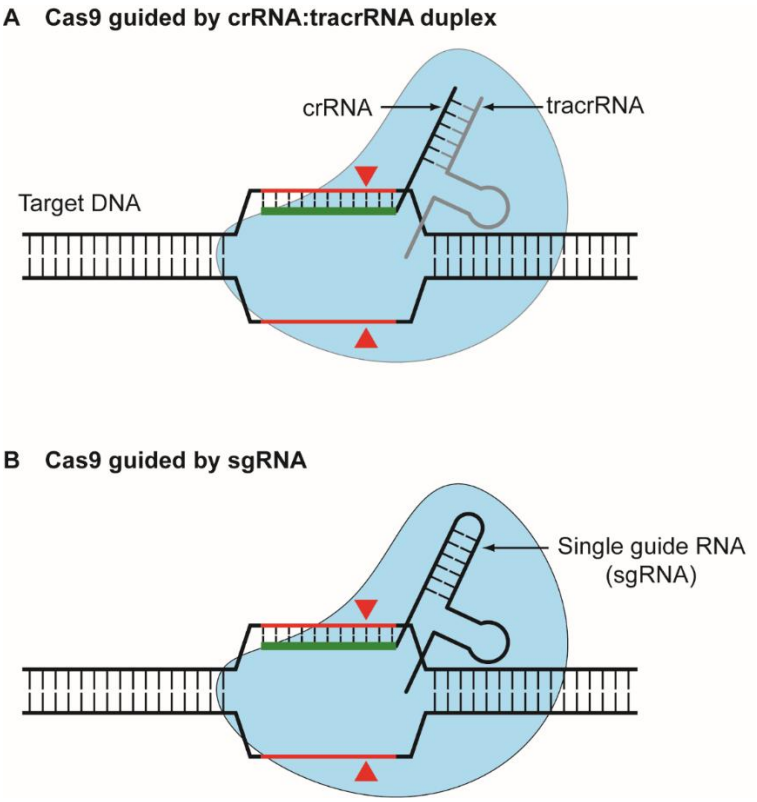


Figure 1.2 Cas9 targeting using crRNA-tracrRNA or a single guide RNA chimera.

In order to introduce chromosomal mutations at the desired locations, the chromosome should be repaired after cleavage. The CRISPR/Cas system is only responsible for sequence-specific cleavage, but there is no repair system. Therefore, in order to use the CRISPR/Cas system for genetic engineering, it is necessary to have a repair system such as a recombination system. There are two major pathways for repair of DNA double-strand breaks (DSBs) such as homologous recombination (HR) and non-homologous end joining (NHEJ) (Bortesi and Fischer 2015) **(Figure 1.3)**. The HR can repair with accurate sequence mediated by donor DNA, while NHEJ introduce random sequence at the cleavage site without the need for donor DNA. *Escherichia coli* has been reported to prefer repairs using HR, and in particular has a HR based SOS response system (Janion 2008). However, repairing the cleavage by Cas9 through the SOS response system is impossible due to the low efficiency of the donor DNA-mediated homologous recombination by the SOS response system. Therefore, the bacteriophage lambda red system reported to have high HR efficiency is essential to repair the chromosomal DNA using donor DNA (Mosberg et al. 2010). In conclusion, donor DNA-mediated recombination by the lambda red recombination system must be performed after sequence-specific cleavage by Cas9 in order to introduce mutations at desired positions on the chromosome using the CRISPR/Cas system.

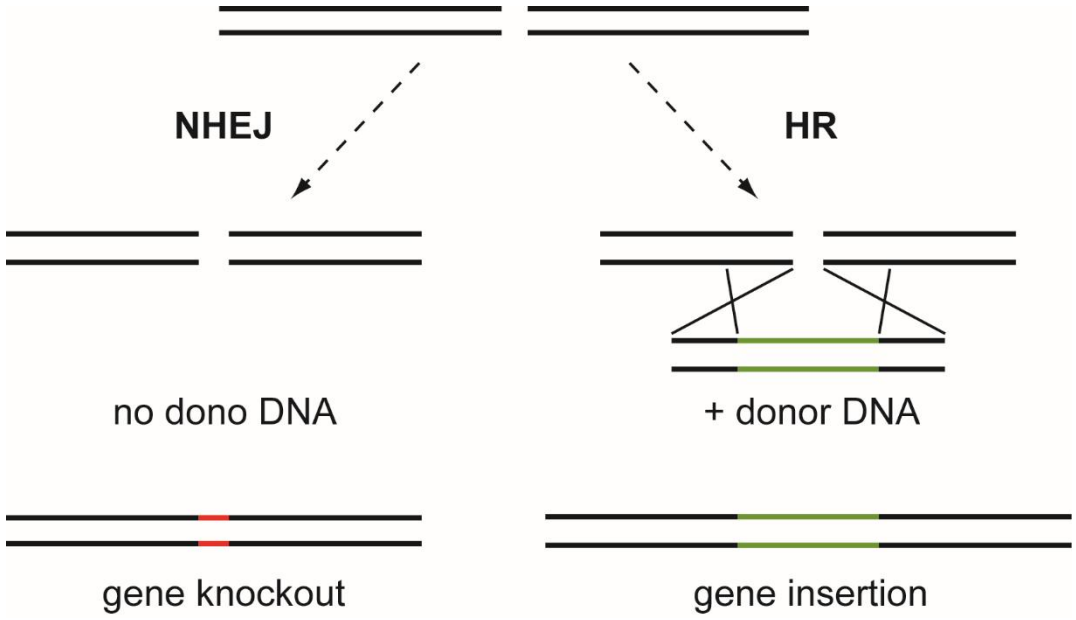


Figure 1.3 Two major repair system for repair of DNA double-strand breaks (DSBs).

(L Bortesi 2015)

1.1.2 CRISPRi system

The CRISPR system using Cas9 is a technique for introducing an accurate mutation into a target using sgRNA and donor DNA, whereas the CRISPR system using dCas9 is a technique for regulating the transcription of a target gene using only sgRNA (Larson et al. 2013). The dCas9 is a catalytically inactive version of Cas9 that lacks endonuclease activity. Therefore, the dCas9-sgRNA complex bind to the target region and interferes with the transcriptional elongation by RNA polymerase, resulting in the repression of the gene expression **(Figure 1.4)**. In the past, genetic modification to achieve transcriptional activation or repression was required to verify the gene function. However, this approach is a burdensome task for manipulating multiple targets. Since dCas9-based CRISPR interference (CRISPRi) can effectively control gene expression levels without chromosomal modification, it can be useful for studying gene function or demonstrating the effect of a target gene on the metabolic pathway by overcoming the drawbacks of conventional approach.

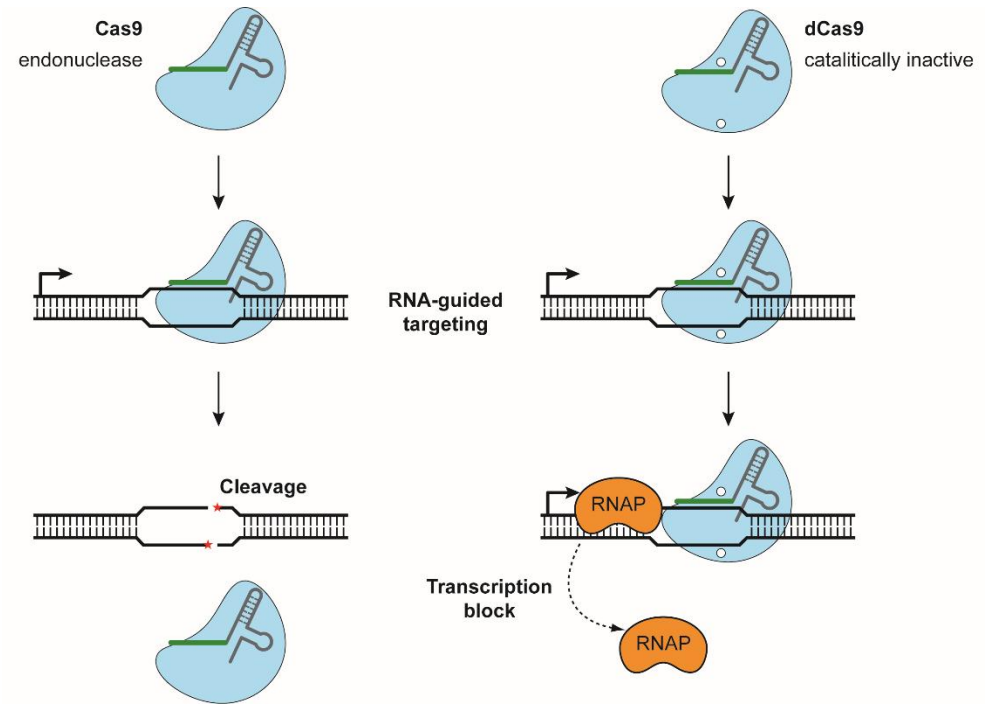


Figure 1.4 Gene editing versus gene regulation using Cas9 and dCas9

Recently, the CRISPR/Cas system for using as a transcriptional regulator to repress or activate gene expression was developed (Deaner et al. 2017). It has been reported that the omega subunit of RNA polymerase has a function as a transcriptional activator (Gunnelius et al. 2014). Therefore, some researchers tried to convert dCas9 into a transcriptional activator by fusing omega subunit to the dCas9 (Bikard et al. 2013). The transcriptional activation was observed when the dCas9-ω was binding to the promoter region **(Figure 1.5)**. On the other hand, it was also observed that dCas9-ω could inhibit transcription when binding to an open reading frame **(Figure 1.5)**. The dual mode of dCas9-ω can provides a simple and efficient method for global regulation of gene expression. Furthermore, if this technology is combined with multiple sgRNA generation strategy, it can provide easy and rapid way to find the most efficient combination of genetic modulation for metabolic pathway optimization.

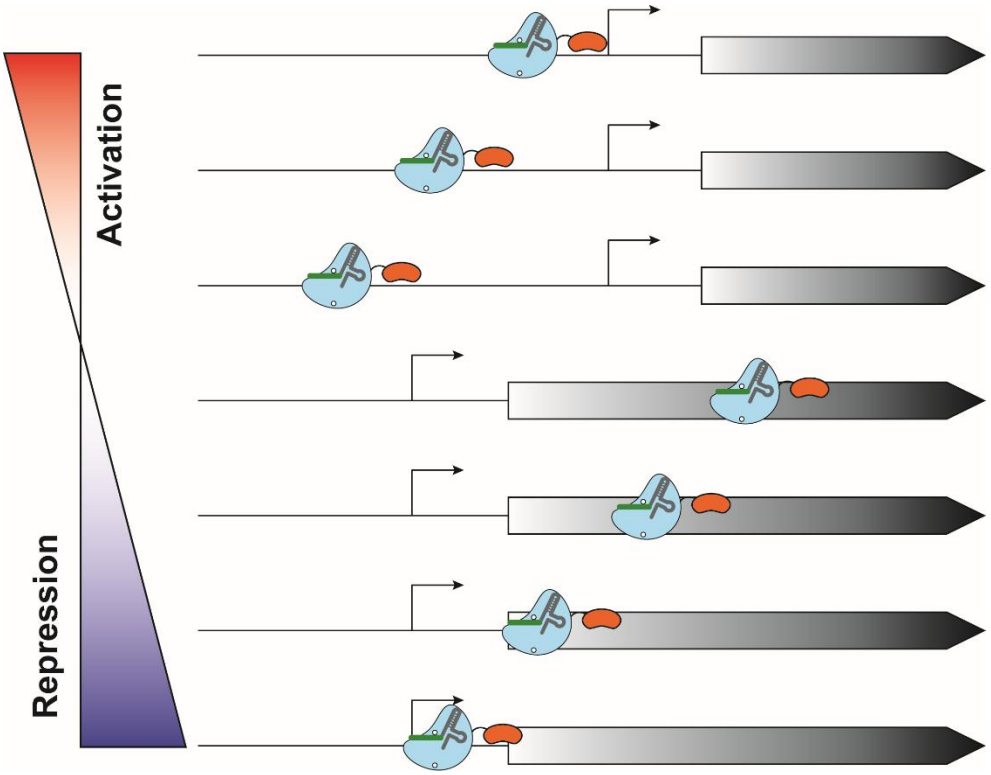


Figure 1.5 Dual mode of dCas9-activator

(M Deaner, 2017)

1.2 Multiple sgRNA generation strategy

1.2.1 Conventional methods to generate multiple sgRNAs

In order to modify the bacterial strain using CRISPR/Cas9 system into the desired phenotype through metabolic engineering, several genes have to be modulated. This system uses single guide RNAs (sgRNAs) to direct Cas9 endonuclease to distinct genome sites in a sequence-specific manner. Therefore, CRISPR/Cas9 system has capability to modulate multiple genes at the same time by expressing Cas9 along with multiple sgRNAs.

The most representative strategy to generate multiple sgRNAs is using individual promoters for each sgRNA (Wu et al. 2015). In order to produce multiple sgRNAs with this strategy, each sgRNA expression cassette consisting of a promoter, sgRNA and terminator should be arranged in series in one plasmid. However, it is inconvenient to create sgRNA expression cassette for each sgRNA, and each cassette has a promoter and a terminator in addition to sgRNA, so there is a limitation in the size of putting several plasmids into one plasmid. To overcome these drawbacks, methods for producing multiple sgRNAs from a primary transcript using RNA processing mechanisms have been studied. One of the representative method is using tRNA processing machinery to excise a primary transcript **(Figure 1.6)** (Xie et al. 2015). Since the tRNA is one of the most abundant component in the cell, the tRNA processing machinery supposed to be robust. Primary transcript is processed by RNase P and RNase E (or RNase Z in eukaryotes) in bacterium to produce mature tRNA by removing 5’ and 3’ extra

sequences. Therefore, when a primary transcript composed of tandemly repeating tRNA-gRNA architecture is prepared, it can produce multiple functional sgRNAs. Another strategy to produce multiple sgRNAs from a primary transcript is using Csy4 (Kabadi et al. 2014) **(Figure1.7)**. Csy4 is an endoribonuclease from Pseudomonas aeruginosa that recognized the 28-nt sequence and can cleave pre-crRNA transcript to yield multiple crRNAs. Therefore, by introducing the 28-nt sequence between sgRNAs, a primary transcript can produce multiple sgRNAs. However, the limitations of these methods are that RNA processing machinery is required to produce multiple sgRNAs. Thus, in order to apply the tRNA-processing system or Csy4-mediated processing for generating multiple sgRNAs at desired strain, additional studies are necessary.

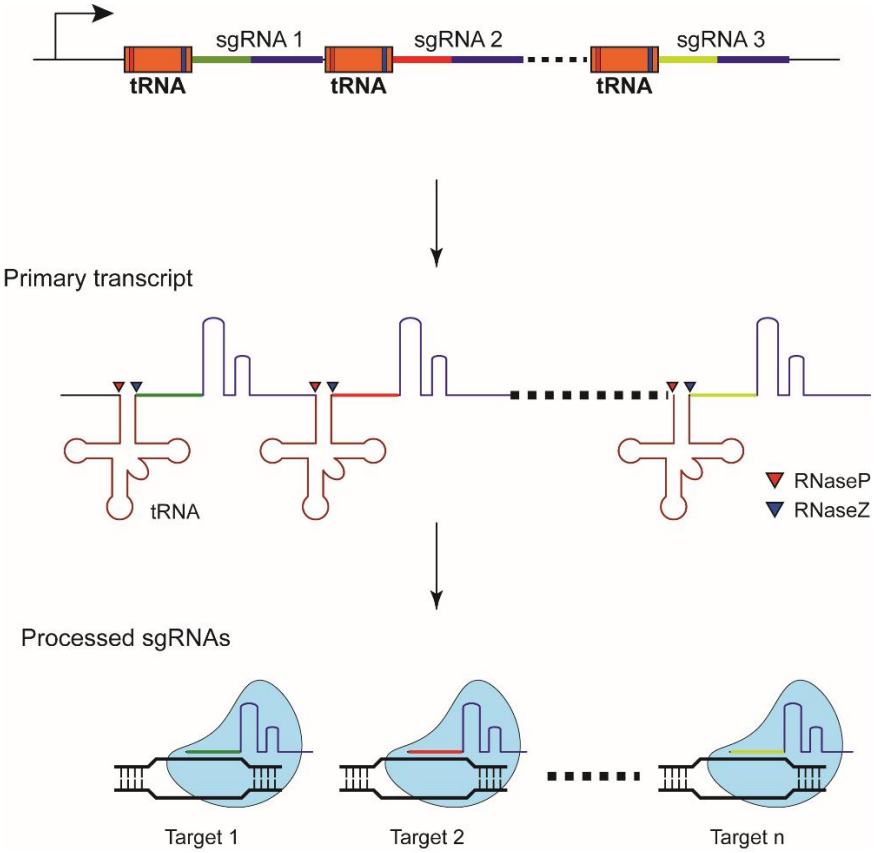


Figure 1.6 Engineering the endogenous tRNA system for multiplex genome editing with CRISPR/Cas.

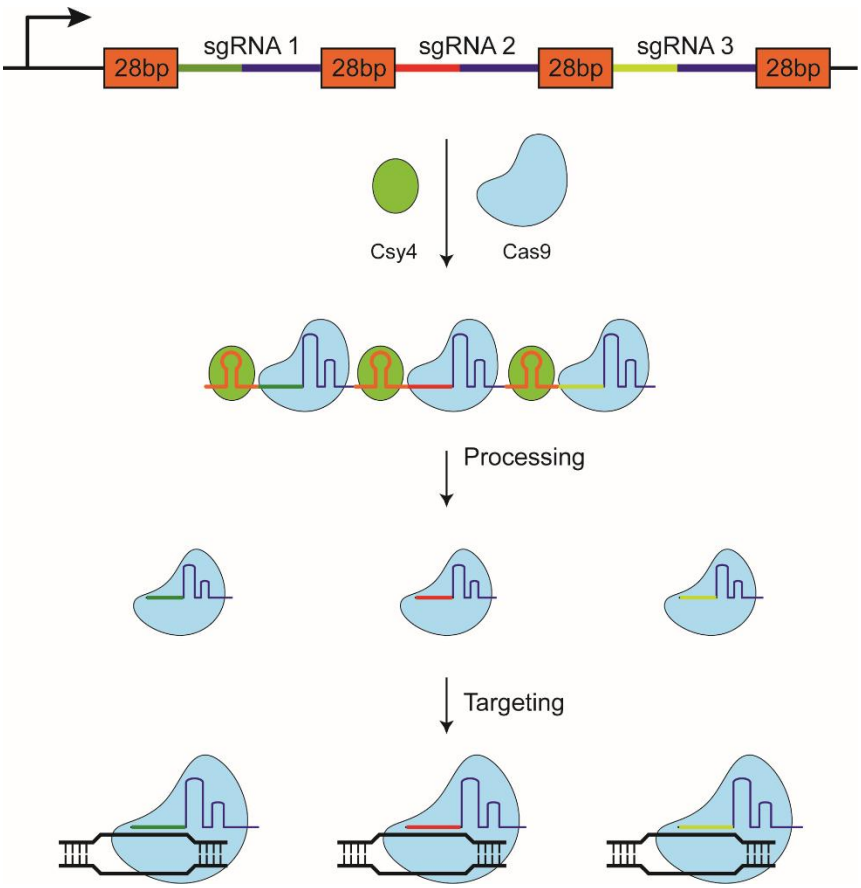


Figure 1.7 Multiplexed gRNA expression from a single transcript using Csy4.

1.2.2 Multiple sgRNA generation by using Rz

Recently, a number of studies to generate multiple sgRNAs from a primary transcript using ribozyme (ribonucleic acid enzyme) have been reported (Gao and Zhao 2014). Ribozymes (Rzs) are RNA molecules having self-cleavage activity without the aid of any RNA processing machinery such as endonuclease. And also, it shows sequence-specific RNA catalytic activity when forming a tertiary structure. Therefore, the strategy of generating multiple sgRNAs using ribozyme Rz has the advantage of overcoming the aforementioned limitations.

1.2.3 Methods to calculate the Rz efficiency

In order to use HHRz to generate multiple sgRNAs, the intracellular cleavage activity of HHRz should be sufficiently high and accurately reproducible. Nuclease protection assay is a representative method for analyzing intracellular cleavage activity of Rz (Donahue and Fedor 1997) **(Figure 1.8)**. In this method, RNA containing Rz sequence was extracted from the cell and annealed with a radiolabeled transcript which has complementary sequence to the Rz. The hybridized RNA was treated with a nuclease to digest unhybridized sequence, and protected fragment was fractionated on denaturing gel. The intracellular cleavage activity of Rz could be calculated by quantifying the cleavage product and uncleaved RNA. However, since this method uses cumbersome radiolabeled RNA, the measured Rz efficiency is not accurate, which becomes its disadvantage. Reporter gene expression assay is another way to evaluate the intracellular activity

of Rz (Deshler et al. 1995) **(Figure 1.9)**, which is a radiolabeling-free and qPCR-based method based on the fact such that Rz leads to rapid degradation of RNA, resulting in loss of protein expression level. Thus, the activity of Rz could be evaluated by measuring the reduced reporter gene expression or the remaining concentration of its mRNA. However, since the value measured by this method changes depending on the degree of RNA degradation, which may vary according to RNA extraction time. Because of the limitation, this method cannot be used to determine absolute intracellular catalytic activity of Rz, but can be used to evaluate the relative intracellular catalytic activity of a particular Rz among multiple Rzs, i.e. screening of good Rzs.

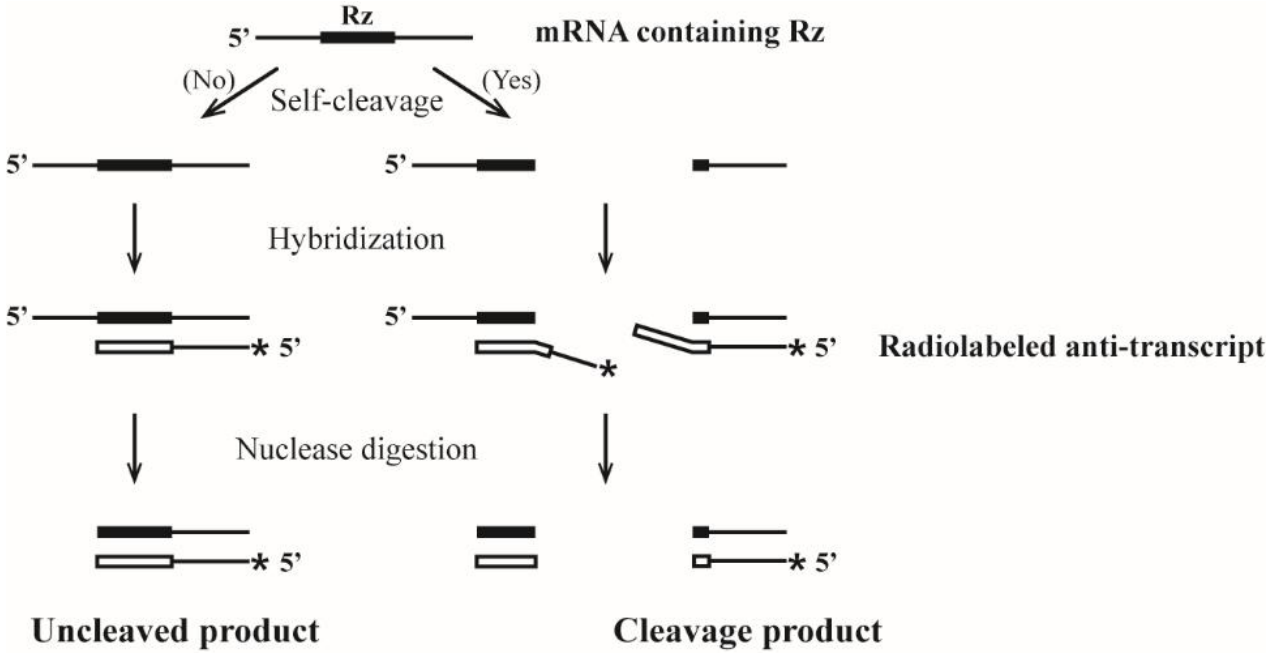


Figure 1.8 Graphical description of nuclease protection assay

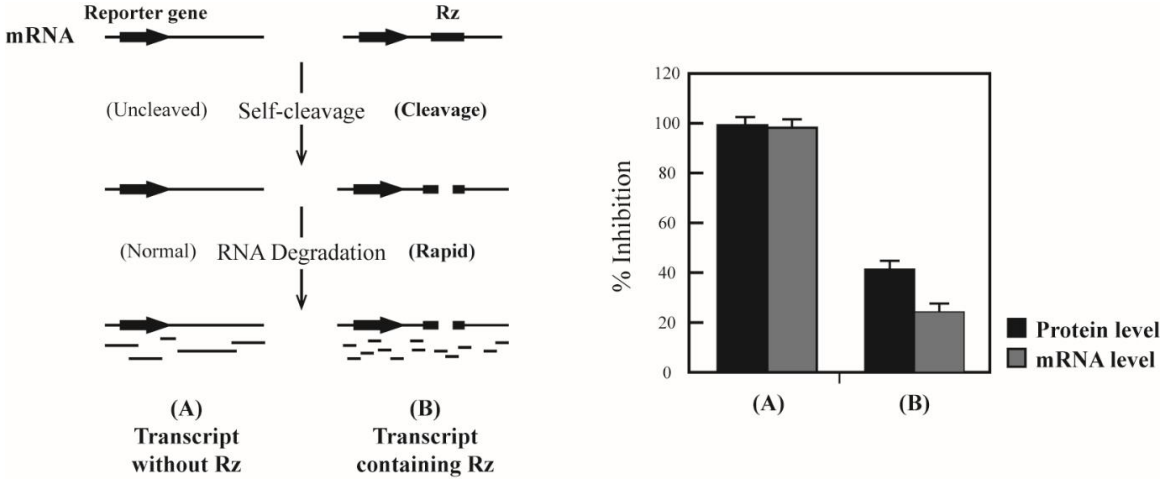


Figure 1.9 Graphical description of reporter gene expression assay

In this work, therefore, we present a radiolabeling-free qPCR-based method to more accurately evaluate the intracellular cleavage activity of HHRz. Since the intracellular activity of HHRz is highly correlated with its catalytic efficiency, we determined to evaluate the intracellular activity of HHRz by calculating “intracellular cleavage efficiency” (Khvorova et al. 2003). In order to ensure that any cleavage products detected from extracted RNA were generated only from intracellular cleavage, the *in vitro* activity of HHRz, the activity of HHRz during cDNA synthesis after cell lysis, has to be deactivated. HHRz is known to be activated and stabilized when it forms a complete tertiary structure in the presence of Mg2+ (Denesyuk and Thirumalai 2015). Therefore, in our quenching method, antisense oligonucleotide complementary to HHRz was to prevent the formation of tertiary structures of HHRz and to deactivate the *in vitro* activity of HHRz. In addition, in order to accurately measure the cleavage efficiency of HHRz, qPCR was used for the quantification of the amount of the cleaved mRNA by HHRz and the total amount of mRNA containing HHRz. Hereby, this method is called as "Quenching-qPCR (Q-Q) assay". Since Q-Q assay does use intact (or undegraded) mRNA samples for the calculation of “intracellular cleavage efficiency” of HHRz, this method gives a constant value of “intracellular cleavage efficiency” irrespective of the degree of RNA degradation. This method will enable an accurate calculation of “intracellular cleavage efficiency” of HHRz, and hence provides an accurate tool to compare *in vivo* activity of HHRz to screen better HHRzs among multiple HHRzs.

1.3 Multiplex genome editing

1.3.1 MAGE

The multiplex automated genome engineering (MAGE) is a method that facilitates genome-scale engineering using oligonucleotide mediated recombination (Wang et al. 2009). MAGE can simultaneously introduce mutations at multiple locations on the chromosome, thus producing combinatorial genomic diversity. This method utilizes donor DNA to recombine with high efficiency with assistance from lambda Red recombinase **(Figure 1.10)**. To increase the replacement efficiency, MAGE repeats a series of donor DNA addition and recombination cycles. Since the replacement efficiency is proportional to the recombination efficiency, high recombination efficiency is required for a successful introduction of mutations at multiple locations. In order to increase the recombination efficiency, single-stranded oligonucleotide that was shown to recombine with high efficiency with assistance from the lambda red protein Bet have been used in mutS deficient strain. However, due to the mutS is a gene responsible for the native methyl-directed mismatch repair (MMR) system, the inactivation of the MMR system cause nearly two orders-of-magnitude increase in the background mutation rate, leading to the accumulation of off-target mutations. Additionally, although mutations at multiple location can be introduced simultaneously through MAGE, a large number of library and a high-throughput screening method are required to obtain a recombinant clone. Therefore, in order to apply the MAGE method in the metabolic engineering field, the drawbacks should be overcome.

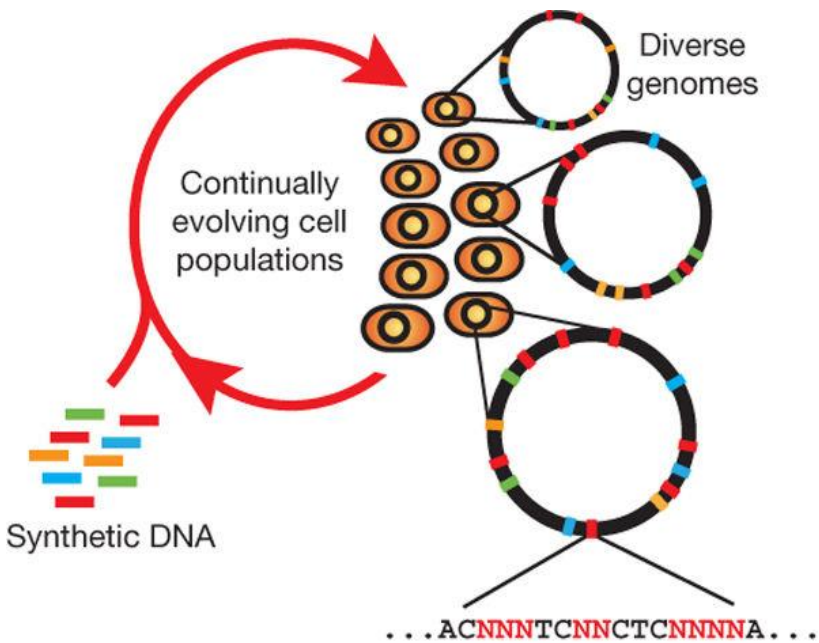


Figure 1.10 Multiplex automated genome engineering enables the rapid and continuous generation of sequence diversity at many targeted chromosomal locations.

(HH Wang., 2009)

1.3.2 CRMAGE

CRISPR/Cas9 and lambda Red recombination based MAGE technology (CRMAGE) was developed, which is an improved version of MAGE with capability of negative selection (Ronda et al. 2016). All populations of cells generated through MAGE contains parent strain as well as variants. In the CRMAGE system, CRISPR/Cas9 system against wild type sequence was performed after the MAGE cycle to remove parent strain. In addition, a transient mutS inactivation system was introduced through overexpression of the Dam methyltransferase to obtain the mutS deficient phenotype without modifying mutS.

1.4 The scope of thesis

The purpose of this thesis is the development of an efficient genome editing method for metabolic pathway optimization in *Escherichia coli*. To achieve the goal, I used the CRISPR/Cas with improved replacement efficiencyand MAGE systems, and as a result it was able to introduce mutations at several locations on the chromosome.

In Chapter 3, the properties of ribozymes (Rzs) that can be used for generating multiple sgRNAs from a primary transcript was examined. In order to use Rz for generating sgRNAs, Rz must have highly active and specific catalytic activity. However, current methods for assessing the intracellular activity of Rz have limitations such as difficulty in handling and inaccuracies in the evaluation of correct cleavage activity. Therefore, in order to confirm whether it is appropriate to

apply Rz to generate multiple sgRNAs, a method of calculating the intracellular cleavage activity of Rz has been proposed.

In Chapter 4, the multiple sgRNA generation strategy was developed to optimize the metabolic pathway using the CRISPR/Cas system. The CRISPR/Cas system is a powerful tool for the accurate and rapid modification of endogenous genes. This system uses sgRNA to direct Cas9 endonuclease (or inactivated Cas9, dCas9) to the target sequence, allowing for precise modulating. Therefore, multiple genomic region can be modulated simultaneously, by generating multiple sgRNAs. Here, I developed a strategy to produce multiple sgRNA from a primary transcript in *E. coli* using the self-cleavage property of Rzribozyme. And also, the capability of multiple repression using the multiple sgRNA generation strategy coupled with dCas9 was also demonstrated.

In Chapter 5, we conducted a study to improve the efficiency of obtaining recombinants through the CRISPR/Cas system. In order to perform metabolic engineering through a combinatorial approach, it is necessary to obtain a sufficient library size. To do this, the CRISPR/Cas system should be able to produce recombinants with high efficiency. Therefore, a plasmid capable of producing highly efficient recombinant using CRISPR/Cas system was constructed. Finally, it was confirmed that a sufficient size of mutant library can be generated by sequentially introducing a mutation using this plasmid.

It is expected that the method suggested in this thesis can be useful for metabolic pathway optimization in the field of metabolic engineering.

Chapter 2.

Materials and methods

2.1 Bacterial strains and culture conditions

For the experiment of Q-Q assay, the *E. coli* BL21(DE3) strain was used as host strain for transcribing the mRNA containing HHRz. Luria-Bertani (LB) medium was used for cell growth. Kanamycin was added at concentration of 10mg/mL as needed. After one loopful inoculation from a single colony on LB agar plate, the cells were grown in 2ml LB in 18- by 150-mm borosilicate glass test tube with shaking (200 rpm) at 37°C. When the OD600 reached to ca. 0.6, 0.1mM of final IPTG (isopropyl-β0D-thiogalactopyranoside) was added to the media and incubated for another 3 hrs with shaking (200 rpm) at 30°C.

For the multiple repression experiments, *E. coli* DH5α was used as a cloning host, and *E. coli* BL21(DE3) strain was used for testing of *in vivo* activity of Rz and for characterization of CRISPR/Cas mediated genome editing. *E. coli* BW25113 strain was used to observe auxotrophic phenotype through the repression of gene expression associated with amino acid biosynthesis by the CRISPRi system. *E. coli* MG1655 strain containing pAC-LYC was used to confirm the effect of metabolic pathway re-balancing by CRISPRi system on lycopene production. pdCas9-bacterial (it was a gift from Stanley Qi (Addgene plasmid # 44249)) or pdCas9K, which is modified from pdCas9-bacteria, were introduced into the above mentioned strains for expressing dCas9 protein.

E. coli BL21 (DE3) strain was used for multiple genome editing and for calculating the endonuclease activity of Cas9 protein and measuring HR efficiency using ssDNA. *E. coli* MG1655 strain was used to confirm the editing efficiency of MAGE. All the strains used in this study are listed in **Table1**.

LB media was used in all cases unless otherwise noted. Chloramphenicol, kanamycin and ampicillin were added at concentrations of 30mg/ml, 10mg/ml and 100mg/ml, respectively. Anhydrotetracycline (aTc) was added at a final concentration of 100ng/L to induce dCas9 and IPTG (isopropyl-β-D-thiogalactopyranoside) was added at concentrations of 0.1mM for transcription of sgRNAs. X-gal for blue/white selection was added at concentration of 40μg/ml. In case of the auxotroph experiment, *E. coli* BW25113, harboring pdCas9-bacteria and sgRNA transcribing plasmid, was pre-cultured in LB liquid medium with appropriate antibiotics at 37°C and 220 rpm for 12h. A volume of 100μl was centrifuged at 5,000g for 3min to harvest the cells and washed twice with M9 minimal media and diluted to an OD600 nm of 0.1. This diluted culture was inoculated in 2ml of M9 minimal media supplemented with 3mM of the focal amino acid. The cell growth was determined after 18h incubation.

Table 1 Strains used in this study

Strains	Relevant information	Source or reference
E.coli BL21(DE3)	F- *ompT hsdS*$_{B}$ (r_B-m_B-) *gal dcm rne*131 λ(DE3)	Invitrogen
E. coli DH5α	F^{-} *endA1 glnV44 thi-1 recA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ^{-}	TaKaRa
E. coli BW25113	F^{-}, DE(*araD-araB*)567, *lacZ*4787(del)::*rrnB*-3, LAM^{-}, *rph*-1, DE(*rhaD-rhaB*)568, *hsdR*514	Invitrogen
E. coli MG1655	F^{-}, lambda^{-}, *rph*-1	Invitrogen
BLd	*E. coli* BL21(DE3) containing pdCas9	In this study
BWd	*E. coli* BW25113 containing pdCas9	In this study
B1	*E. coli* BL21(DE3) containing pGGA1_AL	In this study
BLd1	BLd containing pGGA1_*adhE*	In this study
BLd2	BLd containing pGGA1_ *lacZ*	In this study
BLd3	BLd containing pGGA1_ LA	In this study
BLd4	BLd containing pGGA1_ AL	In this study
BLd213	BLd containing pGGA1_L13	In this study
BLd215	BLd containing pGGA1_L15	In this study
BLd235	BLd containing pGGA1_L35	In this study
BLd255	BLd containing pGGA1_L55	In this study
BWd_*argH*	BWd containing pET28_*argH*	In this study
BWd_*lysA*	BWd containing pET28_ *lysA*	In this study
BWd_*metB*	BWd containing pET28_ *metB*	In this study
BWd_*proC*	BWd containing pET28_ *proC*	In this study
BWd_ALM	BWd containing pGGA1_ALM	In this study
Mld1	Mld containing pgRNA	In this study
Mld2	Mld containing pgRNA_*gdhA*	In this study
Mld3	Mld containing pgRNA_aceE	In this study
Mld4	Mld containing pgRNA_*gpmA*	In this study
Mld5	Mld containing pgRNA_*icdA*	In this study
Mld6	Mld containing pgRNA_*pfkA*	In this study

Mld7	Mld containing pgRNA_*ytjC*	In this study
Mld_AIY	Mld containing pGGA2_AIY	In this study

2.2 Plasmid construction

2.2.1 Design of transcripts containing HHRzs

The class of HHRz is satellite RNA of tobacco ringspot virus (sTRSV) (Khvorova et al. 2003). To confirm the activity of HHRz, two 150bps nucleotides were linked to both sides of HHRz **(Figure 2.1A)**, and the ligated DNA nucleotides were transcribed under the T7 promoter. HHRz is consisted of stem loops I, II, and III, and the catalytic activity of HHRz is determined by the tertiary interactions between stem loops I and II (Khvorova et al. 2003). In particular, the cleavage site of the RNA can be precisely determined by designing the six nucleotides of the stem loop I to be complementary to the six nucleotides immediately adjacent to the position cleaved (Gao and Zhao 2014). In this work, three transcripts were designed for comparison; the first transcript containing a natural HHRz, the second transcript containing a mutant HHRz, and the third transcript without HHRz. The design of a transcript containing HHRz is typically shown as follows: Left RNA-(N6)-HHRz-(N6)-Right RNA.

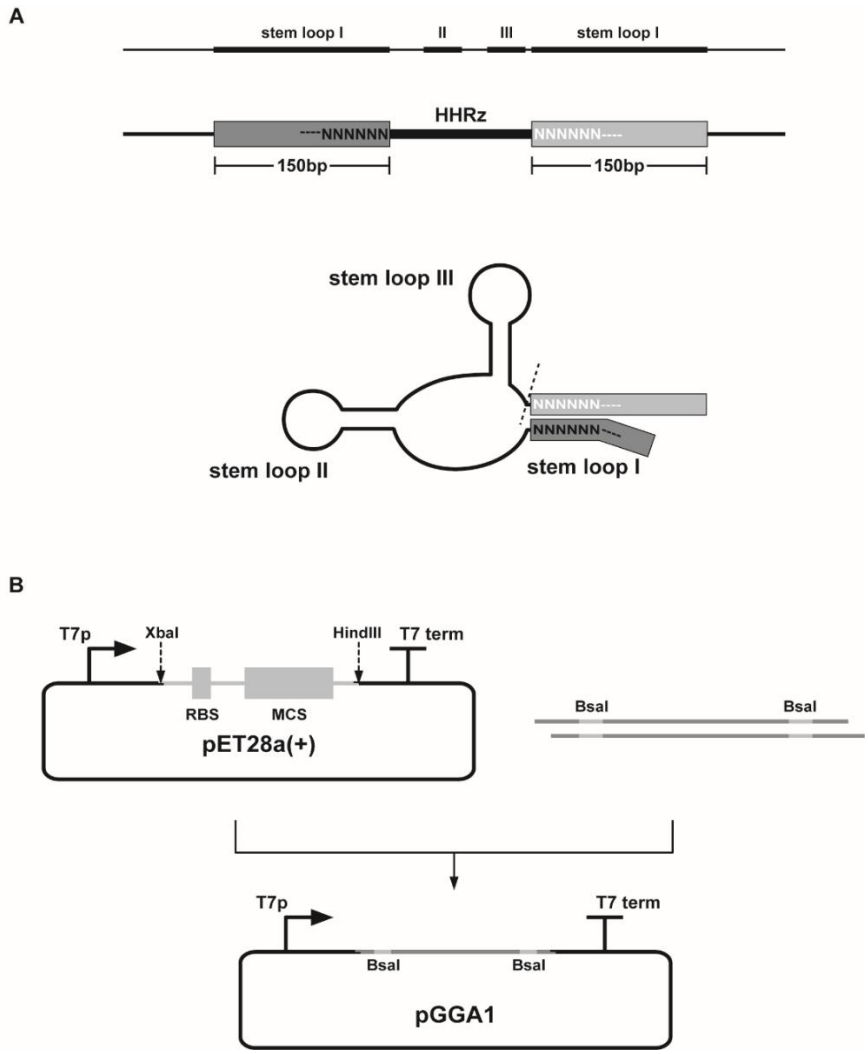


Figure 2.1 Graphical description of HHRz and plasmid construction

2.2.2 Plasmid construction to express transcripts containing HHRzs

The pGGA1 plasmid was constructed to carry out the Golden Gate Assembly based on the pET28a(+) plasmid (Novagen) **(Figure 2.1B)**. The vector was prepared by digesting the pET28a(+) plasmid with restriction enzymes XbaI and HindIII. The insert was prepared by hybridization of two complementary oligonucleotides (Oligo_GGA1/Oligo_GGA2). Both ends of the hybridized insert were designed to be sticky ends that could bind to XbaI/HindIII and two BsaI restriction sites were included in the insert.

To construct plasmids expressing any transcripts containing HHRzs, the amplicons for Left and Right RNAs were amplified from pgRNA backbone using the set of corresponding primers (LRNA_F/LRNA_R and RRNA_c, n, m /RRNA_R). In particular, the HHRz sequence and 12 nucleotides to form the stem loop I were synthesized to the Right RNA primer. Both amplicon were assembled into pGGA1 based on the principle of Golden Gate Assembly (Engler et al. 2008). Three kinds of plasmids, i.e. pHHRz-c, pHHRz-n and pHHRz-m were used in this study, and they produced a transcript without HHRz, a transcript containing the natural HHRz, and a transcript containing a mutant HHRz, respectively. The detailed design of the primers and plasmids are given in **Table2** and **Table3**.

Table 2 Primers used in this study (chapter 3)

no.	Description	Sequence (5' – 3')
1	LRNA_F	ATATGGTCTCTCTAGAACTTTCAGTTTAGCGGTCTGTTTTAGAGCTAGAAATAGC
2	LRNA_R	ATATGGTCTCTCAATAAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCACTT
3	RRNA-c	ATATGGTCTCTATTGATGTCGGTTTCCGCGAGGTGGTTTTAGAGCTAGAAATAGC
4	RRNA-n	ATATGGTCTCTATTGCGACAT***CTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC***ATGTCGGTTTCCGCGAGGTGGTTTTAGAGCT AGAAATAGC
5	RRNA-m	ATATGGTCTCTATTGAAGACAT***CTGATGAGTCCCTGAAATGGGACGAAACGAGTAAGCTCGTC***ATGTCGGTTTCCGCGAGGTGGTTTTA GAGCTAGAAATAGC
6	RRNA_R	ATATGGTCTCTAGCTAAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCACTT
7	Oligo_GGA1	CTAGAGAGACCTCCGCTAGCTAGGGTCTCA
8	Oligo_GGA2	AGCTTGAGACCCTAGCTAGCGGAGGTCTCT
9	Frag_a_F	CTAGTCCGTTATCAACTTGA
10	Frag_a_R	TAACTTGCTATTTCTAGCTC
11	Frag_b_F	ATGTCGGTTTCCGCGAGGTG
12	Frag_b_R	TTTCGGGCTTTGTTAGCAGC
13	rssA_F	GAAATTGATATCGTTGCAGG
14	rssA_R	GGCATTATTTCGCGGTATTG

The underlined sequence represents the six nucleotides that form the stem loop I, the italic bold sequence represents the stem loop II, III and catalytic core region of ribozyme.

Table 3 Plasmids used in this study

Plasmid	Relevant information	Source or reference
pET28a(+)	Kan^r, F1 origin, His tag	Novagen
pdCas9-bacteria	dCas9 expressing plasmid, Cm^R	Addgene
pgRNA-bacteria	sgRNA expressing plasmid	Addgene
pdCas9K	dCas9 expressing plasmid, Km^R	In this study
pGGA1	pET28 based plasmid for Golden Gate Assembly	In this study
pHHRz-c	pET28 based plasmid for expression of transcript without HHRz	In this study
pHHRz-n	pET28 based plasmid for expression of transcript containing natural HHRz (sTRSV)	In this study
pHHRz-m	pET28 based plasmid for expression of transcript containing mutant HHRz (sTRSV+PL1&2)	In this study
pGGA1_*adhE*	pGGA1 based plasmid for expression of sgRNA_*adhE*	In this study
pGGA1_*lacZ*	pGGA1 based plasmid for expression of sgRNA_ *lacZ*	In this study
pGGA1_AL	pGGA1 based plasmid for expression of sgRNA_*adhE* and *lacZ*	In this study
pGGA1_ALc	pGGA1 based plasmid for expression of sgRNA_*adhE* and *lacZ* (both sgRNA scaffolds are conneted without Rz)	In this study
pGGA1_LA	pGGA1 based plasmid for expression of sgRNA_*lacZ* and *adhE*	In this study
pGGA1_ L13	pGGA1 based plasmid, sgRNA_*lacZ* is transcribed at 1/3 position from primary transcript	In this study
pGGA1_ L15	pGGA1 based plasmid, sgRNA_*lacZ* is transcribed at 1/5 position from primary transcript	In this study
pGGA1_ L35	pGGA1 based plasmid, sgRNA_*lacZ* is transcribed at 3/5 position from primary transcript	In this study
pGGA1_ L55	pGGA1 based plasmid, sgRNA_*lacZ* is transcribed at 5/5 position from primary transcript	In this study
pGGA1_ *argH*	pGGA1 based plasmid for expression of sgRNA_*argH*	In this study
pGGA1_ *lysA*	pGGA1 based plasmid for expression of sgRNA_ *lysA*	In this study
pGGA1_ *metB*	pGGA1 based plasmid for expression of sgRNA_ *metB*	In this study

pGGA1_ *leuB*	pGGA1 based plasmid for expression of sgRNA_ *leuB*	In this study
pGGA1_ *proC*	pGGA1 based plasmid for expression of sgRNA_ *proC*	In this study
pGGA1_ ALM	pGGA1 based plasmid for expression of sgRNA_*adhE, leuB, metB*	In this study
pGGA2	pgRNA based plasmid for Golden Gate Assembly	In this study
pgRNA_*gdhA*	pGGA2 based plasmid for expression of sgRNA_*gdhA*	In this study
pgRNA_ *aceE*	pGGA2 based plasmid for expression of sgRNA_ *aceE*	In this study
pgRNA_ *gpmA*	pGGA2 based plasmid for expression of sgRNA_ *gpmA*	In this study
pgRNA_ *icdA*	pGGA2 based plasmid for expression of sgRNA_ *icdA*	In this study
pgRNA_ *pfkA*	pGGA2 based plasmid for expression of sgRNA_ *pfkA*	In this study
pgRNA_ *ytjC*	pGGA2 based plasmid for expression of sgRNA_ *ytjC*	In this study
pgRNA_ AIY	pGGA2 based plasmid for expression of sgRNA_*aceE, icdA, ytjC*	In this study

2.2.3 Construction of sgRNA transcribing plasmid

The pGGA1 and pGGA2 were prepared as plasmids for golden gate assembly to assemble multiple inserts simultaneously. pGGA2 was constructed by PCR amplification from pgRNA with primers BbsI_F and BbsI_R, followed by ligation, which was digested with HindIII.

To construct the multiple sgRNAs transcribing plasmids, several set of primers were used to amplify the sgRNA scaffold from pgRNA backbone. The 20bp of spacer sequence and 43bp of Rz sequence for each target was synthesized in primers. The PCR products were assembled using Golden Gate Assembly. The detailed design of the primers is given in **Table4**.

Table 4 Primers used in this study (chapter 4)

no.	Description	Sequence (5' – 3')
1	BbsI_F	CTTCAAGCTTGAAGACCCCCATGCGAGAGTAGGGAA
2	BbsI_R	CTTCAAGCTTGAAGACAAACTAGTATTATACCTAGGAC
3	lacZ_F (pGGA1_*lacZ)*	ATATACTAGTTTGGGAAGGGCGATCGGTGCGTTTTAGAGCTAGAAATAGC
4	adhE_F (pGGA1_*adhE)*	ATATACTAGTGAAAAACTGTCCCCGACTCGTTTTAGAGCTAGAAATAGCA
5	lacZ1_F (pGGA1_LA) (pGGA1_L13) (pGGA1_L15)	ATATGGTCTCTCTAGTCCCAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTTGGGAAGGGCGATCGGTGCG TTTTAGAGCTAGAAATAGC
6	lacZ1_R (pGGA1_LA)	ATATGGTCTCAAAAATTTTTTTTTTTTTTTTTTTGCACCGACTCGGTGCCACTT
7	adhE1_F (pGGA1_LA)	ATATGGTCTCTTTTTTTTTTCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGAAAAACTGTCCCCGACTCGTTT TAGAGCTAGAAATAGC
8	adhE1_R (pGGA1_LA)	ATATGGTCTCAAGCTTTTTTTTTTTTTTTTTTTTTGCACCGACTCGGTGCCACTT
9	adhE2_F (pGGA1_*AL*)	ATATGGTCTCTCTAGTTTTTCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGAAAAACTGTCCCCGACTCGTT TTAGAGCTAGAAATAGC
10	adhE2_R (pGGA1_*AL*)	ATATGGTCTCAGGGATTTTTTTTTTTTTTTTTTTGCACCGACTCGGTGCCACTT

11	lacZ2_F (pGGA1_*AL*)	ATATGGTCTCTTCCCAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTTGGGAAGGGCGATCGGTGCGTTTTA GAGCTAGAAATAGC
12	lacZ2_R (pGGA1_*AL*)	ATATGGTCTCAAGCTTTTTTTTTTTTTTTTTTTTTGCACCGACTCGGTGCCACTT
13	lacZ3_F (pGGA1_L35)	ATATGGTCTCTATTGTCCCAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTTGGGAAGGGCGATCGGTGCGT TTTAGAGCTAGAAATAGC
14	lacZ5_F (pGGA1_L55)	ATATGGTCTCTTACTTCCCAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTTGGGAAGGGCGATCGGTGCGT TTTAGAGCTAGAAATAGC
15	Dummy1_F (pGGA1_L13 ~ L55)	ATATGGTCTCTCTAGAAAAGTTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAACTTTCAGTTTAGCGGTCTG TTTTAGAGCTAGAAATAGC
16	Dummy2_F (pGGA1_L13 ~ L55)	ATATGGTCTCTGCACAAAAGTTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAACTTTCAGTTTAGCGGTCTG TTTTAGAGCTAGAAATAGC
17	Dummy3_F (pGGA1_L13 ~ L55)	ATATGGTCTCTATTGAAAAGTTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAACTTTCAGTTTAGCGGTCTG TTTTAGAGCTAGAAATAGC
18	Dummy4_F (pGGA1_L13 ~ L55)	ATATGGTCTCTCCAGAAAAGTTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAACTTTCAGTTTAGCGGTCTG TTTTAGAGCTAGAAATAGC
19	Dummy5_F (pGGA1_L13 ~ L55)	ATATGGTCTCTTACTAAAAGTTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAACTTTCAGTTTAGCGGTCTG TTTTAGAGCTAGAAATAGC
20	R1 (pGGA1_L13 ~ L55)	ATATGGTCTCAGTGCAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
21	R2 (pGGA1_L13 ~ L55)	ATATGGTCTCACAATAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC

22	R3 (pGGA1_L13 ~ L55)	ATATGGTCTCACTGGAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
23	R4 (pGGA1_L13 ~ L55)	ATATGGTCTCAAGTAAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
24	R5 (pGGA1_L13 ~ L55)	ATATGGTCTCAAGCTAAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
25	argH_F (pGGA1_*argH*)	ATATTCTAGATTCAAGGATTTGTTGTGGCCGTTTTAGAGCTAGAAATAGC
26	lysA_F (pGGA1_*lysA*)	ATATTCTAGATCGGTGCTGAACAGTGAATGGTTTTAGAGCTAGAAATAGC
27	metB_F (pGGA1_*metB*)	ATATTCTAGACTACCGCCGTAGCAGTCGTGGTTTTAGAGCTAGAAATAGC
28	leuB_F (pGGA1_*leuB*)	ATATTCTAGAAGCGCCTGGGTCATCACTTCGTTTTAGAGCTAGAAATAGC
29	proC_F (pGGA1_*proC*)	ATATTCTAGATTAACCAGTGCGGGAGTGTTGTTTTAGAGCTAGAAATAGC
30	R (pGGA1_*argH~proC*)	ATATAAGCTTGCACCGACTCGGTGCCACTT
31	argH_auxo_F (pGGA1_ALM)	ATATGGTCTCTCTAGCTTGAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTTCAAGGATTTGTTGTGGCC GTTTTAGAGCTAGAAATAGC
32	metB_auxo_F (pGGA1_ALM)	ATATGGTCTCTGCACCGGTAGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCTACCGCCGTAGCAGTCGTG GTTTTAGAGCTAGAAATAGC

33	leuB_auxo_F (pGGA1_ALM)	ATATGGTCTCTATTGGGCGCTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAGCGCCTGGGTCATCACTTC GTTTTAGAGCTAGAAATAGC
34	R1 (pGGA1_ALM)	ATATGGTCTCAGTGCAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
35	R2 (pGGA1_ALM)	ATATGGTCTCACAATAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
36	R3 (pGGA1_ALM)	ATATGGTCTCAAGCTAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
37	gdhA_F (pgRNA_*gdhA*)	ATATACTAGTTTTGATTCGGGTCGCGCTTTGTTTTAGAGCTAGAAATAGC
38	aceE_F (pgRNA_*aceE*)	ATATACTAGTAGCCAGTCGCGAGTTTCGATGTTTTAGAGCTAGAAATAGC
39	gpmA_F (pgRNA_*gpmA*)	ATATACTAGTAGTTTCCAGGATTTCTCAACGTTTTAGAGCTAGAAATAGC
40	icdA_F (pgRNA_*icdA*)	ATATACTAGTGTGATCTTCTTGCCTTGTGCGTTTTAGAGCTAGAAATAGC
41	pfkA_F (pgRNA_*pfkA*)	ATATACTAGTCGAATTGCGGCGTTCATGCCGTTTTAGAGCTAGAAATAGC
42	ytjC_F (pgRNA_*ytjC*)	ATATACTAGTGCGTTTCACCGTGGCGGACTGTTTTAGAGCTAGAAATAGC
43	R (pgRNA_*gdhA~ytjC*)	AATTACTAGTATTATACCTAGGACTGAGCT

44	aceE_F (pGGA2_AIY)	TATGAAGACAATAGTCTGGCTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAGCCAGTCGCGAGTTTCGATG TTTTAGAGCTAGAAATAGC
45	icdA_F (pGGA2_AIY)	TATGAAGACAAGCACGATCACCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGTGATCTTCTTGCCTTGTGCGT TTTAGAGCTAGAAATAGC
46	ytjC_F (pGGA2_AIY)	TATGAAGACAAGACTAAACGCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGCGTTTCACCGTGGCGGACTG TTTTAGAGCTAGAAATAGC
47	R1 (pGGA2_AIY)	ATATGAAGACAAGTGCAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
48	R2 (pGGA2_AIY)	ATATGAAGACAAAGTCAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
49	R3 (pGGA2_AIY)	ATATGAAGACAAATGGAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
50	cRT_F	CTGCCACCGCTGAGCAATAA
51	cRT_R	TAGCTCTAAAACGCACCGAT

2.2.4 plasmids construction for CRISPR/Cas system

pCas_v2.0 was constructed to express only Beta protein in lambda-red recombinase of pCas (Jiang et al. 2015). Using a primer set capable of amplifying the Beta protein, fragment 1 constituting pCas_v2.0 was prepared from pCas. The rest of the region except the recombinant system portion of pCas was divided into two fragments, and each fragment was amplified using two sets of primers. The prepared three fragments were assembled through circular polymerase extension cloning (CPEC) to construct pCas_v2.0 (Quan and Tian 2011).

pCASRec was constructed to have three functions: homologous recombination, endonuclease of Cas9, and plasmid curing. To achieve high HR efficiency, pCASRec should have a recombinant system that uses only Beta protein. Therefore, a recombinant system of pCRMAZ_2.0, which expresses only Beta protein among the three proteins by arabinose induction, was introduced into pCASRec (Ronda et al. 2016). Especially, in addition to the Beta protein, the Dam protein is additionally constituted in the recombinant system of pMAZCR_2.0. Due to the Dam protein can inhibit the activity of MutS, which has the function of repairing discrepancies that may occur in the replication process, overexpression of Dam protein can promote the generation of desired mutations in chromosomes (Acharya et al. 2003). Therefore, the recombination system of pMAZCR_2.0 was amplified by pCR_frag1_F/pCR_frag1_R and prepared as fragment 1 constituting pCASRec. **(Figure2.2)**

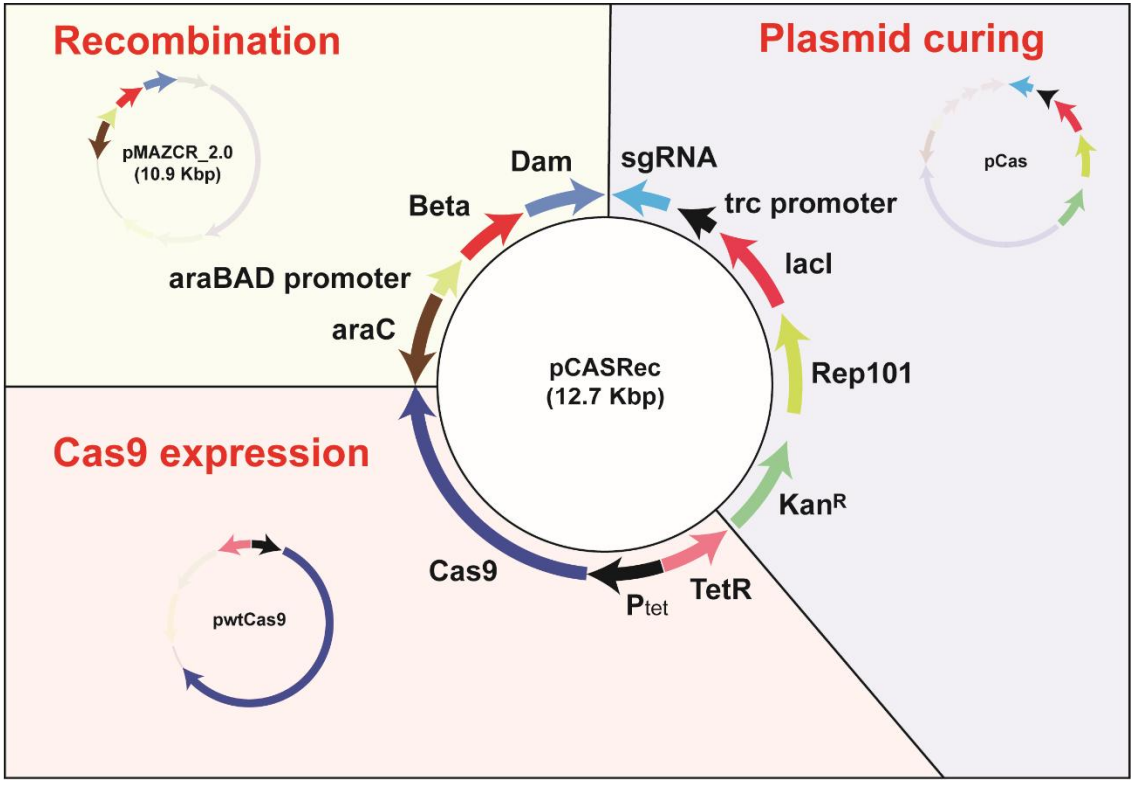


Figure 2.2 Graphical description of the configuration of pCASRec

For achieving the high level of endonuclease activity using CRISPR/Cas system, high concentration of Cas9 protein and sgRNA are required. In General, in order to express Cas9 at high concentration in *E. coli*, the system which induces Cas9 protein by tetracycline is used. Therefore, the Cas9 protein expression system was amplified from pwtCas9 (a gift from Stanley Qi (Addgene plasmid # 44250)) by pCR_frag2_F/pCR_frag2_R and prepared as fragment 2 constituting pCASRec (Qi et al. 2013).

After completion of genome editing using the CRISPR/Cas system, the plasmids used for editing should be removed. We used two plasmids to express Cas9 protein and sgRNA, and therefore needed a system to remove both plasmids. Previously reported pCas has a well-established curing system for the sgRNA expression plasmid and pCas itself (Jiang et al. 2015). The pCas uses a CRISPR/Cas system to remove pgRNA, which is a plasmid that expresses sgRNA, and uses a temperature sensitive replication origin for the self-curing. Therefore, the curing system to remove two plasmids was amplified from pCas by pCR_frag3_F/pCR_frag3_R and prepared as fragment 3 constituting pCASRec. The pCASRec was finally constructed through CPEC with three fragments prepared above. The detailed design of the primers is given in **Table5**.

Table 5 Primers used in this study (chapter 5)

Plasmid	Description	Sequence (5' – 3')
pCas_v2.0	v2_frag1_F	AAAGCCGCAGAGCAGAAGGTGGCAGCATGACGCATCCTCACGATAATATC
	v2_frag1_R	GCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGA
	v2_frag2_F	GCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGT
	v2_frag2_R	TTAAGAAATAATCTTCATCTAAAATATACTTCAGTCACCTCCTAGCTGAC
	v2_frag3_F	ATTGATTTGAGTCAGCTAGGAGGTGACTGAAGTATATTTTAGATGAAGAT
	v2_frag3_R	AATATCAACCTGGTGGTGAGCAATGGTTTCATGGAGAAACAGTAGAGAGT
	v2_frag4_F	TTTTATCGCAACTCTCTACTGTTTCTCCATGAAACCATTGCTCACCACCA
	v2_frag4_R	CGCCTACCCGGATATTATCGTGAGGATGCGTCATGCTGCCACCTTCTGCT
pCASRec	pCR_frag1_F	TCACCTTCGGGTGGGCCTTTCTGCGAGCTC TTATCATCGATGCATAATGT
	pCR_frag1_R	CGCCTACCCGGATATTATCGTGAGGATGCGTTATTTTTTCGCGGGTGAAA
	pCR_frag2_F	GAGTAGAATAGAAGTATCAAAAAAAGAGCTCACATTTCCCCGAAAAGCAT
	pCR_frag2_R	CAGGCACATTATGCATCGATGATAAGAGCTCGCAGAAAGGCCCACCCGAA
	pCR_frag3_F	CCAGGAGTCGTTTCACCCGCGAAAAAATAACGCATCCTCACGATAATATC
	pCR_frag3_R	TAAGCATGCTTTTCGGGGAAATGTGAGCTCTTTTTTTGATACTTCTATTC
pgRNA_*lacZ*	lacZ_F	ATATACTAGTATGTCGGTTTCCGCGAGGTG GTTTTAGAGCTAGAAATAGC
pgRNA_dummy	Dummy_F	ATATACTAGTAACTTTCAGTTTAGCGGTCTGTTTTAGAGCTAGAAATAGC
pgRNA_*dxr*	dxr_F	ATATACTAGTTTTTTATTCTGTCTCAACTCGTTTTAGAGCTAGAAATAGC
pgRNA_*ispA*	ispA_F	ATATACTAGTCGGTGTAAAAGGGGTTAGAGGTTTTAGAGCTAGAAATAGC
pgRNA_*idi*	idi_F	ATATACTAGTGTTGTCGATAAACGCTCACTGTTTTAGAGCTAGAAATAGC
pgRNA_*edd*	edd_F	ATATACTAGTTCACAACGCGTTTCATTCAGGTTTTAGAGCTAGAAATAGC

pgRNA_*zwf*	zwf_F	ATATACTAGTAGAAAATTACAAGTATACCCGTTTTAGAGCTAGAAATAGC
pgRNA_*gnd*	gnd_F	ATATACTAGTATCACCGCGCTGAATGCTCGGTTTTAGAGCTAGAAATAGC
pgRNA_*lacZ~gnd*	Reverse	AATTACTAGTATTATACCTAGGACTGAGCT
pGGA2_DIE	GGA_ispA_F	ATATGAAGACAATAGTACACCGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCGGTGTAAAAGG GGTTAGAGGTTTTAGAGCTAGAAATAGC
	GGA_ispA_R	ATATGAAGACAAGTGCAAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
	GGA_dxr_F	ATATGAAGACAAGCACTAAAAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTTTTTATTCTGTC TCAACTCGTTTTAGAGCTAGAAATAGC
	GGA_dxr_R	ATATGAAGACAACAATAAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC
	GGA_edd_F	ATATGAAGACAAATTGTTGTGACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTCACAACGCGTTT CATTCAGGTTTTAGAGCTAGAAATAGC
	GGA_edd_R	ATATGAAGACAAATGGAAAAAAAAAAAAAAAAAAAAAGCACCGACTCGGTGCCAC

2.3 Extraction of RNA from *E. coli*

Total RNA from *E. coli* cultures grown as described above was isolated with the RNAsnap protocol (Stead et al. 2012). The isolated RNA was equilibrated in 350µl of buffer RLT containing β-mercaptoethanol (1%), and vortex vigorously. Then, 250µl EtOH was added to the sample. The sample was vortexed, and applied to an RNeasy Mini column, and RNA extracted according to manufacturer's instructions (Qiagen).

2.4 *In vitro* transcription

Each plasmid that expresses a transcript containing HHRzs was digested with the restriction enzyme SmaI to linearize the DNA. After purification, 100ng of linearized plasmid was used as a template for *in vitro* transcription using T7 High Yield RNA Synthesis Kit (New England BioLabs, cat. Number #E2040S). The buffer composition was: 2 µl ATP (100 mM), 2 µl GTP (100 mM), 2 µl CTP (100 mM), 2 µl UTP (100 mM), 2 µl reaction buffer, 2 µl T7 RNA polymerase mix, 4 µl template. The total volume of the reaction mixture was 20 µl. The reaction was incubated 2hr at 37◦C.

2.5 Reverse transcription

For RT-PCR, first-strand cDNA was synthesized from 1ug of total RNA with a random hexamer and M-MLV reverse transcriptase (Promega).

2.6 Quantification of transcript

The qPCR was performed with TOPreal SYBR Green PCR Kit (Enzynomics) on the Roche LightCycler® 480 real-time platform to quantify the expression of specific mRNAs. Gene expression was normalized relative to rssA (Zhou et al. 2011). The Fragment-a was amplified using primers Frag_a_F and Frag_a_R and the Fragment-b was amplified using primers Frag_b_F and Frag_b_R. The fragment of rssA was amplified using primers rssA_F and rssA_R. PCR reactions were as follows: 95°C for 5min, followed by 40 cycles of 95°C for 20sec, 55°C for 20sec and 72°C for 20sec, followed by a final extension of 5min at 72°C. All samples were amplified in triplicates technically. All primer sets generated 100bp amplicon. **Table 2** shows the primers used for these experiments.

2.7 Circularized RT-PCR

Total RNA was isolated from the BL21 (DE3) strain containing a plasmid transcribing a primary transcript composed of Rz and sgRNA. The RNAs processed by Rz leaves 5'-hydroxyl and 2';3'-cyclic phosphates as cleavage products. In order to circularize the cleaved RNA, the 5' end of cleaved RNA have to contains a phosphate group. Thus, the isolated total RNA was treated with T4 polynucleotide kinase (New England BioLabs) before circularization with the T4 RNA ligase1 (NEB). After circularization, reverse transcription was performed by M-MLV reverse transcriptase (Promega) with random hexamer. Then, PCR amplification was perform with cRT_F and cRT_R **(Table 4)**. The PCR products

were analyzed with 1.5% agarose gel electroporation and sequencing was subsequently performed.

2.8 Genome editing using CRISPR/Cas system

Single-stranded oligonucleotides (ssDNA) and double-stranded DNA (dsDNA) were used as editing templates for homologous recombination. The dsDNA was prepared by assembly PCR and the ssDNA was synthesized. The detailed sequence and design can be found in **Table 6**.

The *E. coli* strains (BL21(DE3), and MG1655 harboring pCas or pCas_v2.0 or pCASRec were overnight cultured in 2ml LB liquid media with 50mg/ml kanamycin. The following day, 10ml of LB liquid media with 50mg/ml kanamycin and 10mM of arabinose was inoculated with 0.1ml of the overnight culture in a 100ml flask. When the OD600 reached to ca. 0.6, the culture was put in an ice-water bath and left to cool for 5 minutes. The culture was pelleted by centrifugation and the supernatant was discarded. The pellet was washed twice with 10ml of 10% glycerol. For electroporation, 50ul of cell was mix with 100ng of pgRNA which can transcribe sgRNA and editing template (50uM of ssDNA or 500ng of dsDNA). The electroporation was done in a 2-mm Gene Pulse cuvette (Bio-Rad) at 2.5kV, and 1ml of ice-cold LB liquid media was added immediately. Cells were recovered at 30°C for 1h before being spreaded onto LB agar plate with 50mg/ml kanamycin and 100mg/ml ampicillin. The cells were incubated at 30°C or 33°C.

Table 6 Single-stranded oligonucleotides used in this study (chapter 5)

Description	Sequence (5' – 3')
ET_dxr	T*G*C*T*TCATGAAACATCCAGAGTTGAGACAGGCTAGCACTGTACCTAGGACTGAGCTAGCCGTCAAAAAGCAAAACGCCGCCAGCCGATCCGCAAG
ET_ispA	T*A*C*A*AATTCTGCTGTCTGACAATGAAGACGTTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCTAACCCCTTTTACACCGGACAATGAGTAAT
ET_idi	T*T*T*A*CCTGTCGGCATCCGCTCAAAACGGGCTTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCGTGATCAGAATTACATGTGAGAAATTATGC
ET_edd	A*A*A*G*GCTCCTGAAATTGAGTTGTCAGAGCAGGATGGCTAGCACTGTACCTAGGACTGAGCTAGCCGTCAACAGAGGATTTATGACTGAAACGCCTGTAACCGGAG
ET_zwf	G*T*C*A*TTCTCCTTAAGTTAACTAACCCGGTACTTAAGCTAGCACTGTACCTAGGACTGAGCTAGCCGTCAATATACTTGTAATTTTCTTACGGTGCACTGTACTGC
ET_gnd	T*A*C*A*TACTCCTGTCAGGTGTGATCACCGCGCTGAAGCTAGCACTGTACCTAGGACTGAGCTAGCCGTCAATCGCGTCCTGGAATGTTCGCAAATAAGTATACAAAGTACTT
ET_lacZ	A*A*C*T*TAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTG

(note: * indicates phosphorothioate bond)

2.9 Measurement of the endonuclease activity of Cas9 protein

The endonuclease activity of Cas9 (Ec) was calculated through comparison of the CFU of a strain expressing both Cas9 and targeting sgRNA without providing editing template, and the CFU of a strain expressing both Cas9 and non-targeting sgRNA without providing editing template.

$$E_C = \left(1 - \frac{CFU_{targeting\ sgRNA}}{CFU_{non-targeting\ sgRNA}}\right) \times 100$$

2.10 Plasmid curing

For the curing of pCASRec and pgRNA, the edited colony harboring both plasmids was inoculated into 2ml of LB liquid media with 50mg/ml kanamycin, 100mg/ml ampicillin, and 0.1mM of IPTG (isopropyl-β0D-thiogalactopyranoside) at 30°C. After several rounds of cultivation, the culture was inoculated into 2ml of LB liquid media with 50mg/ml kanamycin, and 2ml of LB liquid media with 50mg/ml kanamycin, 100mg/ml ampicillin in order to confirm the curing of pgRNA. The cell with the removal of pgRNA were incubated overnight at 42 ° C to allow self-curing of pCASRec.

2.11 Lycopene measurement

Lycopene producing strains were quantitatively assessed for lycopene production after 18 h of growth in 5ml of LB liquid medium containing chloramphenicol, ampicillin and kanamycin at 30°C. Then, 1ml of cultures were pelleted and washed

once with distilled water. The pellets were resuspended with 200µl of acetone and incubated at 55°C for 15 min in the dark and vortexed every 5 min. After the extraction, the solutions were centrifuged at 13,000 rpm for 5 min to remove cell debris. The lycopene concentration was measured at an absorbance of 475nm using 100µl of supernatant. The values of the measured lycopene concentration were normalized to the dry cell weight.

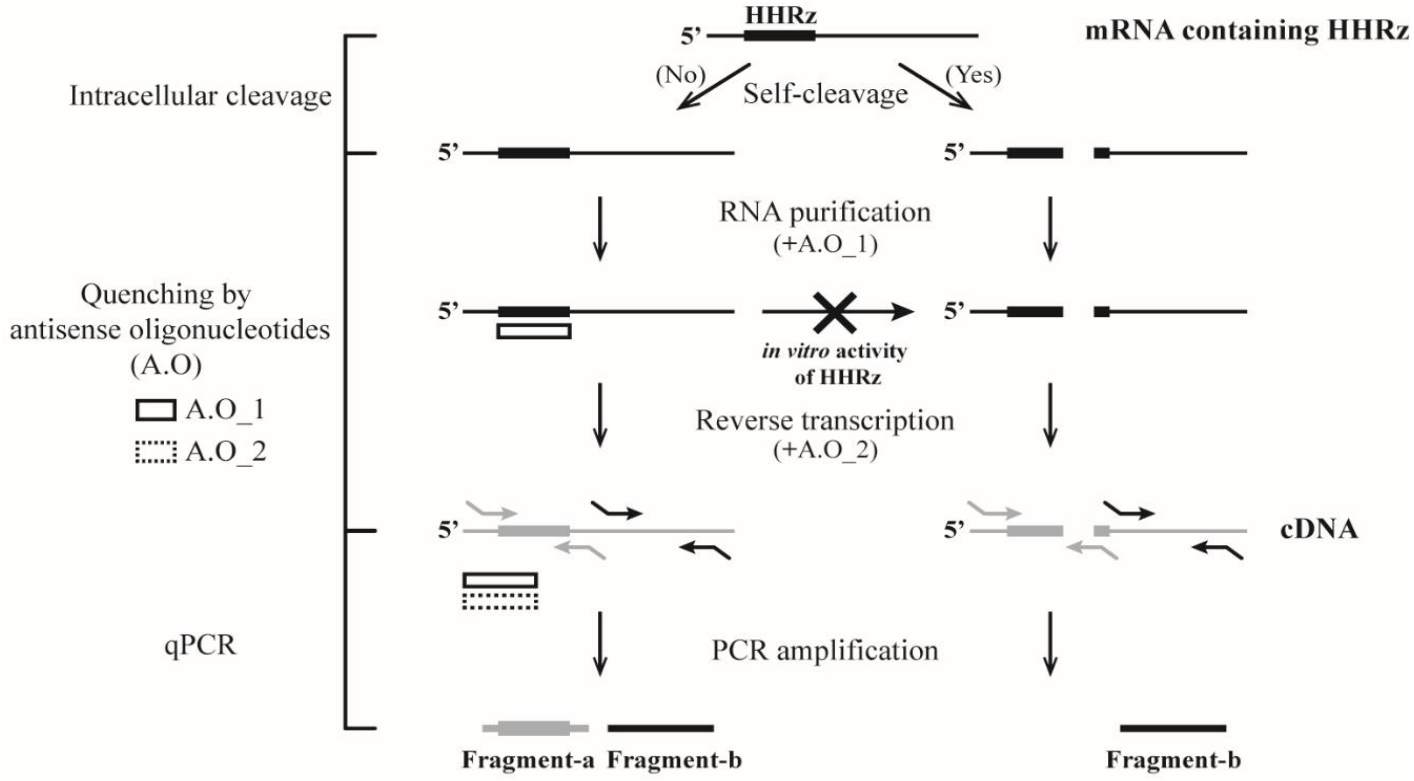
Chapter 3.

Development of Quenching-qPCR (Q-Q) assay for measuring absolute intracellular cleavage efficiency of Rz

3.1 Development of calculation method of absolute cleavage efficiency of HHRz

To calculate the cleavage efficiency of HHRz, mRNA containing HHRz was used as a model transcript. Primers were designed to amplify Fragment-a and Fragment-b, and their PCR was done using the cDNA generated from the transcript described above **(Figure 3.1)**. The Fragment-a cannot be amplified when the mRNA is cleaved by HHRz, but can be amplified when the corresponding mRNA is not cleaved. Thus, the concentration of Fragment-a ([A]) corresponds to the amount of the remaining mRNA not cleaved by HHRz among the total amount of the mRNA transcripts. Due to the Fragment-b does not contain HHRz, the concentration of Fragment-b ([B]) corresponds to the total amount of mRNA expressed under the T7 promoter. As a result, it is possible to calculate the absolute cleavage efficiency of HHRz by comparison of the concentrations of Fragment-a and Fragment-b as follows:

$$Cleavage\ efficiency\ = \left(1 - \frac{[A]}{[B]}\right) \times 100$$



Concentration of Fragment-a [A]: Uncleaved mRNA

Concentration of Fragment-b [B]: Total mRNA

Figure 3.1 proposed method to evaluate the cleavage efficiency of HHRz.

Antisense oligonucleotide_1(A.O_1) was used to deactivate the *in vitro* activity of HHRz and antisense oligonucleotide_2(A.O_2) was used to remove the A.O_1 from HHRz during the reverse transcription. The concentrations of Fragment-a and Fragment-b are measured by qPCR, which represent the concentration of uncleaved mRNA ([A]) and total mRNA ([B]), respectively. Gray bended arrows are primer sets for amplifying Fragment-a, and black bended arrows are primer sets that amplify Fragment-b. [A], the amount of mRNA not cleaved by HHRz; [B], the amount of total mRNA.

3.2 Quenching method to deactivate the *in vitro* cleavage activity of HHRz by using asDNA and anti-asDNA

To calculate absolute "intracellular cleavage efficiency" of HHRz using the above equation (1), it is important that the extracted mRNA from the sampled cell is maintained as almost the same as that in intracellular state. Therefore, complete deactivation of *in vitro* cleavage activity of HHRz is essential for absolute calculation of "intracellular cleavage efficiency", since HHRz catalytic activity may still be remained during the mRNA purification step and the reverse transcription after cell lysis. Since the cleavage reaction of HHRz takes place by forming its tertiary structure in the presence of Mg2+ (Denesyuk and Thirumalai 2015), suppression of the *in vitro* cleavage activity of HHRz would be possible by preventing the formation of a tertiary structure of HHRz. Thus, an oligonucleotide complementary to the catalytic core of HHRz was designed to prevent the formation of the tertiary structure of HHRz, **(Figure 3.2)** and this nucleotide is hereinafter referred to as "asDNA". To confirm the role of asDNA for deactivation of *in vitro* cleavage activity of HHRz, t-pHHRz-n transcript was synthesized through *in vitro* transcription from pHHRz-n with/without (i.e. control) adding asDNA, and subsequently, cDNAs for each transcript were synthesized. In result, the calculated cleavage efficiency of the t-pHHRz-n without addition of asDNA was almost 100% **(Figure 3.3)**, indicating that the *in vitro* cleavage activity of HHRz is still very high. However, unlike the expectation, calculated cleavage efficiency of t-pHHRz-n treated with asDNA was also close to 100% **(Figure 3.3)**.

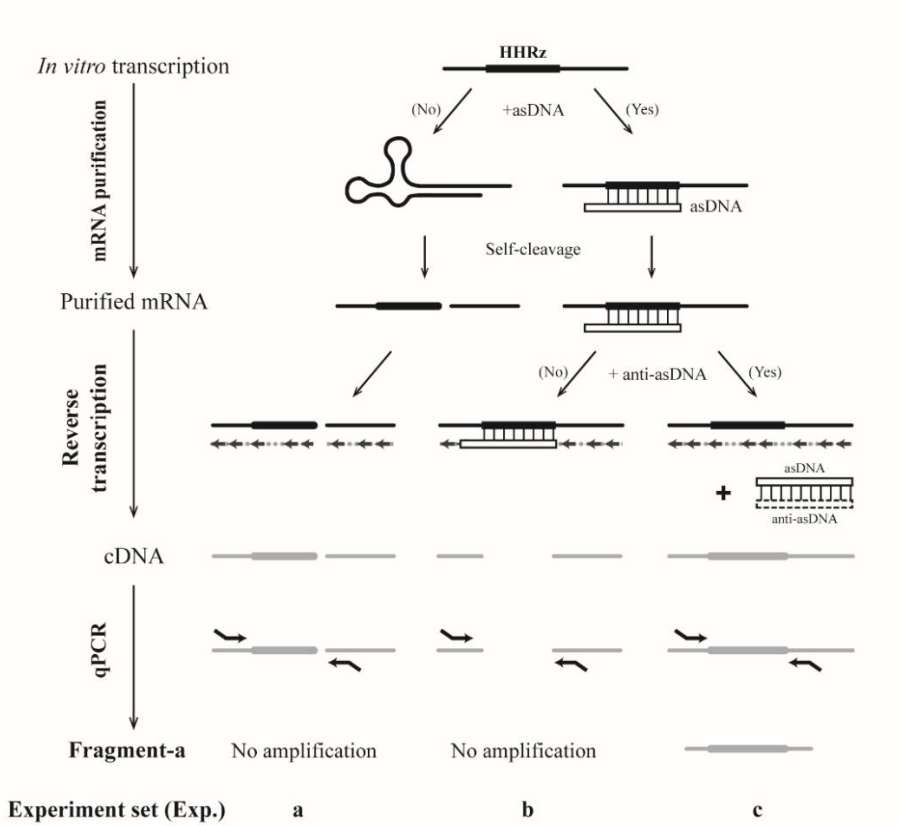


Figure 3.2 Expected functions of asDNA and anti-asDNA.

asDNA binds to the catalytic core of HHRz and prevents the formation of its tertiary structure. Anti-asDNA removes the asDNA attached to the HHRz, by binding with asDNA. The *in vitro* activity of HHRz can be completely deactivated by using asDNA and anti-asDNA, whereby the accurate concentration of Fragment-a can be measured.

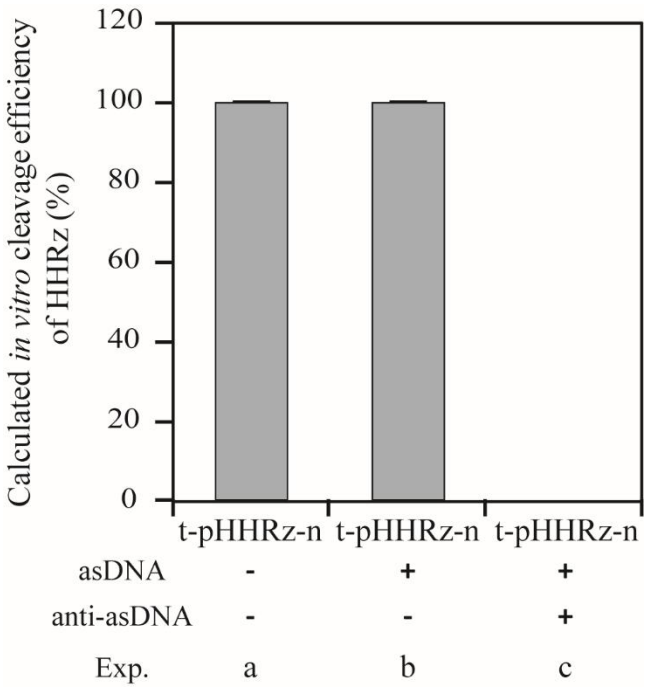


Figure 3.3 Complete deactivation of *in vitro* activity of HHRz by asDNA and anti-asDNA to accurately measure the concentration of Fragment-a.

The *in vitro* cleavage efficiency of the transcript containing HHRz was 0% when the asDNA and anti-asDNA were added (c), otherwise it was 100% (a, b). Results are representatives of three biological replicates. The arrows used in the reverse transcription step indicate random hexamer. The gray dotted line means cDNA synthesis. The bended arrows used in the qPCR step indicate primer sets for amplifying Fragment-a.

This result suggests that strong binding of asDNA and HHRz might prevent the binding of random hexamer used in the process of reverse transcription to HHRz, and cause incomplete cDNA synthesis. Therefore, another oligonucleotide was designed to remove asDNA from HHRz during the reverse transcription. **(Figure 3.2)**. This nucleotide is hereinafter referred to as "anti-asDNA". In order to remove asDNA from HHRz through anti-asDNA, the binding energy between anti-asDNA and asDNA should be higher than that between HHRz and asDNA. Since binding energy of nucleotides increases according to the length of hybridization (Suzuki et al. 2007), the hybridized sequence lengths of anti-asDNA and asDNA were designed to be longer than that of HHRz and asDNA. Thus, an extended nucleotide with a length of 10 bps that do not bind to HHRz, but can bind to anti-asDNA should be added to the 5 'end of asDNA. In this case, the calculated *in vitro* cleavage efficiency of HHRz was turned into almost zero, demonstrating the effective role of anti-asDNA **(Figure 3.3)**. These results showed that the quenching method to deactivate the *in vitro* cleavage activity of HHRz by using asDNA and anti-asDNA can be used for accurate calculation of the absolute "intracellular cleavage efficiency" of HHRz.

3.3 The calculated "intracellular cleavage efficiency" of Rz was constant regardless of the degree of mRNA degradation

After confirming that the *in vitro* activity of HHRz can be effectively deactivated by using asDNA and anti-asDNA, the "intracellular cleavage efficiency" of HHRz was calculated. Wild type sTRSV and its mutant type sTRSV + PL1 & 2 were used as a model systems to confirm the reproducibility of the above quenching method, since the catalytic activity of the mutant HHRz was reported to be higher than that of wild type sTRSV (Khvorova et al. 2003). Transcripts containing each HHRz were generated from the pHHRz-n and pHHRz-m plasmids under T7 promoter in BL21 (DE3) strain. The strain containing pHHRz-c plasmid was used as a negative control. The transcripts extracted from each strain were named as T-pHHRz-n, T-pHHRz-m, and T-pHHRz-c. Corresponding asDNA and anti-asDNA were added in the beginning of RNA purification step, and subsequent reverse transcription step, respectively. The RNA samples from each strain were extracted at five time points (10, 30, 90, 150, and 210 min after IPTG induction) and each sample was used to calculate the total amount of mRNA and the amount of uncleaved mRNA. First, total amounts of mRNA obtained from each strain were compared to determine the degree of mRNA degradation according to the type of HHRz. For this, the fold change ratio was calculated using the RNA samples extracted at 10 min and 210 min after IPTG induction. As a result, the fold change ratio of T-pHHRz-c was ca. 420, whereas the fold change ratios of both T-pHHRz-n and T-pHHRz-m were ca. 40 **(Figure 3.4A)**. These results indicate that the amounts of T-pHHRz-n and T-pHHRz-m are ten times lower than that of T-pHHRz-c, suggesting that the

corresponding HHRzs present in mRNA trigger rapid degradation of mRNA as reported previously (Donahue and Fedor 1997). Then, we wanted to find out how the calculated "intracellular cleavage efficiency" of HHRz by Q-Q assay is affected by the degree of degradation of mRNA. The calculated "intracellular cleavage efficiency" values of HHRz consistently showed a fixed value regardless of the degree of mRNA degradation. The calculated cleavage efficiencies of T-pHHRz-c, T-pHHRz-n and T-pHHRz-m were 4.7%, 89.3% and 93.1%, respectively **(Figure 3. 4B)**. This result is consistent with the previous report showing that mutant HHRz activity was higher than its wild type HHRz activity (Suzuki et al. 2007).

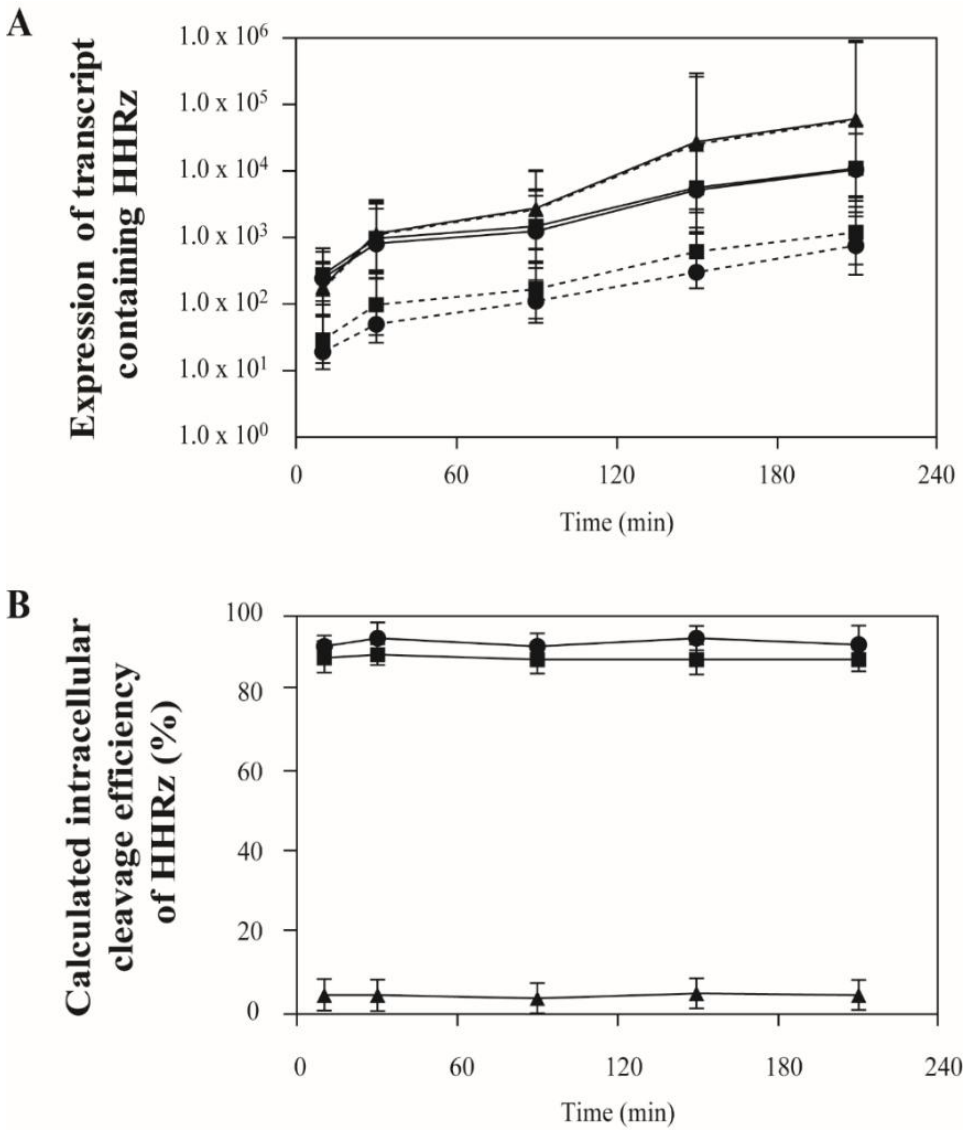


Figure 3.4 The intracellular cleavage efficiency of HHRz can be calculated irrespective of the degree of mRNA degradation.

(A) Comparison of the amount of total mRNA containing HHRz (solid line) and the amount of uncleaved mRNA containing HHRz (dashed line). The sampling was performed at five time points (10, 30, 90, 150, and 210 min) after IPTG induction. There are three types of transcript, T-pHHRz-c (triangle), T-pHHRz-n (square), and T-pHHRz-m (circle). The increase ratio of the total amount of T-pHHRz-c was ten times higher than that of T-pHHRz-n and T-pHHRz-m. This result indicates that mRNA degradation is prompted by the cleavage reaction of HHRz. (B) Measurement of the intracellular cleavage efficiency of HHRz according to the change in time. As expected, constant values were obtained irrespective of the degree of mRNA degradation. Results are representative of the three biological replicates.

3.4 Conclusion

Our study established a simple and accurate method called "quenching-qPCR (Q-Q) assay" to accurately evaluate the intracellular cleavage efficiency of HHRz. The Q-Q assay completely deactivate the *in vitro* cleavage activity of HHRz by quenching method during the cDNA synthesis, and calculate the accurate intracellular cleavage efficiency of HHRz by measuring the concentrations of cleavage products released from HHRz. This method can simultaneously overcome the two disadvantages of conventional analytical methods, i.e. requirement of radiolabeled RNA and inaccurate calculation of intracellular cleavage activity of Rz. Using this method, now we can precisely compare different HHRzs and screen better HHRzs with higher absolute catalytic activity among various HHRzs even *in vivo* system. The proposed method was applied to measure "intracellular cleavage efficiency" of sTRSV, a representative Rz, and its mutant, and their intracellular cleavage efficiencies were calculated as 89% and 93%, respectively.

Chapter 4.

Multiple engineering in *Escherichia coli* by using CRISPR/cas system couple with Rz

4.1 Multiple sgRNA generating strategy by using Rz

In order to generate the multiple sgRNAs from a single promoter, a primary transcript has to be truncated to generate individual sgRNAs and the individual sgRNAs have to maintain their function as a navigator of dCas9 protein. The sgRNA consists of 20bp spacer sequence which recognize the target gene, and 76bp sgRNA scaffold which is necessary for dCas9 binding. Since the extension or truncation of the spacer sequence significantly affects the activity of dCas9 (Larson et al. 2013; Ren et al. 2014), the accurate cleavage activity is required to generate sgRNAs from a primary transcript. Rzs are RNA molecules with self-cleavage activity at a specific site and meet the aforementioned requirements (Serganov and Patel 2007). HHRz (hammerhead Rz), one of typical ribozyme, is consisted of stem loops I, II, and III, and the catalytic activity of HHRz is determined by the tertiary interactions between stem loops I and II (Khvorova et al. 2003). In particular, the cleavage site of the RNA can be precisely determined by designing the six nucleotides of the stem loop I to be complementary to the six nucleotides immediately adjacent to the cleavage site **(Figure 4.1A)**. Therefore, we hypothesized that the Rz can be used to cleave the primary transcript for generating multiple sgRNAs.

To construct a primary transcript capable of generating multiple sgRNAs simultaneously, pre-long sgRNAs was designed in which each sgRNAs are connected by Rz **(Figure 4.1B)**. When the pre-long sgRNAs is transcribed in the presence of Cas9 protein, each sgRNAs cleaved by Rz can form a complex with Cas9 and can affect the target genes **(Figure 4.1C)**.

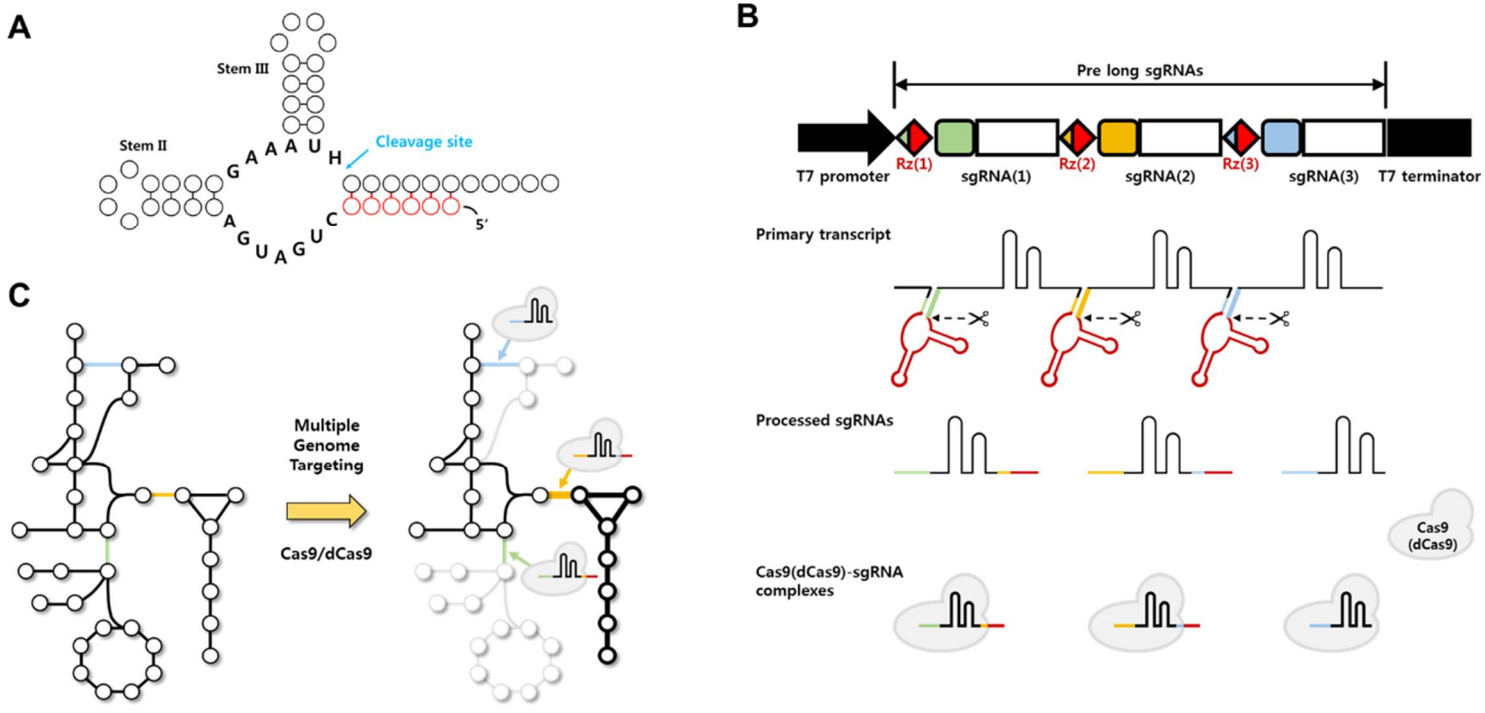


Figure 4.1 Multiple sgRNA generation strategy using Rz for multiplex genome editing with CRISPR/Cas

(A) The secondary structure of ribozyme. (B) Design of primary transcripts capable of expressing three different sgRNAs. (C) Multiple modulation of metabolic pathway through the CRISPR/Cas system

4.2 *In vivo* cleavage assay of Rz

Mismatched extension or truncation of 5' end of sgRNA significantly reduce the activity of dCas9/sgRNA complex as aforementioned. Therefore, in order to use the Rz to generate several sgRNAs, the sequence specific cleavage activity of Rz in intracellular has to be confirmed. We constructed a pGGA1_AL capable of transcribing a primary transcript composed of two sgRNAs (each sgRNA targets *adhE* and *lacZ*) connected by Rz. The plasmid pGGA1_ALc, which expresses the primary transcript with two sgRNAs connected without Rz, was used as a control. Both plasmids were transformed in BL21(DE3) strain and the primary transcript were transcribed under the T7 promoter by induction with IPTG. Then, cRT-PCR was performed to confirm the sequence-specific cleavage activity of Rz in intracellular (Slomovic and Schuster 2013) **(Figure 4.2A and B)**. As a result, a PCR amplicon that could be amplified when the primary transcript was cleaved by Rz was detected in the BL21(DE3) strain harboring pGGA1_AL and the accurate sequence of 5' end of mature sgRNA was confirmed **(Figure 4.2C)**. This result indicates that the Rz has sequence-specific cleavage activity in intracellular.

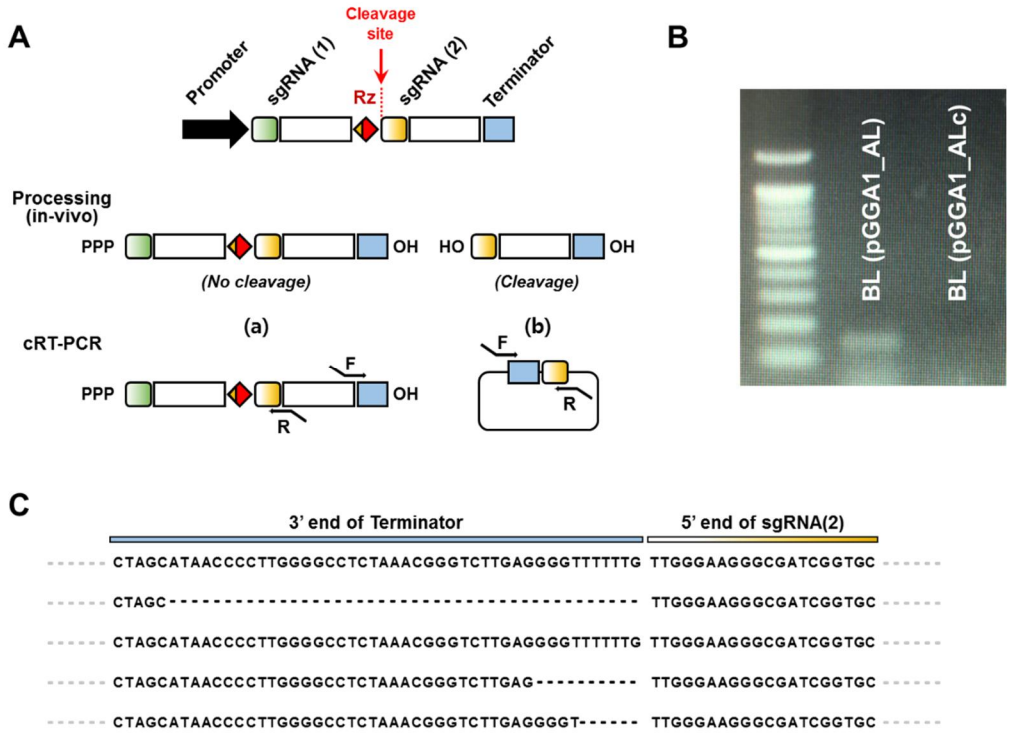


Figure 4.2 Precise intracellular cleavage activity of Rz

(A) The process of cRT-PCR. (B) Verification of cleavage activity of Rz with cRT-PCR (BL(pGGA1_AL): primary transcript containing Rz, BL(pGGA1_ALc): primary transcript without Rz), (C) Verification of the sequence specific cleavage activity of Rz

4.3 Mature sgRNAs cleaved by Rz have function as a navigator of dCas9

After confirming the accuracy of intracellular cleavage activity of Rz, the experiment to confirm the function of mature sgRNA was performed. Since the repression of *lacZ* expression by CRISPRi system is expected to be easily distinguished by blue/white colony assay, the *lacZ* gene was selected as a target for functional testing of sgRNA. At first, we have conducted experiments to confirm the expectation mentioned above. Plasmids for transcribing sgRNA targeting each of *adhE* and *lacZ* were prepared, respectively (pGGA1_*adhE* and pGGA1_*lacZ*). Each plasmid was transformed into BLd (BL21(DE3) strain harboring pdCas9-bacteria), and the blue/white colony assay was performed for both strains. As a result, the expectation that the repression of *lacZ* gene by CRISPRi system can be screened by blue/white colony assay was confirmed by observing the presence of white colony only in the BL21(DE3) strain containing pGGA1_*lacZ* **(Figure 4.3)**. After that, we examined the function of sgRNAs generated by truncating the primary transcript by Rz. When the primary transcript is cleaved by Rz, two sgRNAs can be generated on the left and right of Rz, respectively. In order to confirm the function of both mature sgRNAs, two plasmids capable of transcribing two sgRNAs were designed. One is a plasmid in which sgRNA targeting *lacZ* is placed before Rz (pGGA1_LA), and the other is that sgRNA targeting *lacZ* is place after Rz (pGGA1_AL). Subsequently, the blue/white colony assay was performed using strains prepared by transformation of each plasmid into BLd strains. As a result, white colonies were observed with efficiency of 100% in both strains

(Figure 4.3). This results indicate that the sgRNAs cleaved by Rz are functional regardless of its location. From now on, the multiplexing experiment using CRISPR/Cas system and Rz system is hereinafter referred to as "CRISPRz" system.

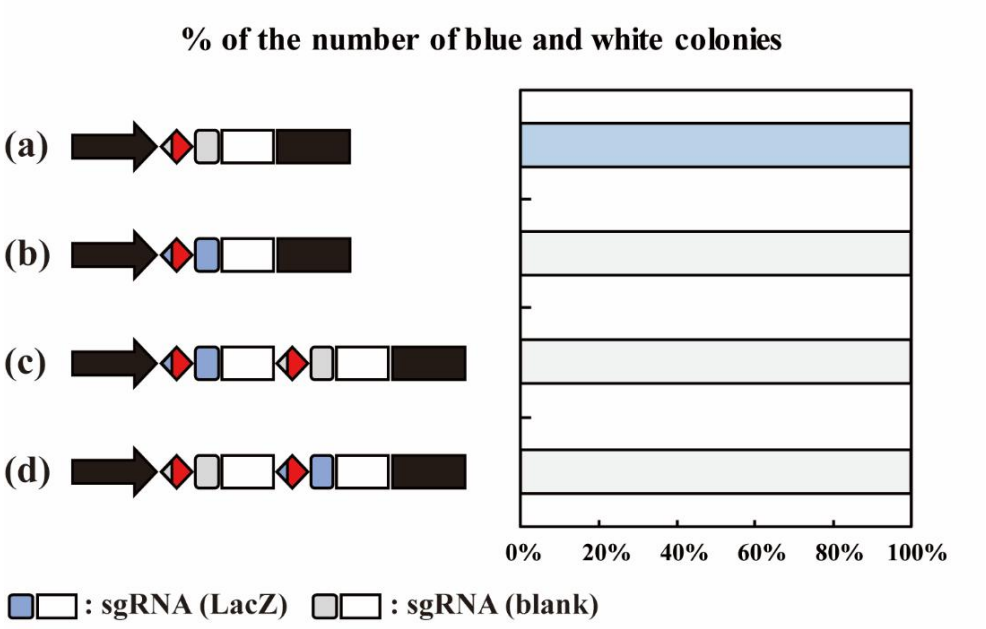


Figure 4.3 Functional test of matured sgRNA.

The red diamonds indicate the ribozyme (reft). The blue bar represents the percentage of the blue colony, and gray bar represents the percentage of white colony (right).

4.4 The number of functional sgRNAs was increased up to five by transcribing a primary transcript using strong promoter.

In order to use Rz as a molecular scissor to generate multiple sgRNAs from a primary transcript, it is necessary to confirm that the function of mature sgRNA is not affected by its location and number within primary transcript. We decided to observe changes in the degree of *lacZ* gene repression by altering the location of the sgRNA targeting *lacZ* gene in various lengths of primary transcripts. Therefore, five plasmids, each producing different primary transcript (PT1-PT5), were constructed and transformed into BLd strain. PT1 - PT3 were used to investigate the changes in the function of sgRNA as the number of sgRNAs constituting the primary transcript increased and PT3-PT5 were used to investigate the changes in the function of sgRNA according to the altered location of sgRNA in the same length of primary transcripts. When the blue / white colony assay was carried out for each strain transcribing PT1-PT3, only white colonies were observed with 100% efficiency from all the strains **(Figure 4.4)**. These results indicate that even if the number of sgRNAs constituting the primary transcript increases from one to five, the function of specific sgRNA is not reduced. However, as a result of performing the same experiments using each strain transcribing PT3-PT5, it was observed that blue colonies appeared as the sgRNA targeting *lacZ* was transcribed far from the promoter **(Figure 4.4)**. These results mean that the ability of sgRNA can be reduced as the transcribed position of sgRNA in primary transcript is away from the promoter. As mentioned before, the primary transcript is consisted of several

sgRNAs and each sgRNA is connected by Rz. In general, however, Rzs promote the RNA degradation (Donahue and Fedor 1997). Therefore, it can be predicted that as the sgRNA position moves away from the promoter, RNA degradation is promoted due to accumulation of the effect of Rz and the amount of sgRNA is decreased.

However, to perform the multiple engineering with multiple sgRNAs coupled with dCas9 protein, the ability of sgRNA should not be affected by changes in transcribing position within primary transcript. Since the cause of reduced ability of sgRNA was predicted to be the decreased amount of sgRNA, we tried to overcome the problem through increase the transcription level of primary transcript by using T7 promoter. When the blue / white colony assay was carried out for each strain transcribing PT3-PT5 under the T7 promoter, it was observed that the blue colonies were disappeared completely for all strains **(Figure 4.4)**. Through these results, it was confirmed that when the primary transcript composed of multiple sgRNAs was transcribed by the strong promoter, the function of each sgRNA is maintained irrespective of the change in the position and number of sgRNA.

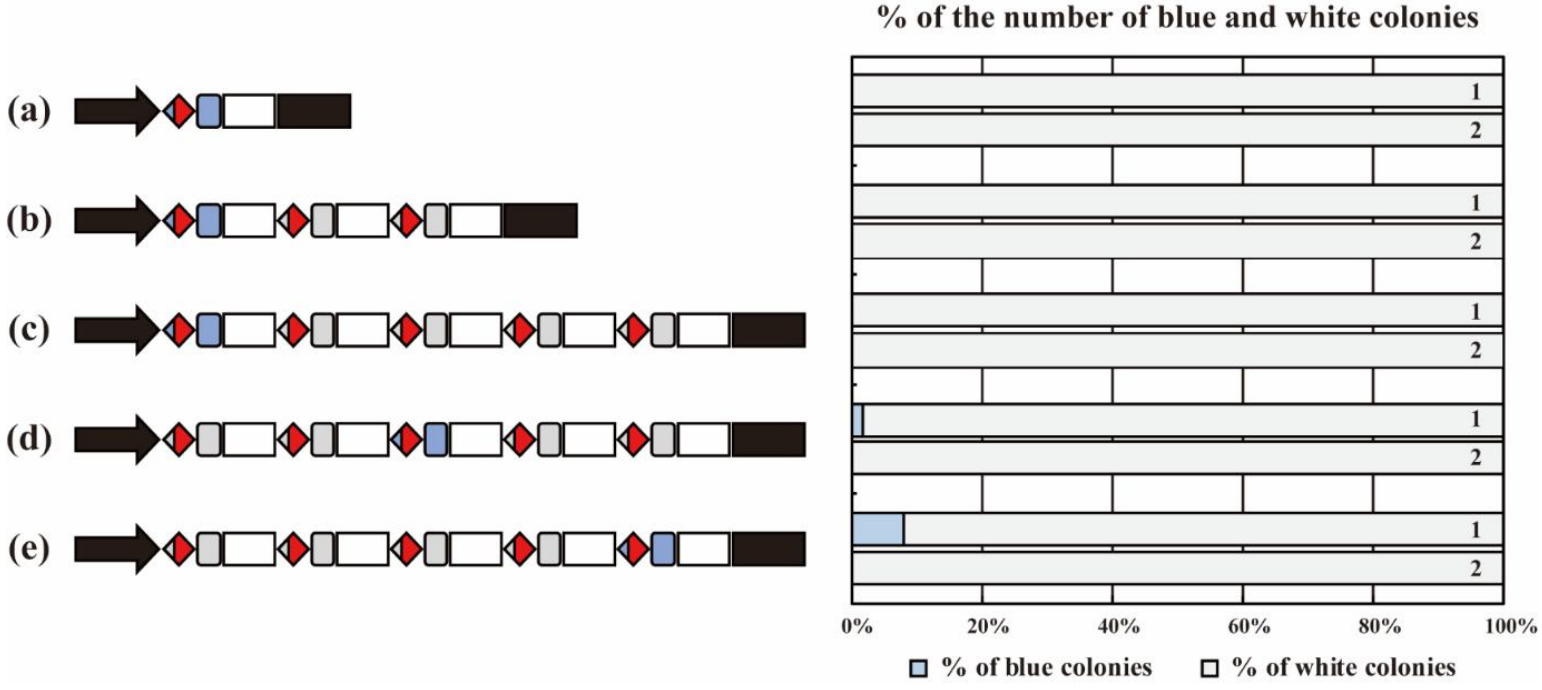


Figure 4.4 The number of functional sgRNAs was increased up to five by transcribing a primary transcript using strong promoter. The number 1 in the barchart represents the strains with constitutive expression (J23119p) of sgRNAs, and the number 2 represents the strains with strong expression (T7p) of sgRNAs.

4.5 Efficient multiplex genome engineering in *Escherichia coli* via CRISPRz

Next, we conducted experiments to ascertain whether several sgRNAs can bind to dCas9 protein and simultaneously repress each gene. To demonstrate the potential application in metabolic engineering, we used our system to repress several genes related with amino acid biosynthesis at the same time. The degree of repression was confirmed by measuring the cell growth in the minimal medium lacking focal amino acid. Five genes (*argH, lysA, metB, leuB, proC*) were selected for amino acid auxotrophy and sgRNAs targeting each gene were designed. Plasmids producing different sgRNAs were prepared and then transformed into BWd strain (BW25113 harboring pdCas9-bacteria), and it was confirmed whether each strain showed amino acid dependent growth. As a results, amino acid dependent growth was observed in three strains out of five strains **(Figure 4.5A)**. From these results, it was confirmed that three sgRNAs targeting *argH, leuB, metB* properly repress each gene, and therefore it was decided to use these sgRNAs for multiple repression. First, pGGA1_ALM producing primary transcript that simultaneously expresses those three sgRNAs was constructed. Subsequently, pGGA1_ALM was transformed into BWd strains and cell growth was measured in a minimal medium supplemented with various combinations of amino acids. As a result, the strain showed weak growth when cultured in a medium supplemented with one or two amino acids out of three amino acids **(Figure 4.5B)**. However, cell growth was completely restored only in medium supplemented with all three amino acids.

From these results, we demonstrated that multiple sgRNA generating strategy using Rz can efficiently produce functional sgRNAs.

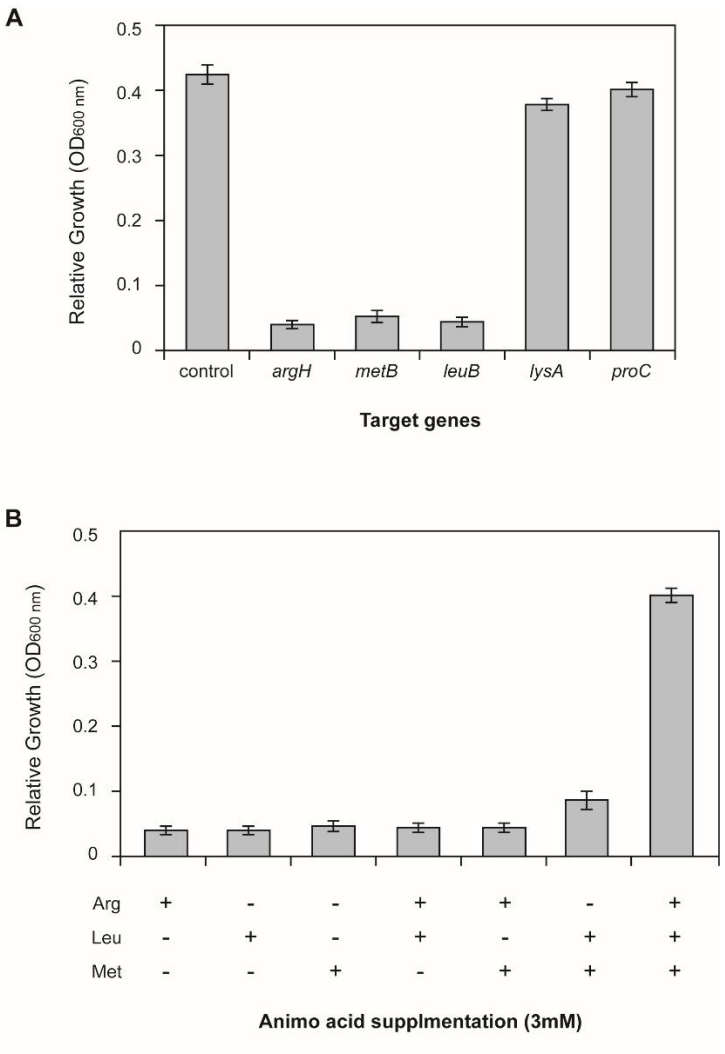


Figure 4.5 Confirmation of the multiple repression by CRISPRz

(A) Confirmation of the degree of repression for target genes by auxotrophy, (B) demonstration of the ability of multiple repression by CRISPRz

4.6 Application of CRISPRi system in lycopene production

After confirming that multiple sgRNA generating strategy could be used to repress the expression of multiple targets simultaneously by binding to the dCas9 protein, this strategy was applied to regulate the lycopene biosynthetic pathway. Lycopene is made by glyceraldehyde 3-phosphate (G3P) and pyruvate (PYR), which are synthesized through glycolysis pathway. A number of studies have been reported to block the competitive pathway through gene knockout strategy, for balanced accumulation of G3P and PYR. However, in the case of strains in which some genes have been eliminated through the knockout strategy, there is a problem that the growth rate generally decreases. Since the repression strategy using CRISPRi can overcome these drawbacks, four genes (gdhA, aceE, gpmA, gpmB) which increased the production of lycopene through the knockout strategy were selected (Alper et al. 2005) as repression targets to confirm the effect of redirection by CRISPRi. And also, two genes (icdA, pfkA) known as rate-limiting step in TCA cycle and glycolysis were additionally selected as repression targets. In order to repress each of the six genes, six plasmids capable of transcribing six sgRNAs were prepared, respectively, and transformed into Mld strain (MG1655 strain harboring pAC-LYC and pdCas9K). Non-targeting sgRNA expression plasmid (pgRNA) was also transformed into the Mld to construct Mld1 to use as control. The repression effect of single gene by CRISPRi was confirmed through measurement of lycopene production and cell growth of each strain. Three strains with the largest increase in lycopene production were selected from the six strains, and each strain was a strain in which the expression of aceE, icdA and ytjC was

repressed **(Figure 4.6)**. The production of lycopene in each strain was increased 1.49-fold, 1.54-fold and 1.22-fold, respectively, compared to Mld1 strain. In particular, growth defects were not observed for all the strains constructed above. These results demonstrate that the CRISPRi system is effective for modulating metabolic pathways to increase lycopene production.

Next, we tried to combine the effects of individual gene repression for lycopene production by applying multiple sgRNA production strategies. A pGGA2_AIY capable of simultaneously producing three sgRNAs targeting aceE, icdA, ytjC was constructed and transformed into Mld. As a result, the Mld_AIY strain (Mld harboring pGGA2_AIY) had no defect in cell growth and the production of lycopene was increased 1.93-fold compared to Mld1. In addition, it was confirmed through the measurement of transcription level that the increase of lycopene production was caused by the repression of the expression of three genes at the same time **(Figure 4.6)**.

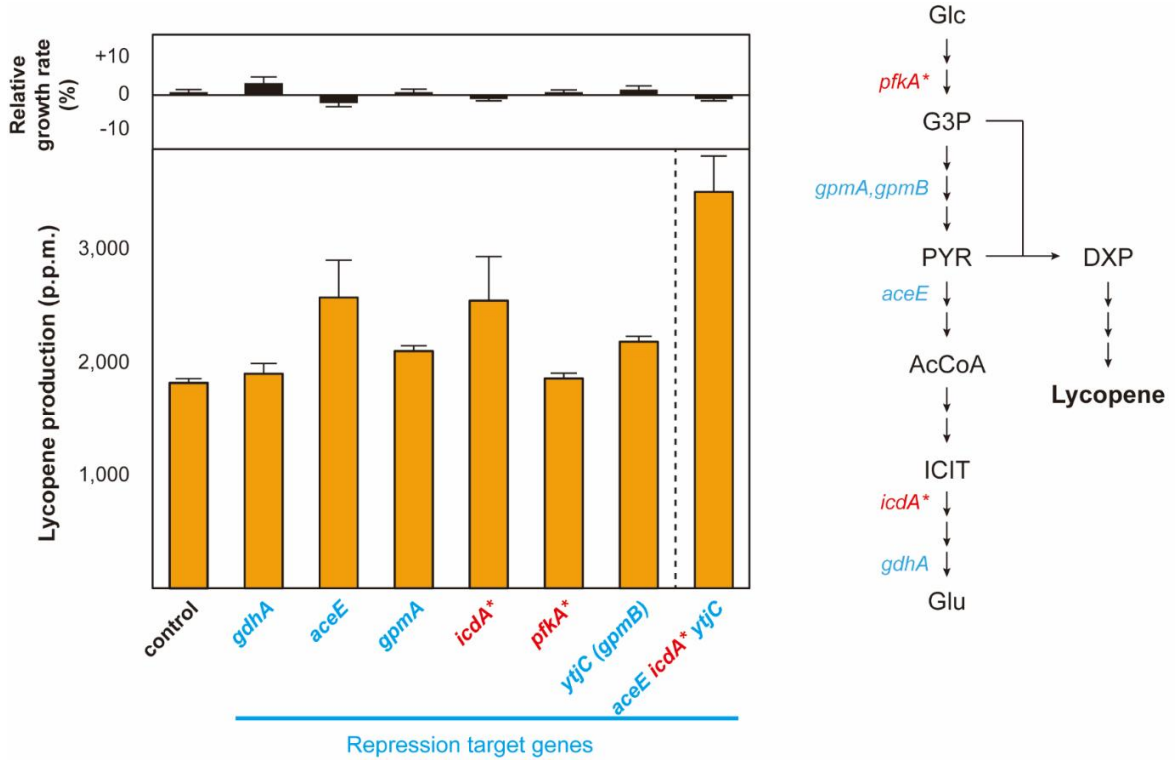


Figure 4.6 Lycopene biosynthesis pathway optimization by multiple repression using CRISPRz

4.7 Conclusion

In the CRISPR/Cas-mediated genome engineering, the sgRNA leads the dCas9 (or Cas9) protein to the specific sequence to regulate the expression of the target gene. Therefore, if multiple sgRNAs could be prepared, multiple genes can be regulated simultaneously. In this work, we developed a general strategy to produce multiple sgRNAs from a primary transcript using Rz in *E. coli*. Since Rz has sequence-specific cleavage activity, it is suitable for generating multiple functional sgRNAs. The precise cleavage activity of Rz in intracellular and the functionality of cleaved sgRNA were confirmed, consequently demonstrated the possibility of multiple repression along with dCas9 expression.

We employed CRISPRi system using dCas9 to verify the function of mature sgRNAs. When confirming the function of sgRNA by mediating Cas9 endonuclease, it is necessary to confirm the accurate recombination of editing template (ET) to the target region after double strand break by Cas9 protein guided by sgRNA. However, in this case, the function of sgRNA can be confirmed only when the ET is recombined with high efficiency by homologous recombination. The advantage of the CRISPRi system is that the function of sgRNA can be easily verified by measuring the degree of repression of target gene expression without the need for a homologous recombination system. Therefore, the function of sgRNA was confirmed with the CRISPRi system.

The strategy of generating multiple sgRNAs can be applied to modulate multiple genes simultaneously or to identify potential target genes to be modulated (Wu et al. 2015). As shown in this study, when the multiple sgRNA generating

strategy is applied to the CRISPRi system, the expression of several genes can be repressed simultaneously. This property is useful for fine tuning the metabolic pathway through endogenous gene regulation. In general, the gene-knockout strategy has been used to remove competition or unnecessary pathways for the efficient production of desired metabolites. However, the number of genes that can be modulated through the gene knockout strategy is limited, especially when target genes are essential for growth. The CRISPRi system can overcome the limitations of the knockout strategy because metabolic fluxes can be readjusted to produce the desired metabolite efficiently without defects in cell growth. The multiple sgRNA generation strategies can also be applied to identify potential target genes that need to be regulated to optimize metabolic pathways in combination with dCas9 mediated transcriptional regulation strategies. According to a previous report, activator fused dCas9 acts as a dual mode regulator (activator/repressor) and their mode could be determined depending on the sgRNA target location (Deaner et al. 2017). It is also reported that fusion of the omega subunit of the RNA polymerase to dCas9 can achieve programmable transcription activation in *E. coli* (Bikard et al. 2013). Thus, using multiple sgRNA generation strategies together with dual mode dCas9 activator systems allows for rapid testing of a variety of target gene combinations, making it easy to find the optimal combination. When this strategy is coupled with Cas9 endonuclease, multiplex genome editing is possible. Multiplex automated genome engineering (MAGE) is a representative method that can efficiently and rapidly create insert/deletion (indel) mutations for many genomic locations simultaneously (Wang et al. 2009). However, since the size of

the mutant library obtained through MAGE is too large, it is very difficult to discover a strain having the desired phenotype without high-throughput screening method. CRMAGE is an advanced technology that combines MAGE and CRISPR/Cas systems to enable negative selection against wild type (Ronda et al. 2016). This property has the advantage of reducing the library size for discovering the strain having the desired phenotype. Thus, applying the multiple sgRNA generation strategy to CRMAGE is essential for screening a strain that contain modifications in several genomic regions at the same time.

In summary, we developed the multiple sgRNA generation strategy using Rz and confirmed that the number of functional sgRNAs that can be generated from a primary transcript through RNA processing is at least five. It has also been demonstrated that this strategy can be employed to optimize the metabolic pathways. The capability of multiple editing based on the multiple sgRNA generation strategy is expected to facilitate more complicated CRISPR/Cas applications, such as evaluations of combinatorial effects of various genetic interventions, identification of potential genetic perturbation targets, transcriptional regulation of multiple genes and protein engineering.

Chapter 5.

Enhancement of the CRISPR/Cas9 editing by improving the homologous recombination efficiency

5.1 Attempts to enable multiple genome editing through improved HR efficiency have not been successful.

We have attempted to edit several genes simultaneously by combining multiple sgRNA production strategies with the pCas system (Jiang et al. 2015)), which has been reported to have multiple genome editing capability. Six genes (*dxr, ispA, idi, edd, zwf, and gnd*) that were predicted to have a positive effect on lycopene production when the expression level was increased while associated with the lycopene biosynthetic pathway were selected as target genes to be edited **(Table7)**. In order to perform multiple genome editing, the HR efficiency of each target gene must be verified. Therefore, the endonuclease activity of Cas9 and editing efficiency for six target genes were verified and three (*dxr, ispA*, and *edd*) out of six genes were selected to be edited by multiple genome editing **(Figure 5.1)**. At first, a plasmid which simultaneously expressed sgRNA targeting three genes was constructed by using gold gate assembly. Then, the prepared plasmid and three editing templates were transformed to BL21(DE3) strain harboring pCas. However, the multiple genomes editing experiment has failed. To be able to edit multiple genes at the same time, it is necessary to be able to repair the cleavages generated at different locations on the chromosome simultaneously. Therefore, we predicted that the failure of multiple genome editing experiments was due to low HR efficiency and we conducted experiments to improve the HR efficiency.

Table 7 List of genes that were predicted to have a positive effect on lycopene production

Gene	Enzyme	Function
dxr	1-deoxy-D-xylulose 5-phosphated reductoisomerase	2-C-methyl-D-erythritol 4-phosphate + $NADP^+$ ↔ 1-deoxy-D-xylulose 5-phosphate + NADPH + H^+
ispA	geranyl diphosphate/farnesyl diphosphate synthase	dimethylallyl diphosphate + isopentenyl diphosphate → geranyl diphosphate + diphosphate
idi	isopentenyl-diphosphate Δ-isomerase	isopentenyl diphosphate ↔ dimethylallyl diphosphate
edd	phosphogluconate dehydratase	D-gluconate 6-phosphate → 2-dehydro-3-deoxy-D-gluconate 6-phosphate + H_2O
zwf	$NADP^+$-dependent glucose-6-phosphate dehydrogenase	D-glucopyranose 6-phosphate + $NADP^+$ → 6-phospho D-glucono-1,5-lactone + NADPH + H^+
gnd	6-phosphogluconate dehydrogenase	D-gluconate 6-phosphate + $NADP^+$ → D-ribulose 5-phosphate + CO_2 + NADPH

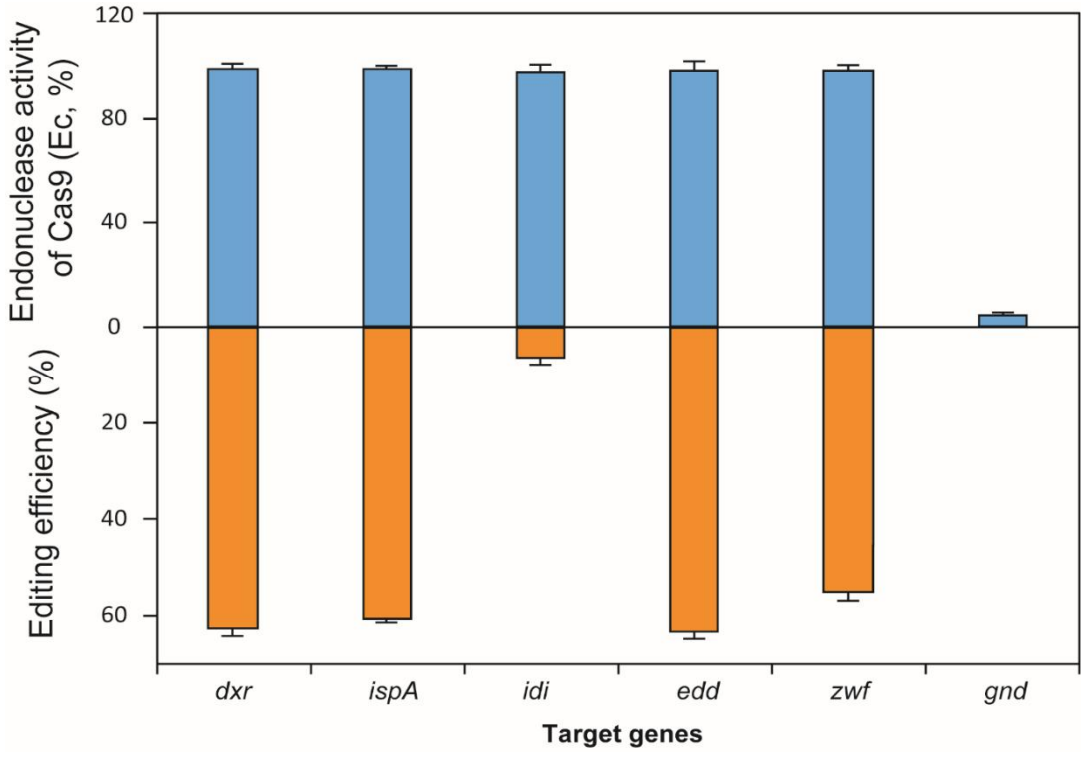


Figure 5.1 Measurement of the endonuclease activity of Cas9 and editing efficiency for six target genes using CRISPR/Cas system.

The homologous recombination system of pCas is using the lambda-red recombinase. The phage lambda-derived red recombination system has been used as a powerful technique for generation of insertion/deletion (indel) mutations in *Escherichia coli*, using either a linear double-stranded DNA (dsDNA) or a synthetic single-stranded oligonucleotide (ssDNA). According to previous reports, it is more efficient to use single-stranded oligonucleotide (ssDNA) than double-stranded DNA (dsDNA) for genetic modification through homologous recombination (Li et al. 2015). Therefore, dsDNA and ssDNA targeting the same region of *lacZ* were prepared as editing template, respectively, and then compared their HR efficiencies **(Figure 5.2A)**. The HR efficiency was confirmed through CFU and editing efficiency. Each editing template (dsDNA and ssDNA), along with pgRNA_*lacZ*, was transformed into BL21 (DE3) strain containing pCas. As a result, in the experiment using dsDNA, the CFU was around ten and the editing efficiency was more than 95%, whereas in the experiment using ssDNA, the CFU was around 70 and the editing efficiency was more than 99% **(Figure 5.2B)**. Therefore, all subsequent experiments have decided to use ssDNA for high HR efficiency.

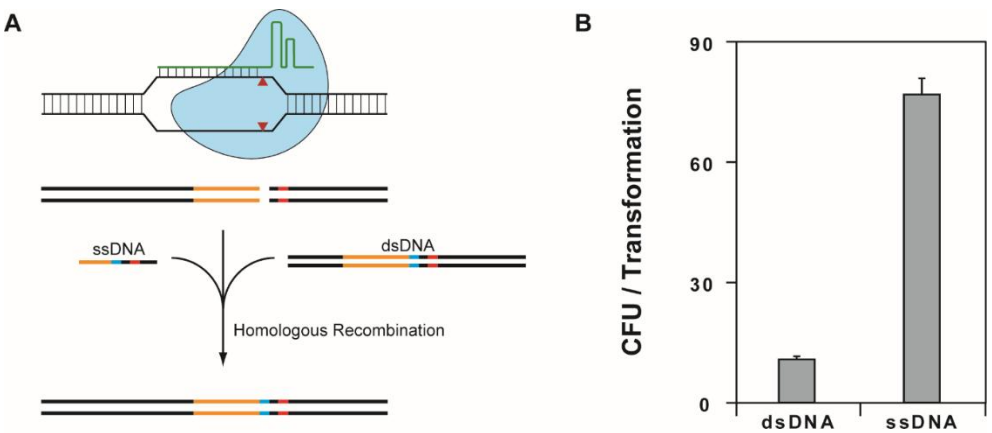


Figure 5.2 Comparison of the HR efficiency using dsDNA or ssDNA.

There are three phage-derived lambda red proteins for carrying out homologous recombination: Gam, Exo, and Beta. However, Beta is the only one that is required for ssDNA-mediated recombination (Mosberg et al. 2010). Therefore, we constructed pCas_v2.0 which can express only Beta protein among three proteins mentioned above, and compare the HR efficiency of pCas_2.0 with that of pCas **(Figure 5.3A)**. As a result, the editing efficiency was almost equal to 99%, but the CFU of pCas_v2.0 was about 15 times higher than that of pCas **(Figure 5.3B)**. From these results, we confirmed that it is more efficient to express beta protein only when homologous recombination using ssDNA is performed.

After confirming that HR efficiency can be improved by using pCas_v2.0 and ssDNA, multiple genome editing was performed again on the three genes that were previously targeted. However, it was confirmed that the multiple genome editing experiment failed by confirming that all the colony formed on the plate was a false positive colony. Based on these results, we conclude that multiple genome engineering through cleavage and repair at the same time in multiple locations is not an efficient way. To overcome these shortcomings and enable multiple genome editing, we adopted the CRMAGE method of introducing mutations at multiple chromosomal locations through MAGE, and then performing negative selection through the CRISPR/Cas system.

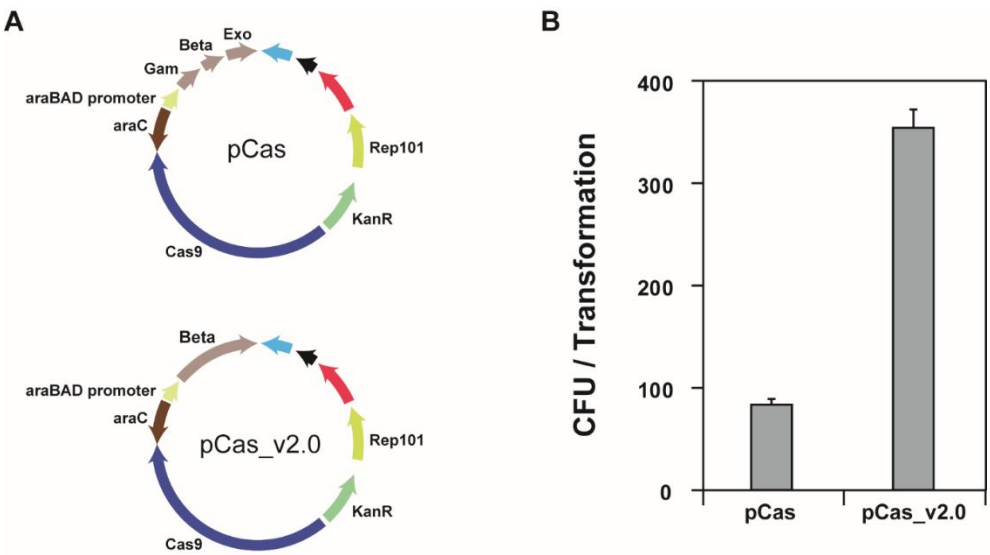


Figure 5.3 Comparison of the HR efficiency of pCas and pCas_v2.0.

In order to simultaneously edit several parts of the chromosome with the CRMAGE system, both of the HR efficiency and the endonuclease activity of the Cas9 protein should be high. In addition, the plasmids used to edit the strain must be able to be removed to perform additional experiments using the resulting strain. However, among the existing plasmids used for activating the CRISPR/Cas system in *Escherichia coli*, there are no plasmids having all of the above functions. Therefore, we constructed plasmid pCASRec having all three functions mentioned above.

5.2 Verification of endonuclease activity of Cas9 of pCASRec

After constructing pCASRec, the endonuclease activity of Cas9 expressed in pCASRec was verified. The efficiency of the CRISPR/Cas system is determined by the sequence-specific cleavage activity of the Cas9/sgRNA complex and the homologous recombination activity using editing template. Therefore, if an editing template is not provided after DNA cleavage by Cas9, the cell cannot grow because the cleaved chromosome cannot be repaired, and the potential for cell growth will decrease as Cas9 activity increases. Using this property, the endonuclease activity of Cas9 (Ec) was calculated as mentioned in the method section.

A gene to be targeted by Cas9 was determined as *lacZ*, and pgRNA_*lacZ* expressing sgRNA targeting *lacZ* was constructed. In addition, pgRNA_dummy expressing sgRNA that does not target any position in the chromosome was constructed as the control of pgRNA_*lacZ*. Furthermore, in order to confirm the effect of Cas9 concentration on endonuclease activity, pCas, which constitutively

expresses Cas9, was used as a control for pCASRec. In order to confirm the function and activity of Cas9 expressed in pCASRec, the following four strains were prepared; BLcl (BL21 harboring pCas and pgRNA_*lacZ*), BLcd (BL21 harboring pCas and pgRNA_dummy), BLil (BL21 harboring pCASRec and pgRNA_*lacZ*), and BLid (BL21 harboring pCASRec and pgRNA_dummy), and compared the number of colony forming units (CFU). As a result, the CFU of the strain harboring pgRNA_dummy (BLcd and BLid) were more than a thousand, but the CFU of the strain harboring pgRNA_*lacZ* (BLcl and BLil) were less than 50 **(Figure 5.4)**. These results indicate that the Cas9 protein expressed from pCASRec has endonuclease activity. Furthermore, it was confirmed that the endonuclease activity of Cas9 of pCASRec was higher than that of Cas9 of pCas by comparing Ec of a strain harboring pCASRec and Ec of a strain harboring pCas.

In order to simultaneously edit several parts of the chromosome with the CRMAGE system, both of the HR efficiency and the endonuclease activity of the Cas9 protein should be high. In addition, the plasmids used to edit the strain must be able to be removed to perform additional experiments using the resulting strain. However, among the existing plasmids used for activating the CRISPR/Cas system in *Escherichia coli*, there are no plasmids having all of the above functions. Therefore, we constructed plasmid pCASRec having all three functions mentioned above.

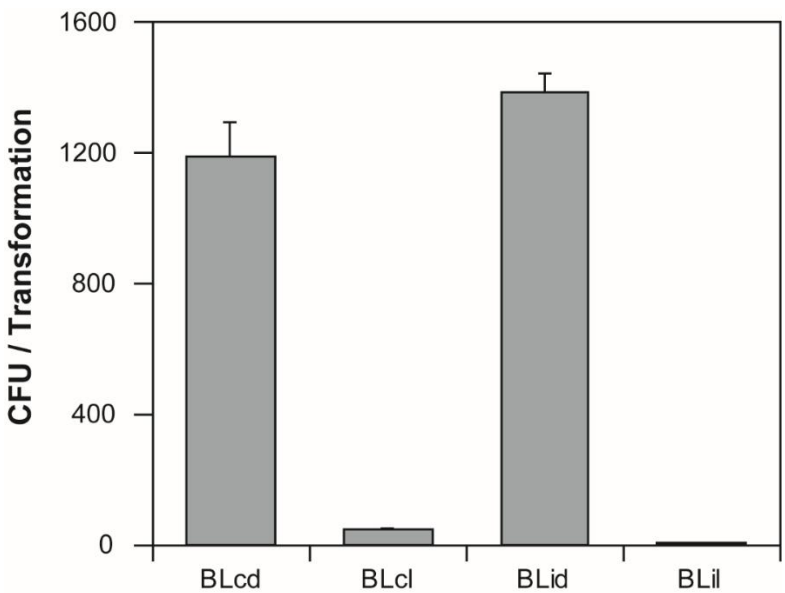


Figure 5.4 Examination of endonuclease activity of Cas9 expressed from pCASRec.

BLcd (BL21 strain harboring pCas and pgRNA_dummy); BLcl (BL21 strain harboring pCas and pgRNA_*lacZ*); BLid (BL21 strain harboring pCASRec and pgRNA_dummy); BLil (BL21 strain harboring pCASRec and pgRNA_*lacZ*)

5.3 Verification of HR efficiency of pCASRec

After confirming the endonuclease activity of Cas9 of pCASRec, the editing efficiency of pCASRec was verified. The editing efficiency of pCASRec was compared with the editing efficiency of the MAGE experiment using EcNR2 strain (Wang et al. 2009). The *lacZ* gene was selected as the target gene, and the editing efficiency was calculated through the blue/white colony assay. The MAGE was carried out for 3, 5, and 10 cycles using the ssDNA targeting *lacZ* in the EcNR2 strain, and the same experiment was carried out in the MG1655 strain containing pCASRec. As a result, it was confirmed that the editing efficiency in MG1655 containing pCASRec was higher than the EcNR2 strain **(Figure 5.5)**. These results indicate that using pCASRec can quickly and easily edit multiple genes at the same time.

After confirming the functions of pCASRec (the endonuclease activity of Cas9 and the recombination efficiency of recombinase), multiple genome editing experiment was performed with MAGE using pCASRec in BL21 (DE3) strain. The target genes were identical to that of the multiple genome editing experiment using pCas in the BL21 (DE3) strain. Therefore, the MAGE cycle was performed by repeatedly introducing three editing templates into the BL21 (DE3) strain containing pCASRec. The MAGE cycle was terminated by transforming pgRNA_*lacZ* to eliminate strains in which no mutation was introduced. However, by confirming that all colonies were false positive, it has been confirmed that multiple genome editing experiments have failed again.

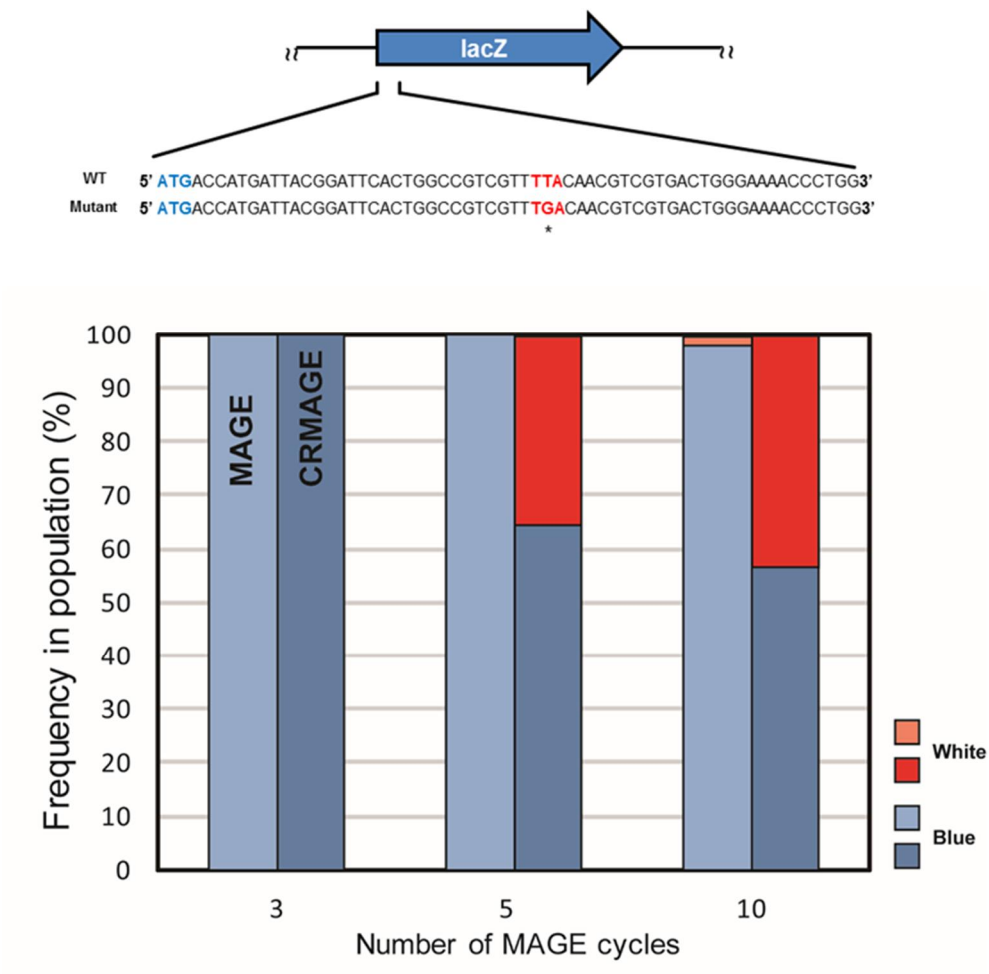


Figure 5.5 Examination of HR efficiency of pCASRec

5.4 Optimization of experimental conditions using pCASRec: Temperature

After completing the construction of pCASRec, optimizations of the experimental condition were performed. The pCASRec is replicated with temperature-sensitive replication origin. Therefore, we assumed that the function of pCASRec will be affected by temperature. To confirm our hypothesis, Cas9 activity and recombination efficiency of pCASRec were tested at various temperatures (30, 33, 35, 37, and 42°C). The BL21(DE3) strain was used as an experimental strain and *dxr* was used as a target gene to be edited. As a result, the endonuclease activity of Cas 9 (Ec) was found to be 98% or more at 30°C ,33°C and 35°C **(Figure 5.6)**. In case of 37°C and 42°C, the Ec could not be measured because no colonies appeared. These results indicate that the optimal temperature range for Cas9 is 30°C to 35°C and the pCASRec cannot be replicated above 37°C. The editing efficiency according to the temperature change was confirmed by colony PCR. The editing efficiency was more than 95% at 33°C, 35°C and 63% at 30°C **(Figure 5.6)**. In particular, in the case of CFU, it was about 10 at 30 °C but about 1,000 at 33°C and 35 °C. These results indicate that the HR efficiency is significantly high at 33°C to 35°C rather than 30°C. Therefore, based on the results for Cas9 activity and the recombinant activity of the recombinase, it was confirmed that the temperature suitable for using pCASRec was 33°C and 35°C instead of 30°C.

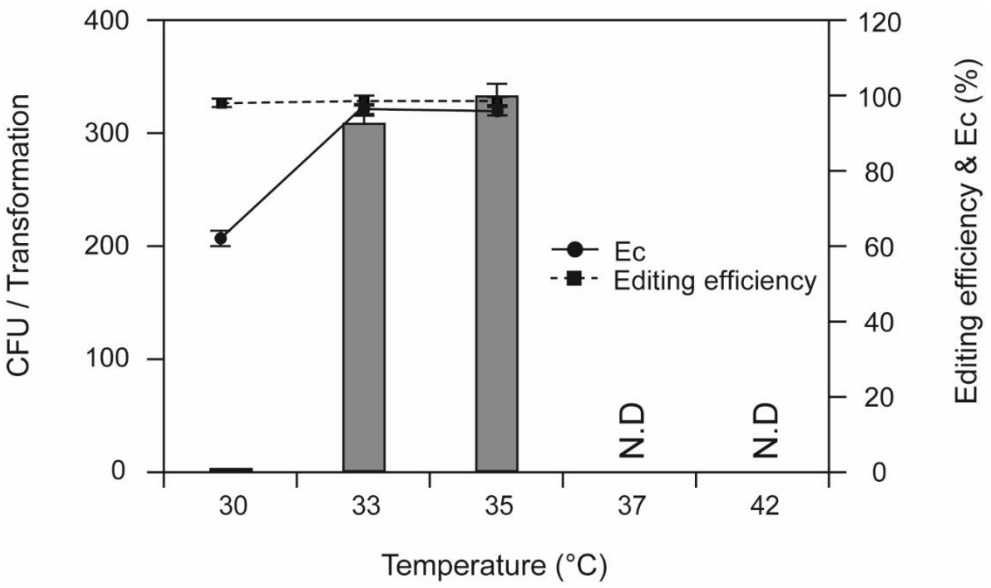


Figure 5.6 Determination of pCASRec activity at various temperature.

N.D: Not Determined.

Since pCASRec is sensitive to temperature variations, the temperature stability of pCASRec should be considered to determine the optimal temperature for its use. Here, we attempted to determine the stability of pCASRec by measuring the curing efficiency of the pgRNA after several generations of continuous cultivation of the BL21 strain harboring pCASRec and pgRNA at 30°C, 33°C and 35 °C. As a result, the curing efficiency was not reduced at 30°C and 33°C, but the curing efficiency was significantly reduced at 35°C. Therefore, all subsequent experiments were performed at 33 °C.

5.5 Optimization of experimental conditions using pCASRec: Arabinose induction

After the experimental temperature using pCASRec was determined at 33°C, the optimization of the arabinose induction condition was performed to maximize the HR efficiency. To do this, the HR efficiency of BL21 strain harboring pCASRec provided with sgRNA and ET were compared at various concentrations of arabinose treatment (0, 10, 20, 30, 40, and 50mM of final concentration). The *dxr* was used as a target gene to be modified. In the strains treated with arabinose, the CFU was similar regardless of the increase of arabinose concentration and the ER was around 96% in all strains. However, the colonies were not appeared in the strain not treated with arabinose **(Figure 5.7A)**. These results indicate that the homologous recombination activity is dependent on the lambda red recombination system and 10 mM of arabinose is sufficient for ssDNA mediated recombination.

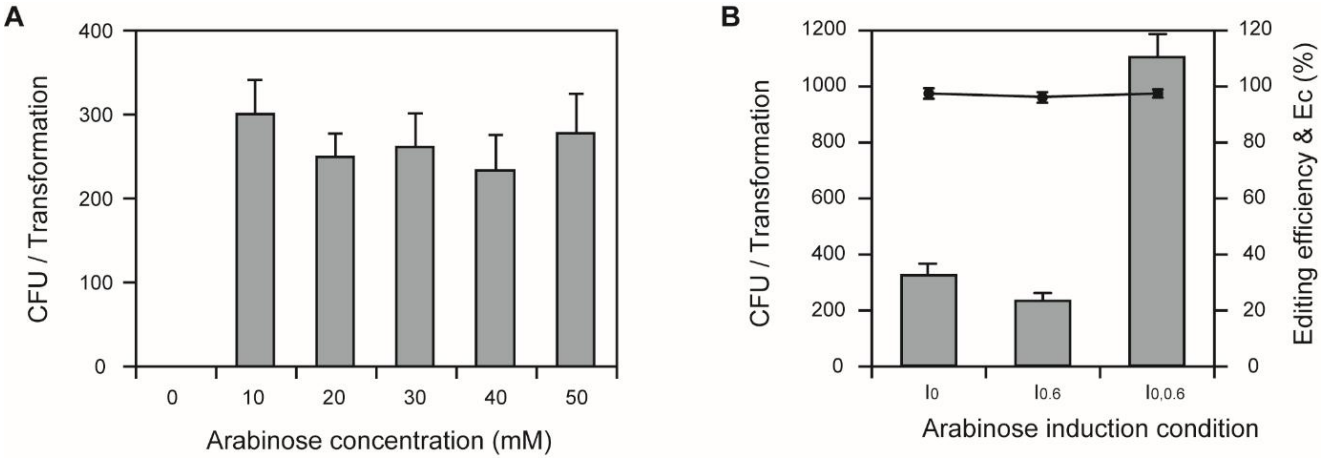


Figure 5.7 Determination of pCASRec activity at various arabinose induction concentration.

In addition, experiments were conducted to optimize the arabinose induction conditions for high HR efficiency. The arabinose was treated in three different conditions (treatment at the point of cell inoculation (I_0), treatment when the OD_{600} reached to ca. 0.6 ($I_{0.6}$), and treatment at both time points ($I_{0,0.6}$)). As a result, under the conditions of I_0 and $I_{0.6}$, the CFU of each strain was similar, but three times higher on $I_{0,0.6}$. However, in the case of editing efficiency, the strain under the condition of $I_{0.6}$ and $I_{0,0.6}$ were twice that of $I_{0.6}$ **(Figure 5.7B)**. From these results, we concluded that arabinose should be applied under conditions of $I_{0,0.6}$ to obtain high HR efficiency.

In particular, colonies formed on plates had two types of colonies that showing different growth. The number of fast-growing colonies was less than 1% of all colonies. As a result of confirming the replacement efficiency, it was 0% for the fast-growing colonies and 100% for slow-growing colonies. Thus, it was possible to select the recombinants by only confirming the size of the colonies.

5.6 Sequential mutagenesis using pCASRec allows multiple editing

Multiple editing was attempted on the three target genes (*dxr, ispA*, and *edd*) selected above using pCASRec. The editing efficiency of pCASRec was evaluated while increasing the number of target genes. As a result, HR efficiency is greatly reduced when two or more simultaneous edits are performed, even though HR efficiency for each target is very high. Therefore, we concluded that sequential mutagenesis is necessary for multiple editing.

Sequential mutagenesis is a method of sequential editing of several genes by repeating the process of editing the gene. To this end, the plasmid producing the target specific sgRNA must be removed every cycle. We compared the curing capabilities of pCas_v2.0 and pCASRec. As a result, in the case of pCas_v2.0, the reduced activity of Cas9 was observed from the second cycle, while in the case of pCASRec, the activity of Cas9 remained high in the second cycle. **(Figure 5.8A,B)** These results confirmed that pCASRec is a suitable plasmid for sequential mutagenesis. Therefore, sequential mutagenesis was performed on three target genes using pCASRec. As a result, it was confirmed that all the genes were edited with 100% efficiency.

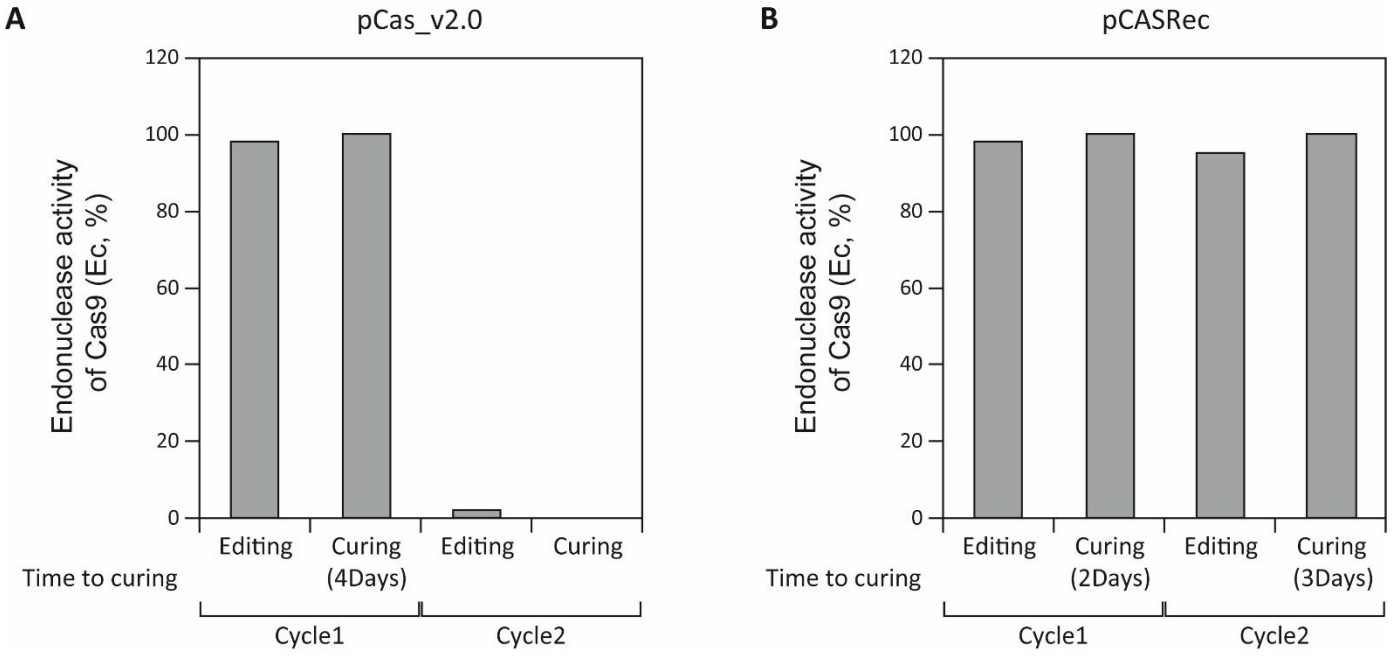


Figure 5.8 Comparison of the curing capabilities of pCas_v2.0 and pCASRec

5.7 Conclusion

In order to rebalance the metabolic flux, the expression level of target genes should be optimized. We decided to use a combinatorial approach to fine tune the metabolic pathway. A combinatorial approach is a method to screen strain showing the best phenotype among all possible mutation libraries. Thus, the ability to generate sufficient size of library is required. When the CRISPR/Cas system was performed using the previously reported plasmid pCas, the ability to produce recombinant was very low. This the reason of the low efficiency to generate the recombinants was confirmed to be due to the low HR efficiency resulting from the use of dsDNA as the donor DNA. Therefore, in order to produce many recombinants, plasmids with high HR efficiency should be used. In the pMA7CR_2.0 used in the previously reported CRMAGE system, HR efficiency is high because ssDNA can be used as donor DNA. In this case, however, an additional plasmid (pZS4Int-tetR) is required for the expression of the TetR protein, an essential component of Cas9 expression through tetracycline induction. Further, since there is no plasmid curing system, the strain constructed by this method has a limitation that it cannot be used as a strain producing a desired product. Thus, pCASRec, a plasmid with high HR efficiency and high Cas9 activity and easily removable, has been constructed. As a result of the optimization of the experimental conditions, it was confirmed that the HR efficiency in the experiment using pCASRec was 100 times higher than that in the experiment using pCas.

However, although the high HR efficiency of pCASRec was achieved, it was not possible to edit multiple targets simultaneously. Therefore, we performed

multiple editing through sequential mutagenesis using pCASRec and confirmed that multiple editing is possible with 100% efficiency.

Chapter 6.

Overall Conclusion and Further Suggestions

6 Overall conclusion and further suggestions

In this thesis, I suggest a strategy for metabolic engineering through a combinatorial approach using the CRISPR/Cas system. This requires a multiple sgRNA generation strategy and a CRISPR/Cas system with high HR efficiency and Cas9 activity. Since Rz has self-cleavage activity, it has been applied to the strategy of generating multiple sgRNAs. In order to use the Rz as a RNA scissor, the intracellular cleavage activity of Rz should be high. Therefore, Q-Q assay method was developed to measure the intracellular activity of Rz. As a result, it was confirmed that Rz can cleave RNA with an efficiency of 90% or more.

Then, the function of sgRNA generated by cleavage of the primary transcript by Rz was confirmed by observing the reduced expression of target gene using dCas9. It was confirmed that the number of functional sgRNAs that can be generated from a primary transcript through RNA processing is at least five. It was also confirmed that several genes related to the metabolic pathway can be regulated simultaneously. This suggests that metabolic engineering through a combination approach can be performed using the dCas9-ω system and multiple sgRNA generation strategy.

In order to modulate the expression of the target genes discovered through the combinatorial approach described above, a technique for multiple genome editing was required. Debottlenecking one rate limiting step often introduce another constraint in pathway engineering. Therefore, we sought to simultaneously optimize the expression levels of all target genes. To do this, libraries with

different expression levels for all target genes should be producible. Since the high HR efficiency is required to generate a sufficient number of libraries, pCASRec was designed and constructed. As a result of confirming the HR efficiency, it was confirmed that the HR efficiency in the experiment using pCASRec was 100 times higher than the experiment using pCas. Multiple genome editing was attempted using pCASRec with high HR efficiency, but it was not possible to introduce more than three mutations at the same time. Thus, we conducted experiments that sequentially introduce three mutations. As a result, it was confirmed that mutations could be introduced with high efficiency for three target sites. In particular, since correct mutant can be distinguished by the size of colonies in sequential mutagenesis, libraries without false positive colonies can be generated.

Next, multiple sgRNA assembly and production strategies will be used coupled with dCas9-ω to identify the optimal combinations of the multiple genes for improving the production of the target molecule. Creating all combinations of multiple genes is not efficient in screening strains with the best phenotype because the library is too large. Therefore, efficient design of experiments is required to implement desired phenotypes in the minimal size of library. We have designed an experiment called “modular approach” that divides the metabolic pathway into several modules (upstream module, downstream module, and competition modules) based on major precursor, and regulates the expression of one gene per module. Therefore, the number of genes to be combined is identical to the number of modules. This method is based on the assumption that one or more of the genes that make up each module need to be modulated to increase the desired product. It

is more efficient than the MAGE method, which produces all genetic variants because the number of libraries to implement the desired genotype is very small. Furthermore, by controlling the number of modules to be separated and the number of genes constituting the module, optimization of metabolic pathway can be performed on target molecules for which the high-throughput screening method has not been developed. The disadvantage of "modular approach" is that it is based on the assumption that all modules have the same effect on producing the desired product because they control one gene for each module. However, it is impossible to produce a strain in which all the metabolic fluxes are optimized in a single experiment. Therefore, it is expected to be able to overcome the deficiencies of the modular approach by repeating the pathway optimization process. Metabolic pathways of lycopene and GDP-fucose will be optimized using the proposed metabolic engineering method.

References

Acharya S, Foster PL, Brooks P, Fishel R (2003) The Coordinated Functions of the E. coli MutS and MutL Proteins in Mismatch Repair. Molecular Cell 12(1):233-246 doi:https://doi.org/10.1016/S1097-2765(03)00219-3

Alper H, Jin YS, Moxley JF, Stephanopoulos G (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab Eng 7(3):155-64 doi:10.1016/j.ymben.2004.12.003

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315 doi:10.1126/science.1138140

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Research 41(15):7429-7437 doi:10.1093/nar/gkt520

Bortesi L, Fischer R (2015) The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnol Adv 33(1):41-52 doi:10.1016/j.biotechadv.2014.12.006

Deaner M, Mejia J, Alper HS (2017) Enabling Graded and Large-Scale Multiplex of Desired Genes Using a Dual-Mode dCas9 Activator in *Saccharomyces cerevisiae*. ACS Synthetic Biology 6(10):1931-1943 doi:10.1021/acssynbio.7b00163

Denesyuk NA, Thirumalai D (2015) How do metal ions direct ribozyme folding? Nature Chemistry 7:793 doi:10.1038/nchem.2330

https://www.nature.com/articles/nchem.2330#supplementary-information

Deshler JO, Li H, Rossi JJ, Castanotto D (1995) Ribozymes expressed within the loop of a natural antisense RNA form functional transcription terminators. Gene 155(1):35-43 doi:https://doi.org/10.1016/0378-1119(94)00889-Z

Donahue CP, Fedor MJ (1997) Kinetics of hairpin ribozyme cleavage in *yeast*. Rna 3(9):961-73

Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One 3 doi:10.1371/journal.pone.0003647

Gao Y, Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. J Integr Plant Biol 56(4):343-9 doi:10.1111/jipb.12152

Gibson GJ, Yang M (2017) What rheumatologists need to know about CRISPR/Cas9. Nature Reviews Rheumatology 13:205 doi:10.1038/nrrheum.2017.6

Gunnelius L, Hakkila K, Kurkela J, Wada H, Tyystjärvi E, Tyystjärvi T (2014) The omega subunit of the RNA polymerase core directs transcription efficiency in *cyanobacteria*. Nucleic Acids Research 42(7):4606-4614 doi:10.1093/nar/gku084

Janion C (2008) Inducible SOS Response System of DNA Repair and Mutagenesis in *Escherichia coli*. International Journal of Biological Sciences 4(6):338-344

Jiang Y, Chen B, Duan C, Sun B, Yang J, Yang S (2015) Multigene editing in the

Escherichia coli genome via the CRISPR-Cas9 system. Appl Environ Microbiol 81(7):2506-14 doi:10.1128/aem.04023-14

Kabadi AM, Ousterout DG, Hilton IB, Gersbach CA (2014) Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. Nucleic Acids Res 42(19):e147 doi:10.1093/nar/gku749

Khvorova A, Lescoute A, Westhof E, Jayasena SD (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Mol Biol 10(9):708-712

Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nature Protocols 8:2180 doi:10.1038/nprot.2013.132

Li Y, Lin Z, Huang C, Zhang Y, Wang Z, Tang YJ, Chen T, Zhao X (2015) Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing. Metab Eng 31:13-21 doi:10.1016/j.ymben.2015.06.006

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV (2011) Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9(6):467-77 doi:10.1038/nrmicro2577

Mosberg JA, Lajoie MJ, Church GM (2010) Lambda Red Recombineering in *Escherichia coli* Occurs Through a Fully Single-Stranded Intermediate. Genetics 186(3):791-799 doi:10.1534/genetics.110.120782

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-

specific control of gene expression. Cell 152(5):1173-83 doi:10.1016/j.cell.2013.02.022

Quan J, Tian J (2011) Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 6(2):242-51 doi:10.1038/nprot.2010.181

Ren X, Yang Z, Xu J, Sun J, Mao D, Hu Y, Yang S-J, Qiao H-H, Wang X, Hu Q, Deng P, Liu L-P, Ji J-Y, Li Jin B, Ni J-Q (2014) Enhanced Specificity and Efficiency of the CRISPR/Cas9 System with Optimized sgRNA Parameters in Drosophila. Cell Reports 9(3):1151-1162 doi:https://doi.org/10.1016/j.celrep.2014.09.044

Ronda C, Pedersen LE, Sommer MOA, Nielsen AT (2016) CRMAGE: CRISPR Optimized MAGE Recombineering. Scientific Reports 6:19452 doi:10.1038/srep19452

https://www.nature.com/articles/srep19452#supplementary-information

Serganov A, Patel DJ (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nat Rev Genet 8(10):776-790

Slomovic S, Schuster G (2013) Circularized RT-PCR (cRT-PCR): analysis of the 5' ends, 3' ends, and poly(A) tails of RNA. Methods Enzymol 530:227-51 doi:10.1016/b978-0-12-420037-1.00013-0

Stead MB, Agrawal A, Bowden KE, Nasir R, Mohanty BK, Meagher RB, Kushner SR (2012) RNAsnap: a rapid, quantitative and inexpensive, method for isolating total RNA from *bacteria*. Nucleic Acids Res 40(20):e156

doi:10.1093/nar/gks680

Suzuki S, Ono N, Furusawa C, Kashiwagi A, Yomo T (2007) Experimental optimization of probe length to increase the sequence specificity of high-density oligonucleotide microarrays. BMC Genomics 8(1):373 doi:10.1186/1471-2164-8-373

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894 doi:10.1038/nature08187

https://www.nature.com/articles/nature08187#supplementary-information

Wu J, Du G, Chen J, Zhou J (2015) Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in *Escherichia coli*. Sci Rep 5:13477 doi:10.1038/srep13477

Xie K, Minkenberg B, Yang Y (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proceedings of the National Academy of Sciences 112(11):3570-3575 doi:10.1073/pnas.1420294112

Zhou K, Zhou L, Lim QE, Zou R, Stephanopoulos G, Too H-P (2011) Novel reference genes for quantifying transcriptional responses of *Escherichia coli* to protein overexpression by quantitative PCR. BMC Molecular Biology 12:18-18 doi:10.1186/1471-2199-12-18

국문 초록

미생물의 대사 흐름을 조절하여 목적 물질의 생산성을 증가시키기 위해서는 미생물의 게놈에 대한 광범위한 조절이 필요하다. 대사 흐름의 재조정을 위한 유전자 변형을 도입하기 위해서는 조절해야 하는 표적 유전자를 결정하고, 표적 유전자의 발현을 최적화 해야 한다. 그러나, 대사 경로에 관여하는 유전자들의 발현 변화가 목적 물질의 생산에 미치는 영향을 모두 확인하는 것은 매우 노동 집약적이기 때문에, 표적 유전자를 정확히 선정하는 것은 어렵다. 또한 몇 개의 유전자가 표적 유전자로 선택되었다 하더라도 조절의 조합이 예측할 수 없는 상호 작용으로 인해 균주에 해로운 영향을 줄 수 있기 때문에 대사 경로를 최적화 하는 것은 어렵다. 따라서, 전술한 결점을 극복하기 위해서, 조합 접근법을 사용하여 대사 공학을 수행하는 것이 합리적이다. 조합 접근법은 원하는 표현형을 향상시킬 것으로 기대되는 다양한 돌연변이 라이브러리 중에서 가장 좋은 표현형을 가진 균주를 선별하는 것이다. 따라서 돌연변이 라이브러리를 생성하기 위해서는 게놈의 여러 위치에 돌연변이를 일으키는 도구가 필요하다.

이 논문에서는 CRISPR/Cas 시스템을 사용한 대사 경로 최적화를 가능하게 하기 위해서 다중 sgRNA 생성 전략을 구축했다. 불활성화된 cas9 (dCas9)에 전사활성인자도메인을 융합시킨 dCas9-ω

는 sgRNA 의 디자인에 따라 표적 유전자 발현을 활성화 시키거나 억제할 수 있다. 따라서 여러 개의 sgRNA 가 생성 될 수 있다면 유전자 변형없이 여러 유전자의 발현을 동시에 조절할 수 있다. 이러한 이유로 여러 개의 sgRNA 를 생성하는 전략이 필요했고, 자체 절단 특성이 있다고 보고된 라이보자임을 사용하여 하나의 1 차 전사체로부터 여러 개의 sgRNA 를 생산하는 전략을 세웠다. 여러 sgRNA 생산 전략을 dCas9-ω 시스템과 결합함으로써 표현형 증대에 강한 영향을 미치는 표적 유전자를 효율적으로 동정 할 수 있다.

조절하고자 하는 표적 유전자를 확인한 후 유전자 발현을 바꾸기 위해 유전자 편집을 수행해야한다. 대사 흐름을 미세하게 조정하기 위해 단일 유전자의 발현 수준을 조절하면 전체 플럭스가 변경된다. 따라서 원하는 표현형을 얻기 위해 표적 유전자의 발현 수준을 최적화하는 대사 공학을 진행할 때, 조합 접근법이 순차적 접근법보다 합리적이다. 이를 위해서는 게놈 내의 여러 위치에 다양한 조합을 유전자 변형이 도입된 충분한 크기의 라이브러리를 구축 할 수 있어야 한다. 그러나, 문헌에 보고된 다중 편집 방법들은 편집 효율이 매우 낮고, 위양성 균주 제거가 쉽지 않기 때문에 조합 접근법을 적용하기 위한 돌연변이 라이브러리를 생산하는데 한계가 있다. 따라서 우리는 CRISPR/Cas 시스템을 사용하여 한 주기에 한 위치에 대한 돌연변이 라이브러리를 도입하고, 이 주기를 반복함으로써 여러 위치에

대한 돌연변이 라이브러리를 축적함으로써 조합 접근법을 허용하는 시스템을 설계했다. 모든 돌연변이 변이체를 포함하는 라이브러리를 생성하기 위해서는 CRISPR/Cas 시스템에서 얻은 클론의 수가 충분해야 하며 편집 효율이 높아야 한다. 따라서 우리는 CRISPR/Cas 시스템을 통해 높은 편집 효율로 많은 수의 콜로니를 얻을 수 있는 플라스미드를 구축 하였고, 3 개의 상이한 유전자에 돌연변이가 도입된 균주가 고효율로 생성 될 수 있다는 것을 확인했다.

주요어: CRISPR/Cas system, 라이보자임, MAGE, 다중 편집 기술, 대사 공학, 조합 접근법

학번: 2010-23350