

Doctoral Thesis

Genetic engineering of *Escherichia coli* for protein production
for functional and NMR structural study

Yojiro Ishida

March 2017

Table of contents	
List of Figures	3
List of Tables	5
Abbreviations.....	6
Abstract	7
Acknowledgements	12
General introduction	13
I. Toxin-Antitoxin (TA) system	13
II. mRNA interferase	18
III. The Single-Protein Production (SPP) system	24
IV. Dual inducible cSPP systems	27
V. T7 expression system	30
VI. Protein S tag	33
Rationale.....	36
Section I Antimicrobial peptide Bac7(1-35) production using the PST-SPP system in <i>E. coli</i>	42
Introduction	43
Materials and Methods	46
Results	49
Discussion	60

Section II. Replacement of Arg residues in MazFbs with canavanine alters its specificity.....	64
Introduction	65
Materials and Methods	70
Results	75
Discussion	90
Section III Construction of a residue- and stereo-specific methyl labeling method by engineering <i>E. coli</i>	94
Introduction	95
Section III-I Stereo- and residue-specific methyl labeling using amino acid precursors.....	99
Materials and Methods	100
Results	107
Section III-II Stereo-and residue-specific labeling using SAIL-Val and –Leu and auxotrophic strains.....	122
Materials and Methods	123
Results	128
Discussion	143
Summary and conclusions.....	147

List of Figures

Figure 1	The Toxin-Antitoxin (TA) systems	16
Figure 2	The Toxin-Antitoxin system, MazF-MazE system	20
Figure 3	Sequence alignments of MazFbs and its homologues	23
Figure 4	The Single-Protein Production (SPP) system	26
Figure 5	A comparison of the SPP system and the His/Trp inducible SPP System	29
Figure 6	The T7 expression system and pET vectors	32
Figure 7	The NMR structure of Protein S	35
Figure 8	A schematic description of the PST-SPP system	50
Figure 9	The production of SUMO-Bac7(1-35) and PrS ₂ -Bac7(1-35) in <i>E. coli</i> BL21(DE3)	51
Figure 10	Purification and identification of Bac7(1-35)	54
Figure 11	Functional assays of Bac7(1-35).....	57
Figure 12	The structure of Canavanine, Arginine, and MazFbs with RNA..	68-9
Figure 13	Schematic procedures for the production of MazFbs(<i>arg</i>) and MazFbs(<i>can</i>)	77
Figure 14	Total mass measurement by MALDI-TO of MazFbs(<i>arg</i>) and MazFbs(<i>can</i>)	80
Figure 15	Alteration of cleavage specificity from 5-base to 6-base recognition in MazFbs(<i>can</i>)	83
Figure 16	Identification of a change of RNA cleavage specificity in MazFbs(<i>can</i>)	85
Figure 17	The structural analysis of MazFbs(<i>arg</i>) and MazFbs(<i>can</i>)	89
Figure 18	The biosynthetic pathway of Ile, Leu and Val	102

Figure 19	Optimization for residue specific methyl labeling by precursors using strain A	111
Figure 20	The ^{13}C -HMQC spectra of residue- and stereo specifically methyl labeled PPD samples.....	113
Figure 21	The ^{13}C -HMQC spectra of stereo- and residue specific methyl labeled MSG	116
Figure 22	^{13}C -HMQC spectra recorded on Val- $^{13}\text{CH}_3$ - <i>proS</i> MSG samples	117
Figure 23	The application of residue- and stereo- specific labeling system for CAP	118
Figure 24	Isotope incorporation efficiency using ^{13}C -aceto-hydroxybutanoate and ^{13}C - <i>proS</i> -acetolactate	121
Figure 25	Comparison of expression levels using BL21(DE3) $\Delta ilvD\Delta leuB$ strain and either standard strain BL21(DE3) pLys S or BL21(DE3) ..	131
Figure 26	Comparison of NMR signals of γ 1-Val MSG samples	136
Figure 27	The NMR spectrum (2D methyl TROSY) using δ_1 , γ_2 -Ile, δ_2 -Leu, γ_1 -Val methyl labeled MSG prepared using the BL21(DE3) $\Delta ilvD\Delta leuB$ Strain.....	138
Figure 28	Methyl-TROSY spectra of MSG samples labeled by Leu and Val along with $[3-^{13}\text{CH}_3;3,4,4,4-^2\text{H}_4]$ - α -ketoisovalerate and δ_2 -Leu	140

List of Tables

Table 1	Kinetic constants for MazFbs(<i>arg</i>) and MazFbs(<i>can</i>).....	87
Table 2	Relative endoribonuclease activity of MazFbs(<i>can</i>) using cleavable and uncleavable substrates	87
Table 3	Isotope labeling efficiency using BL21(DE3) $\Delta ilvD\Delta leuB$ and different concentrations of amino acids	132

Abbreviations

ALS	Acetolactate synthase
AMP	antimicrobial peptides
Bac7	bactenecin 7
CAP	catabolite activator protein
<i>E. coli</i>	<i>Escherichia coli</i>
GST	Gluthathione S-Transferase
HSQC	heteronuclear single quantum coherence
HMQC	heteronuclear multiple-quantum correlation
IPTG	isopropyl β -D-1-thiogalactopyranoside
MBP	Maltose Binding Protein
MSG	Malate synthase G
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect
pI	isoelectric point
PPD	Peptidyl prolyl isomerase domain
ppGPP	guanosine tetraphosphate
PrS ₂	Protein S tag
RNA	ribonucleic acid
SAIL	stereo array isotope labeling
SPP	single-protein production
SUMO	Small Ubiquitin-like Modifier protein
TA	toxin-antitoxin

Abstract

Antimicrobial peptide Bac7(1-35) production using the PST-SPP system in *E. coli*

There are a number of toxin-antitoxin (TA) systems discovered in *E. coli* and many of these TA systems have been characterized. MazF, one of the toxins from *E.coli*, functions as an ACA-specific endoribonuclease in cells. Taking advantage of MazF endoribonuclease activity in *E. coli*, a new expression system, the Single-Protein Production (SPP) system was developed.

In the SPP system, all *E. coli* cellular mRNAs are eliminated when MazF is induced. However, the mRNA of a target protein is engineered without ACA sequences while conserving the amino acid sequence so that it remains singularly intact when MazF is induced. Therefore *E. coli* is converted into a bioreactor producing only target proteins, making it especially effective for toxic proteins such as membrane proteins and antimicrobial peptides.

One of the difficult to express proteins in *E. coli* is an antimicrobial peptide (AMP) because of its toxicity to bacteria. Since the first discovery of the defensin peptide, a number of AMPs have been isolated. One of the reasons why AMPs are attractive to researchers is that while conventional antibiotics create their resistance, leading to multi-drug resistant bacteria and causing serious problems in treating infectious diseases, AMPs do not create drug-resistant bacteria because there are multiple intra-cellular targets for AMPs. In addition, it is relatively easy to modify the peptide sequences and possibly enhance the selectivity of its activity to bacterial cells with such changes.

Bac7(1-35) is a bovine AMP of 35 amino acid residues and is a Pro- and Arg-rich peptide. Previous studies using X-ray structure analysis have revealed one of the mechanisms of its toxicity to bacteria. Since Bac7(1-35) inhibits protein synthesis by binding to 70S ribosome, it is difficult to express in *E. coli* in its functional form.

While chemical synthesis methods have been applied to most AMP production, some recombinant expression systems such as *E. coli* and yeast systems have been successful by fusing AMP with large expression tags such as Maltose Binding Protein (MBP), Gluthathione S-Transferase (GST) and Small Ubiquitin-like Modifier protein (SUMO). However, no effective method to produce Bac7(1-35) has yet been developed.

Here I developed a novel expression system in *E. coli*, combining Protein S, a spore coat protein from *Myxococcus xanthus* with the SPP system for antimicrobial peptide production. This system produces a fusion protein which functions the same as the protein without the Protein S tag (PST). After overexpression of Bac7(1-35) using the PST-SPP system, PST-Bac7(1-35) is not only soluble, but also it functions as an antimicrobial peptide without cleaving the protein S tag from the fusion protein. This technology enables us to obtain a large amount of antimicrobial peptide in *E. coli* in a cost effective way.

Replacement of Arg residues in MazFbs with canavanine alters its specificity

I also explored expanding the capabilities of the SPP system to incorporate amino acid analogues, in particular the toxic arginine analogue, canavanine, in order to observe how protein function is altered with such substitutions. Canavanine is originally extracted from jack beans, and well known as a toxic amino acid analogue to cells. The mechanism of the toxicity is possibly due to the pKa of guanidino group in canavanine being 7 while that of arginine is 11. This charge difference causes an abnormal charge network once canavanine is incorporated into the protein leading to an eventual loss of the proper protein structures in the cells. To understand protein functions, site-directed mutagenesis has been commonly used. In addition, site-directed mutagenesis using amino acid analogues and amber codons together with modified tRNA ligase was developed to further characterize protein function. Although these systems are highly useful in studying residue specific function, a change in the protein conformation might not necessarily result in a functional change when the effect of site-directed mutagenesis is subtle. To better characterize protein function, not only site-directed residue replacement but also global replacement is necessary.

Here, the SPP system is combined with an arginine auxotrophic strain, allowing canavanine to be incorporated into protein efficiently without showing its toxicity to the cells. This is because when MazF is expressed and cell growth is arrested, the cells use canavanine for only target protein production. As an example, MazFbs, a MazF homologue from *Bacillus subtilis* and a UACAU specific endoribonuclease, is used as a model protein. Incorporating canavanine into

MazFbs caused MazFbs(*can*) to become more helical in structure but less stable in comparison to MazFbs because the conformation of the protein was changed. This is considered to be due to the change in the pI of MazFbs(*can*), altering the recognition sequence for cleavage to UACAUA rather than the original MazFbs UACAU recognition site. This is the first example of alteration of the RNA restriction enzyme recognition site by incorporating a toxic amino acid analogue.

Construction of a residue- and stereo-specific methyl labeling method by engineering *E. coli*

Thirdly, since the auxotrophic strain is highly useful for specific amino acid labeling, I established a cost effective labeling system for NMR structural studies. Large molecular weight proteins have some dynamics, and their function and dynamics have been characterized by NMR spectrometry. However, deuteration of proteins larger than 20-kD proteins is necessary and methyl specific protonation of Ile, Leu and Val residues is commonly used to study its dynamics. For 80 kDa or larger proteins, Stereo Array Isotope Labeling (SAIL) amino acids, in which the amino acid is stereo-specifically labeled, are used as it can dramatically reduce proton density compared to proteins labeled with common labeling precursors such as α -ketoisovalerate and α -ketoisobutyrate. However, since SAIL amino acids are extremely expensive, they have not been widely used in the NMR community. Thus, the establishment of an economical method for a high expression system using the minimum amount of SAIL amino acids is highly desirable. Here, I engineered the *E.coli* strain BL21(DE3), a standard expression host for residue-, stereo- and methyl-specific labeling systems, to use minimal SAIL amino acids. As a result, I was able to reduce the usage of SAIL

amino acids up to 10% compared to the standard method while maintaining protein production efficiency.

Lastly, I developed an alternative expression/labeling system for residue-stereo- methyl-specific labeled sample preparation for NMR using the common precursor, 2-acetolactate. In this system, the stereo specifically isotope-labeled 2-acetolactate is combined with genetically engineered *E. coli*, which allows proteins to be labeled in residue specific manners. Using a standard strain, Val-specific labeling is possible but Leu-specific labeling is difficult when using 2-acetolactate. To circumvent this, I engineered a biosynthetic pathway in *E. coli* to allow the cells to use 2-acetolactate for either Leu or Val synthesis so that either Leu or Val in a target protein can be labeled in a residue- and stereo- specific manner. Since auxotrophic strains were applied, the usage of the isotope labeled 2-acetolactate was reduced from 300 mg/L to 40 mg/L. To demonstrate the effectiveness of this system, I applied this method to Malate synthase G (MSG), which consists of 723 amino acid (83 kDa), Catabolite Activator Protein (CAP), a 47 kDa homodimer protein, and the peptidyl-prolyl isomerase domain of Trigger Factor, a 101 residue protein.

Acknowledgements

Firstly, I would like to thank my thesis advisor, Dr. Masayori Inouye, for his valuable guidance and support, as well as all members of the Inouye laboratory for their assistance. He always gave me suggestions whenever I needed them, and there were many discussions on science. During past seven years, what I learned from Dr. Inouye is not only science but also life.

I would also like to thank Dr. Tadashi Shimamoto from Hiroshima University for his guidance and assistance. Dr. Shimamoto has provided me with the wonderful opportunity to complete my thesis and obtain my Ph.D.

I would like to thank Dr. Babis Kalodimos from the University of Minnesota. During our collaboration work, they gave me great suggestions to complete my project.

I would also like to thank Dr. Masatsune Kainosho and the members of his laboratory from Nagoya University for our collaboration.

Last but not least, I would like to express my strong gratitude to my family, especially my parents for their support, encouragement and unconditional love.

Introduction

I. Toxin-Antitoxin (TA) system

The toxin-antitoxin (TA) system in bacteria consists of a stable toxin and an unstable antitoxin which is sensitive to stress-induced bacterial conditions such as proteases like Lon. Under normal growth conditions, the toxin forms a stable complex with its antitoxin, neutralizing the toxicity to cells. The genes for the toxin and antitoxin are located in the genome as an operon, and its transcription is negatively regulated by the antitoxin itself or the TA complex. The TA system located on the *ccd* operon was originally discovered about 30 years ago on the F-plasmid of *E. coli*, and was found to play a role in maintaining plasmids in the host bacteria (Ogura and Hiraga 1983). Since the first discovery of TA systems, a number of different TA systems have been identified from bacteria and archaea (Pedersen et al. 2003; Zhang et al. 2003b; Zhang et al. 2005b; Motiejūnaite et al. 2007; Schmidt et al. 2007; Jørgensen et al. 2009; Yamaguchi et al. 2009; Yoshizumi et al. 2009; Zhang and Inouye 2009; Christensen-Dalsgaard et al. 2010; Yamaguchi et al. 2012), and subsequently classified into three types.

The Type I TA system

In the Type I TA system, the gene expression of a toxin gene is regulated by a transcribed antisense RNA from the same toxin region in the reverse orientation (Kawano et al. 2002; Vogel et al. 2004; Kawano et al. 2007) (Fig. 1A).

Two mechanisms of toxin gene regulation by antisense RNA have been proposed. One mechanism suggests that the antitoxin RNA binds to the SD sequence of the mRNA of the toxin, preventing ribosome binding and resulting in translation inhibition (Kawano et al. 2007). The other mechanism proposes that

antitoxin RNA binds to the toxin gene, which induces mRNA degradation and results in translation inhibition (Jahn et al. 2012). In terms of physiological function, one of the type I TA systems, *hokB-sokB* is involved with persistence (Verstraeten et al. 2015). When the cells are exposed to nutrient limited conditions, the stringent response is triggered where RelA increases the alarmone ppGpp levels in the cells. The ppGpp binds to one of the GTPase Obg, and Obg induces HokB, expression. This results in the disruption of membrane potential since HokB is a small membrane protein and eventually leads to persistence by inducing a dormant state in the cells (Verstraeten et al. 2015).

The Type II TA system

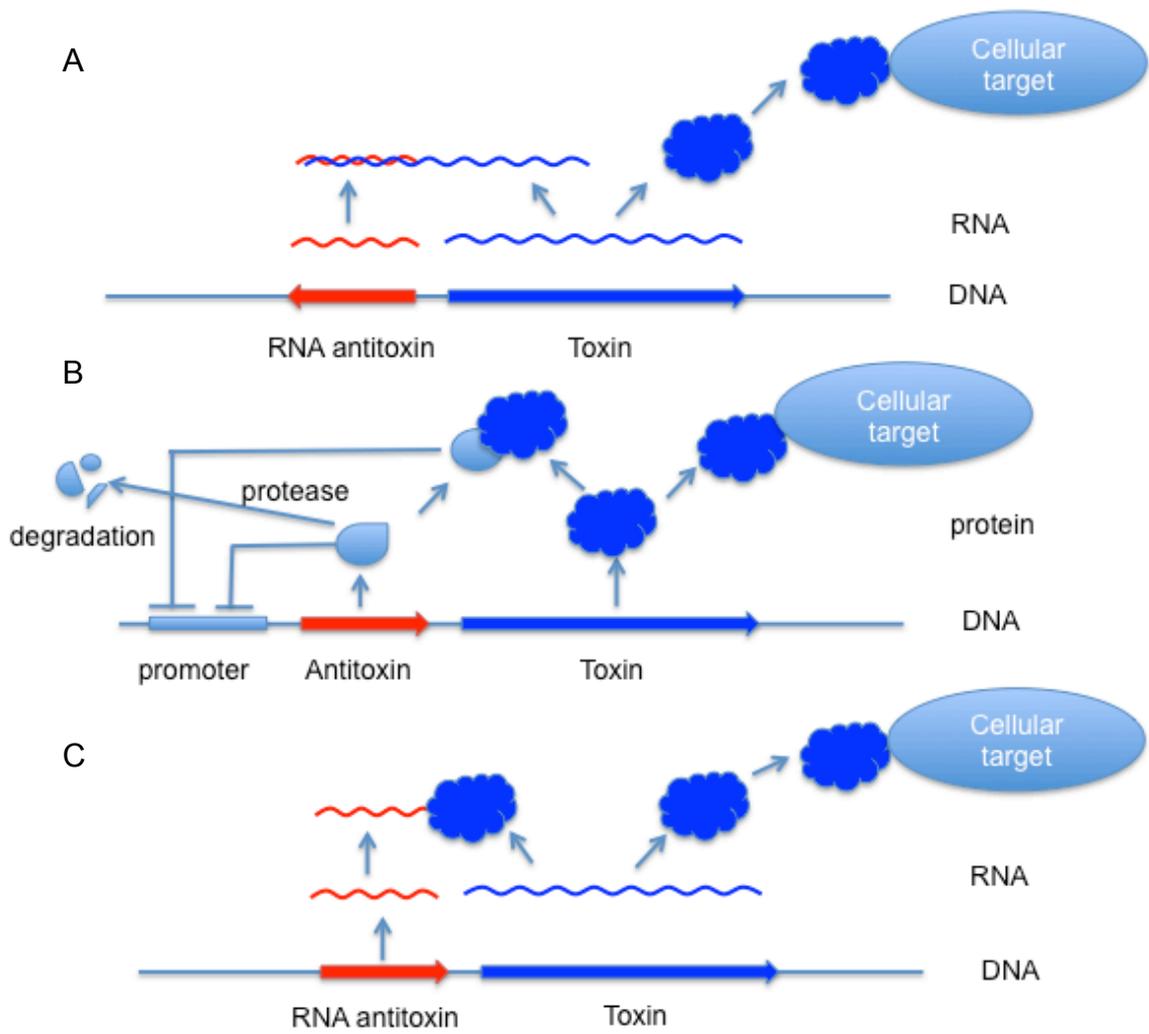
Among the TA systems, the type II TA system has been most extensively studied

(Aizenman et al. 1996; Pedersen et al. 2003; Yamaguchi et al. 2009; Zhang and Inouye 2009; Christensen-Dalsgaard et al. 2010). There are some common features shared in the Type II TA system (Yamaguchi and Inouye 2009) (a) each TA system is encoded by an operon consisting of two small genes usually overlapping by a few bases or apart by only a few bases (Yamaguchi and Inouye 2009); (b) a toxin and its antitoxin form a stable complex to block the function of the toxin (Fig. 1B); (c) the pI values of the toxin proteins tend to be basic, on the other hand, their antitoxin proteins are acidic (Yamaguchi and Inouye 2009). These opposite pI values are considered to stabilize the formation of TA complexes; (d) the antitoxin is unstable compared to the toxin in the cell, therefore, the antitoxin has to be expressed continuously to prevent the toxicity of the toxin (Aizenman et al. 1996); (e) under stress conditions, the degradation of

antitoxins by induced proteases frees toxins from the TA complexes in the cells, causing cell death or cell growth arrest (Aizenman et al. 1996; Christensen-Dalsgaard et al. 2010); (f) the TA complex binds to a palindromic sequence in the promoter region of the TA operon, and as a result, the TA complex functions as the repressor for the transcription of the TA operon (Marianovsky et al. 2001; Kedzierska et al. 2007)

The Type III TA system

A distinguishing feature of the type III system is that the antitoxin RNA sequence is composed of 34-36 nucleotides repeated 2-6 times. This antitoxin RNA binds to the toxin to neutralize the toxicity (Blower et al. 2009; Fineran et al. 2009; Bélanger and Moineau 2015). Recently, the X-ray structure of one of the Type III systems, ToxI-ToxN was determined (Blower et al. 2011), revealing that the RNA antitoxin, ToxI-RNA (36 nucleotides) interacts with three ToxN monomers to generate a trimeric ToxN-ToxI complex. In addition, ToxN proteins have a sequence-specific ribonuclease activity responsible for their toxicity and for the processing of their RNA antitoxins (Short et al. 2013). Besides its TA function, some of the TA systems have been shown to have anti-phage activities (Fineran et al. 2009; Bélanger and Moineau 2015).



See the following page for figure legend

Figure 1. The Toxin-Antitoxin (TA) systems.

Regulation of TA systems. The figure is modified from Yamaguchi and Inouye, 2011. **(A)** Regulation of type I TA systems. The toxin and antitoxin (antisense RNA) genes are transcribed separately. Antitoxin RNA binds to the toxin mRNA to form a complex, inhibiting translation of the toxin. **(B)** Regulation of type II TA systems. mRNAs of the antitoxin and toxin are synthesized from the same promoter and translated to produce proteins. The toxin forms a complex with its antitoxin, inhibiting the toxicity of the toxin to auto-regulate the TA module. Antitoxins are subjected to cleavage by ATP-dependent proteases under stress conditions, releasing the toxin from the TA complex to bind its cellular target. The activity of toxin leads to bacterial cell growth inhibition and to eventual cell death. **(C)** Regulation of type III TA systems. RNA antitoxin inhibits toxin protein by binding to the toxin.

II. mRNA interferase

MazF, an ACA-specific endoribonuclease

A toxin gene, *mazF* is localized on the chromosome as an operon with its cognate antitoxin, *mazE* (Fig. 2A). Among all the TA systems, the MazE-MazF system has been the most extensively studied. MazF functions as a toxin with a stable form while the antitoxin MazE is unstable. Together they form a complex consisting of one MazE homo dimer and two MazF homo dimers (Zhang et al. 2003a), which was substantiated by the X-ray structure of the MazE-MazF complex (Fig. 2C) (Kamada et al. 2003). When cells were exposed to stress conditions, MazE was rapidly degraded by a stress-induced serine protease ClpA (Aizenman et al. 1996), releasing MazF from the MazE-MazF complex to exert its cellular toxicity (Fig. 2A) (Inouye 2006). Furthermore, the negative regulation by the MazE homodimer and the MazE-MazF complex is lost due to degradation of MazE (Marianovsky et al. 2001; Zhang et al. 2003a). This initiates the transcription of *mazE-mazF*, producing both MazE and MazF proteins. Since the cell is under stress conditions, MazE is rapidly degraded to release free MazF in the cells to promote cell death (Fig. 2A). It has been shown that MazF functions as an ACA-specific endoribonuclease, cleaving ACA sites in single-stranded mRNA but not in single-stranded DNA (Fig. 2B) (Zhang et al. 2003b; Zhang et al. 2005a). It was also found that the 2'-OH group is important and required for cleavage by MazF. Cleavage generates a free OH group at the newly formed 5'-end and a cyclic 2' 3'-phosphate at the end of the previous residue (Zhang et al. 2005a).

Various stress condition can trigger the activation of MazF. Those conditions are; (1) exposure to antibiotics such as chloramphenicol, rifampicin and

spectinomycin, which inhibit translation and/or transcription (Sat et al. 2001); (2) amino acid starvation which causes ppGpp production (Aizenman et al. 1996); (3) exposure to thymine-less conditions (Sat et al. 2003); (4) exposure to a toxin, Doc, which inhibits the translation of MazE (Hazan et al. 2001); (5) some other stress conditions such as high temperature, oxidative conditions, and DNA damage (Hazan et al. 2004).

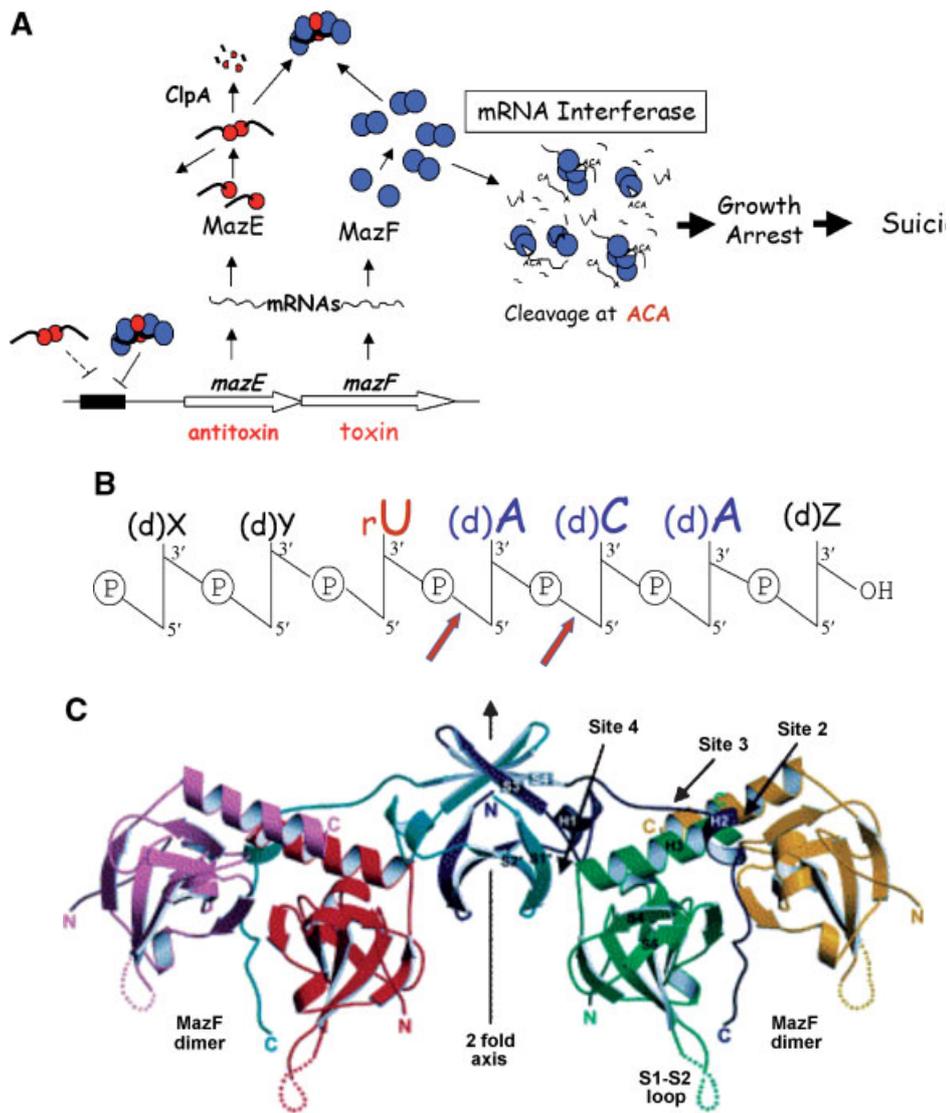


Figure 2. The Toxin-Antitoxin system, MazF-MazE system

The figure is adapted from Inouye, M 2006 and Kamada K. *et al.* The figure legend is in the following page.

Figure 2. The Toxin-Antitoxin system, MazF-MazE system

The figure is adapted from Inouye, M. 2006 and Kamada et al., 2003. (A) Schematic representation of the regulation pathway of the *mazE-mazF* operon. The upstream gene MazE and the downstream gene MazF are co-transcribed from a single promoter upstream of the *mazE* gene. Both MazE and MazF form a stable dimer, and the MazE dimer forms a complex with two MazF dimers so that the toxicity of MazF is suppressed. The MazE dimer and the MazE-MazF complex also bind to a palindromic sequence in the promoter region of the *mazE-mazF* operon so that MazE and MazF production is negatively auto-regulated by either MazE or the MazE-MazF complex. MazE is much more unstable than MazF as it is specifically degraded by ClpA, a stress-induced serine protease. Under stress conditions such as amino acid starvation or exposure to antibiotics, free MazF in the cell cleaves mRNAs at ACA sequences, resulting in cell growth arrest and eventually promoting cell death (B) The cleavage mechanism of MazF on the ACA sequence. The 2'-OH group of the rU and A residues are required for MazF cleavage on each site as indicated by arrows (Zhang et al. 2005a). Thus, MazF cleaves only RNA and not single stranded DNA. (C) X-ray structure of the MazE-MazF complex (Kamada et al. 2003). MazE homodimer, blue/light blue. MazF homodimers, yellow/green and pink/red.

MazFbs from *Bacillus subtilis*, a UACAU specific endoribonuclease

Since the discovery of MazF, a number of sequence-specific endoribonucleases in the genome from bacteria and archaea have been identified and characterized (Zhang et al. 2005b; Yamaguchi et al. 2009; Zhu et al. 2009; Yamaguchi et al. 2012). Those enzymes are termed mRNA interferases, one of which seems to be involved in developmental programmed cell death (Nariya and Inouye 2008; Yamaguchi and Inouye 2009), while the physiological roles of other mRNA interferases in bacteria are still not clear.

Bacillus subtilis contains a MazF homologue, MazFbs (EndoA) (Pellegrini et al. 2005), which has been shown to cleave mRNA at U[^]ACAU (Park et al. 2011). MazFbs has high similarity (79.3-96.6%) to MazF homologues from various Gram-positive bacteria but not from Gram-negative bacteria (Fig. 3A and 3B). The crystal structure of MazFbs has been determined (Gogos et al. 2003) showing high homology to the structures of MazFec (Kamada et al. 2003) and Kid (Hargreaves et al. 2002). Although there is a very high structural similarity between MazFbs and MazFec, MazEec could not neutralize MazFbs toxicity, suggesting that each antitoxin has been co-evolved with its cognate toxin (Pellegrini et al. 2005; Simanshu et al. 2013). The crystal structure of MazFbs with its substrate was determined and it was shown that the MazEbs dimer binds to the MazFbs dimer, which prevents RNA binding and cleavage within both subunits of the MazFbs dimer (see Figure 2). Since MazE binds to MazF with higher affinity than the substrate of MazFbs, MazFbs cannot cleave substrate RNA in the presence of MazEbs (Simanshu et al. 2013).

A

```

MazFec      MVSRYVPDMDGLIWVDFDPTKGSEQAGHRPAVVLSPFMYNNKTGMC-LCVP--CTTQSKG
MazFmt3     -----MRPIHIAQLDKARPVLILTREVVRPH-LTNVTVAPITTTV--RG
MazFmt6     ---MVISRAEIIYWADLGPPSGSQPAKRRPVLVIQSDPYNASRLATVIAAVITSNTALAA
MazFbs      ---MIVKRGDVYFADLSPVVGSEQGGVRLVLIQNDIGNRFSPTA-IVAAITAQIQKAK
              : *   ::   **, :::   .   .

MazFec      YPFEVV----LSGQERDGVADLADQVKSIAWRARGATKKGTVAPEELQLIKAKINVLIG--
MazFmt3     LATEVPVDA-VNGLNQPSVVSCDNTQTIPVCDLGR-QIGYLLASQEPALAEAIGNAFDLD
MazFmt6     MPGNVFLPATTTRRLPRDSVVNVTAIVTLNKTDLTD-RVGEVPASLMHEVDRGLRRVLDL-
MazFbs      LPTHVEIDAKRYGFERDSVILLEQIRTIDKQRLTD-KITHLDDMMDKVDEALQISLALI
              . *           : . *           ::           :   :   .   :   :   :

MazFec      ----
MazFmt3     WVVA
MazFmt6     ----
MazFbs      DF--

```

B

```

MazFsa      ---MIRRGDVYLADLSPVVGSEQGGVRLPVVLIQNDTGNKYSPTVIVAAITGRINKAKIP
MazFcb      MTQQIVKRGDIFYADLSPVVGSEQGGIRPVVLIQNNVGNKYSPTVIAAITSQINKAKLP
MazFbs      ---MIVKRGDVYFADLSPVVGSEQGGVRLVLIQNDIGNRFSPTAIVAAITAQIQKAKLP
MazFlm      ---MVKRGDVYYADLSPVVGSEQGGIRPVLIQNDIGNRFSPTVIVAAITAKIQKAKLP
              :::***:  ***** *****:**:***: **:***: *:****: ;*:***:*

MazFsa      THVEIEKKKYKLDKDSVILLEQIRTLDKRRLKEKLTYSDDKMKEVDNALMISLGLNAVA
MazFcb      THVEISSDYGLNKDSVLLLEQIRTLDKRRLKEKIGHMIDEDMKKVDTAILVSMALN---
MazFbs      THVEIDAKRYGFERDSVILLEQIRTIDKQRLTDKITHLDDMMDKVDEALQISLALIDF-
MazFlm      THVEAT-RKDGFERDSVILLEQIRTIDKQRLTDKITHLDEDLMAKVNKALEVSLGVVEF-
              ****   .   :::***:*****:**:**: * : : : * : * : * : * : .

MazFsa      HQKN
MazFcb      ----
MazFbs      ----
MazFlm      ----

```

Figure 3. Sequence alignments of MazFbs and its homologues

Sequence alignments of MazFbs with other MazF homologues in Gram-negative bacteria (A) and in Gram-positive bacteria (B) (ec: *Escherichia coli*, mt: *Mycobacterium tuberculosis*, bs: *Bacillus subtilis*, sa: *Staphylococcus aureus*, cb: *Clostridium botulinum*, lm: *Listeria monocytogenes*)

III. The single-protein production (SPP) system

MazF from *E. coli* is an ACA-specific endoribonuclease, and cleaves at all ACA sequences in mRNA (Zhang et al. 2003b). Its induction in *E. coli* causes degradation of almost all cellular mRNAs, blocking cellular protein synthesis and leading to total cell growth arrest since almost all *E. coli* mRNAs contain ACA sequences (Zhang et al. 2003b). However, when a gene of interest was engineered without ACA sequences while conserving the amino acid sequence, the cells were able to produce the protein from the ACA-less mRNA without producing any other cellular proteins under cold shock conditions (Fig.4A) (Suzuki et al. 2005; Suzuki et al. 2006).

MazF induction creates a novel physiological state called 'quasi-dormancy' in which cells are metabolically active and capable of producing proteins from ACA-less mRNA sequences. These dormant cells are fully capable of amino acid and nucleotide synthesis, RNA and protein synthesis as well as ATP production (Suzuki et al. 2005).

To co-express a gene of interest and MazF, *mazF* was cloned into the pACYC vector which is under a lactose inducible T7 promoter, and an ACA-less target gene was synthesized and cloned into the pCold vector which is a cold-shock inducible vector (Qing et al. 2004). The human eotaxin gene encoding a 74-residue cytokine was cloned into the pCold vector and shown as an example.

After 2 hrs of MazF induction, background cellular protein synthesis was almost completely blocked, however, eotaxin was still produced with ³⁵S-methionine incorporated into only eotaxin (Fig. 4B) (Suzuki et al. 2006)

A number of other proteins including standard proteins such as Calmodulin, EnzvB (a histidine kinase domain) and CspA (a major cold shock protein) (Suzuki

et al. 2006; Vaiphei et al. 2010) were also specifically expressed using the combination of MazF and the SPP system. Since cell growth was arrested without affecting protein synthesis, some toxic proteins such as membrane proteins and antimicrobial peptides were effectively produced using the SPP system (Mao et al. 2009; Vaiphei et al. 2011). In addition, it was also found that the culture for the SPP system can be condensed up to 100 times without affecting the final protein yield. Using this condensed SPP (cSPP) system, a one-liter culture can be condensed to 25 or 10 ml, resulting in a 95% to 99% cost savings for isotope labeling in NMR structural studies (Suzuki et al. 2006; Mao et al. 2009). Furthermore, a non-natural amino acid was incorporated into a protein (Suzuki et al. 2006) using this system. This capability applied together with an auxotrophic strain, allowed toxic amino acids such as canavanine to be incorporated into a protein efficiently because the amino acid analogue was only used for specific protein production under the quasi-dormant state (Ishida et al. 2013).

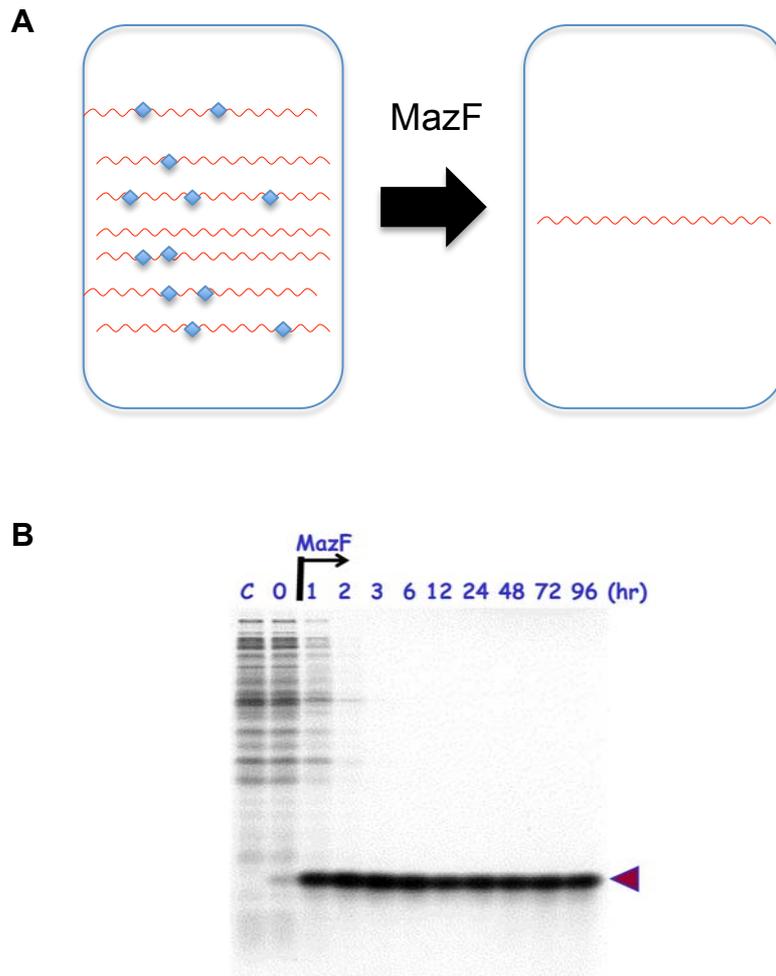


Figure 4. The Single-Protein Production (SPP) system

(A) The quasi-dormant cells are still fully metabolically active, and therefore if an ACA-less mRNA is induced, the cells are capable of producing a protein from that mRNA. Thus, that is the only protein produced in the cell resulting in the single-protein production (SPP) system. (B) The gene for eotaxin was synthesized without ACA-sequences and inserted in pColdI vector. At the times indicated after MazF induction, the cells were pulse-labeled with [³⁵S]-methionine (Suzuki et al. 2005).

IV. Dual inducible cSPP systems

One of the great advantages of the cSPP system is that the cell culture can be condensed after MazF is fully induced. The culture condensation does not affect the protein yield, therefore, this leads to substantial cost savings for NMR sample preparation using isotope-labeled chemicals such as D₂O, ¹³C-glucose, and amino acids. To obtain good signals from NMR analysis, a highly labeled and large amount of sample is required. One drawback of the cSPP system is the use of IPTG as an inducer for both MazF and the target protein which results in the target protein being produced at the same time as MazF. Since culture condensation with isotope labeled chemicals can be done 2 hours after the addition of IPTG to avoid the incorporation of isotope into cellular proteins, non-isotope labeled target protein is also produced during this pre-incubation period. This can result in a higher background of unlabeled target protein, which may be as high as 20% of the final yield of the target protein production (Schneider et al. 2009).

To circumvent this, a system was developed in which the target protein production is induced using amino acid. In this gene expression system MazF and a target protein can be induced separately by IPTG in combination with amino acid auxotrophs (Vaiphei et al. 2010). Previously, It has been shown that His auxotrophic *E. coli* cells can still produce a protein which does not contain His residues in the absence of histidine in the medium without producing any other cellular proteins (Hirashima and Inouye 1973). Therefore, when His residues in MazF are removed by mutagenesis, MazF(Δ H) can be induced in the absence of histidine by using a His auxotroph (Fig. 5) (Vaiphei et al. 2010). In this system the target protein is induced by the addition of histidine in the

medium 2 hours after MazF(Δ H) induction so that background production of the protein is eliminated (Fig. 5) (Vaiphei et al. 2010; Ishida et al. 2013). In the same manner, a MazF(Δ W) system in which Trp residues were removed from MazF was developed together with Trp auxotrophic cells (Fig. 5) (Vaiphei et al. 2010). These systems are highly useful to achieve protein production with high labeling efficiency.



Figure 5. A comparison of the SPP system and the His/Trp inducible SPP system BL21(DE3) cells harboring MazF and pCold-target genes were used. Under the normal IPTG induced system both MazF and the target protein are simultaneously induced with IPTG. During the 2 hrs of MazF induction that is required before condensing the culture into the isotope labeled medium, unlabeled target protein was observed (up to 20%). On the other hand, when the His auxotroph strain was used with pMazF(Δ H), only MazF(Δ H) was induced in the presence of IPTG and in the absence of His in the medium. After 2 hrs of MazF(Δ H) induction, the culture was condensed into the isotope labeled medium. Thus, unlabeled background was significantly reduced. In the same manner, when the Trp auxotroph strain was used with pMazF(Δ W), only MazF(Δ W) was induced in the presence of IPTG and in the absence of Trp in the medium. After 2 hrs of MazF(Δ W) induction, the culture was condensed into the isotope labeled medium. Therefore, isotope incorporation was significantly improved compared to the basic SPP system (Vaiphei et al. 2010).

V. T7 expression system

The T7 expression system is the most well-known recombinant protein production system in *E. coli* and many modifications to this system have allowed for different expression conditions. After infection by phage λ DE3, the DNA sequence coding T7 polymerase was incorporated into the chromosome to construct expression host cells such as BL21(DE3), which is under the *lacUV5* promoter, a variant of *lac* promoter (Studier and Moffatt 1986) (Fig. 6). For protein production, a gene of interest is cloned into a vector which contains the T7 promoter (Studier and Moffatt 1986). Thus, after induction of T7 polymerase by IPTG, the transcription of the target gene is initiated when T7 polymerase binds to the T7 promoter. Since it was found that the T7 system had basal level expression without adding IPTG, a new variant was developed by adding the *lac* operator downstream of the T7 promoter sequence, which blocks the elongation initiated by T7 polymerase complex with its promoter (Dubendorff and Studier 1991). To control the gene expression more stringently, pLys or pLysE was incorporated into host cells to produce T7 lysozyme which binds and inactivates T7 RNA polymerase in uninduced cells (Studier 1991). More recently a T7 system was combined with a host strain that has very weak RNaseE activity which avoids the degradation of mRNA and increases mRNA stability, leading to higher protein expression (Grunberg-Manago 1999).

All these technology developments were applied to the sample preparation for structural studies using X-ray crystallography and ^{15}N , ^{13}C -labeling for Nuclear Magnetic Resonance (NMR), which requires large amounts of protein (Burley et al. 1999; Acton et al. 2011). Since the induction of the target gene can be strictly controlled, this T7 expression system was applied to toxic proteins such as

membrane proteins, and also evaluated for in-cell NMR sample preparation in which signal-to-noise ratio is critical (Serber et al. 2001).

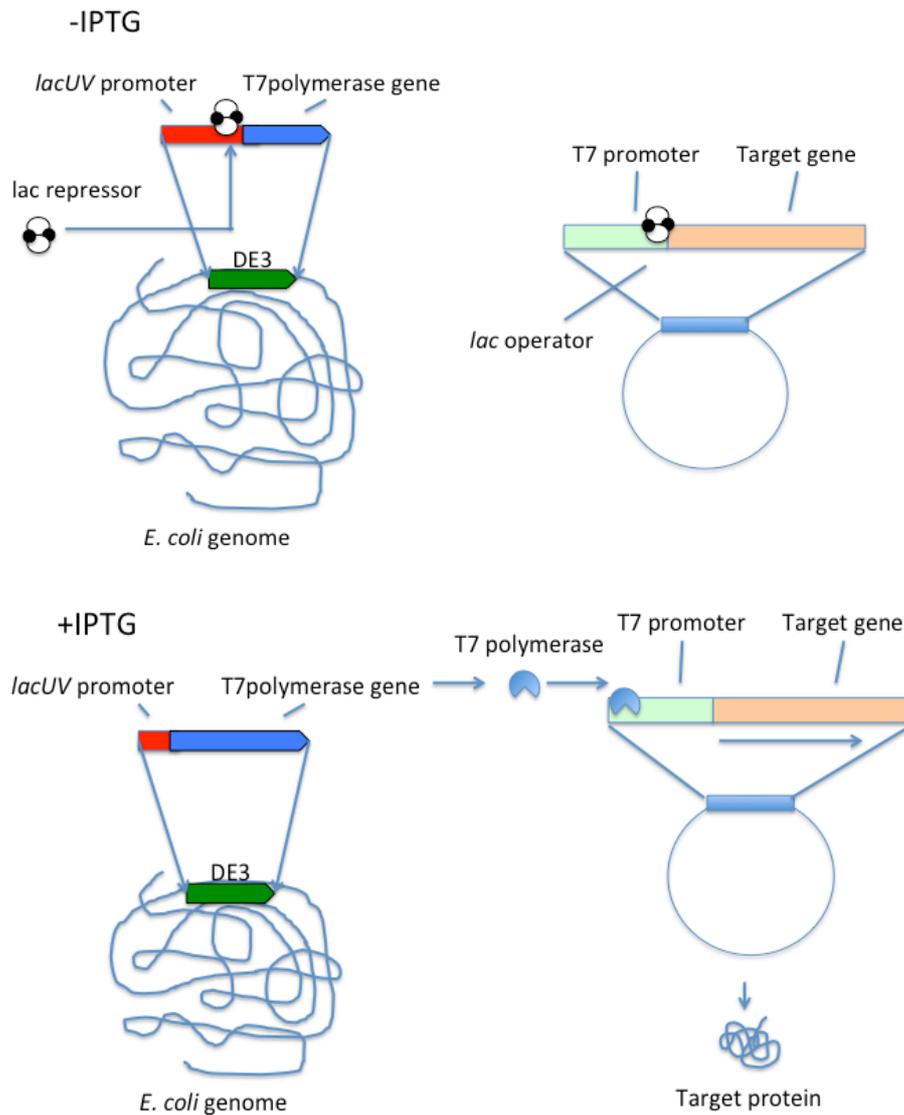


Figure 6. The T7 expression system and pET vectors

The figure is modified from <http://www.rnursingschool.biz/restriction-enzyme/the-t7-expression-system.html>. The plasmid harboring the target gene under the control of T7 promoter is transformed into the host cell, which bears the T7 RNA polymerase gene on its chromosome, the λ (DE3) as indicated. T7 RNA polymerase is induced by addition of IPTG, also inducing the expression of the target gene upon binding of T7 RNA polymerase to the T7 promoter region.

VI. Protein S tag

Protein S from *Myxococcus xanthus*, consisting of 173 amino acid residues (18.8 kDa), is a major spore coat protein detected during fruiting body formation (Inouye et al. 1979). The function of protein S in *M. xanthus* is still not clear however, and in the gene deletion mutant, fruiting body formation is delayed and reduced by about 70% compared with the wild-type strain. In the gene deletion mutant, when myxospore formation is induced by the addition of guanosine 3'-di-5'-(tri)di-phosphate nucleotides[(p)ppGpp], fruiting bodies are not formed (Singer and Kaiser 1995). These results suggest that protein S plays a role in the formation of fruiting bodies and cell-cell interactions during fruiting body formation (Inouye et al. 1979). Protein S shows high sequence homology to γ B-crystallin, and since both protein S and γ B-crystallin belong to the $\beta\gamma$ -crystallin super family, they share typical features of the $\beta\gamma$ -crystallin super family such as two β -sheet rich domains and high solubility. Structurally, protein S consists of four internal homologous motifs, in which the second and fourth motifs are composed of a regular Greek key motif (four β -strands), while the first and third motifs contain α -helices in addition to normal Greek key motif (Fig. 7). These four motifs produce two domains, an N-terminal domain (NTD; 92 residues) and a C-terminal domain (CTD; 81 residues), with both domains containing a calcium binding site (Bagby et al. 1994b). The structure was determined by solution NMR (Bagby et al. 1994a) (Fig. 7). Protein S itself is stable at high temperatures and against denaturants in the presence of calcium ions (Wenk and Mayr 1998). The NTD has a higher stability than the CTD (Wenk et al. 1999).

Taking advantage of high expression and solubility of the NTD, a new expression vector pCold-PST, encoding two tandem repeated NTDs (PrS₂) was

developed (Kobayashi et al. 2009). By using this vector, a number of human proteins that were expressed at low levels or in insoluble forms by a pET vector were expressed not only at high levels but also in soluble forms. It was also shown that a target protein fused with PrS₂ fully retained its function, indicating that it folded independently from the tag.

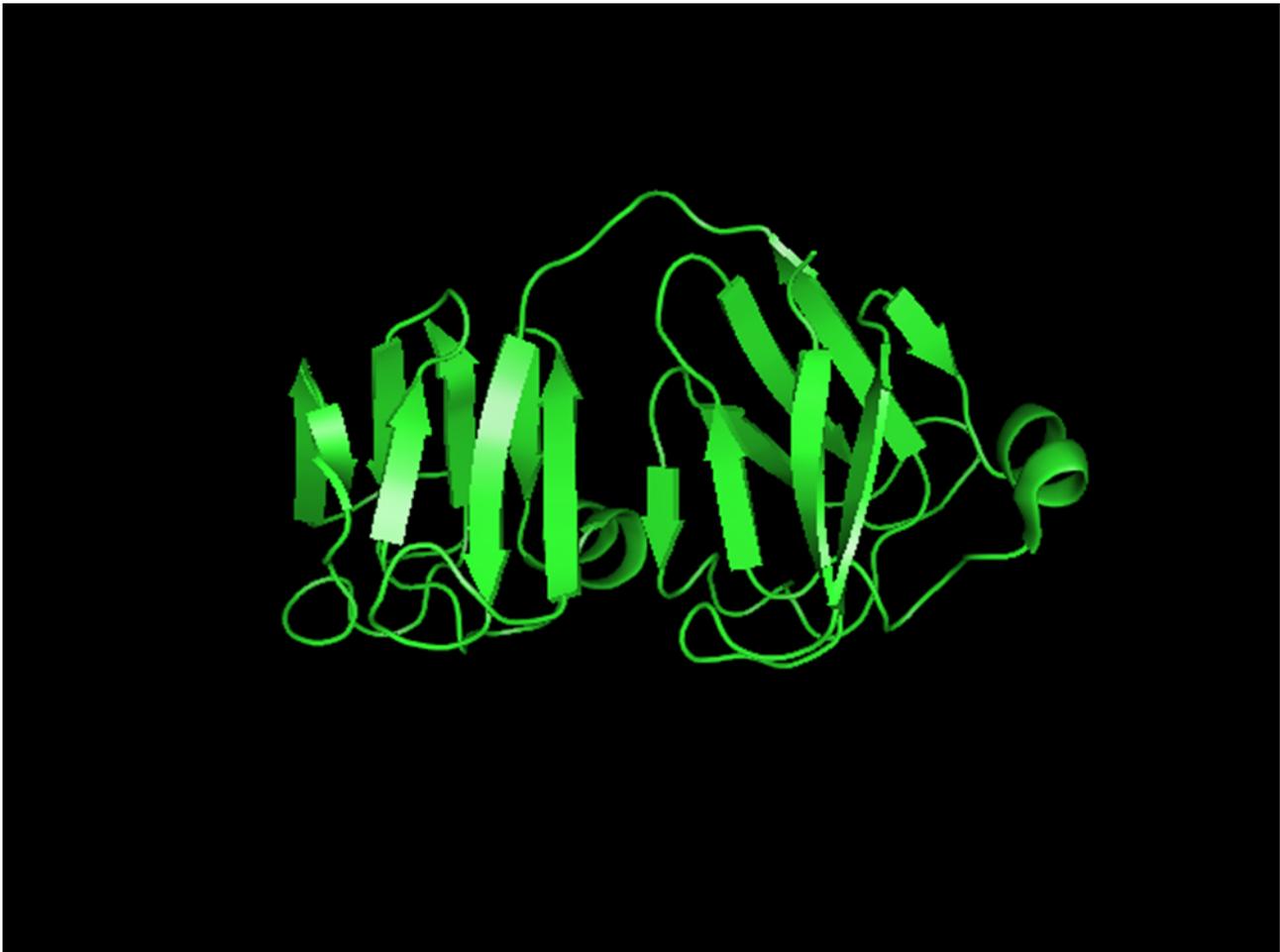


Figure 7

The NMR structure of Protein S (PDB ID;1PRS). This figure is adapted from (Bagby et al. 1994a)

Rationale

Bac7(1-35) production by PST-SPP system in *E. coli*

Despite the number of antibiotics available to date, the emergence of their drug resistant bacteria continues to be a serious issue for treatment against infectious diseases. While antibiotic development is still ongoing, antimicrobial peptides (AMPs) have been isolated and characterized as an alternative antimicrobial agent. Some of them are in clinical trials against some infections (Fox 2013) and most of them have lytic activity by disrupting membranes. On the other hand, however, the mechanisms of antibacterial activity are not fully understood especially for non-lytic peptides. One type of non-lytic peptide is the Pro-rich peptide which has been discovered from the hemolymph of insects and bovine neutrophils (Scocchi et al. 2011). Because Pro-rich AMPs have a wide range of antimicrobial activity especially for Gram-negative bacteria, these peptides are being considered for potential clinical use. Among Pro-rich AMPs, bovine batenecin-7 (Bac7), a 60 residue polypeptide which is also found in sheep and goats, is Pro- and Arg rich. It has been shown that Bac7 inhibits protein synthesis by binding 70S ribosomes (Seefeldt et al. 2016), but is non-toxic to human cells. Interestingly, the C-terminal truncated construct of Bac7 retains its antimicrobial activity (Guida et al. 2015). While functional analysis of Bac7 has been carried out, there is still no effective production method except by chemical synthesis. Increasing medical costs are due to the time-consuming and expensive development of new antibiotics as well as their chemical synthesis. Thus, establishing a method for cost effective production of AMPs is necessary for clinical use. In fact, some AMP production have been successful using Maltose Binding Protein (MBP) (Miller et al. 1998) or Glutathione S Transferase

(GST) (Feng et al. 2010; Chen et al. 2015) or Small Ubiquitin Modifier (SUMO) protein (Li et al. 2011; Zhang et al. 2016) as fusion partners since those AMPs do not have activities as protein synthesis inhibitors. Since Bac7 is a protein synthesis inhibitor, there was no effective method of production in *E. coli*. Thus, the PrS₂ tag fusion expression system was coupled with the single-protein production (PST-SPP) system for the production of Bac7(1-35), as described in Section I. This takes advantage of the dormant cells created by the induction of MazF, an ACA-specific endoribonuclease. The PST-SPP system offered the best expression and more importantly, the fusion protein retained its protein synthesis inhibition activity and Bac7(1-35) was readily purified (Ishida and Inouye 2016). Since the fusion protein retains its antibacterial activity, it will be helpful in identifying intracellular targets.

Canavanine incorporation into MazFbs by SPP system

Protein engineering using amino acid analogues including D-amino acids can possibly enhance or change the function of proteins (Oddo et al. 2015). Thus, various methods to incorporate amino acid analogues have been developed. In order to understand the role of specific residues in a protein, site-directed mutagenesis or single amino acid replacement using amber codon has been carried out. When amino acid analogues are not recognized by tRNAs, engineered orthogonal tRNA from other bacteria have been cloned into *E. coli* to incorporate the amino acid analogues into amber codon positions (Wang et al. 2002; Minaba and Kato 2014). Recently, amino acid analogue incorporation was demonstrated in mammalian cells (Peng and Hang 2016). When amino acid analogues are recognized by tRNA in *E. coli*, they can be incorporated into

proteins, however, since the affinity of amino acid analogues to their tRNAs is lower than the natural amino acid (Igloi and Schiefermayr 2009), incorporation efficiency is expected to be low. In addition, global amino acid replacement by its analogue causes structural differences in entire protein networks, thus showing toxicity to the cells. For example, azetidine-2-carboxylic acid (a Pro analogue) (Bessonov et al. 2010) and canavanine (Can) (an Arg analogue) (Schachtele et al. 1968; Rosenthal et al. 1989) are known as toxic amino acid analogues. These amino acid analogues showed toxicity to *E. coli*, possibly because they caused abnormal protein structures after incorporation. To circumvent these issues, an amino acid auxotroph together with the SPP system was applied, shown in section II. To demonstrate the dual system, MazFbs from *Bacillus subtilis*, a UACAU specific endoribonuclease was chosen. After the first discovery of MazF in 2003 (Zhang et al. 2003b), it was demonstrated that this highly-sequence specific mRNA interferase may be used for regulation of gene expression by eliminating a specific set of genes (Yamaguchi et al. 2012). This raises an intriguing possibility of using mRNA interferases as a novel alternative method for mRNA interference to silence specific gene expression in addition to the well-established RNA interference using RNAs such as antisense RNA and siRNA. However, the number of MazF homologues for which the target RNA sequences are known is quite limited. Recently, the crystal structure of MazFbs from *Bacillus subtilis* together with its uncleavable substrate was determined (Simanshu et al. 2013), however, there have been no successful studies of extending the recognition site of MazFbs. In section II, I demonstrated that the replacement of Arg residues with Can led to the altering of function from UACAU cleaving activity to a UACAUA recognition enzyme. There are only a few established protein

modification methods by amino acid analogue incorporation, thus this global replacement by SPP together with an auxotrophic strain provides an additional option to create new proteins.

Stereo- and residue specific labeling system using deletion mutants

NMR presents the best tool to observe protein-ligand interactions, in particular the dynamics involved, as dynamics cannot be observed by using X-ray crystallography or Cryo-EM. On the other hand, in comparison with X-ray crystallography and Cryo-EM, it is very evident that protein structure determination by NMR has serious size limitations. To overcome the size limitations, selective protonation of methyl groups in Ile, Leu and Val at perdeuterated background using ^2H -glucose and D_2O has been applied (Goto et al. 1999). Using perdeuterated background silences the signals but Ile, Leu and Val signals can be observed by using selectively methyl protonated α -ketoisovalerate and α -ketobutyric acid, which are Leu/Val and Ile precursors, respectively. However, since methyl labeled α -ketoisovalerate compounds are in a racemic mixture, the signal intensities of Val and Leu were 50%. Recently, a stereo specifically labeled precursor, acetolactate was developed, and the labeling patterns of Val-*proS* and Leu-*proS* or Val-*proR* and Leu-*proR* became feasible. These signals are clearly observed and very helpful for assignments (Gans et al. 2010; Kerfah et al. 2015). Although stereo specific labeling can be done, residue specific labeling using precursors is still problematic especially for Leu labeling since α -ketoisovalerate and acetolactate are common precursors for Leu and Val. A new Leu precursor, α -ketoisocaproate was developed but this compound is still not commercially available, and there is no effective method to

prevent the conversion from α -ketoisovalerate and acetolactate to Val. Since the establishment of a method for residue- and stereo specific labeling is needed for NMR structural study, in section III-I, Leu or Val selective labeling using stereo specifically labeled acetolactate was demonstrated by constructing the acetolactate auxotrophic strains. This allows Leu and Val signals to be assigned without errors since the spectrum was simplified (Monneau et al. 2016).

On the other hand, in order to observe NMR signals from large molecular weight proteins, using stereo-specific labeled amino acids is the most straightforward method. For this purpose, Stereo-Array Isotope Labeled (SAIL) amino acids were developed (Kainosho et al. 2006). Since the labeling patterns of amino acid can be controlled, SAIL amino acids are thought to be ideal for NMR structural studies. However, endogenous amino acid production in *E. coli* dilutes the labeling efficiency by diluting exogenous SAIL amino acids, limiting the application of SAIL amino acids to a cell-free expression system (Takeda and Kainosho 2012). Recently, a SAIL-Leu labeling method was developed by adding Leu in 2 steps allowing the cells to incorporate SAIL-Leu into the protein effectively (Miyanoiri et al. 2013). However, Val labeling was unsuccessful possibly because efficiently produced pyruvate in the cells dilutes the SAIL-Val. Since stereo- and residue specific labeling using SAIL amino acids is highly important, in section III-II, a cost effective Val or Leu labeling system using Ile, Leu and Val auxotrophic strains was described. The minimum requirement for Val high-labeling efficiency using an auxotrophic strain was 10 mg/L while the standard strain requires more than 100 mg/L, while retaining 80% labeling efficiency. Since SAIL amino acid is highly expensive and thereby not widely

used, our labeling system will provide the opportunity to other NMR groups to use the SAIL system for their structural studies.

Section I

The antimicrobial peptide Bac7(1-35) production using PST-SPP system in *E. coli*

Introduction

Although quite a few antibiotics have been developed after penicillin was discovered, the emergence of drug resistant bacteria is still a serious problem in the treatment of infectious diseases. As an alternative antibacterial agent, relatively small peptides of 10 to 50 amino acids from insects to mammals harboring some antimicrobial activities have been designated as antimicrobial peptides (AMPs) (Daher et al. 1988; Xi et al. 2014; Jayamani et al. 2015). Some of the advantages to using AMPs are 1; good efficacy and tolerability 2; high selectivity and potency 3; predictable metabolism 4; shorter time to market 5; lower attrition rates 6; standard synthetic protocols can be applied (Fosgerau and Hoffmann 2015). On the other hand, some weaknesses are known 1; chemically and physically unstable 2; tendency for aggregation 3; short half-life and fast elimination 4; low membrane permeability (Fosgerau and Hoffmann 2015).

Although AMPs are a unique and diverse group of peptides, they are classified into 4 categories based on their amino acid composition and structure, 1) α -helical, 2) extended, 3) β -hairpin or loop due to the presence of the disulfide bond, 4) β -stranded due to multiple disulfide bonds (Wang 2015). Most AMPs function by either disrupting membranes or inhibiting cytoplasmic targets, however, most of the specific targets have not been identified. Thus, elucidating the antimicrobial mechanism will enable us to design a better drug for treatment. The disadvantages of AMPs are their cytotoxicity to the host, their stability in the cells, and the cost of chemical synthesis. Among AMPs, proline-rich antimicrobial peptides (PR-AMPs) have been isolated from mammalian neutrophils and from haemolymph of some invertebrate species (Scocchi et al. 1994; Schnapp et al. 1996; Anderson and Yu 2003). Bac7(1-35) is a PR-AMP isolated from bovine and belongs to the cathelicidin family (Podda et al. 2006). Bac7(1-35) inhibits

protein synthesis at 0.5 μM , and disrupts the membrane structure at 64 μM or higher concentration (Podda et al. 2006). Recently, Bac7(1-35) was reported to bind to 70S ribosome, with the N-terminal 16mer of Bac7(1-35), Bac7(1-16) (Seefeldt et al. 2016) required for protein synthesis inhibition. In addition, the crystal structure of Bac7(1-16) with 70S ribosome was determined, and it was found that Bac7(1-16) bound within the exit tunnel of 70S ribosome (Roy et al. 2015). Despite the high toxicity to Gram-negative bacteria such as *E. coli*, *Klebsiella pneumoniae*, *Salmonella typhimurium*, and *Enterobacter cloacae* at 1 to 10 μM , Bac7(1-35) has remarkably low cytotoxicity to the host mammalian cells (not toxic even at 50 μM) (Pelillo et al. 2014). Therefore, Bac7(1-35) has been extensively studied because of its potential use for clinical applications. A method to stabilize Bac7(1-35) by glycosylation has been developed to reduce its renal clearance by which Bac7(1-35) still retains its antibacterial activity as well as cell penetration activity (Benincasa et al. 2015). For pharmaceutical applications, development of a method for efficient AMP synthesis is important. The production of some AMPs has been successfully carried out with a yeast system when these AMPs are not toxic to yeast (Mulder et al. 2015; Zhang et al. 2015b; Chahardooli et al. 2016). On the other hand, some AMPs have been produced using an *E. coli* system in combination with fusion tags such as thioredoxin (Herbel et al. 2015), Glutathione S-transferase (GST) (Feng et al. 2014), Maltose-Binding Protein (MBP) (Miller et al. 1998) and Small Ubiquitin-like Modifier (SUMO) (Yi et al. 2015; Zhang et al. 2016). To enhance protein production, inclusion body preparation procedures have been applied (Panteleev and Ovchinnikova 2015), however, denaturing, refolding and subsequent cleaving of the fusion tag require labor extensive and time consuming work.

However, it is necessary to cleave the tags off from the AMPs to generate active AMPs. Eliminating the necessity to cleave the tag from the fusion construct while retaining AMP activity would considerably facilitate expression, purification and toxicity assay. For this purpose, protein S (PrS), a major spore-coat protein from *Myxococcus xanthus* is used as a fusion tag [called Protein S technology (PST)], enhancing the expression as well as the solubility of the target protein (Kobayashi et al. 2009). Importantly, protein S fused to the target protein not only does not interfere with the function of the target protein but also does not affect the structure of the target protein (Kobayashi et al. 2009; Kobayashi et al. 2012). To further improve protein production, I used the Single-Protein Production (SPP) system for the production of the PrS-Bac7(1-35) fusion protein, in which MazF, an ACA-specific endoribonuclease from *E. coli* is induced at low temperature (Suzuki et al. 2006).

Materials and methods

Construction of pColdPrS₂Bac7(1-35) and pColdSUMOBac7(1-35)

A codon optimized ACA-less Bac7(1-35) gene was synthesized from IDT and the synthetic gene was cloned into ACA-less pColdPrS₂ vector (Takara Bio) by using an infusion cloning system (Clontech), generating pColdPrS₂-Bac7(1-35). In order to obtain pColdSUMOBac7(1-35), a codon-optimized ACA-less SUMO-Bac7(1-35) was synthesized (IDT) and cloned into pColdIII vector by infusion cloning system, generating pColdSUMOBac7(1-35).

Bac7(1-35) production with fusion tags in BL21(DE3)

In order to produce Bac7 (1-35) as a fusion protein, BL21(DE3) cells transformed with either pColdSUMO-Bac7(1-35) or pColdPrS₂-Bac7(1-35) were inoculated into 10 ml of LB medium and the culture was incubated at 37 °C. When O.D₆₀₀ reached 0.8, the culture was transferred to 15°C and 1mM IPTG was added to induce the fusion proteins. The culture was further incubated overnight.

The production of Bac7(1-35) by PST-SPP system

BL21(DE3) co-transformed with pColdPrS₂-Bac7(1-35) and pACYCmazF was inoculated into 10 ml of LB medium and the culture was incubated at 37°C overnight. Next day, the culture was diluted into 1L LB and incubated at 37°C. When the O.D₆₀₀ reached 0.8, the culture was placed on ice for 5 min, followed by incubation at 15 °C for another 1 hr. Subsequently, MazF and PrS₂-Bac7(1-35) were induced by the addition of 1 mM IPTG, and the culture was incubated at 15°C overnight. The cells were collected by centrifugation and stored at -80 °C.

The purification of the PrS₂-Bac7(1-35)

The frozen cells were thawed on ice and suspended into 20 ml of binding/washing buffer consisting of 20 mM Tris-HCl (pH8.0), 500 mM NaCl, and 20 mM imidazole-HCl (pH8.0). After breaking the cells by passing through a French press, the unbroken cells were removed by centrifugation at 14,000 rpm for 20 min. The supernatant after centrifugation was subjected to further centrifugation at 50,000 rpm for 30 min to isolate the soluble fraction. The supernatant was mixed with 1 ml of Ni-resin equilibrated in binding/wash buffer and the mixture was incubated for 1 hr at 4°C with shaking. The Ni-resin was washed twice with 10 ml of washing buffer, and PrS₂-Bac7(1-35) was eluted with elution buffer consisting of 20 mM Tris-HCl(pH8), 500 mM NaCl and 300 mM imidazole-HCl(pH8.0). After collecting the eluted fractions, the purity was examined by SDS-PAGE. After the PrS₂-Bac7(1-35) was dialyzed against 20 mM Tris-HCl(pH8.0) and 100 mM NaCl, the protein solution was concentrated to 3 mg/ml by using a protein concentrator (Millipore) and stored at -80 °C.

Cleavage by Factor Xa and identification of Bac7(1-35)

30 µg of PrS₂-Bac7(1-35) was digested with 4 µg of Factor Xa in 50 µl of Factor Xa buffer containing 20 mM Tris-HCl (pH 8.0), 50 mM NaCl and 2 mM CaCl₂, and the reaction mixture was incubated for 4 hr at 37°C. The solution was analyzed by 19% SDS-PAGE followed by Coomassie blue staining. The reaction mixture was analyzed by mass spectrometry.

***In vitro* translation inhibition assay**

PURExpress In Vitro Protein Synthesis kit and dihydrofolate reductase (DHFR) (New England BioLabs) was used as a positive control. The reaction mixture containing buffer A and buffer B supplied from NEB were mixed with 20 U of RNase inhibitor (Roche), linearized DHFR DNA (125 ng) and chemically synthesized Bac7(1-16) (GenScript) or purified PrS₂-Bac7(1-35) or water, and the reaction mixture was incubated at 37°C for 2 hr. The DHFR production was examined by SDS-PAGE followed by Coomassie blue staining.

Growth inhibition test when PrS₂-Bac7(1-35) is induced in *E. coli*

The *E. coli* strain, BL21(DE3) harboring either pColdPrS₂ or pColdPrS₂-Bac7(1-35) was grown in the M9-glucose medium. When O.D₆₀₀ reached 0.2, PrS₂ and PrS₂-Bac7(1-35) were induced by the addition of 1 mM IPTG. The culture medium in the absence of IPTG was also incubated as a negative control and the O.D₆₀₀ was monitored every 30 min.

The antimicrobial activity of purified Bac7(1-35) in *E. coli*

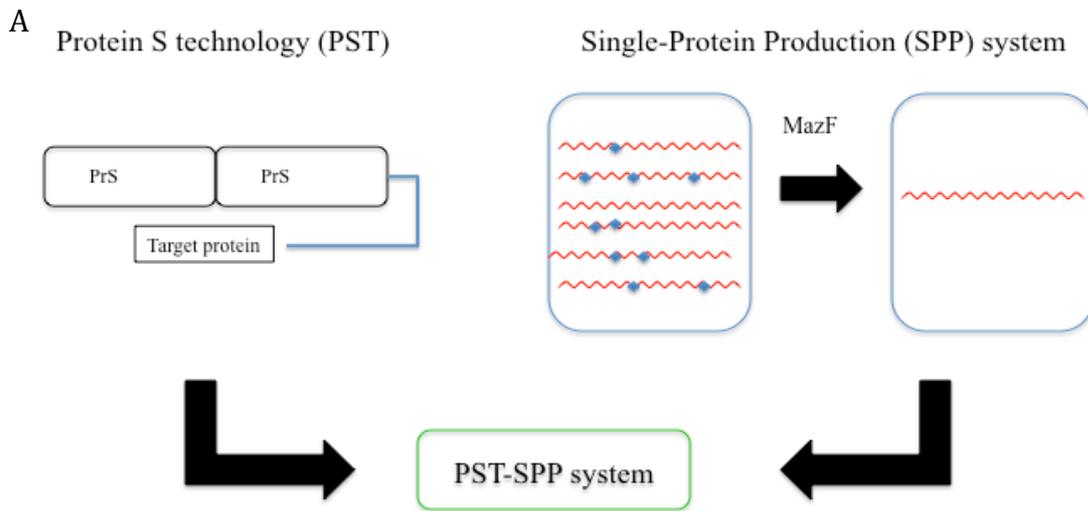
The *E. coli* strain BL21(DE3) was grown in the M9-glucose medium and when O.D₆₀₀ reached 0.2, the purified Bac7(1-35) was added at the final concentration of 2 µM. O.D₆₀₀ was monitored every 30 min.

Results

Development of expression systems for Bac7(1-35) in *E. coli*

In order to obtain Bac7(1-35), three expression systems (SUMO, PST, PST-SPP) were applied to reduce the toxicity of Bac7(1-35) and also to enhance its expression level in *E. coli*. To take advantage of the dormant state of cells created by inducing MazF (Fig. 8A), I synthesized the codon-optimized ACA-less Bac7(1-35) gene (Fig. 8B), which was then cloned into pColdPrS₂ vector. The resultant pColdPrS₂-Bac7(1-35) was co-expressed with pACYCmazF at 15°C (Fig. 8A).

At first, I compared PST and SUMO expression systems which showed that the PST system was able to enhance PrS₂-Bac7(1-35) production while the SUMO tag was not effective to express SUMO-Bac7(1-35) (Fig. 9A). However, expression level was relatively low, possibly because the protein S tag does not interfere with the toxicity of Bac7(1-35) completely. Next, the PST-SPP system was examined, and as shown in Figure 9A, I observed a slightly better expression level than when using the PST system. Furthermore, when I examined the localization of PrS₂-Bac7(1-35), it was found that PrS₂-Bac7(1-35) was exclusively localized in the soluble fraction (Fig. 9B).



B

```

AGGAGAATTCGTCCTCGCCCACCACGTTTGCCAAGGCCAAGGCCAAGGCCATTGCCATTCCCACGGCCTGGG
R R I R P R P P R L P R P R P R P L P F P R P G

ccaagggccaattccaaggccaactgccattecca
P R P I P R P L P F P

CGTCGCATTTCGTCCTCGCCCACCACGTTTACCCCGTCCCCGCCCCCGTCCCTTACCTTTTCCCAGCCCGGA
R R I R P R P P R L P R P R P R P L P F P R P G

CCACGCCCTATTCCACGCCCTTACCATAA
P R P I P R P L P *

```

Figure 8. A schematic description of the PST-SPP system. A. This figure is adapted from Ishida *et al.*, (Ishida and Inouye 2016). The N-terminal 88-residues of PrS, a major-coat protein from *Myxococcus xanthus*, was directly repeated (PrS₂) and cloned into pCold vector, generating pCold-PST. A codon-optimized ACA-less target gene can be expressed together with MazF, an ACA-specific endoribonuclease. B. Codon optimization for the Bac7(1-35) gene. The Bac7(1-35) gene (accession number: BC134776) (upper gene) was codon-optimized for *E. coli* expression system (lower gene) using codon optimization tool from IDT. The amino acid sequence was shown under the DNA sequence.

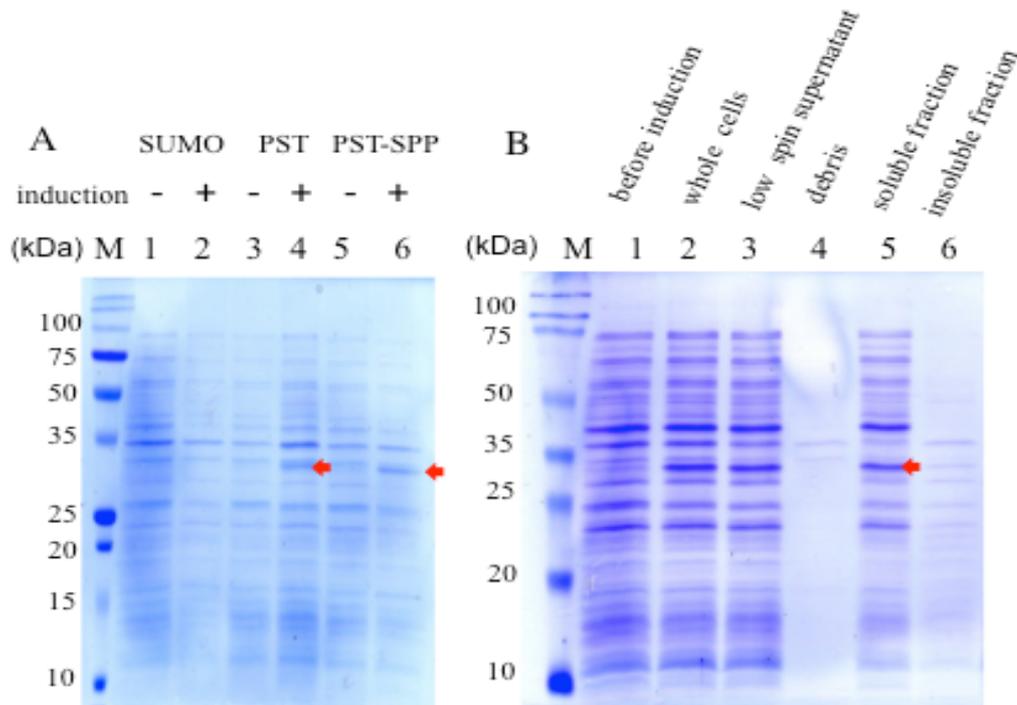


Figure 9. The production of SUMO-Bac7(1-35) and PrS₂-Bac7(1-35) in *E. coli* BL21(DE3).

This figure is adapted from Ishida *et al.*, (Ishida and Inouye 2016). A. BL21(DE3) harboring either pColdSUMO-Bac7(1-35) or pColdPrS₂-Bac7(1-35) was inoculated in 5 ml of LB medium. After O.D₆₀₀ reached 0.8, 1 mM IPTG was added, and the culture was incubated overnight at 16°C. The position of PrS₂-Bac7(1-35) is indicated by an arrow. BL21(DE3) co-transformed with pACYCmazF and pColdPrS₂-Bac7(1-35) (PST-SPP) was inoculated into 5 ml of LB medium and the culture was incubated at 37 °C. When O.D₆₀₀ reached 0.8, the culture was transferred onto ice for 5 min, followed by 1 hr incubation at 16 °C. After protein induction by the addition of 1 mM IPTG, the culture medium was incubated overnight at 16°C. The 1-ml culture was spun down and re-suspended into 100 µl of 1xSDS-PAGE loading buffer, and 10 µl was loaded into each lane.

Lane 1, SUMO-Bac7(1-35) before induction; lane 2, SUMO-Bac7(1-35) after induction; lane 3, PrS₂-Bac7(1-35) before induction; lane 4, PrS₂-Bac7(1-35) after induction; lane 5, SPP cells before induction; and lane 6, SPP cells after induction. B. The PrS₂-Bac7(1-35) production using the SPP system and its cellular localization. The ACA-less gene for Bac7(1-35) was expressed using pCold-PrS₂ vector together with pACYCmazF at 15°C. Lane 1, before 1mM IPTG was added. After inducing the PrS₂-Bac7(1-35) at 15°C overnight, 2-ml culture was collected by centrifugation and subsequently cellular fractionation was carried out. From each fraction, fractionated solution which corresponds to 100 µl cell culture was loaded into each lane ; lane 2, whole cells; lane 3, cell lysate after low speed centrifugation; lane 4, cell pellets after low speed centrifugation; lane 5, the soluble fraction, and lane 6, the insoluble fraction.

Purification of Bac7(1-35)

In order to purify Bac7(1-35), PrS₂-Bac7(1-35) was purified by Ni-NTA agarose column chromatography. After one-step purification, I was able to obtain 2.5 mg of highly pure PrS₂-Bac7(1-35) from 1 L LB medium. The existence of Bac7(1-35) was confirmed after Factor Xa cleavage reaction by SDS-PAGE analysis. As shown in Figure 10A, a small fragment appeared after the cleavage reaction. This cleavage reaction mixture was further analyzed by mass spectrometry. I observed 3 peaks at 4.2 kDa, 10.8 kDa and 21.6 kDa (Fig.10C). Since 4.2 kDa and 21.6 kDa correspond with the theoretical size of Bac7(1-35) and PrS₂, respectively, the small fragment obtained by Factor Xa treatment was confirmed as Bac7(1-35). The remaining 10.8 kDa was possibly due to the translation from internal Met residue in PrS₂. In order to further purify Bac7(1-35), ion-exchange column chromatography using SP sepharose was carried out. The pI of Bac7(1-35) is 13.0 while the pI of PrS₂ is 5.75, thus highly pure Bac7(1-35) was obtained after this second purification (Fig. 10B). The protein concentration of Bac7(1-35) was determined by using Bradford assay (Bradford 1976). When 360 µg of PrS₂-Bac7(1-35) was used, 55 µg of pure Bac7(1-35) was obtained indicating a recovery rate of about 90%. Therefore, 0.37 mg of highly pure Bac7(1-35) was expected from a 2.5 mg sample of PrS₂-Bac7(1-35).

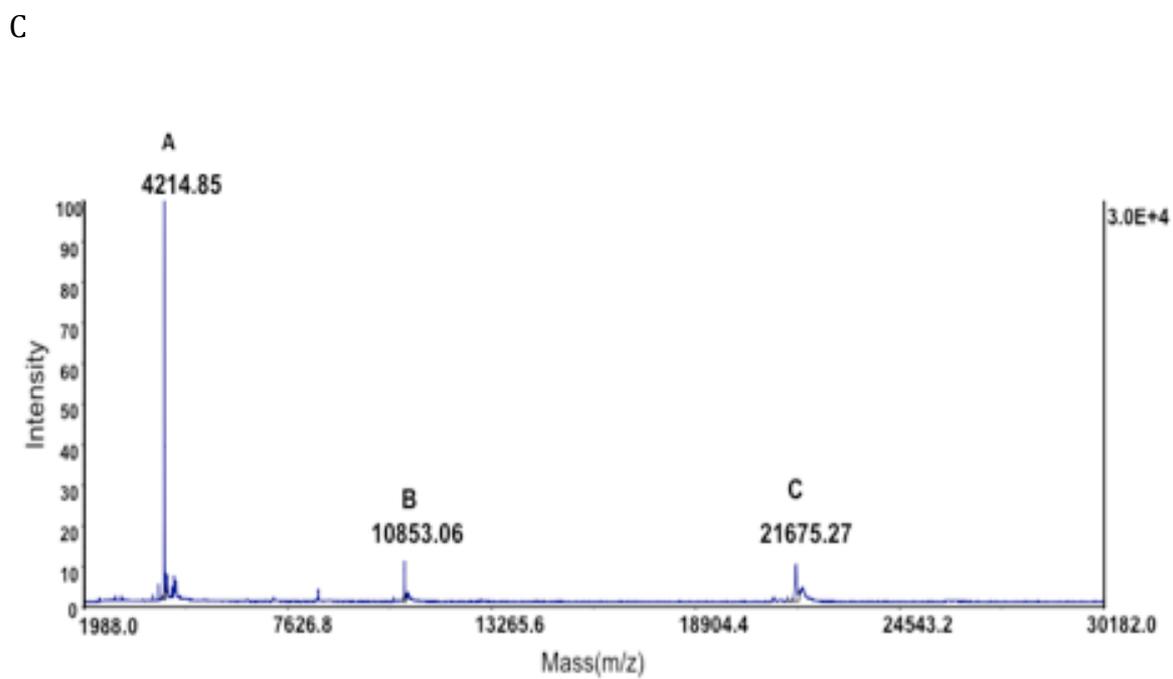
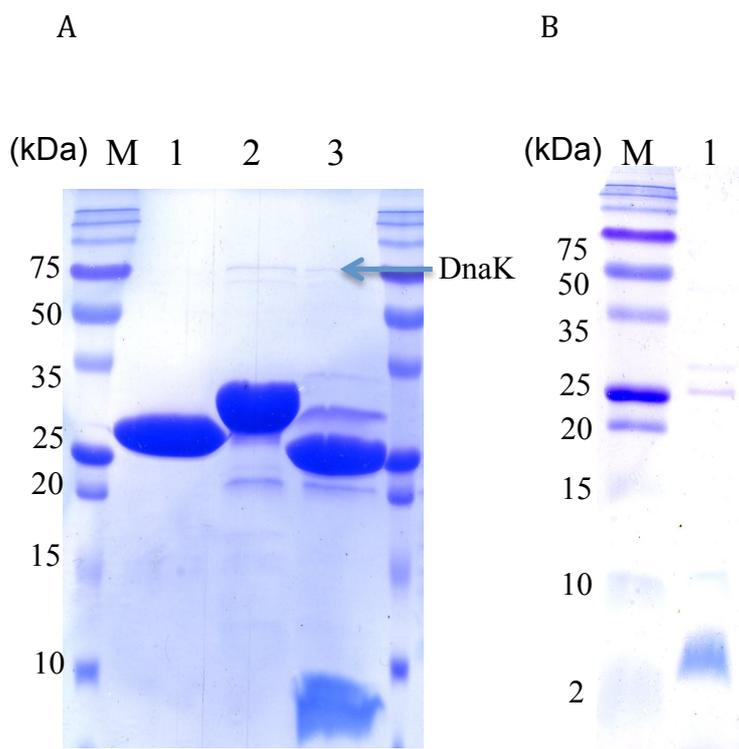


Figure legends are on the following page

Figure 10. Purification and identification of Bac7(1-35)

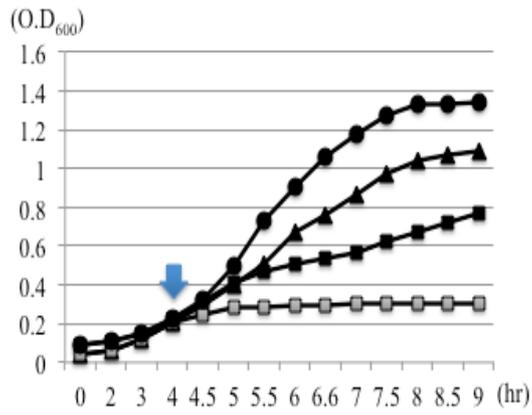
These figures are adapted from Ishida *et al.*, (Ishida and Inouye 2016). A. Cleavage of PrS₂-Bac7(1-35) by Factor Xa protease. 30 µg of purified PrS₂-Bac7(1-35) was mixed with 4 µl of Factor Xa protease in 50 µl of Factor Xa cleavage buffer, and the reaction mixture was incubated at 37°C for 4 hr. After incubation, 10 µl of the reaction mixture was subjected to 20 % SDS-PAGE. Lane 1, PrS₂ only; lane 2, PrS₂-Bac7(1-35) before cleavage by Factor Xa; lane 3, PrS₂-Bac7(1-35) after cleavage by Factor Xa; and lane M, molecular weight markers. DnaK, one of the target proteins of Bac7(1-35), was co-purified, and shown by an arrow. B. The purification of Bac7(1-35) after PrS₂-Bac7(1-35) treatment by Factor Xa followed by ion-exchange column chromatography. The concentration of purified Bac7(1-35) was determined by Bradford reagent (Bradford 1976), and 1.6 µg of Bac7(1-35) was analyzed by 20% SDS-PAGE. C. Identification of Bac7(1-35) by Mass spectrometry analysis. The purified PrS₂-Bac7(1-35) was digested by Factor Xa, and the digestion mixture was analyzed by MALDI-TOF Mass Spectrometry using a positive mid-mass linear mode from 2 to 30 kDa. Peak A represents Bac7(1-35) (4.21 kDa) , B represents PrS fragment (10.85 kDa), and C represents PrS₂ fragment (21.68 kDa), respectively.

Functional assay using PrS₂-Bac7(1-35) and Bac7(1-35)

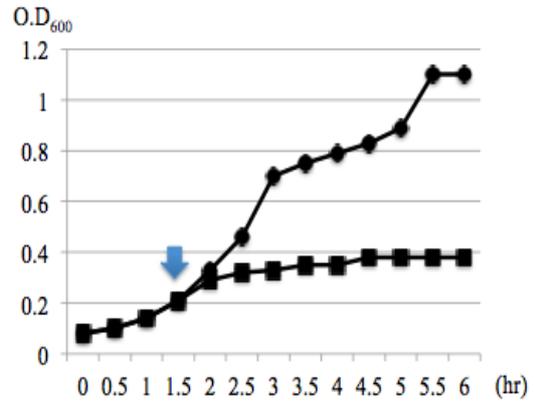
I examined the toxicity of pColdPrS₂-Bac7(1-35) using an M9-glucose medium in the presence or absence of 0.1 mM IPTG. Results showed that PrS₂ by itself did not severely affect cell growth while the induction of PrS₂-Bac7(1-35) completely inhibited cell growth (Fig.11A), indicating that PrS₂ did not severely affect the function of Bac7(1-35). I have also examined the toxicity of purified Bac7(1-35) in *E. coli* cells. As shown in Figure 11B, the bacterial growth was effectively inhibited at 2 μM after 1hr incubation.

In order to confirm PrS₂-Bac7(1-35) activity as a translation inhibitor, expression of dihydrofolate reductase (DHFR) by Cell-Free expression system was examined. As shown in Figure 11C, DHFR was well expressed under standard conditions, and it was confirmed that PrS₂ did not affect DHFR expression. Since chemical synthesis of 35-residue peptides is highly expensive, the N-terminal 16 residues of Bac7(1-35), Bac7(1-16) was used as a control (Guida et al. 2015; Seefeldt et al. 2016). As shown in Figure 11D, PrS₂-Bac7(1-35) as well as Bac7(1-16) were both able to inhibit DHFR synthesis. Therefore, PrS₂-Bac7(1-35) was shown to have antibacterial effect as a translation inhibitor.

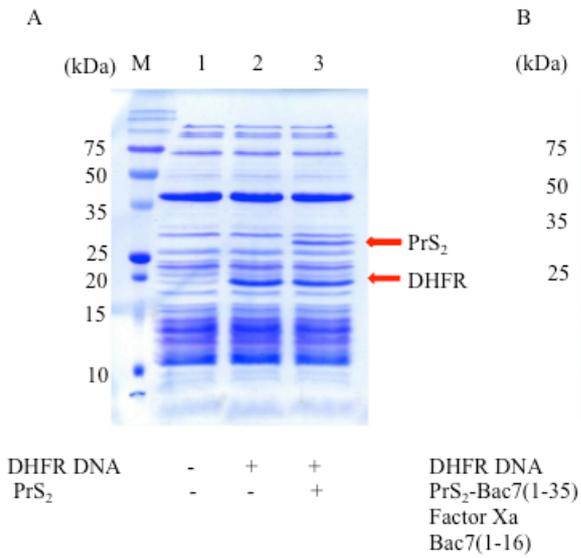
A



B



C



D

Figure legends are on the following page

Figure 11. Functional assays of Bac7(1-35)

These figures are adapted from Ishida *et al.*, (Ishida and Inouye 2016). A. The toxicity assay using pColdPrS₂-Bac7(1-35) in *E. coli*. BL21(DE3). pColdPrS₂ was used as a control vector. The cells were grown in M9 medium, and the target gene was induced by adding 1 mM IPTG when O.D₆₀₀ reached 0.2. The O.D₆₀₀ was monitored every 30 min. The circles represent pColdPrS₂ without IPTG, the triangles pColdPrS₂-Bac7(1-35) without IPTG, the closed squares represent pColdPrS₂ in the presence of IPTG and the open squares represent pColdPrS₂-Bac7(1-35) in the presence of IPTG. The induction time point is shown by an arrow. B. The toxicity of purified Bac7(1-35) in *E. coli*. BL21(DE3) *E. coli* cells were grown in M9-glucose medium. When O.D₆₀₀ reached 0.2, purified Bac7(1-35) was added at the final concentration of 2 μM into the medium, and O.D₆₀₀ was monitored every 30 min. The time of the addition of Bac7(1-35) is shown by an arrow. The circles and squares are in the absence and in the presence of Bac7(1-35), respectively.

C. DHFR expression in the presence or absence of PrS₂ using a cell-free protein expression system. Solutions A and B were mixed according to the manufacturer's protocol (NEB), and the mixture was incubated as a negative control (lane 1). DNA (10 ng/μl) of DHFR was added to mixture A and B and incubated as a positive control (lane 2). DNA of DHFR (10 ng/μl) together with PrS₂ (10 μM) were incubated with mixture A and B (lane 3). All reactions were incubated at 37 °C for 2 hr. After 2hr incubation, 2x SDS loading dye was added and the mixture was subjected to 17 % SDS-PAGE with molecular weight marker shown as lane M; D. DHFR expression in the presence of PrS₂-Bac7 with and without Factor Xa. Solutions A and B from the Cell-Free system (NEB) were mixed, the reaction mixture without DNA was used as a negative control (lane 1).

The reaction mixture in the presence of DHFR DNA (10 ng/ μ l) but in the absence of proteins was used as a positive control (lane 2). Lane 3, the reaction mixture containing DNA (10 ng/ μ l) together with 10 μ M of Bac7(1-16) as a control; lane 4, the reaction mixture containing DNA (10 ng/ μ l) together with 10 μ M of PrS₂-Bac7(1-35) after cleavage by Factor Xa; lane 5, the reaction mixture containing DNA (10 ng/ μ l) together with 10 μ M of PrS₂-Bac7 (1-35) without cleaving by Factor Xa, and lane M, molecular weight marker. All the reactions were carried out at 37 °C for 2hr. The reaction was stopped by 2x SDS loading buffer and the reaction was subjected to 15% SDS-PAGE, followed by Coomassie blue staining. The arrows on lane 2 and 3 indicate DHFR and PrS₂, respectively.

Discussion

The chemical synthesis for AMP production is highly expensive (\$500 for 20 amino acid peptide/mg), thus an economical method to produce large amounts of AMP is essential. AMP production in *E. coli* is cost effective but presents some challenges due to their antimicrobial activity. To suppress their toxicity, relatively large tags such as GST, MBP and SUMO may be used. Currently, for most AMP production, SUMO has been widely applied and many AMPs have been successfully expressed as functional forms (Li et al. 2011; Zhang et al. 2014; Zhang et al. 2015a). The SUMO tag has been shown to improve protein folding and solubility, and can be used for protein detection (Luan et al. 2014). Thus, I attempted to examine if the fusion of Bac7(1-35) to the C-terminal end of SUMO could suppress Bac7 toxicity, however, the production of the fusion protein was not detected. This indicates that the SUMO tag was not able to suppress the toxicity of Bac7(1-35), which is known to inhibit the function of 70S ribosomes (Mardirossian et al. 2014). Therefore, I next tried protein S as a fusion tag for Bac7(1-35). Protein S from *Myxococcus xanthus* is known to function as an intra-molecular chaperone without severely affecting the function of the protein fused to it, and has been applied for the expression of proteins which are insoluble and/or difficult to express (Kobayashi et al. 2009; Kobayashi et al. 2012). In the present study, I fused two 88-residue Protein S N-terminal domains (PrS₂) repeated in tandem to the N-terminal end of Bac7(1-35). In this PST-SPP system, an ACA-less gene encoding His₆-PrS₂ was used as an N-terminal tag for Bac7(1-35) to produce His₆-PrS₂-Bac7(1-35) (Fig. 8A). I also constructed the ACA-less His₆-SUMO-Bac7(1-35) system. However, His₆-PrS₂-Bac7(1-35) was expressed (Fig. 9A) while the expression of SUMO-Bac7(1-35) was not detected,

indicating that the SUMO tag was not able to suppress Bac7(1-35) toxicity even with the use of the SPP system. Notably, however, the expression of His₆-PrS₂-Bac7(1-35) was rather low, possibly because protein S fusion to Bac7(1-35) did not completely suppress the toxicity of Bac7(1-35). Using the SPP system, MazF cleaves at all ACA in mRNA while only the codon-optimized ACA-less gene for His₆-PrS₂-Bac7(1-35) remains intact. Therefore, upon induction of MazF, only His₆-PrS₂-Bac7(1-35) is produced in the cells (Fig. 8A) because in the SPP system, all the cellular mRNAs containing ACA sequences are digested by MazF. This results in cell growth arrest, allowing for the production of only the target protein from the ACA-less mRNA. In this manner, toxic proteins can still be produced as long as they do not inhibit ATP production and protein synthesis. Previously, I demonstrated that it is possible to completely replace all arginine residues in a protein with canavanine, a highly toxic analogue of arginine using the SPP system, since the incorporation of canavanine into any other cellular proteins is well suppressed (Suzuki et al. 2006; Mao et al. 2009; Ishida et al. 2013). In this study, I combine both PST and SPP technologies (PST-SPP technology) to successfully express His₆-PrS₂-Bac7(1-35). Since this technology works for AMP production despite the target protein being a protein synthesis inhibitor, the PST-SPP system may be applicable to the production of other AMPs in the future.

It has been shown that Bac7(1-35) blocks protein synthesis by inhibiting the 70S ribosomes (Mardirossian et al. 2014). To confirm the activity of biosynthesized Bac7(1-35), a cell-free protein expression system was applied together with a synthetic peptide Bac7(1-16) as a control for protein synthesis inhibitor (Seefeldt et al. 2016). As shown Figure 11, the production of DHFR by

the cell-free system was indeed inhibited by both His₆-PrS₂-Bac7(1-35) and Bac7(1-35). Bac7(1-35) was generated from His₆-PrS₂-Bac7(1-35) by Factor Xa treatment which resulted in a small amount of uncleaved His₆-PrS₂-Bac7(1-35) (Fig. 9C). Notably, the cleavage mixture effectively inhibited protein synthesis (Fig. 12). Since the minimum inhibitory concentration of Bac7(1-35) has been reported to be 0.5 μM (Benincasa et al. 2004), it is assumed that there was an excessive amount of Bac7(1-35) in the reaction mixture to inhibit protein synthesis. The PrS₂ tag is known not to interfere with its fusion partner, (Kobayashi et al. 2009); for example, PrS₂ fused at the N-terminal end of OmpR, a phosphor sensory protein, did not inhibit OmpR function at all (Kobayashi et al. 2009). Thus, it is not surprising to see that His₆-PrS₂-Bac7(1-35) still possesses antibacterial activity (Fig. 10A) despite the fact that the N-terminal part of Bac7(1-35) has been shown to be crucial for antimicrobial activity (Guida et al. 2015).

In this study, I demonstrated how to obtain 90% pure His₆-PrS₂-Bac7(1-35) by one-step purification using His-tag column chromatography. In addition, Bac7(1-35) was readily purified from His₆-PrS₂-Bac7(1-35) by treatment with Factor Xa followed by ion exchange column chromatography using SP-sepharose, since the pI value of His₆-PrS₂ is 5.75 while that of Bac7(1-35) is 13.0. Finally, highly pure Bac7(1-35) with approximately 90% yield was obtained. It is also important to note that since Bac7(1-35) does not have any aromatic residues, the protein concentration should be determined by ninhydrin or the Bradford assay (Bradford 1976). While the chemical synthesis of long AMPs such as Bac7(1-35) is highly expensive, the technology developed in this study will greatly reduce the cost of AMP production. Successful AMP production in *E. coli* in its active form will help us identify its target which is highly important since this leads to the development

of new antibiotics. In the case of Bac7(1-35), DnaK was co-purified with Bac7(1-35) and the identification of other targets co-purified with Bac7(1-35) is under investigation.

Section II

Replacement of Arg residues in MazFbs with canavanine alters its specificity

Introduction

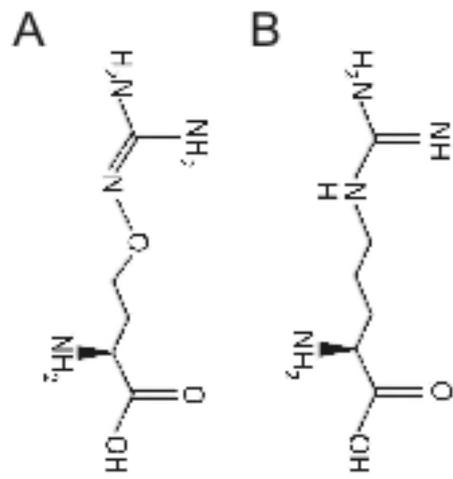
The amino acid analogues are non-natural amino acids which have structural similarity to natural amino acids (Hendrickson et al. 2004). The application of amino acid analogues offers significant potential as some studies have shown that replacement can enhance the activity of antimicrobial peptides (Wang et al. 2014), can be used as a fluorescent probe (Yang et al. 2016), and can change the protein function (Worst et al. 2015). The style of incorporation of amino acid analogues is classified into two groups, site-specific replacement and global replacement. In the case of site-specific replacement, the most common method to incorporate the amino acid analogue into a protein is by using amber codon together with its suppressing orthogonal tRNA (Ogawa et al. 2016; Yang et al. 2016). In this method, incorporating amino acid analogues into protein in both residue-specific and site-specific manners can be achieved. On the other hand, global replacement using amino acid analogues has been studied to engineer enzyme stability and activity. The most frequently used amino acid analogue is selenomethionine, which is a methionine analogue. Selenomethionine is greatly useful for X-ray structural studies because selenium and sulfur are one of the chalcogens and the substitution of methionine with selenomethionine may have only a limited effect on protein structure and function. Global replacement of selenomethionine using *E. coli* has been done using non-auxotrophic strains, however selenomethionine shows modest toxicity to the cells, resulting in an incorporation efficiency of less than 90%. On the other hand, utilizing endoribonuclease MazF together with an ACA-less gene enables us to produce proteins with high efficiency (Suzuki et al. 2006). Another common amino acid analogue is fluorinated amino acid, and some studies using fluorinated amino

acids have been done. For example, aromatic residues in lipase from *Candida antarctica* have been successfully replaced with fluorinated amino acids. Although global fluorination decreased lipase enzyme activity, the half life of lipase activity was prolonged, which is very useful for storage (Budisa et al. 2010). Another example uses fluorotryptophan to enhance the catalytic activity of glutathione transferase by changing its structural conformation (Parsons et al. 1998).

Amino acid analogues showing mild toxicity are likely to be incorporated into proteins in the *E. coli* expression system, however, some amino acid analogues, which show high toxicity to cells, are very difficult to incorporate into proteins with high efficiency. For example, proline consists of five rings while its analogue, azetidine-2-carboxylic acid, consists of four rings. This difference disrupts the protein structure creating high toxicity to the cells, and thereby incorporation of azetidine-2-carboxylic acid is very low (Bessonov et al. 2010).

Likewise, canavanine (Can), an Arg analogue found in leguminous plants, is also known as a very toxic amino acid analogue. The difference between Arg and Can is that the methylene bridge in Arg is changed to an oxa group in Can (Fig.12A). It has been studied using bacterial strains for more than 50 years (Richmond 1959). The mechanism of cell death by canavanine is possibly due to structural changes in the proteins due to abnormal protein networks caused by canavanine replacement (Schachtele et al. 1968). Studies to determine tRNA synthetase discrimination of Arg over Can show that Arg has over 100 times higher affinity to tRNA than Can (Igloi and Schiefermayr 2009).

In the present study, the Single-Protein Production (SPP) system was applied to overcome toxicity from Can(Suzuki et al. 2006). In addition, a high efficiency labeling system using amino acid as an inducer together with an auxotrophic strain was applied (Vaiphei et al. 2010). Lastly, Arg auxotrophic cells were combined with a dual inducible SPP system so that I was able to overcome the lower affinity of Can to tRNA. To demonstrate this high labeling system, I used one of the mRNA interferases from *Bacillus subtilis*, MazFbs which recognizes the sequence UACAU in mRNA (Park et al. 2013). The crystal structure of MazFbs has been solved (Fig. 12C) (Gogos et al. 2003).



C



Figure legends are in the page 69

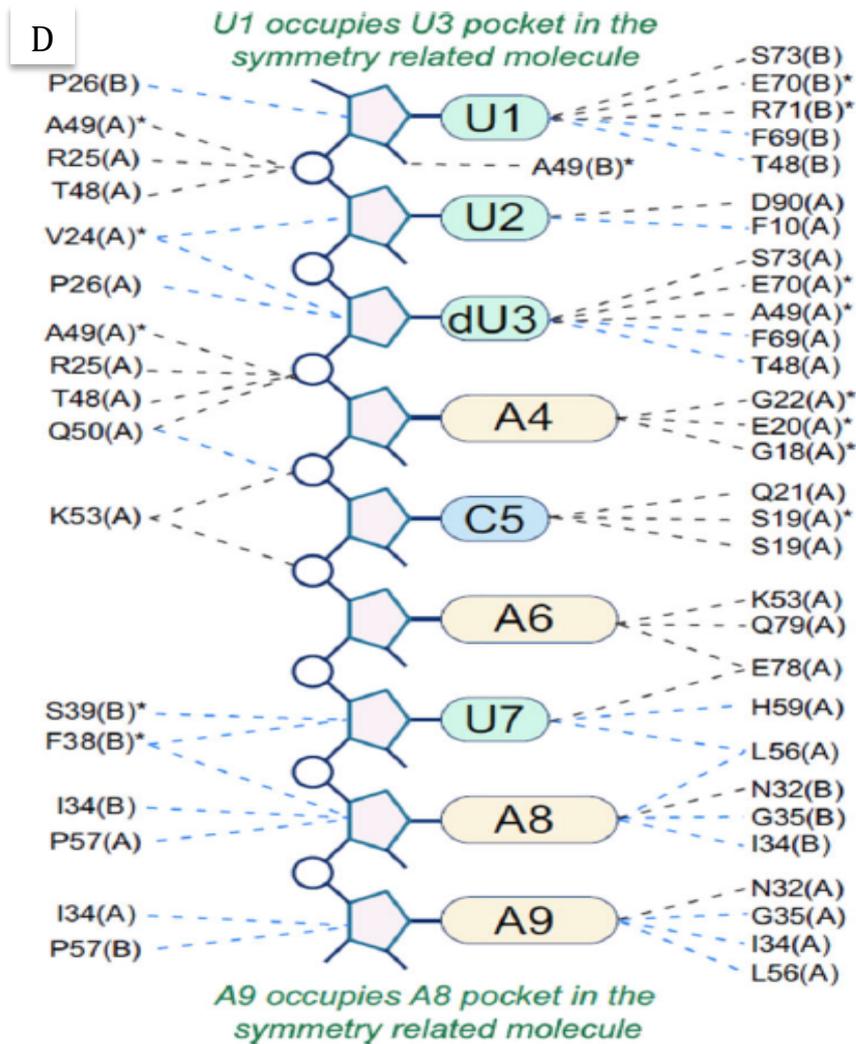


Figure 12 The structure of Canavanine (A), Arginine (B) Overview of complex of MazFbs with substrate RNA (C) The structural model was created by the use of Pymol and PDB ID; 4MDX. Each monomer is indicated in blue and green, respectively, and Arg and substrate RNA are shown in magenta and orange, respectively. (D) The atomic view of molecular interaction between MazFbs and substrate RNA. This figure is adapted from Shimanshu, *et al* (Shimanshu et al. 2013). The uncleavable substrate UUdUACAUA with the deoxyuridine at the 3rd position is shown with MazFbs residues. Interactions are indicated by dashed lines.

Materials and methods

Strain construction

E. coli BL21(DE3) ($\Delta argH\Delta trpC\Delta hisB$) was constructed from *E. coli* BL21(DE3) ($\Delta trpC\Delta hisB$) by P1 transduction using the $\Delta argH$ strain from the Keio collection.

At first, P1 phage lysate using *E. coli* strain, $\Delta argH$ (JW3932) from the Keio-collection was prepared (Baba et al. 2006). Next, BL21(DE3)($\Delta hisB\Delta trpC$) (Vaiphei et al. 2010) was grown in 10 ml of LB medium supplemented with 5 mM $CaCl_2$, and when $O.D_{600}$ reached 0.5, 100 μ l of P1 phage lysate was added. After 30 min incubation at 30°C, the transduction was stopped by the addition of 100 mM sodium citrate (pH5.5). The cells were then collected and washed with 200 mM sodium citrate (pH 5.5). Colonies from LB-Km plates (Km; 50 mg/L) were examined for gene deletion by using M9-glucose medium supplemented with amino acids and also by polymerase chain reactions (PCR).

Plasmid construction

The gene coding for MazFbs with a C-terminal hexa Histidine tag was synthesized (Genscript). The gene was designed for optimal codon usage in *E. coli*, and there are no ACA sequences in the gene. The gene was cloned into pColdIII (SP-4), generating pColdIII*mazFbs* (Suzuki et al. 2006; Vaiphei et al. 2010).

Protein expression and purification

The BL21(DE3) ($\Delta argH\Delta trpC\Delta hisB$) cells were transformed with pColdIII*MazFbs* together with pACYC*mazF*(ΔH) (Vaiphei et al. 2010), and grown

in a 1-liter culture of M9-glucose medium in the presence of Arg (20 mg/L), His (20 mg/L), and Trp (20 mg/L) at 37 °C. When the A_{600} value reached 0.5, the culture was chilled in an ice-water bath for 5 min and incubated at 15 °C for 1 h to acclimate the cells to cold shock conditions. Cells were harvested and washed twice with M9 medium. The cells were re-suspended in 50 ml of M9-glucose medium containing Arg (20 mg/L) and Trp (20 mg/L) but without His. Isopropyl β -D-1-thiogalactopyranoside (IPTG; 0.5 mM) was added to induce only the expression of MazF(Δ H) followed by an additional 2 hr incubation at 15 °C. Cells were harvested and washed twice with M9 medium. The cells were re-suspended in 50 ml of M9-glucose medium containing His (20 mg/L), and Trp (20 mg/L), Can (100 mg/L; Sigma) or Arg (20 mg/ml) and IPTG (0.5 mM) to incorporate Can or Arg into MazFbs, respectively. The cell culture was incubated at 15 °C for an additional 24 h to induce MazFbs(*can*) and MazFbs(*arg*). Cells were collected by centrifugation and subjected to SDS-PAGE followed by Coomassie Blue staining. MazFbs(*arg*) and MazFbs(*can*) were purified from BL21(DE3) (Δ *argH* Δ *trpC* Δ *hisB*) cells carrying pColdIII*MazFbs* with the use of nickel-nitrilotriacetic acid resin (Qiagen) following the manufacturer's protocol. MazFbs(*can*) and MazFbs(*arg*) were further purified by ion-exchange chromatography using DEAE-Sepharose (GE Healthcare).

Circular Dichroism (CD) Analysis

CD analysis was carried out using an Aviv model 62DS spectropolarimeter (Aviv Associates, Inc., Lakewood, NJ). Spectra were recorded in 2.0-nm steps between 260 and 200 nm at 4 °C with an integration time of 4 s at each

wavelength, and the base line was corrected against buffer alone. The α -helical and β -sheet contents were calculated using the K2D2 program (<http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d2/>). Protein melting was examined at 208 nm with increasing temperatures, from 0 to 90 °C, in 0.3 °C steps. Protein solutions were equilibrated at each temperature point for 1.5 min, and the temperature was increased at the average rate of 0.1 °C/min. The path length of the cell used was 0.1 cm, and all measurements were carried out in 10 mM Tris-HCl (pH 7.8).

Cleavage of MS2 Phage RNA by MazFbs(can)

MS2 RNA (70 nM) was incubated with MazFbs mutants (0.5 μ M) for 1, 5, 10 and 30 min. The reaction was stopped by adding 2X stop solution (95% formamide, 10 mM EDTA, 0.1% xylene cyanol and 0.1% BPB). After incubating the samples in boiling water for 2 min, the samples were placed on ice for 5 min. The RNA was analyzed by 1% agarose gel electrophoresis followed by ethidium bromide staining.

Primer Extension Analysis

The annealing mixture consisting of 70 nM of MS2 RNA digested by MazFbs mutants and 0.1 μ M labeled primer was incubated in boiling water for 1 min and placed on ice for 5 min. Then to the annealing mixture, the RT mixture containing 1U RNase inhibitor (Sigma), 1 mM dNTPs, 1 U AMV Reverse Transcriptase (Sigma) and 1 μ M CspA, an RNA chaperone, was added and the final mixture was incubated for 1 hr at 45°C. The reaction was stopped with the addition of 2X stop solution. The sample was then incubated in a boiling water bath for 2 min and placed on ice for 5 min.

For the RNA sequencing ladder, 0.14 μM of MS2 RNA was incubated with RT mixture containing 0.1 μM labeled primer, 1mM dNTPs, 0.5 mM ddG or ddA or ddC or ddT, and 1U AMV Reverse Transcriptase. The reaction mixture was incubated for 30 min at 45°C. The reaction was stopped by addition of the stop solution, followed by incubating in a boiling water bath for 2 min. After placing the reaction tubes on ice for 5 min, the synthesized DNA was analyzed by sequencing gel (6% acrylamide (29:1), 8M Urea, 2xTBE) by running at 45W for 1.5 hr. The DNA was visualized by autoradiography.

Cleavage of Synthetic RNA by MazFbs(*arg*) and MazFbs(*can*)

Four of 13-base ribonucleotides (CUCXUACAUAUCA), where the 4th base (*X*) was A, U, G, or C were synthesized (IDT). Three additional 13-base RNA ribonucleotides (CUCUUAUAUYUCA) were synthesized, where the *Y* position was replaced with U, G, or C. These ribonucleotides were used as substrates. The substrates were labeled (0.2 μM) with [γ -³²P]ATP using T4 kinase (New England Biolabs) and incubated with 0.1 μM MazFbs(*arg*) or MazFbs(*can*) for 10 min at 37 °C in a reaction mixture (10 μl) of 10 mM Tris-HCl (pH 7.8) containing 0.2 μl of the Protector RNase inhibitor. The reactions were stopped with 2X stop solution. To analyze the cleavage of the synthetic RNAs, the products were analyzed by electrophoresis on a 20% polyacrylamide gel containing 8M urea with a molecular mass ladder.

Kinetics Analysis

A 13-base ribonucleotide (CUCAUACAUAUCA) was used as a substrate. The substrate in various concentrations (0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0 μM)

was incubated with 1 nM MazFbs(*arg*) or 5 nM MazFbs(*can*) in a reaction mixture (10 μ l) of 10 mM Tris-HCl (pH 7.8) containing 0.2 μ l of the Protector RNase inhibitor. The reaction with MazFbs(*arg*) was incubated for 5, 10, 15, and 20 min. The reaction with MazFbs(*can*) was incubated for 30, 60, 90, and 120 min. The reaction was stopped by the use of 2X stop solution, and the sample mixtures were incubated at 90 °C for 5 min prior to electrophoresis on a 20% polyacrylamide gel containing 8 M urea. The cleavage products were analyzed by ImageJ software.

Competitive Analysis

A 13-base ribonucleotide (CUCUUACAUAUCA) was used as substrate, and three other 13-base ribonucleotides (CUCUUACA**UU**UCA, CUCUUACA**UC**UCA, and CUCUUACA**UG**UCA) in which only the 10th base is different from the substrate (shown in *bold*) were used to examine whether these ribonucleotides are able to inhibit the cleavage of the substrate. The concentration of the substrate analogues was fixed at 1 μ M, whereas the substrate concentrations were at 1.0 and 4.0 μ M. The substrate with and without the substrate analogues in a 10- μ l reaction mixture containing 10 mM Tris-HCl (pH 7.8) and 0.2 μ l of Protector RNase inhibitor was incubated with 5 nM MazFbs(*can*) at 37 °C for 30, 60, 90, and 120 min, respectively. The reaction was stopped by the use of 2X stop solution, and the reaction mixtures were incubated at 90 °C for 5 min prior to electrophoresis on a 20% polyacrylamide gel containing 8 M urea. The cleavage products were analyzed by ImageJ.

Results

Construction of Arg auxotrophic strain by P1 transduction

In order to combine an auxotrophic strain with the dual inducible SPP system (Fig. 4), *ΔargH* was introduced into the BL21(DE3) *ΔhisBΔtrpC* strain (Datsenko and Wanner 2000; Vaiphei et al. 2010). The resultant strain, BL21(DE3) *ΔhisBΔtrpC ΔargH* strain does not grow without His, Trp and Arg in the medium.

Production of MazFbs (can) using dual inducible SPP system

In order to produce MazFbs(can) by suppressing the toxicity from Can, the Single-Protein Production (SPP) system was applied (Suzuki et al. 2006). The ACA-less MazFbs gene (Fig.13A) was cloned into the pColdIII vector, which has only MNHKV residues, the translation enhancing element. pColdIII*mazFbs* was transformed into BL21(DE3) *ΔhisBΔtrpCΔargH* cells which contains pACYC*mazF(ΔH)*, harboring a lac-inducible *mazF* gene which has mutations in G27L and H28R (Vaiphei et al. 2010). The cells were grown in M9 glucose medium supplemented with His, Trp and Arg. When O.D.₆₀₀ reached 0.5, the culture was placed on ice for 5 min followed by 45 min at 16 °C for cold shock treatment. The cells were collected and washed by 1xM9 buffer to remove His from the medium, and re-suspended into M9 glucose supplemented with Trp and Arg. Only MazF(ΔH) was induced by addition of 0.5 mM IPTG at this time. This is because MazFbs contains two His residues in addition to the C-terminal 6x His residues, thereby, MazFbs(*can*) was not induced in the absence of His in the medium. After 2 hr incubation, the cells were collected again and washed by 1xM9 buffer to remove Arg from the medium, and finally re-suspended into M9 glucose supplemented with His, Trp and Can and 0.5 mM IPTG to induce

MazFbs(*can*) production. Since MazF(Δ H) induction places the cells in a dormant state, Can was only used for MazFbs(*can*) production. In order to produce MazF(*arg*), Arg (20 mg/L) was added into the medium instead of Can.

A

```

1 atgattgtgaaacgtggcagatgtgtattttgcggatctgagcccgggtggggcagcagcagggcgccgctgctccggtgctggtgatt 90
1 M I V K R G D V Y F A D L S P V V G S E Q G G V R P V L V I 30

91 cagaacgatattggcaaccggttttagcccgaccgagattgtggcggcattaccgcgagattcagaagcgaactgcccaccatgtg 180
31 Q N D I G N R F S P T A I V A A I T A Q I Q K A K L P T H V 60

181 gaaattgatgcgaaacgttatggcctttgaacgtgatagcgtgattctgctggagcagattcgtaccattgataagcagcgtctgaccgat 270
61 E I D A K R Y G F E R D S V I L L E Q I R T I D K Q R L T D 90

271 aaaattaccatctggatgatgaaatgatggataaagtggatgaagcgtgcagattagcctggcgtgattgattttcatcatcatcat 360
91 K I T H L D D E M M D K V D E A L Q I S L A L I D F H H H H 120

361 catcattaa 369
121 H H * 122

```

B

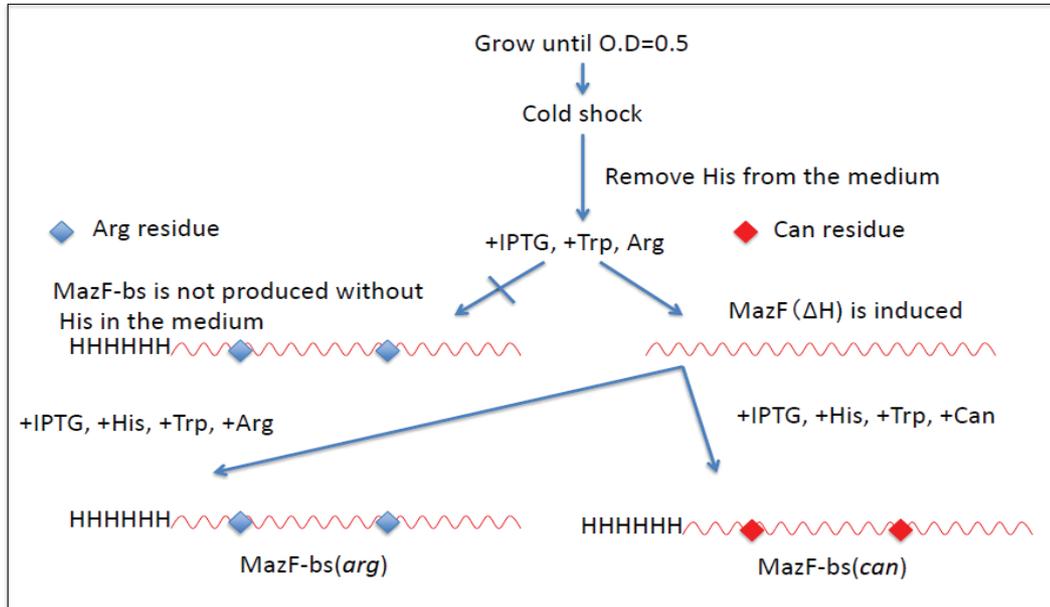


Figure legends are in the next page.

Figure 13. Schematic procedures for the production of MazFbs(*arg*) and MazFbs(*can*). These figures are adapted from Ishida *et al.*, (Ishida *et al.* 2013) A, The DNA sequence and amino acid sequence of MazFbs. The gene is codon-optimized for *E. coli* and designed to be ACA-less. The amino acid sequence of MazFbs is shown under the DNA sequence. B, The dual inducible Single-Protein-Production (SPP) system. The BL21(DE3) ($\Delta argH\Delta trpC\Delta hisB$) cells were transformed with pACYC*mazF*(ΔH) together with pColdIII*MazFbs* and grown in a 1-liter culture of M9-glucose medium supplemented with Arg (20 μ g/ml), His (20 μ g/ml), and Trp (20 μ g/ml) at 37°C. When the A_{600} value reached 0.5, the culture was placed on ice for 5 min and incubated at 15°C for 1 hr for cold shock treatment. Cells were collected and washed twice with 1xM9 medium. The cells were re-suspended in 50 ml of M9-glucose medium supplemented with Arg (20 μ g/ml) and Trp (20 μ g/ml) but without His. Only the expression of MazF(ΔH) was induced by addition of Isopropyl β -D-1-thiogalactopyranoside (IPTG; 0.5 mM) followed by an additional 2 hr incubation at 15°C. Cells were collected and washed twice with 1xM9 medium. The cells were re-suspended in 50 ml of M9-glucose medium supplemented with His (20 μ g/ml), and Trp (20 μ g/ml), Can (100 mg/L) and IPTG (0.5 mM) to incorporate Can into MazFbs. The culture was incubated at 15°C for an additional 24 hr to induce MazFbs(*can*). In order to produce MazF(*arg*), Arg (20 mg/L) was added into the medium instead of Can.

Identification of labeling efficiency after incorporating Can into MazFbs

After overnight incubation using the SPP system in the presence of Can, a new band was induced at 14 kDa position, and this protein, MazFbs(*can*) was purified by Ni-NTA affinity chromatography followed by DEAE ion-exchange column chromatography (Fig. 14A) because the pI of MazFbs(*arg*) is 6.34 while the pI of MazFbs(*can*) is 5.86 by calculation. There are seven Arg residues in MazFbs so the molecular mass should differ by 13.8 Da ($1.97 \text{ Da} \times 7$) to that of MazFbs(*arg*) if all seven Arg residues were replaced with Can. It was shown that MazFbs(*can*) was 13.4 Da larger than MazFbs(*arg*) by mass spectrometry analysis (Fig. 14B), indicating that 97% of Arg residues were replaced with Can .

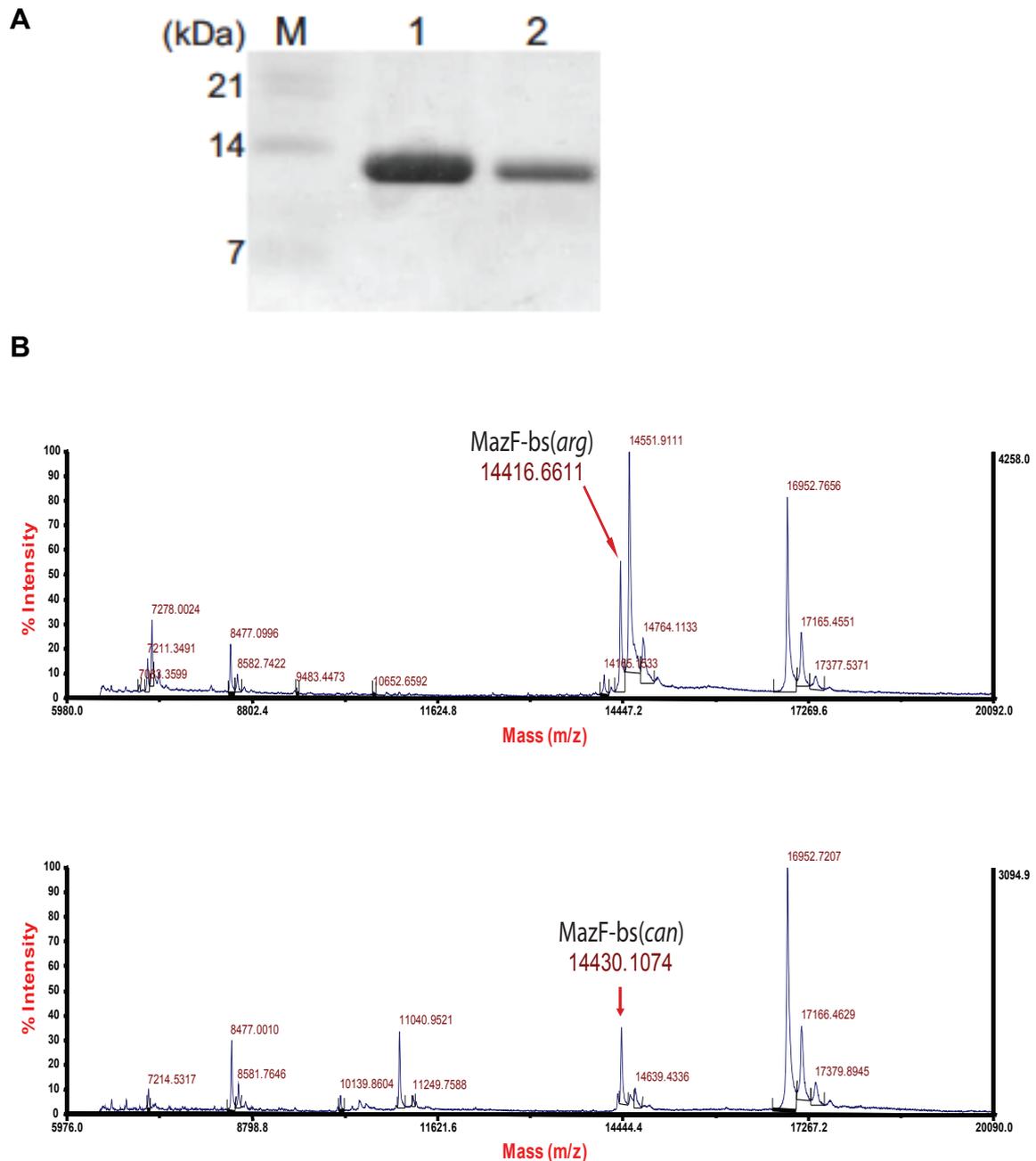


Figure 14. Protein identification of MazFbs(*arg*) and MazF(*can*). (A) After His-tag purification, 1 μ g of MazFbs(*arg*) and 0.5 μ g MazF(*can*) were analyzed by SDS-PAGE gel. (B) Total mass measurement by MALDI-TO of MazFbs(*arg*) and MazFbs(*can*), respectively. As an internal reference, apomyoglobin (8,477 Da and 16,952 Da) was used. This figure is adapted from Ishida *et al.*, (Ishida *et al.* 2013).

RNA cleavage specificity was altered in MazFbs(*can*)

The endoribonuclease activity of MazFbs(*can*) was examined using 3.5-kb MS2 phage RNA as a substrate (Park et al. 2011). To our surprise, the cleavage patterns were found to be very different between the two enzymes (Fig. 15A). Therefore, I attempted to determine the exact cleavage site by *in vitro* primer extension. As a result, U[^]ACAU sites in MS2 RNA cleaved by MazFbs(*arg*) were detected ([^] indicates the cleavage position), while MazFbs(*can*) cleaves the MS2 RNA at U[^]ACAU sites only when these sites contain one extra A residue at the 3' end (Fig. 15B,C,D,E,F). This indicates that the activity of MazFbs(*can*) was altered to have a higher RNA-cleavage specificity, from a five-base to a six-base recognition sequence. To further confirm this, 13-base RNA substrates covering all possible 7-base sequences having an extra base at both sides of UACAU were synthesized and results showed that MazFbs(*can*) specifically cleaves at U[^]ACAUA (Fig. 16A, B). Lanes 2 [MazFbs(*can*)] and 7 [MazFbs(*arg*)] in Figure 5A show an extra band corresponding to the product cleaved after the first C residue in addition to the cleaved product after the fifth U residue, C[^]UCUU[^]ACAUAUCA ([^] indicates the cleavage sites). No extra cleavage products are observed with the three other ribonucleotides (CUCAUACAUAUCA, CUCGUACAUAUCA, CUCCUACAUAUCA) for both MazFbs(*can*) and MazFbs(*arg*) (bases which are replaced are shown in red). Furthermore, lane 2 in Figure 16B using MazFbs(*can*) with CUCUUACAUAUCA shows an extra cleavage product (cleaved after the first C residue) in addition to the product after the fifth U residue. Lanes 7, 8 and 10 in Figure 16B using MazFbs(*arg*) also show an extra cleavage product corresponding to C[^]UCUUACAUUCA, C[^]UCUUACAAUCA, C[^]UCUUACACUCA, respectively. It is still unknown why

these substrates were cleaved by MazFbs(*arg*) and MazFbs(*can*) between U and C at the 3' end.

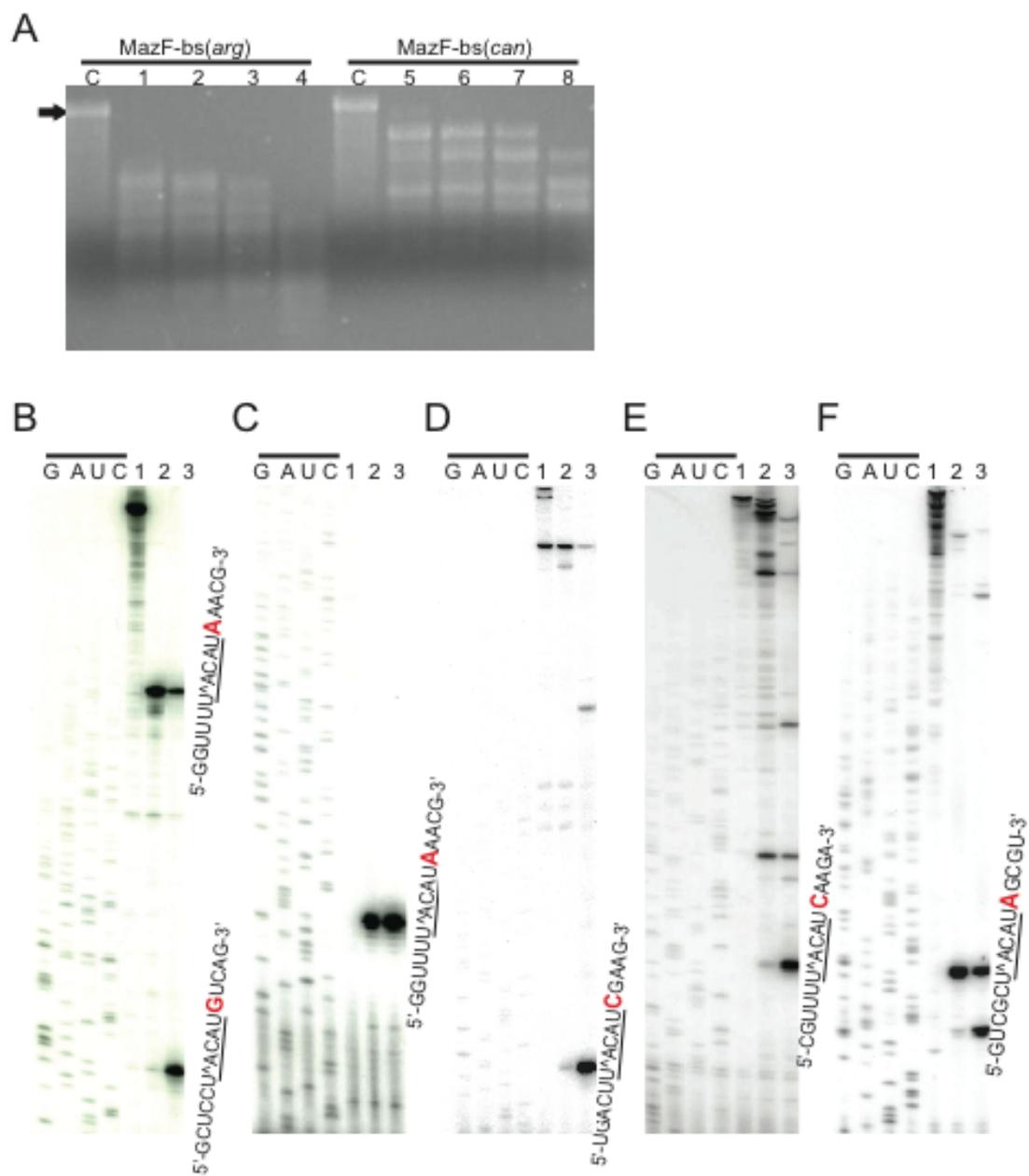


Figure legends in the next page

Figure 15 Alteration of cleavage specificity from 5-base to 6-base recognition in MazFbs(*can*). These figures are adapted from Ishida *et al.*, (Ishida *et al.* 2013). A, Endoribonuclease activity using MazFbs(*arg*) and MazFbs(*can*) with MS2 phage RNA. Lane C represents negative control in the absence of protein, lane 2-4 and 6-8 are the reaction in which MS2 RNA was incubated with MazFbs(*arg*) or MazFbs(*can*) for 1, 5, 10, 30 min, respectively [Lanes 1-4; MazFbs(*arg*), 5-8; MazFbs(*can*)]. A black arrow indicates the full length (3.5 kb) of MS2 phage RNA. B-F, Analysis of MazFbs(*can*) cleavage positions in MS2 phage RNA by *in vitro* primer extension. Each panel represents different UACAU sites in MS2 RNA. Lane 1; MS2 RNA was incubated in the presence of purified CspA as a negative control. Lanes 2 and 3; MS2 RNA was incubated with MazFbs(*can*) and MazFbs(*arg*) in the presence of CspA, an RNA chaperone, respectively. G, A, U, and C with an upper black bar indicate the sequence ladder for each reaction primer. Ribonucleotide sequences in each panel (B-F) represent the cleaved sequences for MazFbs(*can*) and MazFbs(*arg*), respectively. The U^ACAU sequences with an under bar indicate the cleavage sites by MazFbs(*arg*). The extra A at the 3' end is shown in red and is required for the cleavage by MazFbs(*can*).

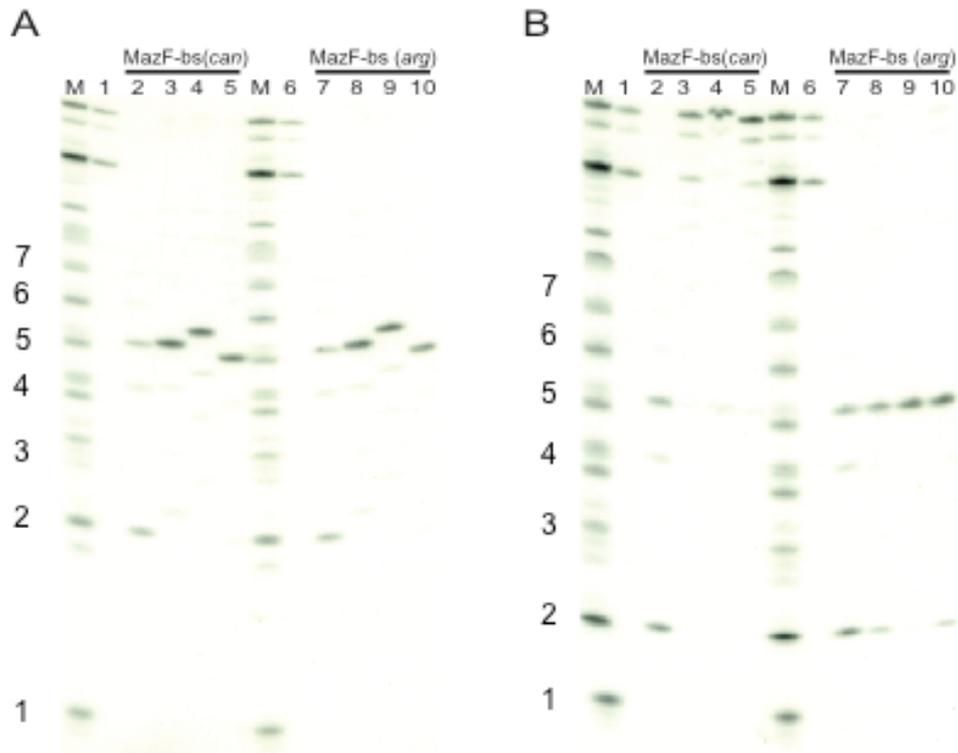


Figure 16. Identification of a change of RNA cleavage specificity in MazFbs(*can*). These figures are adapted from Ishida *et al.*, (Ishida *et al.* 2013). A, The synthetic ribonucleotides consisting of 13-bases ($C_1U_2C_3X_4U_5A_6C_7A_8U_9A_{10}U_{11}C_{12}A_{13}$), where the 4th base (X) was U, A, G, or C, were incubated with MazFbs(*can*) or MazFbs(*arg*) (lanes 2-5, and 7-10, respectively). Lanes 1 and 6 represent control reactions in the absence of MazFbs. B, Synthetic 13-base ribonucleotides ($C_1U_2C_3U_4U_5A_6C_7A_8U_9Y_{10}U_{11}C_{12}A_{13}$), where the Y position was A, U, G or C, were incubated with MazFbs(*can*) or MazFbs(*arg*) (lanes 2-5, and 7-10, respectively) Lanes 1 and 6 represent control reactions in the absence of MazFbs. The numbers next to M indicated the length of RNA were generated by hydrolysis with NaOH.

Kinetic study

The K_m value and the K_{cat}/K_m value of MazFbs(*arg*) were determined to be $2.0 \pm 0.2 \mu\text{M}$ and $1.0 \pm 0.2 \times 10^{-2}$, respectively using UACAUA as a substrate. Although the K_m value for MazFbs(*can*) is almost identical to that of MazFbs(*arg*), the K_{cat}/K_m value of MazFbs(*can*) is approximately 5% of that of MazFbs(*arg*) (see Table 1). Since MazFbs(*can*) became six-base specific, cleaving at U[^]AACAUA, but not UACAUG, UACAUC and UACAUU (Fig.16), I next examined if the cleavage of U[^]AACAUA is inhibited by these substrate analogues having different bases at the sixth position (UACAUU, UACAUC and UACAUG). Results showed that there was no inhibition of the cleavage reaction by UACAUG, UACAUC, and UACAUU, indicating the A residue at the sixth position plays a critical role for the substrate binding to the enzyme (Table 2). It was found that there was no inhibition of the cleavage reaction by UACAUG, UACAUC, and UACAUU, even if the inhibitor-to-substrate ratio increased (Table 2).

Table 1. Kinetic constants for MazFbs(*arg*) and MazFbs(*can*)

	V _{max} ($\mu\text{M}/\text{min}$)	K _m (μM)	K _{cat} (min^{-1})	K _{cat} /K _m ($\mu\text{M}^{-1} \cdot \text{min}^{-1}$)
MazFbs(<i>arg</i>)	$4.2 \pm 1.1 \times 10^{-2}$	2.0 ± 0.2	$2.2 \pm 0.4 \times 10^{-2}$	$1.0 \pm 0.2 \times 10^{-2}$
MazFbs(<i>can</i>)	$8.4 \pm 1.2 \times 10^{-3}$	1.8 ± 0.3	$8.4 \pm 1.2 \times 10^{-4}$	$5.0 \pm 1.0 \times 10^{-4}$

Table 2. Relative endoribonuclease activity of MazFbs(*can*) using CUCUUACAUAUCA as a cleavable substrate in the presence and the absence of substrate analogues having different bases at the sixth position (CUCUUACAUCUCA, CUCUUACAUGUCA and CUCUUACAUUUCA)

	CUCUUACAUAU CA only	+CUCUUACAUCU CA	+CUCUUACAUGU CA	+CUCUUACAUUU CA
CUCUUACAUAUCA : substrate analogue=1:1	1.0	1.1	1.0	1.1
CUCUUACAUAUCA :substrate analogue=1:4	1.0	1.1	1.1	1.2

Structural analysis of MazFbs(*can*) by circular dichroism (CD) spectroscopy The secondary structures of MazFbs(*arg*) and MazFbs(*can*) were analyzed by CD spectroscopy. Minimum peaks were observed at 208 and 222 nm from MazFbs(*arg*), which indicate α -helical structures. MazFbs(*can*) showed a minimum peak at 208 nm, however, the value was higher than that for MazFbs(*arg*), while the signal at 222 nm for MazFbs(*can*) was lower than that for MazFbs(*arg*) (Fig. 17A). We calculated α -helix and β -sheet content using K2D2 program which showed that the β -sheet content decreased from 39.5 to 37.2% while the α -helix content of MazFbs(*can*) slightly increased from 27.5 to 29.7%. Therefore, MazFbs(*can*) is likely folded in a very similar manner as MazFbs(*arg*), however the replacement of Arg with Can appears to affect the α -helical structures. Next, the thermal stability was examined for both proteins between 4 and 90°C by measuring the change in ellipticity at 222 nm in the CD spectra. Notably, the melting temperature of MazFbs(*can*) was 59 °C, which is lower by approximately 4°C than that of MazFbs(*arg*) (Fig. 17B). There is a salt bridge between Arg-87 and Glu-20 and hydrogen bonds between Arg-5 and Ala-112 that have been shown to stabilize the dimer formation (Gogos et al. 2003). Although both Arg and Can contain a guanidino group, the replacement of the methylene group in Arg with oxygen in Can results in the reduction of the pKa value from 12.48 to 7.01 (Rosenthal et al. 1989). Therefore, the salt bridge in MazFbs(*can*) is likely to be weaker when all the Arg residues are replaced with Can, resulting in a less thermo-stable protein.

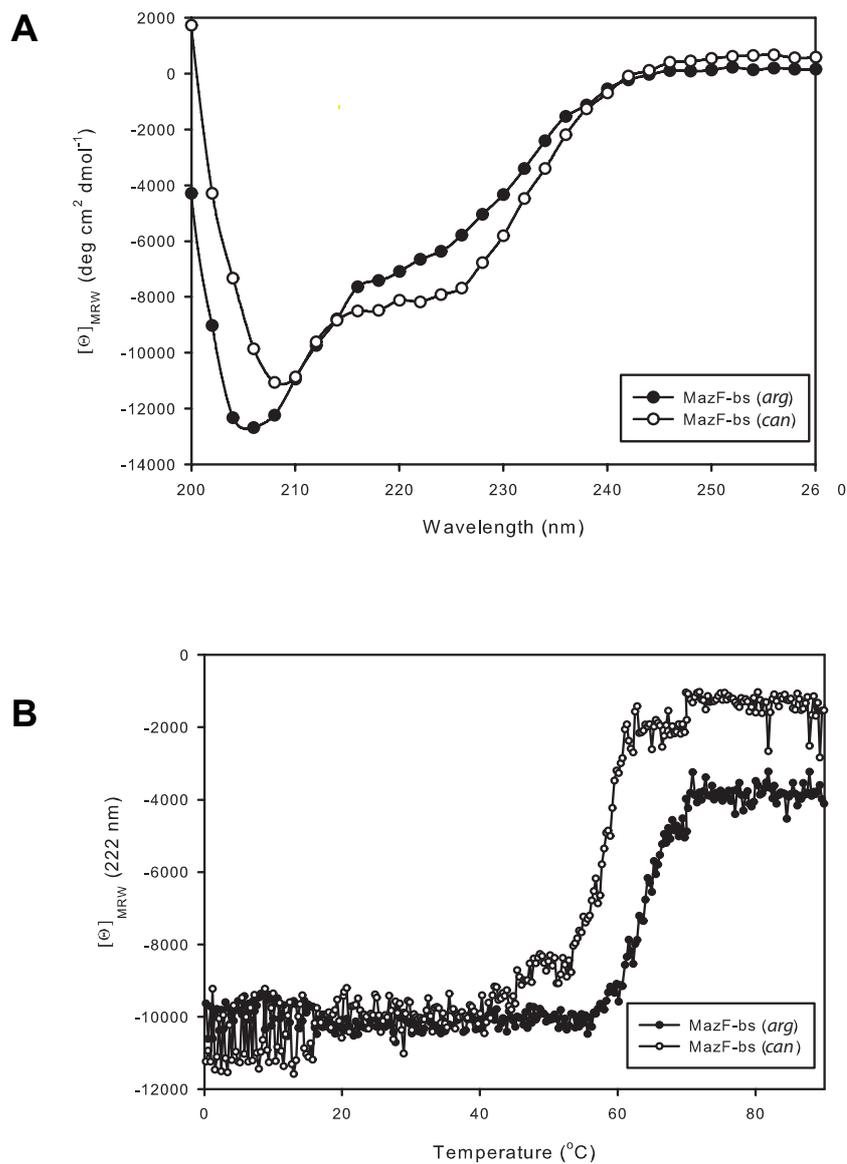


Figure 17 The structural analysis of MazFbs(*arg*) and MazFbs(*can*)

These figures are adapted from Ishida *et al.*, (Ishida et al. 2013). A, CD spectra of the secondary structures of MazFbs(*arg*) and MazFbs(*can*). B, Thermal stabilities of MazFbs(*arg*) and MazFbs(*can*). The open and black circles represent MazFbs(*can*) and MazFbs(*arg*), respectively.

Discussion

RNA interference can be carried out by two different methods; one by sequence-specific mRNA interferases and the other by RNAi using antisense RNA. In contrast to antisense RNA, which can be generally designed against any RNA, the use of mRNA interferases is much more defined. Therefore, expanding the recognition of the RNA restriction enzyme is highly important as it can be applied for gene therapies against viral infections and cancers by regulating specific genes (Zhu et al. 2009; Park et al. 2012; Shapira et al. 2012). To date, some of the MazF homologues in various bacterial genomes have been identified and characterized revealing substrate recognition lengths varying from 3-bases to 7-bases (Zhang et al. 2003b; Zhu et al. 2009; Park et al. 2011; Yamaguchi et al. 2012). In the MazFbs-RNA complex structure, it was shown that the five-base target sequence tightly binds to MazFbs with the bases facing towards the MazFbs interface and the backbone phosphate moieties facing outside and away from the MazFbs interface. Thus, interaction between RNA and MazFbs occurs mainly through interaction with bases, facilitating sequence-specific recognition of RNA substrates (Simanshu et al. 2013). It was suggested that other homologues are likely to bind their RNA substrates within the dimer interface similar to the structure of the MazFbs-RNA complex. While characterizations of new MazF homologues are still ongoing, the efforts to extend recognition sequences by known MazF homologues are also highly important. In one study involving recognition site alteration, the loop-2 region in MazFec from *E. coli* was exchanged with the corresponding loop-2 regions of MazFmx from *Myxococcus xanthus* and MazFmt from *Mycobacterium tuberculosis*, resulting in a change of

cleavage specificity from ACA to other sequences. Exchanging the loop 1 region did not create new cleavage sequences (Park et al. 2013).

In the present study, I have attempted to change the specificity of MazFbs by exchanging Arg to Can. In order to achieve complete replacement, there are some issues to be addressed: 1. Suppressing the toxicity of Can to the cells. 2. Achieving high labeling efficiency. 3. Maintaining the structure as well as function. Can is highly toxic for the cells as it will be incorporated into cellular proteins producing structurally abnormal proteins, resulting in cell growth arrest and eventual cell death. Therefore, the simple addition of Can into a culture medium does not yield a protein in which all the Arg residues in MazFbs are replaced with its analogue. Ideally, the incorporation of an amino acid analogue into any other cellular proteins but the target protein has to be completely prevented. In addition, it is important to suppress Arg biosynthesis to achieve the complete replacement of all the Arg residues with Can because the Arg affinity to tRNA is more than 100 times higher than that of Can (Igloi and Schiefermayr 2009).

In order to suppress Can toxicity to the cells while maintaining the biosynthetic function of the cells, the SPP system was applied so that Can is not incorporated into any cellular proteins except for MazFbs. The second requirement was achieved by using an Arg auxotroph together with a dually inducible system (Vaiphei et al. 2010). The use of the SPP system for the replacement of all Arg residues in MazFbs with Can appears to be essential as Can incorporation into other cellular proteins would likely cause the cells severe inhibitory effects on various biosynthetic reactions including protein synthesis.

Therefore, the present system can be applied for other toxic amino acid analogues as far as they can be recognized by *E. coli* aminoacyl-tRNA

synthases. The second requirement for the present system is the use of an amino acid auxotroph to avoid the incorporation of a natural amino acid into a target protein. Eliminating the cellular Arg is essential since Arg is preferentially used for protein production rather than Can. In addition, the dual inducible system enabled MazFbs to be labeled in a highly efficient manner. The use of the SPP system in combination with amino acid auxotroph strains is an innovative method to create proteins which have novel functions and structures without genetic manipulation of aminoacyl tRNA synthetases and tRNAs.

The most exciting and promising result in the present study is that MazFbs(*can*) became a UACAUA 6-base recognition enzyme and it was no longer able to cleave RNA at the original MazFbs(*arg*) five-base, UACAU site. The cleavage activity of MazFbs(*can*) was reduced to approximately 5% of MazFbs(*arg*), however, the *K_m* value of MazFbs(*can*) using UACAUA as a substrate is almost identical to that of MazFbs(*arg*), indicating that the substrate binding affinity in MazFbs(*can*) was compensated by an extra A residue at the 3' end of the substrate. This specificity alteration may be due to the replacement of Arg to Can changing the interaction between Arg residues in MazFbs(*arg*) and the substrates. The residues Arg-25, Arg-81 and Arg-87 are located in the interface between the monomers (Fig. 12C) (Gogos et al. 2003), thus, these 3 residues may be critical for recognizing the 6th base of the substrates. However, according to the crystal structure of the complex of MazFbs together with substrate, none of the Arg residues interact with the UACAU bases. The residues that interact with the 6th base are Leu-56, Asn-32, Gly-35, and Ile-34, respectively (Simanshu et al. 2013). Since these 4 amino acids are non-charged amino acids, Can residues in MazFbs(*can*) may be able to interact with

substrates directly. Can replacement of Arg residues also causes the secondary structure of MazFbs(*can*) to be more helical than that of MazFbs(*arg*). In addition, the salt bridge between Arg-87 and Glu-20 and hydrogen bonds between Arg-5 and Ala-112, which stabilize MazFbs dimer formation, are possibly lost since there is no charge in the guanidino group of Can. Because residual interactions in a protein are critical for its stability, the thermostability of MazFbs(*can*) is lower than that of MazFbs(*arg*). Considering the difference of secondary structure and thermostability, the conformational change possibly leads to a dimer interface change. Thus Can residues in MazFbs(*can*) may be able to interact with the 6th base position of the substrates directly. On the other hand, it is also possible to hypothesize that Can residues are involved with substrate recognition indirectly. This can be explained by conformational change as well. After Can replacement, overall MazFbs(*can*) structure is different, and thus some of the residues are able to interact with the 6th base of the substrate tightly. At present, it is not known how MazFbs(*can*) recognizes the UACAUA site and it remains to be further elucidated by X-ray crystallography.

Section III

Construction of a residue- and stereo-specific methyl labeling method by engineering *E. coli*

Introduction

Although NMR spectroscopy is a great tool to study the dynamics of proteins and protein-protein interactions in solution state, the applicable range of molecular weights for conventional NMR methodologies are still limited to relatively small proteins in comparison to other methods such as X-ray crystallography and cryo-electron microscopy (CryoEM). However, technology advances to overcome the size limitation under a highly deuterated background have been accomplished in the past two decades (Goto et al. 1999; Ruschak et al. 2010; Velyvis et al. 2012), and it is now possible to observe ^{13}C -labeled methyl signals for protein particles in the 700 kDa range (Ruschak and Kay 2010; Kerfah et al. 2015). Among the methyl-containing amino acids, Ile-, Leu-, and Val-residues are highly useful to obtain structural and dynamic information for larger proteins because these amino acid residues and the aromatic amino acids form the hydrophobic cores of proteins. Important structural information of proteins can be obtained from the ILV-methyl signals, however, the *residue*- and *regio-stereo*-specific assignments of the ILV-methyl signals are problematic. This is because relatively small chemical shift dispersions and overlapping chemical shifts often happen, making it difficult to analyze the overall spectra. In order to obtain detailed structural information by NMR, it is essential to elucidate all of the assignments including the ambiguous peaks from simultaneous labeling of ILV-methyl groups. To simplify the spectra, various methyl specific labeling methods have been developed using keto-precursors (Rosen et al. 1996; Goto et al. 1999; Tugarinov and Kay 2003).

In order for selective Ile labeling, methyl ^{13}C -labeled α -ketobutyrate was developed as a first precursor (Gardner et al. 1997). In this system, the δ_1 methyl group of Ile residues were selectively labeled. Later, a selective labeling method for Leu and Val using γ - α -ketoisovalerate, in which the γ_1 - or γ_2 -methyl groups were isotope-labeled with $^{13}\text{CH}_3$ and $^{12}\text{CD}_3$, respectively, was developed.

In this system, γ_1/γ_2 - and δ_1/δ_2 -methyl Val and Leu residues are labeled, respectively (Goto et al. 1999; Gross et al. 2003). These two precursors are commercially available and most frequently used. However, in order to achieve residue-selective labeling using α -ketoisovalerate, one of the biosynthetic pathways of either Leu or Val needs to be inhibited since α -ketoisovalerate is a common precursor for both Val and Leu biosynthesis. To date, *residue*-selective labeling of Val has been achieved by suppressing its metabolic conversion from α -ketoisovalerate to Leu by adding enough amount of perdeuterated Leu into the culture medium. In order to achieve Leu-specific methyl labeling, a Leu specific methyl ^{13}C -labeling method using α -ketoisocaproate has been proposed but this method has not been widely used as α -ketoisocaproate is not commercially available yet (Lichtenecker et al. 2013).

The disadvantage of using either methyl ^{13}C -labeled α -ketoisovalerate or α -ketoisocaproate precursors is the lack of a method for *stereo*-specific labeling of *prochiral* methyl groups. The inability to *stereo*-specifically label *prochiral* methyl groups causes overlapping between the Leu δ_1/δ_2 and Val γ_1/γ_2 signals and thus 50% lowering of labeling efficiencies. To overcome this, the γ_2 -methyl group of Ile can be *regio*-selectively labeled using a precursor α - ^{13}C -aceto- α -hydroxybutyrate (Ruschak et al. 2010; Ayala et al. 2012). It is more difficult to *stereo*-selectively label *prochiral* methyl groups of Leu and Val residues by the

precursor method. However, since a new compound, α -[^{13}C]-2-acetolactate, was developed, both γ_2 -Val (Val-*proS*) and δ_2 -Leu (Leu-*proS*) can be labeled in a *stereo*-specific manner (Gans et al. 2010). Furthermore, Val-*proS* labeling can only be achieved by suppressing the conversion from 2-acetolactate to Leu (Mas et al. 2013). To date, there are some technologies available for Val selective labeling (Mas et al. 2013; Miyanoiri et al. 2013), but there is no effective method for Leu selective labeling using precursors.

In an alternative approach, stereo-specifically labeled protein samples can be prepared using the stereo-array isotope labeling (SAIL) method, in which specifically labeled amino acids in a deuterated background are used. In this system, structural determination of large proteins can be done very accurately since the proton density of proteins is much lower than that of uniform labeling proteins (Kainosho et al. 2006).

However, due to the amino acid biosynthetic pathway in *E. coli*, the incorporation efficiency of SAIL amino acids into a protein by *E. coli* expression system was not high enough for application to high molecular weight proteins. Since a high isotope incorporation efficiency is essential to obtain NMR structural information, the SAIL method was limited to a cell-free expression system (Takeda et al. 2008). Recently, it was found that the cellular concentration of Leu is not high, thus stereo- and Leu-specific labeling method was developed using *E. coli* cells (Miyanoiri et al. 2013). However, it was also found that the labeling efficiency in cells of Val using SAIL-Val was only 80% even when 100 mg/L was added into the culture medium, possibly due to the high pyruvate concentration in cells. Pyruvate is the major source for Val synthesis, and thereby SAIL-Val was diluted by endogenous pyruvate (Miyanoiri et al. 2013).

In section III-I, I developed a Leu selective labeling system using genetically engineered *E. coli* in which conversion from acetolactate to Val is completely blocked by deletion mutations.

In section III-II, I developed a Val and Leu labeling system using SAIL amino acids together with amino acid auxotrophs.

Section III-I. Stereo- and residue-specific methyl labeling using amino acid precursors

Materials and Methods

Construction of *E. coli* strains for specific methyl labeling

The following gene deletion mutants from the Keio collection (Keio university, Japan) were used for P1 transduction [JW5605($\Delta ilvD$), JW5606($\Delta ilvE$), JW5652($\Delta avtA$), JW5807($\Delta leuB$), JW0076($\Delta ilvI$), JW3704($\Delta ilvG$), JW3646($\Delta ilvB$), and JW0003($\Delta thrC$)] based on analysis of biosynthetic pathways. (Fig. 18). To obtain a single deletion mutant, firstly P1 lysate for each gene was prepared. Next, BL21(DE3)($\Delta hisB\Delta trpC$) (Vaiphei et al. 2010) was grown in 10 ml of LB medium supplemented with 5 mM $CaCl_2$, and when $O.D_{600}$ reached 0.5, 100 μ l of P1 phage lysate was added. After 30 min incubation at 30°C, the transduction was stopped by the addition of 100 mM sodium citrate (pH5.5). The cells were then collected and washed with 200 mM sodium citrate (pH 5.5). Colonies from LB-Km plates (Km; 50 mg/L) were examined for gene deletion by PCR. To confirm auxotrophy, cell growth was tested in M9-glucose medium in the presence or absence of amino acids, in particular, Ile, Leu and Val. In addition, growth with acetolactate and α -ketoisovalerate was also examined. In order to establish a Val specific labeling strain, blocking the Val biosynthesis was confirmed using M9-glucose medium in the presence of Ile and α -ketoisovalerate or acetolactate (Leu and Val precursors) but in the absence of Val. Likewise, blocking Leu was confirmed by the use of M9-glucose medium in the presence of Ile and α -ketoisovalerate or acetolactate but in the absence of Leu.

Removal of the Km cassette from the genome

In order to carry out multiple gene deletion mutations, the inserted Km cassette was removed by using pCP20 which carries the *red* recombinase gene (Datsenko and Wanner 2000). The deletion mutants were grown in LB-Km medium, and pCP20 was electroporated into the cells. The grown colonies were re-streaked on LB plates and incubated at 42°C overnight. The loss of both the Km cassette and pCP20 was tested using three different plates: LB, LB-Km and LB-Amp (Amp; 100 mg/L) plates, respectively.

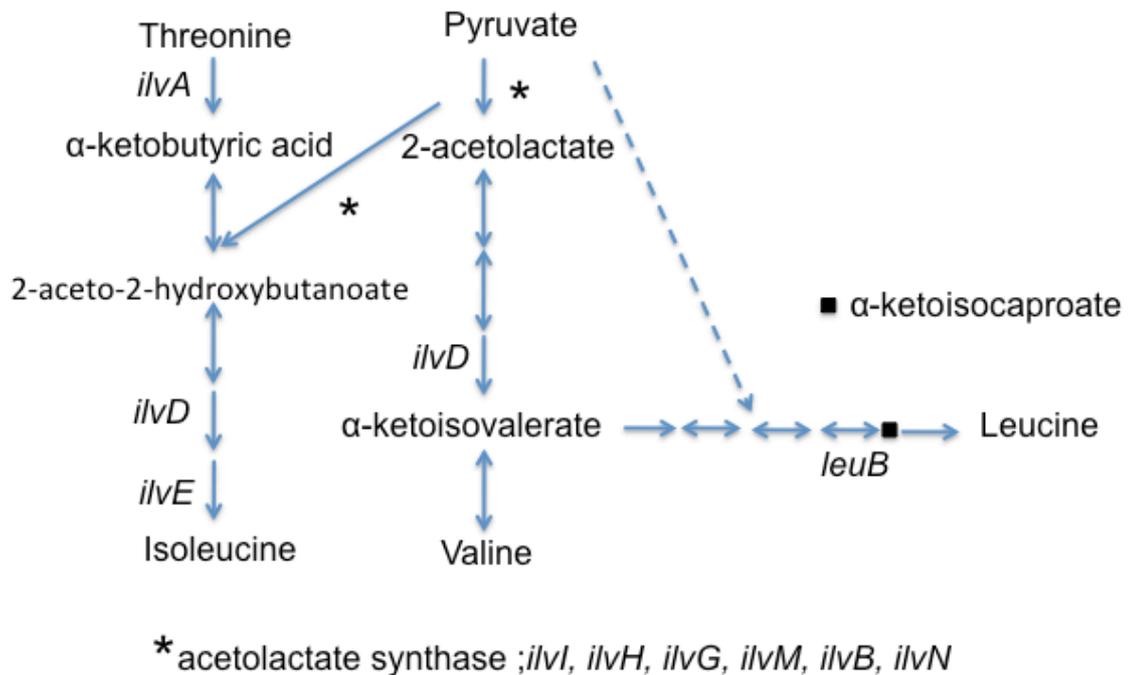


Figure 18. The biosynthetic pathway of Ile, Leu and Val

Simplified pathways for Ile, Leu, Val and threonine biosynthesis depicted starting from glucose as the main carbon source, according to the Kyoto Encyclopedia of Genes and Genomes (<http://www.genome.jp/kegg/>). The commercially available isotope-labeled amino acid precursors [α -ketobutyrate, α -ketoisovalerate, α -ketoisocaproate, 2-aceto-2-hydroxybutanoate and 2-acetolactate] are shown in the figure. Each solid arrow depicts one enzymatic reaction and the related gene encoding for the enzyme is shown in *italics*. Dashed arrows indicate multi-step reactions and double arrows represent reversible enzymatic reactions. Asterisks indicate acetolactate synthase genes.

Sample preparation

Production of methyl-labeled protein samples for Val

In order to label only Val-methyl groups in proteins, either BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$ strain (strain A) or BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$ $\Delta leuB$ (strain B) was used. As standard proteins, peptidyl-prolyl isomerase domain of trigger factor (PPD, 101 residues), Catabolite Activator Protein (CAP), and the full-length malate synthase G plasmid (MSG, 731 residues) were used. Strain A transformed with pET21c-PPD or pET21c-MSG or pColdIII-CAP were grown at 37°C in 50 mL (or 250 ml for MSG) of M9-D₂O medium supplemented with glucose (2 g/L), Trp (30 mg/L), His (30 mg/L), Leu (30 mg/L), Ile (30 mg/L), Val (30 mg/L) and Amp (100 mg/L). When OD₆₀₀ reached 0.3, the cells were collected and washed by M9-D₂O medium twice and re-suspended in 50 ml (or 250 mL for MSG) of M9-D₂O medium containing glucose (2 g/L), Trp (30 mg/L), His (30 mg/L), Met (methyl-¹³C) at 80 mg/L, Amp (100 mg/L), Leu (30 mg/L) and either α -¹³C-ketoisovalerate (30 mg/L) or ¹³C-acetolactate (40 mg/L). After incubating for 1.5 hours at 37°C, 1 mM IPTG was added to induce protein production. The cells were collected and stored at -20°C after 2.5 hours for PPD or after overnight incubation for MSG and CAP.

Production of methyl-labeled protein samples for Leu

The engineered strain BL21 (DE3) ($\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$, $\Delta ilvE$, $\Delta avtA$) (strain C) was used for Leu-methyl labeling. The growth condition medium contained 0.1 g/L of [²H]-celtone base powder (Cambridge Isotope Laboratories) in addition to the growth medium for Val sample preparation. When OD₆₀₀ reached 0.3, the cells were collected and washed by M9-D₂O medium twice and

re-suspended in M9-D₂O medium containing glucose (2 g/L), Trp (30 mg/L), His (30 mg/L), Met (methyl-¹³C) at 80 mg/L, Amp (100 mg/L), Val (30 mg/L), 0.1 g/L of [²H]-celtone base powder, and either α-¹³C-ketoisovalerate (30 mg/L) or ¹³C-acetolactate (40 mg/L). After incubating for 1.5 hours at 37°C, 1 mM IPTG was added and the culture was incubated overnight.

The purification of PPD, MSG and CAP

The purification of isotope-labeled proteins was carried out using Ni-NTA column chromatography. The cells were re-suspended into binding/washing buffer composed of 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 20 mM imidazole-HCl (pH8.0). After breaking the cells by French press, the soluble fraction was prepared using ultra centrifuge at 40,000 rpm for 1 hr. The supernatant was mixed with pre-equilibrated Ni-NTA agarose and incubated for 1 hr at 4°C. The resin was washed 3 times with binding/wash buffer. Finally, the protein was eluted with elution buffer consisting of 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 300 mM imidazole-HCl (pH 8.0). The purity of the eluted protein fractions was analyzed by SDS-PAGE, and after concentrating the sample by protein concentrator (Millipore), the PPD sample was dialyzed against PPD NMR buffer consisting of 50 mM Na-phosphate (pH 6.0) and 10% D₂O and the MSG sample was dialyzed against MSG NMR buffer composed of 20 mM Na-phosphate (pH 7.1), 20 mM MgCl₂, 5 mM DTT and 10% D₂O (Tugarinov and Kay 2003). CAP was dialyzed against NMR buffer consisting of 50 mM Na-phosphate (pH6.0), 500 mM KCl, 1 mM β-mercaptoethanol and 7% (v/v) D₂O.

Optimization of the precursor and amino acid composition

In order to optimize the conditions for Leu-/Val- stereo specific labeling using precursors, protein production and labeling efficiency were examined using various concentrations of precursors. For selective Val labeling, strain A, BL21(DE3)($\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$) was used to optimize U -[1H , ^{12}C], Leu/Val-[$^{13}CH_3$]^{pro-S} PPD samples. Multiple concentrations of ^{13}C -proS-acetolactate (10, 15, 20, 25, 30 or 60 mg/L) were examined for the production of PPD samples. In addition to ^{13}C -proS-acetolactate, the induction medium was supplemented with 30 mg/L of His, Trp and Ile. In order to optimize Leu-specific labeling conditions, U -[1H , ^{12}C], Leu-[$^{13}CH_3/^{12}CD_3$] PPD samples were prepared using strain C, BL21(DE3) ($\Delta hisB$, $\Delta trpC$, $\Delta ilvB$, $\Delta ilvG$, $\Delta ilvI$, $\Delta ilvE$, $\Delta avtA$). The protein production and labeling efficiency were examined using 17, 25, 50, or 120 mg/L of α - ^{13}C -ketoisovalerate. The medium was supplemented with 30 mg/L of His, Trp, Ile and Val, respectively. Protonated glucose and protonated amino acids were used in both the growth and induction steps. The concentration of recombinant proteins and the volume of the protein sample obtained at the end of the purification were measured to assess the protein yield. For all labeled protein production, L-Met (methyl- ^{13}C) at 80 mg/L was used as an internal reference. The relevant methyl signals were measured by recording ^{13}C -HMQC.

NMR spectroscopy

Spectra were recorded at 305 K by using Bruker Avance III 700 MHz or 600MHz equipped with a triple resonance gradient cryo-probe and a standard probe, respectively. Topspin 3.2 (Bruker BioSpin) was used for data collection and NMRPipe (Delaglio et al., 1995) for spectra processing followed by analysis with Sparky 3.115 (T. D. Goddard and D. G. Kneller, University of California, San Francisco, CA).

Results

Engineering the biosynthetic pathway of Ile, Leu and Val

In order to achieve residue specific labeling using commercially available precursors (acetolactate and α -ketoisovalerate), the biosynthetic pathways of Ile, Leu and Val were analyzed (Fig.18). For Leu specific labeling, it is necessary to inhibit Val biosynthesis since both Val and Leu share some of the reactions for their biosynthesis. Therefore, deletion mutations were designed at the last step of Val synthesis. Thus, two transaminase genes, *ilvE* and *avtA*, located at the last step of Val synthesis were deleted by P1 transduction using Keio collection strains. To achieve Val specific labeling, Leu biosynthesis was inhibited. As shown in Figure 18, there are multiple reactions after α -ketoisovalerate synthesis, and *leuB* was chosen as the gene to be deleted.

To optimize labeled acetolactate incorporation efficiency, some of the cell acetolactate synthase (ALS) genes were deleted. There are 3 ALSs in the *E. coli* genome and each ALS consists of one large subunit and one small subunit. Thereby, 6 genes for ALSs, *ilvI*, *ilvG*, *ilvB*, *ilvH*, *ilvM* and *ilvN*, are present in the *E. coli* genome. Since the large subunits have major functions for Ile, Leu and Val biosynthesis, the *ilvI*, *ilvG*, *ilvB* genes encoding the ALS large subunits were deleted (Fig. 18). Notably, although the ALS mutant strain ($\Delta ilvI$, $\Delta ilvB$, $\Delta ilvG$), could not grow in the absence of Ile, Leu and Val, the cells were able to grow in the presence of acetolactate and Ile. In order to achieve stereo- and residue-specific methyl labeling, I incorporated the Val and Leu labeling system into the ALS mutant strain, respectively.

Considering the cost effectiveness, I attempted to construct the deletion mutants consisting of $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$, $\Delta ilvE$, $\Delta avtA$ and $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$,

ΔleuB. However, isolation of these strains was not successful. On the other hand, when I attempted to isolate the same deletion mutations from the BL21(DE3) *ΔhisB, ΔtrpC* strain (Vaiphei et al. 2010), strain A, BL21(DE3) *ΔhisB, ΔtrpC, ΔilvI, ΔilvG, ΔilvB*, strain B, BL21(DE3) *ΔhisB, ΔtrpC, ΔilvI, ΔilvG, ΔilvB ΔleuB*, and strain C BL21(DE3) *ΔhisB, ΔtrpC, ΔilvI, ΔilvG, ΔilvB, ΔilvE, ΔavtA* were successfully isolated.

Optimization for the production of Val/Leu specific methyl labeled PPD

In order to prepare methyl specific labeled samples in a timely manner, two media were used in this study, one was as a growth medium and the other one was as an induction medium. Both media consisted of M9-D₂O glucose media, but the growth medium was supplemented with His, Trp, Ile, Leu and Val while the induction medium to label Leu contained His, Trp, Ile, acetolactate and Val. For Val labeling, the induction medium contained His, Trp, Ile, acetolactate and Leu. In order to determine the minimum concentration of precursors, [¹H,¹³C]-methyl incorporation was examined using 2-oxo-3-[²H]-3-[²H₃]methyl-[4-¹³C]butanoate (¹³C-monomethyl α-ketoisovalerate). The catalytic N-terminal PPI domain from *E. coli* Trigger Factor (PPD) was used as an example. To determine optimum isotope-labeling conditions, several isotope-labeled PPD with different isotope concentrations were prepared, and the signals from methyl labeled samples were analyzed by recording the ¹³C-HMQC spectra (Fig. 19A). As an internal standard, the mean signal intensity of the Met ε-methyl group was used and all the methyl signals were normalized based on this standard (Fig. 19B and C). Val-[¹³CH₃]^{proS} and Leu-[¹³CH₃]^{proS} PPD samples using strain A, BL21(DE3) *ΔhisB, ΔtrpC, ΔilvI, ΔilvG, ΔilvB* with different concentrations of ¹³C-proS-

acetolactate were analyzed. As shown in Figure 19A and 19B, 25-30 mg/L was sufficient to reach the maximal isotope incorporation. This represents a reduction to 1/10 of the 300 mg/L of acetolactate required for the standard strain (Gans et al. 2010), without loss of labeling efficiency by using strain A. Next, I tested the isotope labeling efficiency of Leu using strain C, BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$, $\Delta ilvE$, $\Delta avtA$. For optimal Leu labeling conditions, isotope incorporation levels were examined using ^{13}C -monomethyl α -ketoisovalerate as a Leu precursor. Results showed that 25 mg/L of ^{13}C -monomethyl α -ketoisovalerate was sufficient to reach maximum isotope incorporation, while the standard strain required 88 mg/L of α -ketoisovalerate (Fig. 19C). Protein yields using different conditions were compared to the protein expressed by the standard protocol in which the BL21(DE3) strain with 120 mg/L of ^{13}C -monomethyl α -ketoisovalerate was used to see if the deletion mutation affects protein yield. For this purpose, the protein concentration was determined after purification. The protein yield of Leu/Val- ^{13}C [$^{13}\text{CH}_3$] proS PPD sample produced by using 40 mg/L ^{13}C -*proS*-acetolactate and strain A was the same as that of the standard protocol. Interestingly, when 60 mg/L of ^{13}C -*proS*-acetolactate was used, the protein yield was twice as high as that of the standard protocol. Since 300 mg/L of acetolactate was required for the standard protocol, the use of 60 mg/L resulted in an 80% reduction of acetoacetate while doubling the protein yield.

In the case of Leu labeling, the protein yields using strain C and 50 mg/L of ^{13}C -monomethyl α -ketoisovalerate yields were about the same as the standard strain. However, when 40 mg/L of ^{13}C -*proS*-acetolactate was used for MSG production, the protein yield was about 40% of the standard protocol production.

As shown in Figure 20, residue- and stereo-specific methyl labeling was successful using PPD as an example. Using strains A, B and C along with precursors, PPD was labeled in either a Leu or Val specific manner which simplified the spectra compared to the spectrum generated by using a standard strain with α - ketoisovalerate.

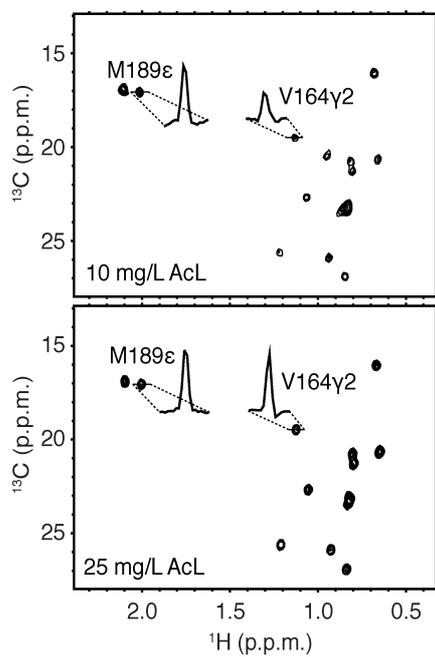
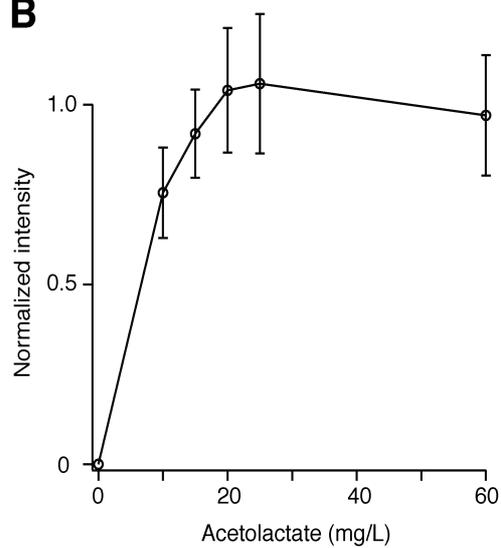
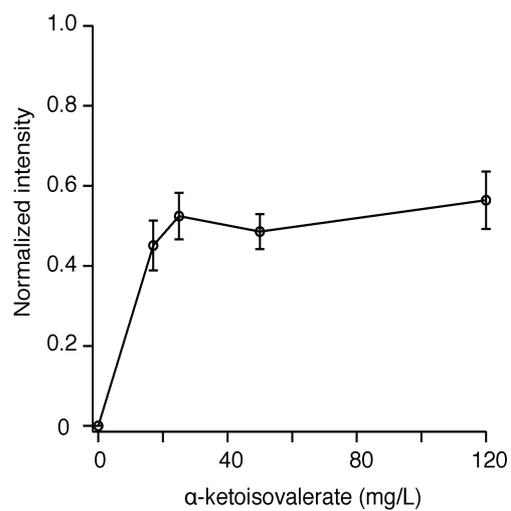
A**B****C**

Figure legends are in the following page

Figure 19. Optimization for residue specific methyl labeling by precursors using strain A. These figures are adapted from Monneau *et al.*, (Monneau et al. 2016). Peptidyl prolyl isomerase domain (PPD) of Trigger Factor was expressed by strain A BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$. (A) Two spectra of U - $[^{12}C, ^1H]$, Leu/Val- $[^{13}CH_3]$ -*proS* PPD are shown with the proton cross-sections depicted for Met189 ϵ -methyl and Val164 γ 2-methyl groups. The signal of Met ϵ -methyl group is used as an internal standard to assess labeling efficiency. (B) and (C); The intensities of peaks of Val/Leu in the ^{13}C -HMQC (heteronuclear multiple-quantum correlation spectroscopy) were examined. In order to normalize the intensities, 9 Leu/Val-*proS* peaks were analyzed (B) and 6 Leu-*proS/proR* peaks were analyzed (C). The normalization was based on the intensities of methionine ϵ -methyl group, and the mean \pm sd is plotted against the concentration of the precursor.

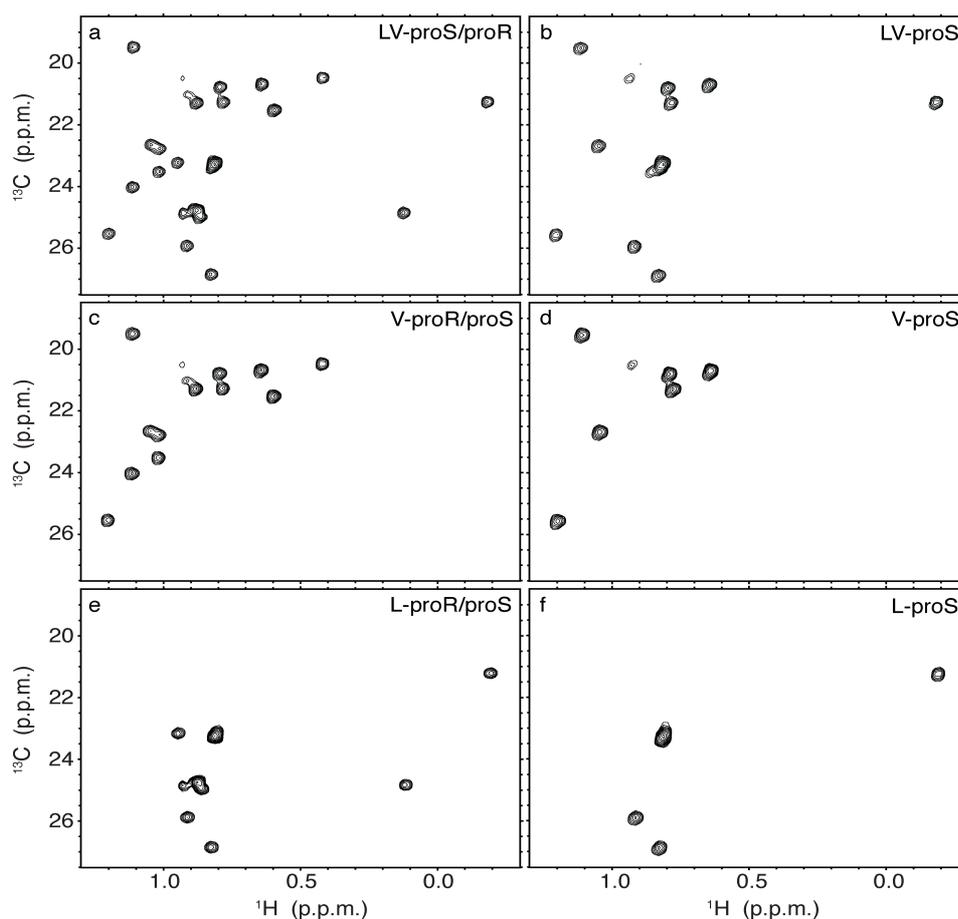


Figure 20. The ^{13}C -HMQC spectra of residue- and stereo-specifically methyl labeled PPD samples. This figure is adapted from Monneau *et al.*, (Monneau *et al.* 2016) The ^{13}C -HMQC spectra were recorded on different samples of PPD, produced using either BL21(DE3) strain and α -ketoisovalerate (a); BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$ strain with *proS*-acetolactate (b); BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$ $\Delta leuB$ strain with α -ketoisovalerate (c); strain B with *proS*-acetolactate (d); BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$, $\Delta ilvE$, $\Delta avtA$ with α -ketoisovalerate (e); BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$, $\Delta ilvE$, $\Delta avtA$ with *proS*-acetolactate (f).

Optimization for the production of Val/Leu specific methyl labeled MSG and application of this labeling technology to Catabolite Activator Protein (CAP)

Various methyl-labeled samples of the 82 kDa malate synthetase G (MSG) were prepared using optimized conditions. MSG, consisting of 731 amino acids, is one of the largest monomeric proteins for NMR structural study and its methyl assignments have been previously determined (Gans et al., 2010; Tugarinov & Kay, 2003). The MSG sequence contains 70 Leu and 46 Val residues and assignment of the overlapping area is very difficult. The Val-*proS* labeled MSG samples were prepared using strain B, BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$ $\Delta leuB$ and 40 mg/L of ^{13}C -*proS*-acetolactate. The Leu-*proS* labeled MSG was prepared using strain C, BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$, $\Delta ilvE$, $\Delta avtA$ and 40 mg/L of ^{13}C -*proS*-acetolactate. As shown in Figure 21, Leu and Val signals were completely separated, and the overlapping area was less crowded than the spectrum generated by using the standard protocol. In order to evaluate the residue- and stereo-specific methyl labeling, the Leu/Val methyl-labeled MSG sample was prepared using BL21(DE3) and 120 mg/L of ^{13}C -monomethyl α -ketoisovalerate. The signal intensity of both Val- γ_2 and Leu- δ_2 was twice as high as the methyl signals from the sample labeled by ^{13}C -monomethyl α -ketoisovalerate since acetolactate produces a stereo specifically labeled compound while α -ketoisovalerate produces a racemic mixture. In order for Val specific labeling, Leu signals need to be suppressed since acetolactate is a common source to produce Leu and Val (Gans et al. 2010). In a previous study, it was reported that 98% of Leu signals can be suppressed by adding deuterated Leu in the culture medium (Mas et al. 2013). I took the same strategy to suppress Leu signals using strain A, however, about 6% of residual Leu methyl signals

were observed (Fig. 22A). Thus, I applied strain B, which has a deletion mutation in the *leuB* gene so that the conversion from acetolactate was completely blocked. As shown in Figure 22B, small signals from Leu were completely erased.

This newly developed Leu and Val specific labeling technology was applied to CAP, a 50-kDa homodimeric gene regulatory protein. In order to create a standard spectrum, CAP was prepared using BL21(DE3) using ^{13}C α -ketoisovalerate (a). Selectively labeled samples were prepared as follows; using strain A with ^{13}C -*proR*-acetolactate (b) strain A with ^{13}C -*proS*-acetolactate (c), strain B with α -ketoisovalerate (d), strain B with ^{13}C -*proR*-acetolactate (e), strain B with ^{13}C -*proS*-acetolactate (f), and strain C with α -ketoisovalerate (g). As shown in Figure 23, stereo-specific labeled precursors simplified the ^{13}C -HMQC spectra (a-c). In addition, the residue specific labeling system using strain B and C further simplified the spectra (d-g). Thus, using stereo-specifically methyl labeled compounds together with our deletion mutants enabled us to simplify the spectra, which is highly useful for assignments of Val and Leu residues.

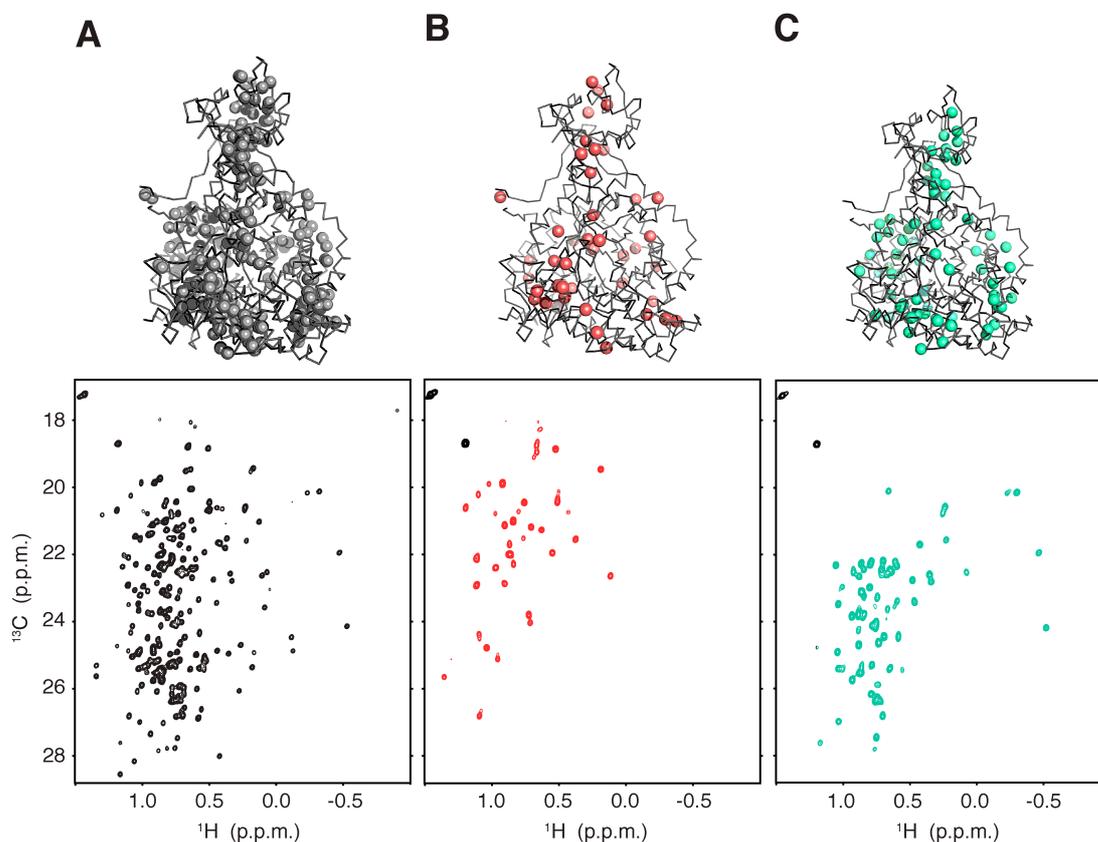


Figure 21. The ^{13}C -HMQC spectra of stereo- and residue-specific methyl labeled MSG. This figure is adapted from Monneau *et al.*, (Monneau *et al.* 2016). Methyl ^{13}C -HMQC spectrum was recorded on samples of MSG expressed with **(A)** BL21(DE3) with 120 mg/L of α - ^{13}C -ketoisovalerate. This spectrum contains Leu- δ_1/δ_2 and Val- γ_1/γ_2 crosspeaks. **(B)** The sample was prepared by using strain B with 40 $\mu\text{g}/\text{m}$ of ^{13}C -acetolactate, containing only Val- γ_2 crosspeaks. **(C)** Strain C with 40 mg/L of ^{13}C -acetolactate, containing only Leu- δ_2 crosspeaks. The crystal structure of MSG shows that the highly spatially dispersed positions of Leu- δ_2 (*green*) and Val- γ_2 (*red*) methyl groups were maintained for each residue-specific labeling scheme compared to the crowded Leu- δ_1/δ_2 and Val- γ_1/γ_2 methyl population (*grey*). The crosspeaks that appear in black on **(B)** and **(C)** panels are from Met ϵ -methyl group used as the internal standard **(D)**

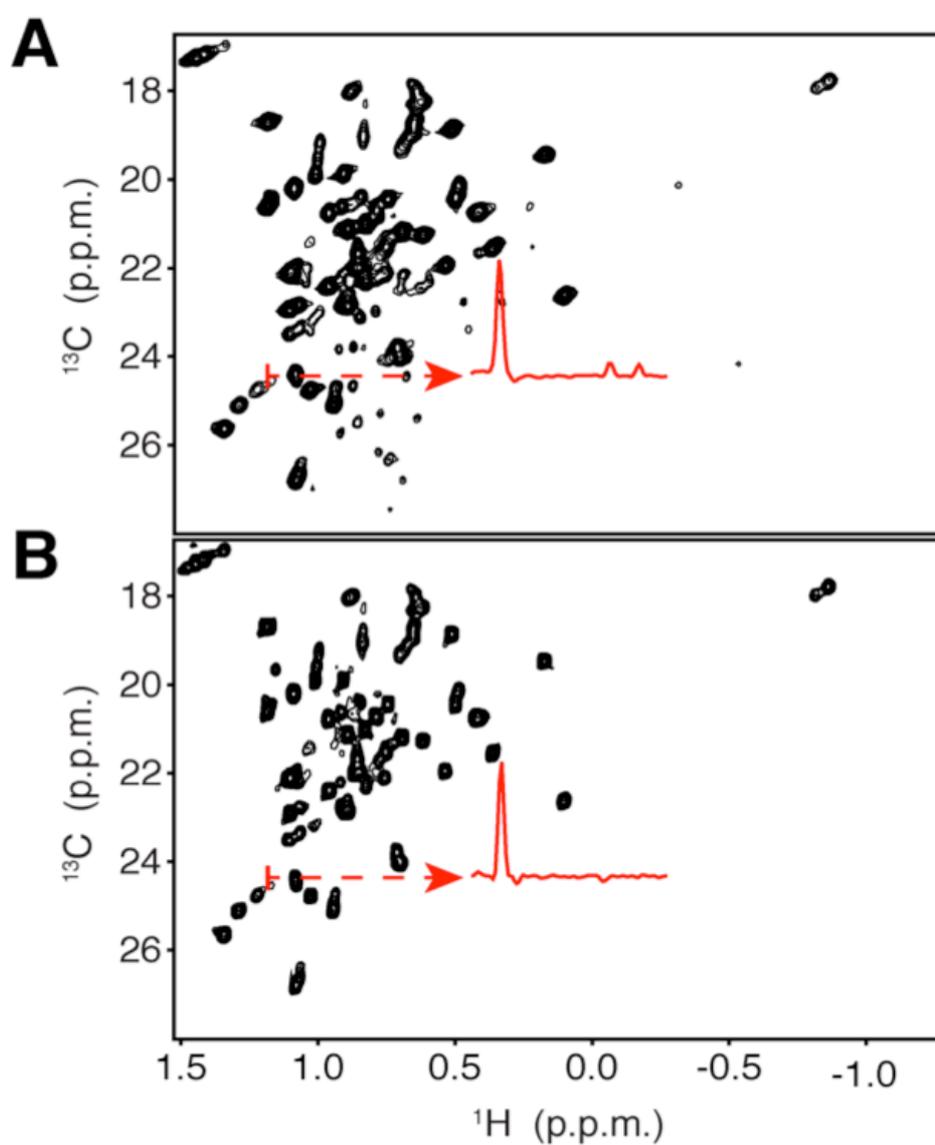


Figure 22. ^{13}C -HMQC spectra recorded on Val- $^{13}\text{CH}_3$ -*proS* MSG samples. This figure is adapted from Monneau *et al.*, (Monneau *et al.* 2016). Samples were prepared with 40 mg/L of ^{13}C -*proS*-acetolactate and 30 mg/L of ^2H -Leu. Strain A was used to generate (A), and strain B was used to generate (B). The range of the cross-section at 24.4 ppm (on ^{13}C axis) highlighted in red is shown by dashed arrows.

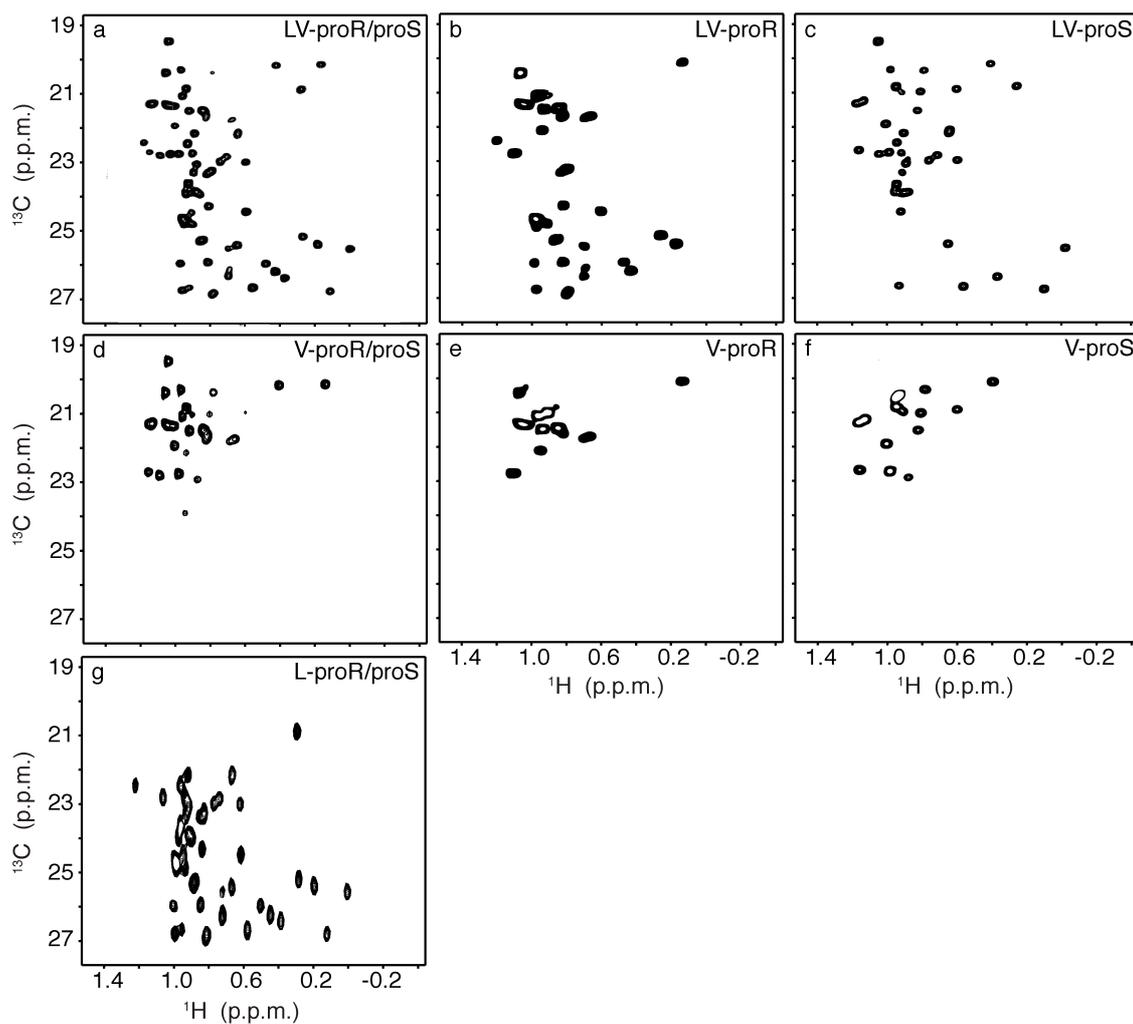


Figure 23. The application of residue- and stereo- specific labeling system for CAP. This figure is adapted from Monneau *et al.*, (Monneau et al. 2016). The ^{13}C -HMQC spectra were recorded on CAP labeled with different precursors, produced using either BL21(DE3) strain with α - ^{13}C -ketoisovalerate (**a**) or ^{13}C -acetolactate [either *proR* (**b**) or *proS* (**c**)]; or strain B strain with α -ketoisovalerate (**d**) or acetolactate [either *proR* (**e**) or *proS* (**f**)]; or strain C with α -ketoisovalerate (**g**).

The possibility of using an Ile precursor together with acetolactate

In order for methyl labeling of Ile in proteins, α -ketobutyrate has been commonly used to label at Ile- δ_1 positions (Fig. 18), however, it cannot be used as an Ile precursor for deletion mutants in the present study because it is located upstream of the acetolactate synthase gene (Fig. 18). Although selective Leu or Val labeling is highly useful for NMR structural studies, combining it with Ile labeling would be additionally useful. The moiety of Ile- δ_1 is a very useful probe in most NMR studies, as it does not overlap with Leu and Val signals and eventually gives intense signals. Recently, ^{13}C -aceto-hydroxybutanoate was developed as an alternative precursor (Kerfah et al. 2015) to label Ile residues. In principle, ^{13}C -aceto-hydroxybutanoate is compatible with ALS deletion mutants, therefore I attempted to combine the Ile labeling system with the Leu or Val labeling system. For Val/Ile labeling, I prepared samples using strain B with different concentrations of ^{13}C -aceto-hydroxybutanoate and ^{13}C -*proS*-acetolactate. As a result, when aceto-hydroxybutanoate and acetolactate were added simultaneously, Leu/Val labeling efficiency was decreased. This is possibly because aceto-hydroxybutanoate and acetolactate are both substrates of the ketol-acid reductoisomerase (KARI), the downstream enzyme after ALS in the Ile, Leu and Val pathways. The affinity of aceto-hydroxybutanoate to KARI is 5-8 times higher than that of acetolactate (Dumas et al. 2001), but in theory, it should be possible to incorporate both precursors to be equally used for amino acid biosynthesis. Therefore, I examined the isotope incorporation efficiency using multiple concentration sets. As shown in Figure 24, when the ratio of aceto-hydroxybutanoate and acetolactate was 1:5, the labeling efficiency of Val- $^{13}\text{CH}_3$ -*proS* was about 50% compared to Met incorporation, while it was greatly

improved when the ratio used was 1 : 8. In the case of Leu labeling, strain C was not able to grow with aceto-hydroxybutanoate, indicating that IlvE has a critical role for Ile biosynthesis.

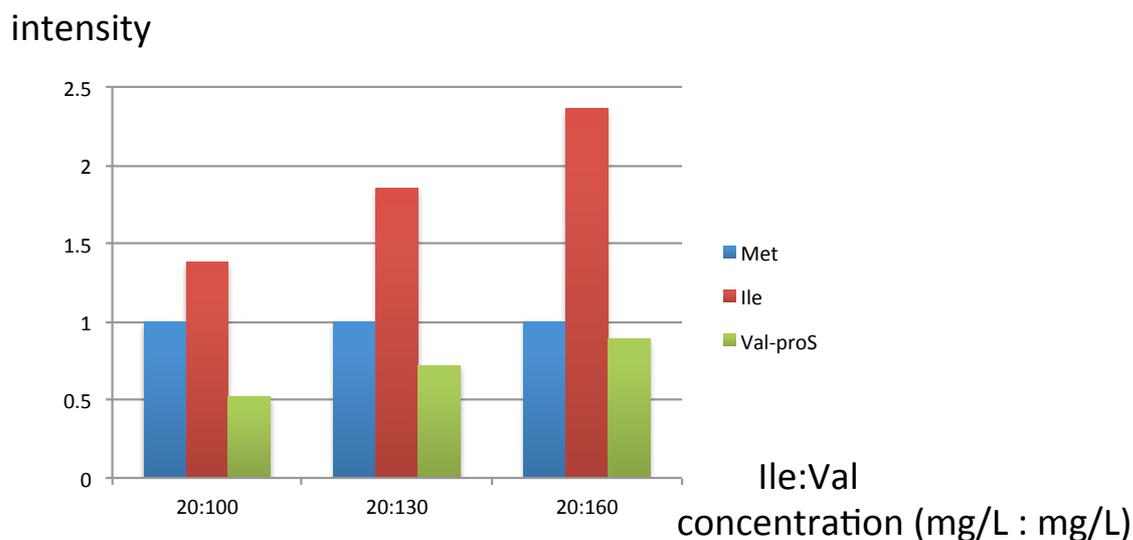


Figure 24. Isotope incorporation efficiency using ^{13}C -aceto-hydroxybutanoate and ^{13}C -*proS*-acetolactate. Strain B transformed with pET21c-MSG was grown in the growth medium as described in Materials and Methods. When O.D_{600} reached 0.3, the cells were collected and washed twice with $\text{M9-D}_2\text{O}$. The cells were then re-suspended into induction medium supplemented with different concentrations of ^{13}C -aceto-hydroxybutanoate and ^{13}C -*proS*-acetolactate. After adding the precursors, the culture was further incubated for 1.5 hrs and MSG production was induced by the addition of 1 mM IPTG. The culture was incubated overnight. After purification of isotope-labeled MSG, the protein was concentrated and dialyzed against NMR buffer, and isotope incorporation efficiency was determined by recording ^{13}C -HMQC spectra using Met as an internal control. The relative intensities are shown in the figure.

Section III-II Stereo-and residue-specific labeling using SAIL-Val and -Leu using auxotrophic strains

Materials and Methods

Construction of Ile/Leu/Val auxotrophic strains of *E. coli*

At first, P1 phage lysate using two *E. coli* strains, $\Delta ilvD$ (JW5605) and $\Delta leuB$ (JW5807), from the Keio-collection were prepared (Baba et al. 2006). In order to delete a gene, the BL21(DE3) strain was grown in 10 ml LB supplemented with 1 mM $CaCl_2$, Km (50 mg/L) and 0.2% glucose. When O.D reached around 0.5, 100 μ l of P1 lysate from the $\Delta leuB$ strain was added into the medium. After 30 min incubation at 30°C, the cells were collected and washed twice with 500 μ l of 0.1 M Na-citrate (pH 5.5), and $\Delta leuB$ cells were selected by using LB-Km plates incubated at 37°C overnight. The grown colonies were examined using M9-glucose medium in the presence and absence of Leu (20 mg/L).

After confirming Leu auxotrophy, the inserted Km cassette was removed using pCP20, carrying red recombinase (Datsenko and Wanner 2000). The electrocompetent cells were prepared using isolated BL21(DE3) $\Delta leuB$, and pCP20 was electroporated into the cells. Transformants were selected by incubating on LB-Amp plates at 30°C. Grown colonies were re-streaked onto LB plates and the plates were incubated at 42°C overnight. The grown colonies were re-plated on LB-Km, LB-Amp and LB plates, respectively. Since pCP20 is a temperature-sensitive plasmid, plasmid loss should occur by the treatment described above. After confirmation of loss of Km and Amp resistance while retaining Leu auxotrophy, *ilvD* gene deletion was carried out using P1 lysate from the $\Delta ilvD$ strain. The gene deletion was confirmed by using M9-glucose medium in the presence or absence of Ile, Leu and Val (20 mg/L each) together with Km. This auxotrophic *E. coli* was named BL21(DE3) $\Delta ilvD\Delta leuB$. After

confirming the auxotrophy of the cells, the Km cassette was removed as described above.

Protein expression using BL21(DE3) $\Delta ilvD\Delta leuB$

In order to examine protein expression using the auxotrophic cells, the MSG gene (Howard et al. 2000) was cloned into pET28a and calmodulin (CaM) and EnvZB, a histidine kinase domain (Vaiphei et al. 2010) were cloned into the pColdI vector (Suzuki et al. 2006). The cells were acclimated to D₂O using the following method; *E. coli* BL21(DE3) $\Delta ilvD\Delta leuB$ transformed with plasmids were grown in 3 ml of LB-H₂O at 37°C for 3 hrs, and the cells were collected and re-suspended into 3 ml LB-D₂O, which was then incubated at 37°C for 5 hrs. The grown cells were collected and inoculated into 10 ml M9-D₂O medium containing 20 mg/L each of ²H-Val, ²H-Leu and ²H-Ile, and Km (50 mg/L) for MSG and Amp (100 mg/L) for CaM and EnvZB. When O.D.₆₀₀ reached 0.4-0.45, 1mM IPTG was added and the culture was incubated overnight. The expression of the proteins was confirmed by SDS-PAGE.

In order to compare the protein expression levels using a standard strain and BL21(DE3) $\Delta ilvD\Delta leuB$, MSG, CaM and EnvZB were examined. For the protein standards, MSG was prepared using the *E.coli* BL21(DE3) pLys S and *E.coli* BL21(DE3) was used to produce CaM and EnvZB. Next, isotope incorporation efficiency using exogenous amino acids was examined. MSG samples using *E. coli* BL21(DE3) $\Delta ilvD\Delta leuB$ with various concentrations were prepared. At first cells were grown in 30 ml of M9-D₂O glucose medium supplemented with 10 mg/L each ²H-Val, ²H-Leu at 37°C. When the O.D₆₀₀ value reached about 0.3, the cells were collected and washed twice with M9-D₂O buffer and re-suspended

into 30 ml of M9-D₂O glucose medium supplemented with various concentrations of Ile, Leu and Val (5 to 20 mg/L). The protein was induced by addition of 1 mM IPTG at O.D₆₀₀=0.4-0.45, and the culture was incubated at 37°C overnight.

NMR sample preparation for MSG using stereo-specifically labeled amino acids

For NMR experiments, (U-²H;¹⁵N;¹³CH₃-labeled Ile, Leu and Val) MSG was prepared using BL21(DE3) *ΔilvDΔleuB*. Cells transformed with pET28a-MSG were cultured at 37°C in M9-D₂O glucose medium supplemented with 10 mg/L each of ²H-Val, ²H-Leu and ²H-Ile. When OD₆₀₀ reached 0.3, the cells were collected and washed with M9-D₂O buffer. The cell pellets were then re-suspended with M9-D₂O glucose medium. In order to produce the Val γ₁ methyl labeled MSG, 10 mg/L of L-(γ₁-¹³C; α,β,γ₂-²H₅;α-¹⁵N)-Val (γ₁-Val), 10 mg/L ²H-Leu and 10 mg/L of ²H-Ile were added simultaneously. In order to prepare γ₁-Val, δ₂-Leu and δ₁,γ₂-Ile methyl labeled MSG samples, 10 mg/L each of γ₁-Val, L-[δ₂-¹³C; α,β,γ,δ₁-²H₇;α-¹⁵N]-Leu (δ₂-Leu) (Miyanoiri et al. 2013) and L-[δ₁,γ₂-¹³C; α,β,γ₁-²H₄;α-¹⁵N]-Ile (δ₁,γ₂-Ile) were used. All the cultures were incubated at 37°C, and MSG production was induced by addition of 1 mM IPTG when O.D₆₀₀ reached 0.4-0.45. The culture was incubated overnight. In order to evaluate the efficiency of amino acid incorporation, [γ₁-Val; U-²H,¹⁵N]-MSG was also prepared by using *E. coli* BL21(DE3) pLys S cells in M9-D₂O glucose medium supplemented with 10 mg/L γ₁-Val, 10 mg/L ²H-Leu and 10 mg/L ²H-Ile. The produced MSG was purified by Ni-NTA column chromatography (Tugarinov et al. 2002). The cells were briefly disrupted by sonication. The cell lysate after centrifugation was passed through a Ni- column and the protein was washed and

subsequently eluted with elution buffer consisting of 20 mM Tris-HCl (pH 8), 300 mM NaCl, 10 mM β -mercaptoethanol, and 500 mM imidazole-HCl (pH 8). The eluted fractions were concentrated by using Vivaspin 2 (GE Healthcare) and the concentrated samples (0.1 mM) were dialyzed against NMR buffer consisting of 20 mM Na-phosphate (pH 7.1), 20 mM MgCl_2 , 5 mM DTT and 10% D_2O (Tugarinov and Kay 2003). The isotope incorporation efficiency was determined using GC/MS.

NMR sample preparation using methyl-labeled α -ketoisovalerate

The BL21(DE3) $\Delta ilvD\Delta leuB$ harboring pET28a-MSG was cultured in 30 ml of M9- D_2O glucose medium supplemented with 10 mg/L each ^2H -Val, ^2H -Ile, ^2H -Leu at 37°C. When OD_{600} reached 0.3, the cells were collected and washed with M9- D_2O buffer. The cell pellets were then re-suspended with M9- D_2O glucose medium supplemented with 10.8 mg/L [$3\text{-}^{13}\text{CH}_3;3,4,4,4\text{-}^2\text{H}_4$]- α -ketoisovalerate, 10 mg/L δ_2 -Leu and 10mg/L ^2H -Ile. After incubating the culture for 1 hr, 1 mM IPTG was added and culture was incubated at 37°C overnight. Protein purification was done using Ni-NTA chromatography as described above. The isotope labeling efficiency was determined using GC/MS.

NMR spectroscopy

For the NMR experiment, a slotted Shigemi tube was used (Takeda et al. 2011), and NMR measurements were carried out at 37°C using an Avance III 900 spectrometer, equipped with a TCI cryoprobe (Bruker Biospin). For the 2D ^1H - ^{13}C methyl TROSY experiments (Tugarinov and Kay 2003), the data size and spectral width were 256 (t_1) \times 2048 (t_2) and 5,400 Hz (ω_1 , ^{13}C) \times 14,400 Hz (ω_2 ^1H), respectively. The ^1H and ^{13}C carrier frequencies were 4.7 and 20 ppm,

respectively. The number of scans/FID was 16, and the repetition time was 2s, giving rise to a net acquisition time of 2.5 hrs. In the case of the ^{13}C -edited 3D NOESY-HMQC experiment, the data size and spectral width were $256 (t_1) \times 24 (t_2) \times 2048 (t_3)$ and $14,400 \text{ Hz } (\omega_3, ^1\text{H}) \times 3,200 \text{ Hz } (\omega_2, ^{13}\text{C}) \times 14,400 \text{ Hz } (\omega_3, ^1\text{H})$, respectively. The ^1H and ^{13}C carrier frequencies were 4.7 and 20 ppm, respectively. The number of scans/FID was 16. The NOE mixing time and repetition time were 300ms and 2s, respectively, giving rise to a net acquisition time of 72 hrs. All NMR spectra were processed with the TopSpin software, version 3.1 (Bruker Biospin).

Results

Isolation of the Ile, Leu and Val auxotroph strain

The Keio collection, which is widely used for the construction of deletion mutants, was used to transfer the individual deletion mutations by P1 transduction to the BL21(DE3) strain. Based on the biosynthetic pathway, the dihydroxy-acid dehydratase gene (*ilvD*) was deleted (Fig.18) to construct BL21(DE3) Δ *ilvD*. After colonies were purified, a few colonies were tested in M9-glucose medium containing Km (50 mg/L) in the presence or absence of Ile, Leu and Val (20 mg/L each). Since the BL21(DE3) Δ *ilvD* strain isolated above contained a Km-resistant gene, next I attempted to remove the Km cassette using pCP20 containing the FRT-recombinase-inducible gene. However, I was not able to remove the Km cassette. Thus, I attempted to block the Val-to-Leu conversion before constructing a Δ *ilvD* strain. For this purpose, the NAD(+)-dependent 3-isopropylmalate dehydrogenase gene (*leuB*) was deleted by P1 transduction (Fig.18). After confirmation of the Leu auxotroph, the Km cassette was removed using pCP20. After incubation at 42°C overnight, most of the deletion mutant cells lost pCP20, which was confirmed by three media; M9-glucose containing Leu in the presence of either Km or Amp and in the absence of both antibiotics. Next, I added the deletion mutation of the *ilvD* gene to the BL21(DE3) Δ *leuB* strain. After P1 transduction, the isolated colonies were tested by M9-glucose containing Leu in the presence or absence of Ile and Val.

Optimization of NMR sample preparation using auxotrophic strains

Stereo-specifically $^{13}\text{C}_3$ -labeled amino acids and isotope labeled chemicals are highly expensive. For example, the cost of ^2H -Ile is \$ 866 /0.25 g (Cambridge Isotope Laboratories, Inc.). Therefore, it is important to optimize the labeling conditions to obtain the maximum protein yield. For this purpose, I examined the lowest concentration of amino acids necessary to be able to obtain a sufficient amount of MSG in M9-D₂O glucose medium. It was found that MSG was highly expressed in BL21(DE3) $\Delta ilvD\Delta leuB$, in the presence of 20 mg/L of each amino acid, and the protein yields using deletion mutants were almost identical as that of the wild-type *E. coli* strain, BL21(DE3) pLys S (Fig. 25a). However, the use of a low concentration of amino acids (10 mg/L each) caused a lower MSG expression, even if the cells were growing well. BL21(DE3) $\Delta ilvD\Delta leuB$ was also able to produce other proteins such as EnvZB and calmodulin at similar yields as the wild-type *E. coli* strain (Fig. 25b and 25c, respectively). Thus, BL21(DE3) $\Delta ilvD\Delta leuB$ required 10-20 mg/L of each amino acid for proliferation as well as for optimum protein expression.

In order to establish the most cost effective isotope labeling method, amino acids were added to the BL21(DE3) $\Delta ilvD\Delta leuB$ cultures in a stepwise manner (see **Material and Methods**). On the basis of the results above, 10 mg/L of ^2H -Ile, ^2H -Leu and ^2H -Val each were added to M9-D₂O glucose medium to support growth. After O.D.₆₀₀ reached 0.3 for MSG expression, the cells were washed twice with M9-D₂O buffer to remove the remaining ^2H amino acids. Then, cells were re-suspended into induction medium containing isotope-labeled amino acids, and MSG expression was induced by the addition of 0.1 mM IPTG. I examined the labeling efficiency by using various amino acid concentrations. In

order to analyze isotope incorporation efficiency by GC/MS, unlabeled MSG was prepared using unlabeled Ile, Leu and Val in M9-D₂O glucose medium. When BL21(DE3) pLys S was used for MSG production with 10 mg/L unlabeled Val, its incorporation efficiency was only 60% (Table 3). In wild-type *E. coli*, Ile, Leu and Val are produced from pyruvate (Fig. 18). In a previous study, Val biosynthesis in *E. coli* was found to occur in a very effective manner so that the Val synthetic pathway could not be completely blocked even by an excess addition of Val. At 110 mg/L, Val incorporation efficiency was still only 60-80% (Miyanoiri et al. 2013). This result clearly indicates that the inhibition of Ile, Leu and Val biosynthesis from pyruvate is the most crucial step for Val labeling. When BL21(DE3) $\Delta ilvD\Delta leuB$, lacking genes of hydroxy-acid dehydrogenase and 3-isopropylmalate dehydrogenase was used (Fig. 18), the labeling efficiency of Val improved to 97% using just 10 mg/L Val without any loss in protein yield (Table 3). The incorporation efficiencies of Ile and Leu were also higher than 93% by adding 10 mg/L each to M9-D₂O glucose medium (Table 3). Notably, only 10 mg/L of each isotope labeled amino acid is necessary for preparing an advanced isotope labeled protein, such as amino acid- and stereo-selectively methyl labeled proteins, by using BL21(DE3) $\Delta ilvD\Delta leuB$.

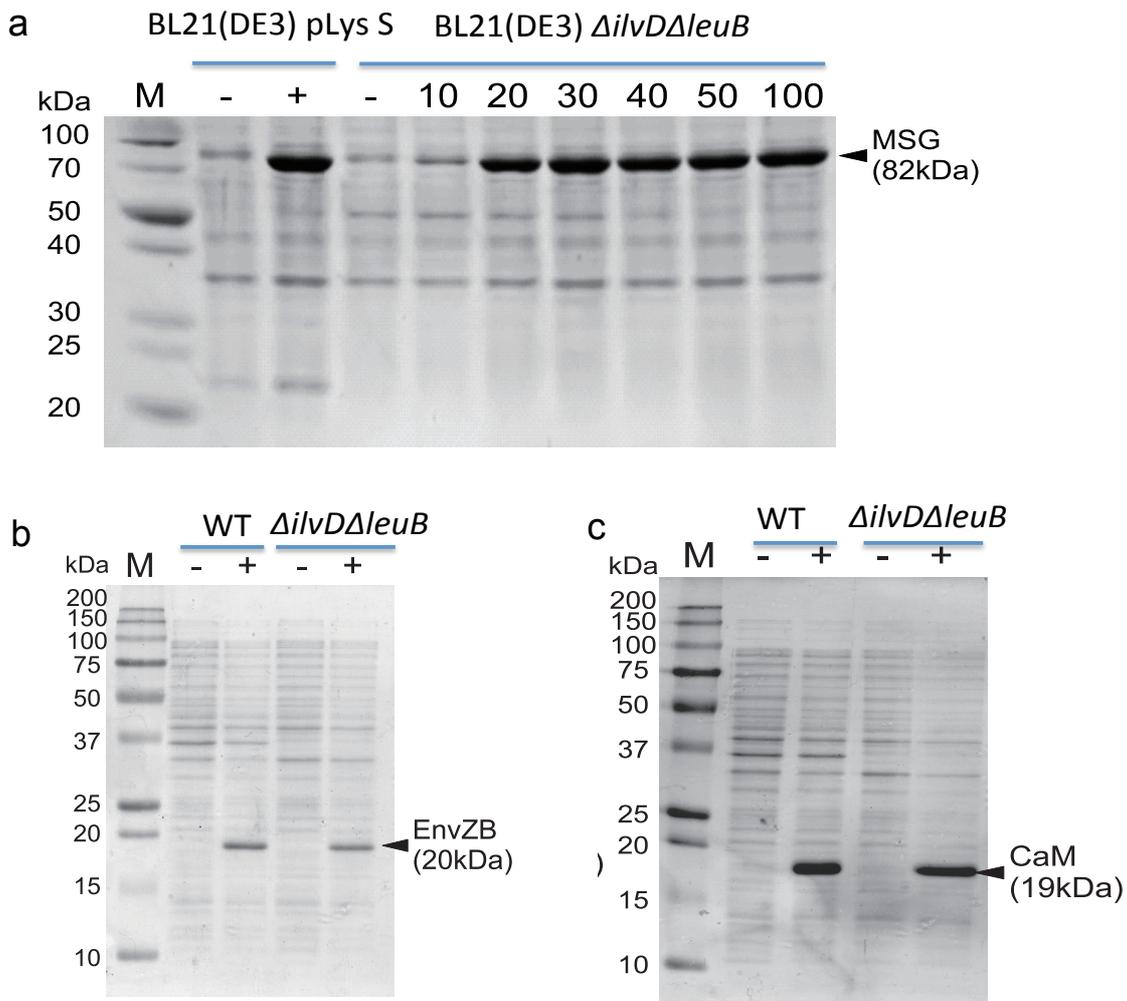


Figure 25. Comparison of expression levels using BL21(DE3) $\Delta ilvD\Delta leuB$ strain and either standard strain BL21(DE3) pLys S or BL21(DE3). These figures are adapted from Miyanoiri *et al.*, (Miyanoiri *et al.* 2016). The cells transformed with either MSG (a) or EnvZB (b) or CaM (c) were grown in M9-D₂O glucose medium supplemented with Ile, Leu and Val (20 mg/L each). When O.D₆₀₀ reached 0.3, the protein production was induced by the addition of 1 mM IPTG. After incubation at 37°C overnight, protein production was analyzed by SDS-PAGE. In order to optimize the isotope labeling conditions for MSG, different concentrations of Ile, Leu and Val (in mg/L) were added (a).

	Ile (mg/L)	Leu (mg/L)	Val (mg/L)	protein yield (mg)	labeling efficiency (%)	strain
Val	10	10	5	0.21	92	
	10	10	10	0.41	97	<i>ΔilvD,ΔleuB</i>
	10	10	20	0.42	98	
	10	10	10	0.41	60	wild-type cells
Leu	10	5	10	0.23	77	
	10	10	10	0.42	93	<i>ΔilvD,ΔleuB</i>
	10	20	10	0.43	94	
Ile	5	10	10	0.22	92	
	10	10	10	0.41	98	<i>ΔilvD,ΔleuB</i>
	20	10	10	0.43	98	

Table 3. Isotope labeling efficiency using BL21(DE3) *ΔilvDΔleuB* and different concentrations of amino acids. Each *E. coli* strain transformed with pET28a-MSG was grown in 30 ml of M9-D₂O glucose medium supplemented with 10 mg/L each ²H-Ile, ²H-Leu and ²H-Val. When O.D₆₀₀ reached 0.3, the cells were collected and washed twice by M9-D₂O buffer and re-suspended into 30 ml of M9-D₂O glucose medium supplemented with various concentrations of amino acids. Red and black colored numbers in the amino acid section indicate labeled (¹H) and unlabeled amino acid (²H), respectively. After incubating for 1-1.5 hr, 1 mM IPTG was added to induce protein production. The culture was further incubated at 37°C overnight. After purification of MSG using Ni-NTA column chromatography, protein yields were measured and labeling efficiency was determined using GC/MS.

Sensitivity enhancement and unambiguous NOE signal analysis by residue- and stereo-specific methyl labeled proteins

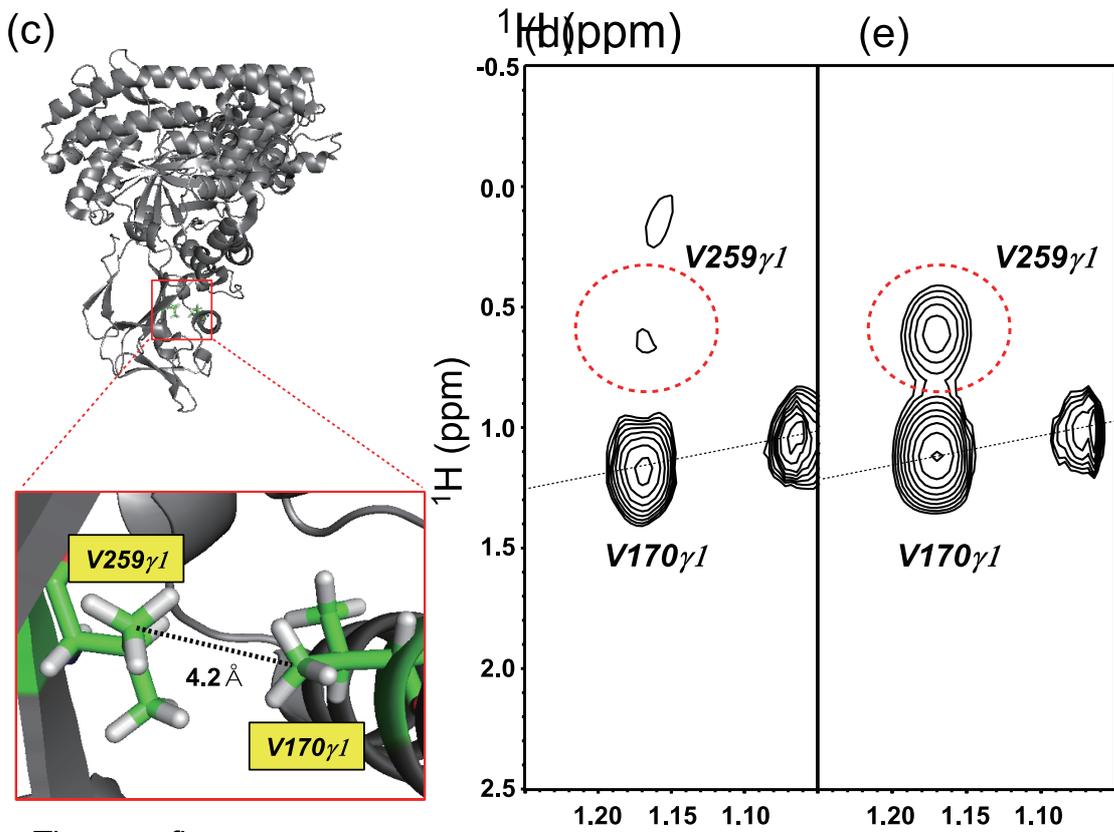
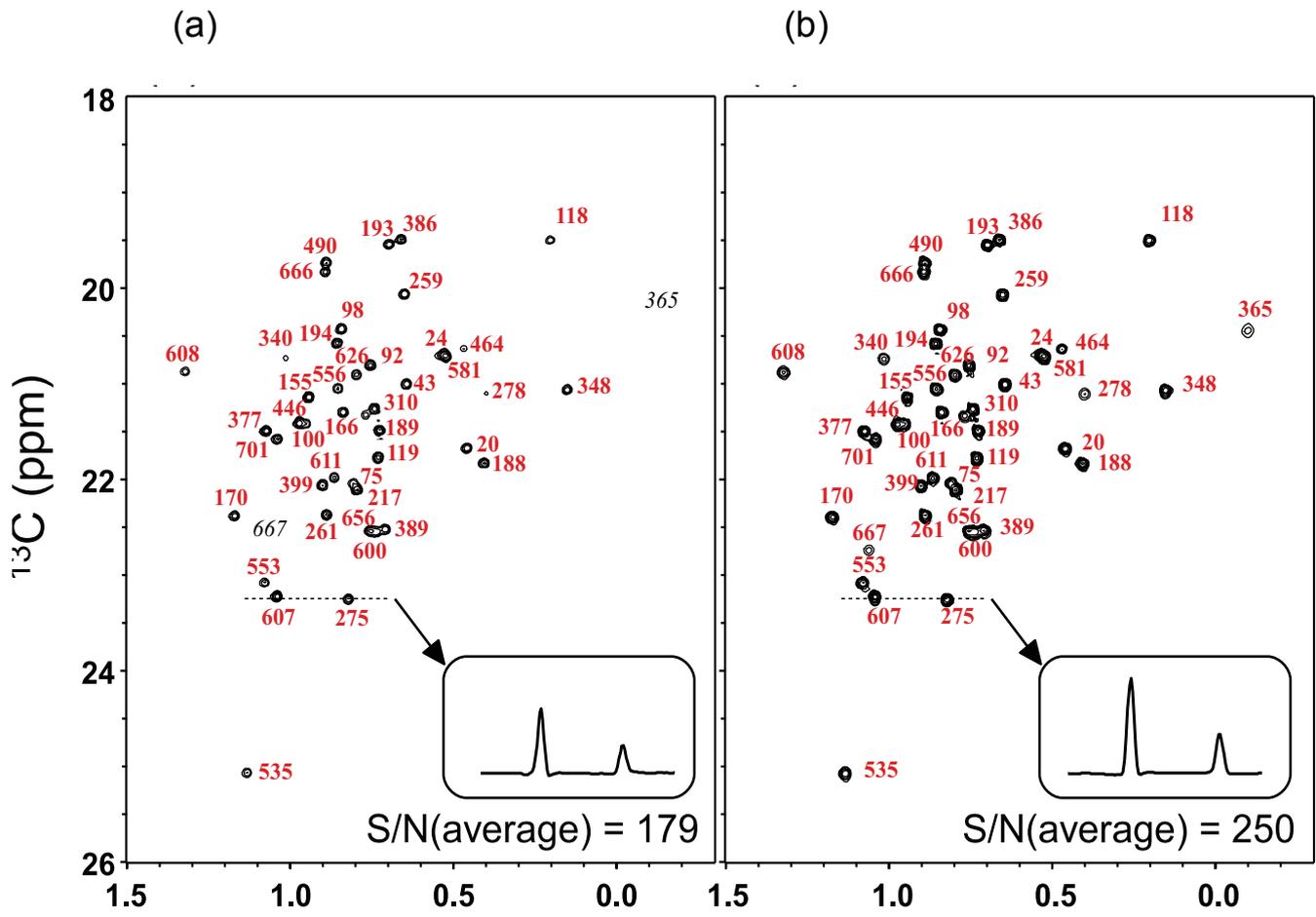
Next, NMR samples (γ_1 -Val; U- ^2H , ^{15}N)-MSG were prepared using BL21(DE3) pLysS and BL21(DE3) $\Delta ilvD\Delta leuB$ strains, respectively, by the procedure described in the **Materials and Methods** section. As a result, 0.1 mM (γ_1 -Val; U- ^2H , ^{15}N)-MSG was obtained from 30 ml M9-D₂O glucose medium from both strains. The ^1H - ^{13}C methyl TROSY spectra for both were measured next. The labeling efficiency of Val was only 60% using 10 mg/L Val with BL21(DE3) pLys S (Table 4), resulting in a relatively low signal intensity of Val- γ_1 methyl group. Thereby, I was not able to observe some of the methyl signals from this sample (Fig. 26). On the other hand, when BL21(DE3) $\Delta ilvD\Delta leuB$ was used under the same conditions, all Val- γ_1 methyl signals in MSG were observed with high sensitivity (Fig. 26b). The average signal to noise ratio (S/N) of all Val γ_1 methyl signals of MSG prepared using BL21(DE3) $\Delta ilvD\Delta leuB$ were about 1.5 times higher than that of the sample prepared using BL21(DE3) pLys S (Fig. 26a and b). These results are consistent with the quantitative analysis of the isotope incorporation efficiency by GC/MS (Table 3).

Since low incorporation efficiency of isotope labeling causes a drastic loss in signal sensitivity of inter-residue NOEs, adjacent residues V170- γ_1 and V259- γ_1 (Fig. 26c) were examined. According to the crystal structure of MSG (Howard et al. 2000), the distance between the two residues is 4.2 Å, thus, strong NOE signals should be observed. In fact, this inter-methyl NOE was very weak from the (γ_1 -Val; U- ^2H , ^{15}N)-MSG sample prepared using BL21(DE3) pLys S and 10mg/L γ_1 -Val (Fig. 26d) while the methyl-methyl NOE signal between V170- γ_1 and V259- γ_1 was clearly observed with a 2.6-time higher sensitivity from the [γ_1 -

Val; U-²H, ¹⁵N]-MSG sample prepared using BL21(DE3) $\Delta ilvD\Delta leuB$ with 10 mg/L γ_1 -Val (Fig. 26e). Since the sensitivity of these NOE is critical for structure determination, this new methyl labeling method is highly useful for structural analysis of low concentration protein samples. In the methyl TROSY and NOESY spectra of [γ_1 -Val; U-²H, ¹⁵N]-MSG prepared using BL21(DE3) $\Delta ilvD\Delta leuB$, I was able to assign all Val- γ_1 methyl signals and no other methyl signals were observed. This indicates that Val biosynthesis from pyruvate was completely suppressed by the deletion mutation.

In this study, stereo-specifically methyl-labeled MSG was prepared using isotope labeled Ile, Leu and Val residues under optimized conditions (Table 3). Indeed, Val- γ_1 , Leu- δ_2 and Ile- δ_1 , γ_2 methyl groups in MSG were successfully labeled and as a result, methyl signals were clearly observed in a methyl-TROSY spectrum. As shown in Figure 27, under optimum conditions consisting of 10 mg/L each Ile, Leu and Val, no amino acid scrambling between these amino acids was detected. Taking advantage of this, it should be possible to produce proteins labeled in different patterns. In addition to the efficient incorporation of isotope-labeled amino acids using BL21(DE3) $\Delta ilvD\Delta leuB$, this strain also has an advantage for effective isotope labeling of Val methyl groups using α -ketoisovalerate, a precursor of Val and Leu. When [3-¹³CH₃;3,4,4,4-²H₄]- α -ketoisovalerate was used with BL21(DE3), the two methyl groups of Leu (ie. δ_1 and δ_2 methyl) and Val (ie. γ_1 and γ_2 methyl) were isotopically labeled without selectively, which likely causes excessive crowding in the methyl-TROSY spectra, especially for large molecular proteins. Moreover, 80-100 mg α -ketoisovalerate was needed for 1L growth medium to maintain up to 90% incorporation efficiency into the prochiral methyl groups of Leu and Val (Goto et

al. 1999). On the other hand, with the new auxotrophic strain constructed in this study, Leu biosynthesis is completely inhibited allowing only Val biosynthesis from α -ketoisovalerate (Fig. 18). Thus, this enables one to perform Val specific-isotope labeling using a low concentration of $[3\text{-}^{13}\text{CH}_3;3,4,4,4\text{-}^2\text{H}_4]\text{-}\alpha$ -ketoisovalerate (Fig. 27). From the methyl-TROSY experiment, the incorporation efficiency of Val was examined using MSG sample prepared with 10.8 mg/L of $[3\text{-}^{13}\text{CH}_3;3,4,4,4\text{-}^2\text{H}_4]\text{-}\alpha$ -ketoisovalerate (78 mM) and BL21(DE3) $\Delta ilvD\Delta leuB$ (Fig. 28a and Fig. 28b), and a sample prepared with BL21(DE3) $\Delta ilvD\Delta leuB$ under optimal conditions, using 10 mg/L of γ_1 -Val (78 μM), respectively (Fig. 28c). The signal to noise (S/N) ratio at V348- γ_1 position in the MSG sample prepared using α -ketoisovalerate was 140 (Fig. 28b) while the S/N ratio from the sample prepared using γ_1 -Val under optimal conditions was 280 (Fig. 28c), indicating that the labeling efficiency of α -ketoisovalerate was about 97%. This indicates that the newly constructed BL21(DE3) $\Delta ilvD\Delta leuB$ strain enables one to carry out Val-specific isotope labeling of proteins utilizing only one-tenth of the amount of α -ketoisovalerate used in the conventional method (Goto et al. 1999). Because commercially available $[3\text{-}^{13}\text{CH}_3;3,4,4,4\text{-}^2\text{H}_4]\text{-}\alpha$ -ketoisovalerate is a racemic mixture of *proS* and *proR*, the sensitivity of Val- γ_1 and Val- γ_2 signal in methyl-TROSY spectra is reduced by 50% (Fig. 28a and Fig. 28b). Nevertheless, it is still highly useful for NMR studies of proteins for their ligand interactions and dynamic analysis.

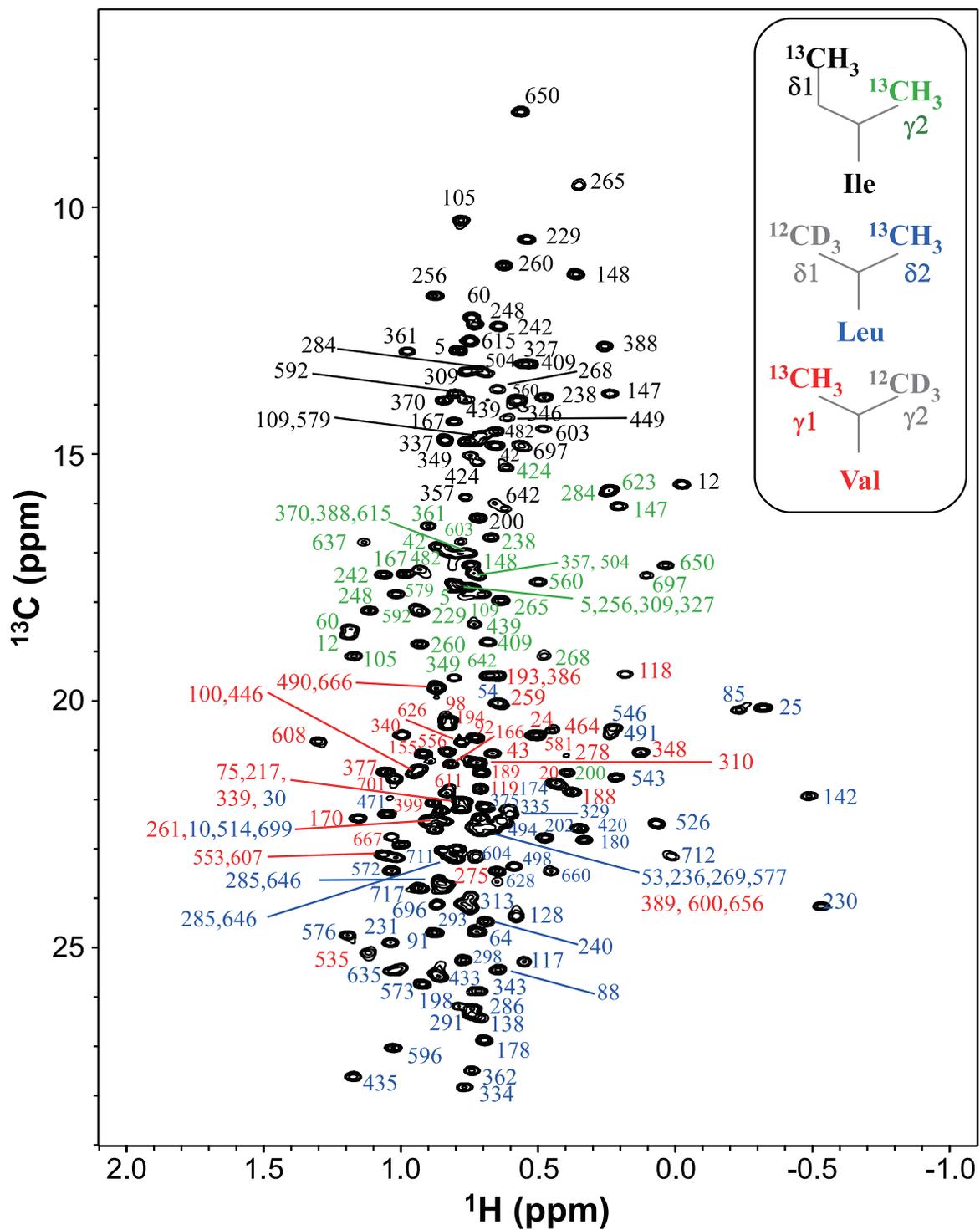


The figure

legends are in the following page

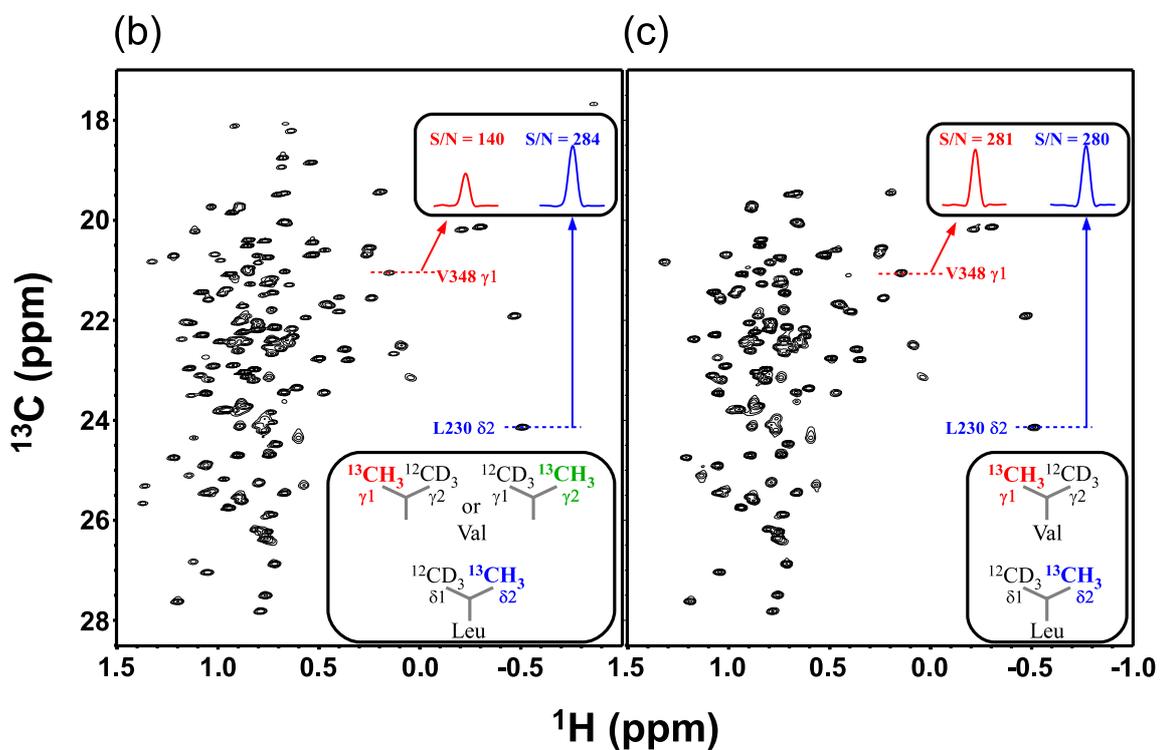
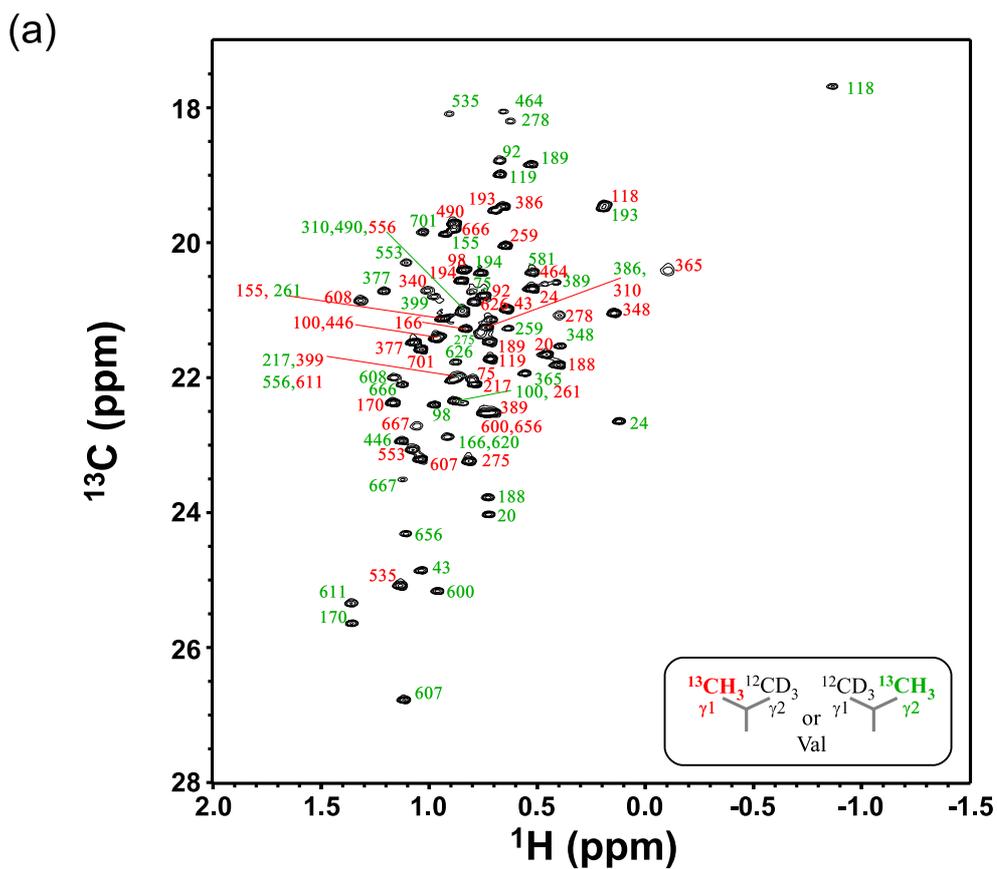
Figure 26. Comparison of NMR signals of γ_1 -Val MSG samples

These figures are adapted from Miyanoiri *et al.*, (Miyanoiri *et al.* 2016). The samples were prepared using standard protocol along with BL21(DE3) pLys S strain or BL21(DE3) $\Delta ilvD\Delta leuB$ strain in M9-D₂O glucose medium supplemented with 10 mg/L each ²H-Leu, ²H-Ile and ¹³C γ_1 -Val. (a) and (b) 2D methyl TROSY spectra observed from 0.1 mM (γ_1 -Val; U-²H, ¹⁵N)-MSG samples prepared using BL21(DE3) pLys S (a) and BL21(DE3) $\Delta ilvD\Delta leuB$ (b). The range of the cross-section, shown as a dotted line, is shown in a box for each spectrum. The signal to noise ratio (S/N) of all Val- γ_1 methyl signals of MSG prepared using BL21(DE3) $\Delta ilvD\Delta leuB$ was about 1.5 times higher than the sample prepared from the BL21(DE3) strain. In each spectrum, signals from Val- γ_1 methyl groups were assigned and shown as red letters (Tugarinov and Kay 2003). The signal from Val- γ_1 667 shown in black was not observed from the sample prepared from BL21(DE3) pLys S due to low signal sensitivities, while the signal was observed from the sample prepared from BL21(DE3) $\Delta ilvD\Delta leuB$. (c) The crystal structure of MSG complexed with magnesium generated from PDB;1D8C (Howard *et al.* 2000). The distance between V170 and V259 is 4.2 Å and shown. (d, e) (¹H, ¹H) plane from the 3D ¹³C edited NOESY-HMQC spectrum at 22.4 ppm in ¹³C position. Signals in (d) are from the (γ_1 -Val; U-²H, ¹⁵N)-MSG samples prepared by using BL21(DE3) pLys S, the signals in (e) are from (γ_1 -Val; U-²H, ¹⁵N)-MSG samples from BL21(DE3) $\Delta ilvD\Delta leuB$. The sensitivity of the methyl-methyl-NOE signals were significantly improved by the sample prepared using BL21(DE3) $\Delta ilvD\Delta leuB$.



The figure legends are in the following page

Figure 27. The NMR spectrum (2D methyl TROSY) using δ_1 , γ_2 -Ile, δ_2 -Leu, γ_1 -Val methyl labeled MSG prepared using the BL21(DE3) $\Delta ilvD\Delta leuB$ strain. This figure is adapted from Miyanoiri *et al.*, (Miyanoiri et al. 2016). The MSG sample was prepared using M9-D₂O glucose medium under optimal conditions. This spectrum was obtained from 0.1 mM MSG solution and data acquisition was 25 hrs. The methyl labeling patterns of Ile, Leu and Val are shown in the box. Assignments for methyl groups of Ile- δ_1 (black), Ile- δ_2 (green), Leu- δ_2 (blue) and Val- γ_1 (red) are shown (Tugarinov and Kay 2003).



The figure legends are in the following page

Figure 28. Methyl-TROSY spectra of MSG samples labeled by Leu and Val along with [3-¹³CH₃;3,4,4,4-²H₄]- α-ketoisovalerate and δ₂-Leu

These figures are adapted from Miyanoiri *et al.*, (Miyanoiri et al. 2016). The *E. coli* strain BL21(DE3) $\Delta ilvD\Delta leuB$ harboring pET-MSG was grown in M9-D₂O glucose medium supplemented with ²H-Ile, ²H-Leu and ²H-Val (10 mg/L each). When O.D₆₀₀ reached 0.3, the cells were collected and washed twice with 1xM9-D₂O buffer and re-suspended into M9-D₂O glucose medium supplemented with (a) ²H-Ile (10 mg/L), [3-¹³CH₃;3,4,4,4-²H₄]- α-ketoisovalerate (10.8 mg/L) and ²H-Leu (10 mg/L), (b) ²H-Ile (10 mg/L), 3-¹³CH₃;3,4,4,4-²H₄]- α-ketoisovalerate (10.8 mg/L) and δ₂-Leu(10 mg/L), (c) ²H-Ile (10 mg/L), γ₁-Val (10 mg/L) and δ₂-Leu(10 mg/L). After incubating for 1-1.5 hr, MSG production was induced by the addition of 1 mM IPTG. After incubating overnight, MSG was purified using Ni-NTA column chromatography and MSG proteins were concentrated and dialyzed against NMR buffer consisting of 20 mM Na-phosphate (pH 7.1), 20 mM MgCl₂, 5 mM DTT and 10% D₂O. All 2D ¹H-¹³C methyl-TROSY spectra were recorded using 0.1 mM MSG samples. The data size and spectra width were 512 (*t*₁) × 2048 (*t*₂) and 2,700 Hz (ω_1 , ¹³C) × 14,400 Hz (ω_2 ¹H), respectively. The ¹³C and ¹H frequencies were 4.7 and 20 ppm, respectively. The number of scans/FID was 32. The repetition time was 1.5s. Although α-ketoisovalerate is a precursor for Leu and Val, selective Val labeling can be done by inhibiting the conversion from α-ketoisovalerate to Leu ($\Delta leuB$) (a). Since 3-¹³CH₃;3,4,4,4-²H₄]- α-ketoisovalerate is a mixture of *proR* and *proS*, signal to noise (S/N) is half in comparison to MSG sample labeled in a stereo-specific manner (b, c). The isotope labeling patterns of the Leu and Val methyl groups are shown in the box

for each spectrum. Methyl assignments for Leu and Val are shown in each spectrum (Tugarinov and Kay 2003).

Discussion

Assignments of methyl groups give important information for the structural study of large molecular weight proteins, however, it becomes exponentially more difficult as the molecular weight of the protein increases. In order to obtain methyl signals, selectively labeled compounds, α -ketobutyrate for Ile labeling and α -ketoisovalerate for Leu and Val labeling, have been commonly used with deuterated backgrounds (Goto et al. 1999; Tugarinov and Kay 2003). Using these compounds, selective labeling of (1) Ile, Val and Leu, or (2) Ile and Val have been successful. However, these compounds are in a racemic mixture, thus signal intensities were 50% compared to a stereo-specific labeling method. Recently, stereo-specifically labeled acetolactate was synthesized from ethyl 3-oxobutanoate (Gans et al. 2010), and has been applied to large molecular weight proteins (Mas et al. 2013; Kerfah et al. 2015). Methyl signals from Val and Leu still cause overcrowded spectra making it very difficult to assign all the methyl signals and consequently difficult to obtain accurate structural information. To overcome this issue, in Section III-I, I developed a new expression system using deletion mutations. The major drawbacks for stereo specific labeling by acetolactate are the high cost and the lack of a selective Leu labeling strategy. Thereby, developing an economical selective labeling system was essential. In this study, I developed acetolactate auxotrophic strains consisting of BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$, $\Delta leuB$ for Val labeling (strain B) and BL21(DE3) $\Delta hisB$, $\Delta trpC$, $\Delta ilvI$, $\Delta ilvG$, $\Delta ilvB$, $\Delta ilvE$, $\Delta avtA$ (strain C) for Leu labeling. Using strain B along with 40 mg/L ^{13}C -*proS*-acetolactate, I was able to obtain only Val-*proS* signals from the ^{13}C -HMQC spectrum. This presents an almost 80% cost saving compared to using the standard strain (Kerfah et al. 2015). Additionally, I

did not observe a significant loss of protein expression using this strain, thus, this system is highly useful for sample preparation. On the other hand, although protein expression levels using strain C along with L¹³C-*proS*-acetolactate were about 40% in comparison to the standard strain, Leu- and stereo-specific methyl labeling were successfully achieved. Next, I examined combining the Ile labeling system with the Leu or Val labeling system using aceto-hydroxybutanoate and acetolactate. Aceto-hydroxybutanoate and acetolactate are both substrates for ketol-acid reductoisomerase (KARI) with aceto-hydroxybutanoate having a 5-8 times higher affinity to KARI than acetolactate, making controlling isotope incorporation efficiency difficult. Thus, I examined various concentrations to equalize the signal intensities, however, aceto-hydroxybutanoate was preferentially used even when a lower concentration of aceto-hydroxybutanoate and a higher concentration of acetolactate were applied. Therefore, the Ile labeling system is incompatible with the selective Leu labeling system. Taking all of this into consideration, this newly developed Leu or Val selective labeling system is still highly useful for the NMR community and will be a standard method for methyl assignments for large molecular weight proteins.

A straightforward method to obtain methyl signals is to use methyl specific labeled amino acids such as SAIL amino acids. However, a limitation of SAIL amino acids is that it requires a cell-free expression system since endogenous amino acids are relatively high and will cause poor isotope incorporation. In the case of Leu labeling, feeding amino acid in stepwise manner was successful for high labeling efficiency (Miyanoiri et al. 2013). The disadvantages of SAIL amino acids were (1) each protein needs an optimum expression protocol for high labeling efficiency, (2) SAIL amino acids are highly expensive (\$5,000/50mg),

and (3) an efficient Val labeling system was not available. To circumvent these issues, in section III-II, I constructed BL21(DE3) $\Delta ilvD\Delta leuB$ to block the biosynthesis of Val from pyruvate. In order for a scramble-free isotope labeling system, firstly, the auxotrophy to Ile, Leu and Val is essential. In addition, it is very critical that only exogenous amino acids are used for protein production, not to be diluted by endogenous amino acids from protein degradation or amino acid pools in the cells. For this purpose, the *ilvD* gene, encoding dihydroxyacid dehydratase was deleted. Using this strain, not only isotope-labeled amino acids but also α -ketoisovalerate can be used for selective labeling. Next, since amino acid scrambling from Val to Leu was detected even with the addition of exogenous Leu in the medium (Mas et al. 2013), the conversion from Val/ α -ketoisovalerate to Leu was blocked using deletion mutation. For this purpose, from the multiple reactions in the conversion of α -ketoisovalerate to Leu (Fig. 18), *leuB* gene encoding 3-isopropylmalate dehydrogenase was deleted. This gene deletion mutation was sufficient to make the cells auxotrophic to Leu, indicating that the conversion from Val/ α -ketoisovalerate was totally blocked. In this study, I developed an efficient Val labeling system where the usage of Val was reduced down to 10 mg/L without losing protein production efficiency as well as labeling efficiency. In many cases, auxotrophic strains have less protein expression levels in comparison to standard strains, however, I did not observe a reduction of protein expression levels between BL21(DE3) $\Delta ilvD\Delta leuB$ and BL21(DE3). In addition, since amino acid conversion from α -ketoisovalerate was blocked, the combination of labeling using isotope labeled Val, Leu and Ile with high labeling efficiency became feasible. The combination of Ile, Leu and Val labeling is only

possible using SAIL amino acids, therefore, our new expression system is highly useful for the NMR community.

Summary and Conclusions

In the present dissertation, firstly, I established an antimicrobial peptide production system using PST-SPP technology, which is potentially useful for the production of many antimicrobial peptides. In some cases, there have been successful antimicrobial peptide productions using fusion tags such as maltose binding protein and small ubiquitin-like modifier proteins, however, there has not been any effective fusion tags to express Bac7(1-35) in its functional form. This is possibly because one of the major targets of Bac7(1-35) is ribosomes, thus protein synthesis would be blocked once Bac7(1-35) is produced. Indeed, I observed fairly low expression levels even when PrS₂ was fused as an expression tag, but I was able to enhance the expression level by using the SPP system, taking advantage of an ACA-specific endoribonuclease, MazF. Since PrS₂ is known not to interfere with the function of the fusion partner, PrS₂-Bac7(1-35) retains its activity without cleaving the PrS₂ tag. Since its target was co-purified during purification, identification of Bac7(1-35) should be possible and is being currently investigated. Likewise, produced antimicrobial peptides should retain their activity, therefore the PST-SPP system is useful not only for producing antimicrobial peptides but also in identifying their targets. Identification of the targets of antibiotics is extremely important for drug development, thus this expression system will be a standard method for antimicrobial peptide production and for further analysis of its mechanisms.

In section II, I established an amino acid analogue incorporation system using an auxotrophic strain together with the SPP system. Incorporation of toxic amino acid analogues such as canavanine (Can) into a protein without losing function is

very challenging because the pKa of the guanidino group in Can is 7 while that of Arg is 11. Although there is very little structural difference, the charge network is drastically changed, leading to lower protein stability. To suppress the toxicity of canavanine, I applied the SPP system so that exogenous amino acid are used for only target protein production since all other mRNA are cleaved by MazF. One of the difficulties of amino acid analogue incorporation is due to endogenous amino acids in the cells, causing competition between these amino acids and its analogues. In addition to that, the higher affinity of amino acids to their tRNA synthetase results in poor amino acid analogue incorporation. To circumvent this, I applied an Arg deletion mutation by P1 transduction. Although the affinity of Arg to tRNA is 100 times higher than that of Can, Can was successfully used for protein production using an Arg auxotrophic strain. As an example, I applied the method to MazFbs, a UACAU specific endoribonuclease. In order for the complete replacement, I applied a dual inducible expression system using a His deletion mutant together with MazF(Δ H), a MazF with no His residues. In this system, MazF(Δ H) was expressed in the absence of His without producing MazFbs. Thus, in combination with Arg auxotrophy, absolutely no endogenous Arg remained before the induction of MazFbs(Can), and the labeling efficiency of MazFbs(Can) was 98% after second step purification. After functional analysis of MazFbs(Can), it was found that MazFbs(Can) requires an extra A residue (UACAUA) at the end of the MazFbs recognition sequence UACAU. This functional alteration of endoribonuclease recognition sequence is highly useful for gene regulations applicable to cancer and antiviral therapies as specific gene interference can be achieved by a tailor-made longer recognition enzyme. For this purpose, expanding the repertoire of recognition sequence is essential and

thus, characterizing MazF homologues from bacteria to archaea is necessary. Modifying protein for higher stability or enhancing the activity is highly important for industrial purposes. Currently, modification using site-directed mutagenesis for residue-specific analysis is used, however, amino acid replacement using auxotrophic cells together with the SPP system will be another method to discover new functional proteins. Additionally, the partnering of auxotrophic cell lines and the SPP system provides further tools to express and study traditionally difficult-to-express proteins. For example, a Proline analogue, 2-azetidine-carboxylic acid (Aze) causes abnormally structured proteins, thus making it difficult to express in *E. coli*. One of the neuron diseases, multiple sclerosis is thought to be caused by Aze incorporation into myelin basic protein which is believed to be important in the process of myelination of nerves in the nervous system. Since producing an Aze-containing protein is difficult, detailed studies have not been done. Using newly developed technology, we may be able to produce Aze-containing myelin protein using a proline auxotrophic strain, which will help us to understand the mechanism of nerve disease

In section III, I established residue- and stereo- methyl specific labeling using two different approaches. In section III-I, I established a residue- and stereo-specific labeling system using acetolactate auxotrophic strains along with stereo-specifically labeled acetolactate, which is one of the Leu and Val precursors. Using this labeling system, selective Leu-*proS* or Val-*proS* labeling was demonstrated using the MSG protein, which is considered a standard protein of a large molecular weight. After optimization, the usage of isotope labeled acetolactate was reduced from 300 mg/L to 40 mg/L, which presents a more than

80% savings. Although stereo-specific labeling is not still commonly used because of the high cost of acetolactate, this new labeling system will provide NMR scientists with an economical approach to preparing samples. α -Ketoisovalerate, a racemic mixture of *proR* and *proS*, which causes overcrowded signal peaks in the spectra, is still commonly used as it is less expensive than acetolactate. Since this compound is still effective to study relatively large proteins of more than 50 kDa, Leu or Val, selective labeling using α -ketoisovalerate used in conjunction with acetolactate will further help us determine methyl assignments. In this study, it was concluded that Ile labeling is incompatible with the Leu labeling system. However, it may be possible using α -ketoisocaproate precursors since this compound is completely different from ^{13}C -aceto-hydroxybutanoate which along with acetolactate is a substrate for KARI. However, the combination of Ile and Leu labeling still needs to be examined when α -ketoisocaproate becomes commercially available.

In section III-II, I established another residue- and stereo-labeling system using -- auxotrophic strains. By using a newly developed strain, Val- and stereo-specific labeling was achieved, which was not previously available. The advantage of this strain is that there is no amino acid scrambling from isotope labeled amino acids, which gives precise structural information. Another advantage of this system is that any possible labeling combination including Ile labeling can be done. Since SAIL amino acids are commercially available, this labeling system will be highly useful. Lastly, this strain has the same protein expression level as the standard strain, which is important for NMR sample preparations. Although SAIL amino acids are known to be extremely expensive (\$5,000/50mg), the labeling system established here is highly economical, enabling the SAIL technology to be used

more widely in the NMR community. Expanding this methodology to gain structural information of large molecular proteins by NMR not just from Ile, Leu and Val, but also aromatic residues such as Trp, Phe and Tyr is also potentially significant. Since SAIL amino acids and precursors for Phe, Tyr and Trp are available, it is possible to combine Ile, Leu and Val auxotrophy with Phe, Tyr and Trp precursors to provide further tools for the NMR community. In addition, yeast systems have been used to express eukaryotic proteins, however, there is no effective Leu or Val methyl-specific labeling system available to date. Thus, this deletion mutation methodology may be applied to yeast systems for NMR study.

References

- Acton, T.B., Xiao, R., Anderson, S., Aramini, J., Buchwald, W.A., Ciccocanti, C., Conover, K., Everett, J., Hamilton, K., Huang, Y.J., Janjua, H., Kornhaber, G., Lau, J., Lee, D.Y., Liu, G., Maglaqui, M., Ma, L., Mao, L., Patel, D., Rossi, P., Sahdev, S., Shastry, R., Swapna, G.V., Tang, Y., Tong, S., Wang, D., Wang, H., Zhao, L. and Montelione, G.T. (2011) Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol* **493**, 21-60.
- Aizenman, E., Engelberg-Kulka, H. and Glaser, G. (1996) An *Escherichia coli* chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc Natl Acad Sci U S A* **93**, 6059-6063.
- Anderson, R.C. and Yu, P.L. (2003) Isolation and characterisation of proline/arginine-rich cathelicidin peptides from ovine neutrophils. *Biochem Biophys Res Commun* **312**, 1139-1146.
- Ayala, I., Hamelin, O., Amero, C., Pessey, O., Plevin, M.J., Gans, P. and Boisbouvier, J. (2012) An optimized isotopic labelling strategy of isoleucine- γ 2 methyl groups for solution NMR studies of high molecular weight proteins. *Chem Commun (Camb)* **48**, 1434-1436.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008.
- Bagby, S., Harvey, T.S., Eagle, S.G., Inouye, S. and Ikura, M. (1994a) NMR-derived three-dimensional solution structure of protein S complexed with calcium. *Structure* **2**, 107-122.
- Bagby, S., Harvey, T.S., Eagle, S.G., Inouye, S. and Ikura, M. (1994b) Structural similarity of a developmentally regulated bacterial spore coat protein to beta gamma-crystallins of the vertebrate eye lens. *Proc Natl Acad Sci U S A* **91**, 4308-4312.
- Benincasa, M., Scocchi, M., Podda, E., Skerlavaj, B., Dolzani, L. and Gennaro, R. (2004) Antimicrobial activity of Bac7 fragments against drug-resistant clinical isolates. *Peptides* **25**, 2055-2061.
- Benincasa, M., Zahariev, S., Pelillo, C., Milan, A., Gennaro, R. and Scocchi, M. (2015) PEGylation of the peptide Bac7(1-35) reduces renal clearance while retaining antibacterial activity and bacterial cell penetration capacity. *Eur J Med Chem* **95**, 210-219.

- Bessonov, K., Bamm, V.V. and Harauz, G. (2010) Misincorporation of the proline homologue Aze (azetidine-2-carboxylic acid) into recombinant myelin basic protein. *Phytochemistry* **71**, 502-507.
- Blower, T.R., Fineran, P.C., Johnson, M.J., Toth, I.K., Humphreys, D.P. and Salmond, G.P. (2009) Mutagenesis and functional characterization of the RNA and protein components of the toxIN abortive infection and toxin-antitoxin locus of *Erwinia*. *J Bacteriol* **191**, 6029-6039.
- Blower, T.R., Pei, X.Y., Short, F.L., Fineran, P.C., Humphreys, D.P., Luisi, B.F. and Salmond, G.P. (2011) A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nat Struct Mol Biol* **18**, 185-190.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-254.
- Budisa, N., Wenger, W. and Wiltschi, B. (2010) Residue-specific global fluorination of *Candida antarctica* lipase B in *Pichia pastoris*. *Mol Biosyst* **6**, 1630-1639.
- Burley, S.K., Almo, S.C., Bonanno, J.B., Capel, M., Chance, M.R., Gaasterland, T., Lin, D., Sali, A., Studier, F.W. and Swaminathan, S. (1999) Structural genomics: beyond the human genome project. *Nat Genet* **23**, 151-157.
- Bélanger, M. and Moineau, S. (2015) Mutational Analysis of the Antitoxin in the Lactococcal Type III Toxin-Antitoxin System AbiQ. *Appl Environ Microbiol* **81**, 3848-3855.
- Chahardooli, M., Niazi, A., Aram, F. and Sohrabi, S.M. (2016) Expression of recombinant Arabian camel lactoferricin-related peptide in *Pichia pastoris* and its antimicrobial identification. *J Sci Food Agric* **96**, 569-575.
- Chen, X., Shi, J., Chen, R., Wen, Y., Shi, Y., Zhu, Z., Guo, S. and Li, L. (2015) Molecular chaperones (TrxA, SUMO, Intein, and GST) mediating expression, purification, and antimicrobial activity assays of plectasin in *Escherichia coli*. *Biotechnol Appl Biochem* **62**, 606-614.
- Christensen-Dalsgaard, M., Jørgensen, M.G. and Gerdes, K. (2010) Three new RelE-homologous mRNA interferases of *Escherichia coli* differentially induced by environmental stresses. *Mol Microbiol* **75**, 333-348.
- Daher, K.A., Lehrer, R.I., Ganz, T. and Kronenberg, M. (1988) Isolation and characterization of human defensin cDNA clones. *Proc Natl Acad Sci U S A* **85**, 7327-7331.

- Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645.
- Dubendorff, J.W. and Studier, F.W. (1991) Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol* **219**, 45-59.
- Dumas, R., Biou, V., Halgand, F., Douce, R. and Duggleby, R.G. (2001) Enzymology, structure, and dynamics of acetohydroxy acid isomeroreductase. *Acc Chem Res* **34**, 399-408.
- Feng, X., Liu, C., Guo, J., Bi, C., Cheng, B., Li, Z. and Shan, A. (2010) Expression and purification of an antimicrobial peptide, bovine lactoferricin derivative LfcinB-W10 in *Escherichia coli*. *Curr Microbiol* **60**, 179-184.
- Feng, X.J., Xing, L.W., Liu, D., Song, X.Y., Liu, C.L., Li, J., Xu, W.S. and Li, Z.Q. (2014) Design and high-level expression of a hybrid antimicrobial peptide LF15-CA8 in *Escherichia coli*. *J Ind Microbiol Biotechnol* **41**, 527-534.
- Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S. and Salmond, G.P. (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A* **106**, 894-899.
- Fosgerau, K. and Hoffmann, T. (2015) Peptide therapeutics: current status and future directions. *Drug Discov Today* **20**, 122-128.
- Fox, J.L. (2013) Antimicrobial peptides stage a comeback. *Nat Biotechnol* **31**, 379-382.
- Gans, P., Hamelin, O., Sounier, R., Ayala, I., Durá, M.A., Amero, C.D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M.J. and Boisbouvier, J. (2010) Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew Chem Int Ed Engl* **49**, 1958-1962.
- Gardner, K.H., Rosen, M.K. and Kay, L.E. (1997) Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* **36**, 1389-1401.
- Gogos, A., Mu, H., Bahna, F., Gomez, C.A. and Shapiro, L. (2003) Crystal structure of YdcE protein from *Bacillus subtilis*. *Proteins* **53**, 320-322.
- Goto, N.K., Gardner, K.H., Mueller, G.A., Willis, R.C. and Kay, L.E. (1999) A robust and cost-effective method for the production of Val, Leu, Ile (δ 1) methyl-protonated ^{15}N -, ^{13}C -, ^2H -labeled proteins. *J Biomol NMR* **13**, 369-374.

- Gross, J.D., Gelev, V.M. and Wagner, G. (2003) A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes. *J Biomol NMR* **25**, 235-242.
- Grunberg-Manago, M. (1999) Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annu Rev Genet* **33**, 193-227.
- Guida, F., Benincasa, M., Zahariev, S., Scocchi, M., Berti, F., Gennaro, R. and Tossi, A. (2015) Effect of size and N-terminal residue characteristics on bacterial cell penetration and antibacterial activity of the proline-rich peptide Bac7. *J Med Chem* **58**, 1195-1204.
- Hargreaves, D., Santos-Sierra, S., Giraldo, R., Sabariego-Jareño, R., de la Cueva-Méndez, G., Boelens, R., Díaz-Orejas, R. and Rafferty, J.B. (2002) Structural and functional analysis of the kid toxin protein from *E. coli* plasmid R1. *Structure* **10**, 1425-1433.
- Hazan, R., Sat, B. and Engelberg-Kulka, H. (2004) *Escherichia coli* *mazEF*-mediated cell death is triggered by various stressful conditions. *J Bacteriol* **186**, 3663-3669.
- Hazan, R., Sat, B., Reches, M. and Engelberg-Kulka, H. (2001) Postsegregational killing mediated by the P1 phage "addiction module" *phd-doc* requires the *Escherichia coli* programmed cell death system *mazEF*. *J Bacteriol* **183**, 2046-2050.
- Hendrickson, T.L., de Crécy-Lagard, V. and Schimmel, P. (2004) Incorporation of nonnatural amino acids into proteins. *Annu Rev Biochem* **73**, 147-176.
- Herbel, V., Schäfer, H. and Wink, M. (2015) Recombinant Production of Snakin-2 (an Antimicrobial Peptide from Tomato) in *E. coli* and Analysis of Its Bioactivity. *Molecules* **20**, 14889-14901.
- Hirashima, A. and Inouye, M. (1973) Specific biosynthesis of an envelope protein of *Escherichia coli*. *Nature* **242**, 405-407.
- Howard, B.R., Endrizzi, J.A. and Remington, S.J. (2000) Crystal structure of *Escherichia coli* malate synthase G complexed with magnesium and glyoxylate at 2.0 Å resolution: mechanistic implications. *Biochemistry* **39**, 3156-3168.
- Igloi, G.L. and Schiefermayr, E. (2009) Amino acid discrimination by arginyl-tRNA synthetases as revealed by an examination of natural specificity variants. *FEBS J* **276**, 1307-1318.

- Inouye, M. (2006) The discovery of mRNA interferases: implication in bacterial physiology and application to biotechnology. *J Cell Physiol* **209**, 670-676.
- Inouye, M., Inouye, S. and Zusman, D.R. (1979) Biosynthesis and self-assembly of protein S, a development-specific protein of *Myxococcus xanthus*. *Proc Natl Acad Sci U S A* **76**, 209-213.
- Ishida, Y. and Inouye, M. (2016) Suppression of the toxicity of Bac7 (1-35), a bovine peptide antibiotic, and its production in *E. coli*. *AMB Express* **6**, 19.
- Ishida, Y., Park, J.H., Mao, L., Yamaguchi, Y. and Inouye, M. (2013) Replacement of all arginine residues with canavanine in MazF-bs mRNA interferase changes its specificity. *J Biol Chem* **288**, 7564-7571.
- Jahn, N., Preis, H., Wiedemann, C. and Brantl, S. (2012) BsrG/SR4 from *Bacillus subtilis*--the first temperature-dependent type I toxin-antitoxin system. *Mol Microbiol* **83**, 579-598.
- Jayamani, E., Rajamuthiah, R., Larkins-Ford, J., Fuchs, B.B., Conery, A.L., Vilcinskis, A., Ausubel, F.M. and Mylonakis, E. (2015) Insect-derived cecropins display activity against *Acinetobacter baumannii* in a whole-animal high-throughput *Caenorhabditis elegans* model. *Antimicrob Agents Chemother* **59**, 1728-1737.
- Jørgensen, M.G., Pandey, D.P., Jaskolska, M. and Gerdes, K. (2009) HicA of *Escherichia coli* defines a novel family of translation-independent mRNA interferases in bacteria and archaea. *J Bacteriol* **191**, 1191-1199.
- Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Mei Ono, A. and Güntert, P. (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**, 52-57.
- Kamada, K., Hanaoka, F. and Burley, S.K. (2003) Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition. *Mol Cell* **11**, 875-884.
- Kawano, M., Aravind, L. and Storz, G. (2007) An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol Microbiol* **64**, 738-754.
- Kawano, M., Oshima, T., Kasai, H. and Mori, H. (2002) Molecular characterization of long direct repeat (LDR) sequences expressing a stable mRNA encoding for a 35-amino-acid cell-killing peptide and a cis-encoded small antisense RNA in *Escherichia coli*. *Mol Microbiol* **45**, 333-349.

- Kedzierska, B., Lian, L.Y. and Hayes, F. (2007) Toxin-antitoxin regulation: bimodal interaction of YefM-YoeB with paired DNA palindromes exerts transcriptional autorepression. *Nucleic Acids Res* **35**, 325-339.
- Kerfah, R., Plevin, M.J., Pessey, O., Hamelin, O., Gans, P. and Boisbouvier, J. (2015) Scrambling free combinatorial labeling of alanine- β , isoleucine- δ 1, leucine-proS and valine-proS methyl groups for the detection of long range NOEs. *J Biomol NMR* **61**, 73-82.
- Kobayashi, H., Swapna, G.V., Wu, K.P., Afinogenova, Y., Conover, K., Mao, B., Montelione, G.T. and Inouye, M. (2012) Segmental isotope labeling of proteins for NMR structural study using a protein S tag for higher expression and solubility. *J Biomol NMR* **52**, 303-313.
- Kobayashi, H., Yoshida, T. and Inouye, M. (2009) Significant enhanced expression and solubility of human proteins in *Escherichia coli* by fusion with protein S from *Myxococcus xanthus*. *Appl Environ Microbiol* **75**, 5356-5362.
- Li, J.F., Zhang, J., Zhang, Z., Kang, C.T. and Zhang, S.Q. (2011) SUMO mediating fusion expression of antimicrobial peptide CM4 from two joined genes in *Escherichia coli*. *Curr Microbiol* **62**, 296-300.
- Lichtenecker, R.J., Coudeville, N., Konrat, R. and Schmid, W. (2013) Selective isotope labelling of leucine residues by using α -ketoacid precursor compounds. *Chembiochem* **14**, 818-821.
- Luan, C., Zhang, H.W., Song, d.G., Xie, Y.G., Feng, J. and Wang, Y.Z. (2014) Expressing antimicrobial peptide cathelicidin-BF in *Bacillus subtilis* using SUMO technology. *Appl Microbiol Biotechnol* **98**, 3651-3658.
- Mao, L., Tang, Y., Vaiphei, S.T., Shimazu, T., Kim, S.G., Mani, R., Fakhoury, E., White, E., Montelione, G.T. and Inouye, M. (2009) Production of membrane proteins for NMR studies using the condensed single protein (cSPP) production system. *J Struct Funct Genomics* **10**, 281-289.
- Mardirossian, M., Grzela, R., Giglione, C., Meinnel, T., Gennaro, R., Mergaert, P. and Scocchi, M. (2014) The host antimicrobial peptide Bac71-35 binds to bacterial ribosomal proteins and inhibits protein synthesis. *Chem Biol* **21**, 1639-1647.
- Marianovsky, I., Aizenman, E., Engelberg-Kulka, H. and Glaser, G. (2001) The regulation of the *Escherichia coli mazEF* promoter involves an unusual alternating palindrome. *J Biol Chem* **276**, 5975-5984.
- Mas, G., Crublet, E., Hamelin, O., Gans, P. and Boisbouvier, J. (2013) Specific labeling and assignment strategies of valine methyl groups for

NMR studies of high molecular weight proteins. *J Biomol NMR* **57**, 251-262.

- Miller, K.W., Schamber, R., Chen, Y. and Ray, B. (1998) Production of active chimeric pediocin AcH in *Escherichia coli* in the absence of processing and secretion genes from the *Pediococcus pap* operon. *Appl Environ Microbiol* **64**, 14-20.
- Minaba, M. and Kato, Y. (2014) High-yield, zero-leakage expression system with a translational switch using site-specific unnatural amino Acid incorporation. *Appl Environ Microbiol* **80**, 1718-1725.
- Miyanoiri, Y., Ishida, Y., Takeda, M., Terauchi, T., Inouye, M. and Kainosho, M. (2016) Highly efficient residue-selective labeling with isotope-labeled Ile, Leu, and Val using a new auxotrophic *E. coli* strain. *J Biomol NMR* **65**, 109-119.
- Miyanoiri, Y., Takeda, M., Okuma, K., Ono, A.M., Terauchi, T. and Kainosho, M. (2013) Differential isotope-labeling for Leu and Val residues in a protein by *E. coli* cellular expression using stereo-specifically methyl labeled amino acids. *J Biomol NMR* **57**, 237-249.
- Monneau, Y.R., Ishida, Y., Rossi, P., Saio, T., Tzeng, S.R., Inouye, M. and Kalodimos, C.G. (2016) Exploiting *E. coli* auxotrophs for leucine, valine, and threonine specific methyl labeling of large proteins for NMR applications. *J Biomol NMR* **65**, 99-108.
- Motiejūnaite, R., Armalyte, J., Markuckas, A. and Suziedeliene, E. (2007) *Escherichia coli* *dinJ-yafQ* genes act as a toxin-antitoxin module. *FEMS Microbiol Lett* **268**, 112-119.
- Mulder, K.C., de Lima, L.A., Aguiar, P.S., Carneiro, F.C., Franco, O.L., Dias, S.C. and Parachin, N.S. (2015) Production of a modified peptide clavanin in *Pichia pastoris*: cloning, expression, purification and in vitro activities. *AMB Express* **5**, 129.
- Nariya, H. and Inouye, M. (2008) MazF, an mRNA interferase, mediates programmed cell death during multicellular *Myxococcus* development. *Cell* **132**, 55-66.
- Oddo, A., Thomsen, T.T., Kjelstrup, S., Gorey, C., Franzyk, H., Frimodt-Møller, N., Løbner-Olesen, A. and Hansen, P.R. (2015) An Amphipathic Undecapeptide with All d-Amino Acids Shows Promising Activity against Colistin-Resistant Strains of *Acinetobacter baumannii* and a Dual Mode of Action. *Antimicrob Agents Chemother* **60**, 592-599.
- Ogawa, A., Namba, Y. and Gakumasawa, M. (2016) Rational optimization of amber suppressor tRNAs toward efficient incorporation of a non-natural

amino acid into protein in a eukaryotic wheat germ extract. *Org Biomol Chem* **14**, 2671-2678.

- Ogura, T. and Hiraga, S. (1983) Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc Natl Acad Sci U S A* **80**, 4784-4788.
- Panteleev, P.V. and Ovchinnikova, T.V. (2015) Improved strategy for recombinant production and purification of antimicrobial peptide tachyplesin I and its analogs with high cell selectivity. *Biotechnol Appl Biochem*.
- Park, J.H., Yamaguchi, Y. and Inouye, M. (2011) *Bacillus subtilis* MazF-bs (EndoA) is a UACAU-specific mRNA interferase. *FEBS Lett* **585**, 2526-2532.
- Park, J.H., Yamaguchi, Y. and Inouye, M. (2012) Intramolecular regulation of the sequence-specific mRNA interferase activity of MazF fused to a MazE fragment with a linker cleavable by specific proteases. *Appl Environ Microbiol* **78**, 3794-3799.
- Park, J.H., Yoshizumi, S., Yamaguchi, Y., Wu, K.P. and Inouye, M. (2013) ACA-specific RNA sequence recognition is acquired via the loop 2 region of MazF mRNA interferase. *Proteins* **81**, 874-883.
- Parsons, J.F., Xiao, G., Gilliland, G.L. and Armstrong, R.N. (1998) Enzymes harboring unnatural amino acids: mechanistic and structural analysis of the enhanced catalytic activity of a glutathione transferase containing 5-fluorotryptophan. *Biochemistry* **37**, 6286-6294.
- Pedersen, K., Zavialov, A.V., Pavlov, M.Y., Elf, J., Gerdes, K. and Ehrenberg, M. (2003) The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell* **112**, 131-140.
- Pelillo, C., Benincasa, M., Scocchi, M., Gennaro, R., Tossi, A. and Pacor, S. (2014) Cellular internalization and cytotoxicity of the antimicrobial proline-rich peptide Bac7(1-35) in monocytes/macrophages, and its activity against phagocytosed *Salmonella typhimurium*. *Protein Pept Lett* **21**, 382-390.
-
- Pellegrini, O., Mathy, N., Gogos, A., Shapiro, L. and Condon, C. (2005) The *Bacillus subtilis* ydcDE operon encodes an endoribonuclease of the MazF/PemK family and its inhibitor. *Mol Microbiol* **56**, 1139-1148.
- Peng, T. and Hang, H.C. (2016) Site-Specific Bioorthogonal Labeling for Fluorescence Imaging of Intracellular Proteins in Living Cells. *J Am Chem Soc* **138**, 14423-14433.

- Podda, E., Benincasa, M., Pacor, S., Micali, F., Mattiuzzo, M., Gennaro, R. and Scocchi, M. (2006) Dual mode of action of Bac7, a proline-rich antibacterial peptide. *Biochim Biophys Acta* **1760**, 1732-1740.
- Qing, G., Ma, L.C., Khorchid, A., Swapna, G.V., Mal, T.K., Takayama, M.M., Xia, B., Phadtare, S., Ke, H., Acton, T., Montelione, G.T., Ikura, M. and Inouye, M. (2004) Cold-shock induced high-yield protein production in *Escherichia coli*. *Nat Biotechnol* **22**, 877-882.
- Richmond, M.H. (1959) Incorporation of canavanine by *Staphylococcus aureus* 524 SC. *Biochem J* **73**, 261-264.
- Rosen, M.K., Gardner, K.H., Willis, R.C., Parris, W.E., Pawson, T. and Kay, L.E. (1996) Selective methyl group protonation of perdeuterated proteins. *J Mol Biol* **263**, 627-636.
- Rosenthal, G.A., Reichhart, J.M. and Hoffmann, J.A. (1989) L-canavanine incorporation into vitellogenin and macromolecular conformation. *J Biol Chem* **264**, 13693-13696.
- Roy, R.N., Lomakin, I.B., Gagnon, M.G. and Steitz, T.A. (2015) The mechanism of inhibition of protein synthesis by the proline-rich peptide oncocin. *Nat Struct Mol Biol* **22**, 466-469.
- Ruschak, A.M. and Kay, L.E. (2010) Methyl groups as probes of supra-molecular structure, dynamics and function. *J Biomol NMR* **46**, 75-87.
- Ruschak, A.M., Velyvis, A. and Kay, L.E. (2010) A simple strategy for ¹³C, ¹H labeling at the Ile-γ2 methyl position in highly deuterated proteins. *J Biomol NMR* **48**, 129-135.
- Sat, B., Hazan, R., Fisher, T., Khaner, H., Glaser, G. and Engelberg-Kulka, H. (2001) Programmed cell death in *Escherichia coli*: some antibiotics can trigger *mazEF* lethality. *J Bacteriol* **183**, 2041-2045.
- Sat, B., Reches, M. and Engelberg-Kulka, H. (2003) The *Escherichia coli mazEF* suicide module mediates thymineless death. *J Bacteriol* **185**, 1803-1807.
- Schachtele, C.F., Anderson, D.L. and Rogers, P. (1968) Mechanism of canavanine death in *Escherichia coli*. II. Membranes-bound canavanyl-protein and nuclear disruption. *J Mol Biol* **33**, 861-872.
- Schmidt, O., Schuenemann, V.J., Hand, N.J., Silhavy, T.J., Martin, J., Lupas, A.N. and Djuranovic, S. (2007) *prfF* and *yhaV* encode a new toxin-antitoxin system in *Escherichia coli*. *J Mol Biol* **372**, 894-905.
- Schnapp, D., Kemp, G.D. and Smith, V.J. (1996) Purification and characterization of a proline-rich antibacterial peptide, with sequence

similarity to bactenecin-7, from the haemocytes of the shore crab, *Carcinus maenas*. *Eur J Biochem* **240**, 532-539.

- Schneider, W.M., Inouye, M., Montelione, G.T. and Roth, M.J. (2009) Independently inducible system of gene expression for condensed single protein production (cSPP) suitable for high efficiency isotope enrichment. *J Struct Funct Genomics* **10**, 219-225.
- Scocchi, M., Romeo, D. and Zanetti, M. (1994) Molecular cloning of Bac7, a proline- and arginine-rich antimicrobial peptide from bovine neutrophils. *FEBS Lett* **352**, 197-200.
- Scocchi, M., Tossi, A. and Gennaro, R. (2011) Proline-rich antimicrobial peptides: converging to a non-lytic mechanism of action. *Cell Mol Life Sci* **68**, 2317-2330.
- Seefeldt, A.C., Graf, M., Pérébasquine, N., Nguyen, F., Arenz, S., Mardirossian, M., Scocchi, M., Wilson, D.N. and Innis, C.A. (2016) Structure of the mammalian antimicrobial peptide Bac7(1-16) bound within the exit tunnel of a bacterial ribosome. *Nucleic Acids Res.*
- Serber, Z., Ledwidge, R., Miller, S.M. and Dötsch, V. (2001) Evaluation of parameters critical to observing proteins inside living *Escherichia coli* by in-cell NMR spectroscopy. *J Am Chem Soc* **123**, 8895-8901.
- Shapira, A., Shapira, S., Gal-Tanamy, M., Zemel, R., Tur-Kaspa, R. and Benhar, I. (2012) Removal of hepatitis C virus-infected cells by a zymogenized bacterial toxin. *PLoS One* **7**, e32320.
- Short, F.L., Pei, X.Y., Blower, T.R., Ong, S.L., Fineran, P.C., Luisi, B.F. and Salmond, G.P. (2013) Selectivity and self-assembly in the control of a bacterial toxin by an antitoxic noncoding RNA pseudoknot. *Proc Natl Acad Sci U S A* **110**, E241-249.
- Simanshu, D.K., Yamaguchi, Y., Park, J.H., Inouye, M. and Patel, D.J. (2013) Structural basis of mRNA recognition and cleavage by toxin MazF and its regulation by antitoxin MazE in *Bacillus subtilis*. *Mol Cell* **52**, 447-458.
- Singer, M. and Kaiser, D. (1995) Ectopic production of guanosine penta- and tetraphosphate can initiate early developmental gene expression in *Myxococcus xanthus*. *Genes Dev* **9**, 1633-1644.
- Studier, F.W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* **219**, 37-44.
- Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-130.

- Suzuki, M., Roy, R., Zheng, H., Woychik, N. and Inouye, M. (2006) Bacterial bioreactors for high yield production of recombinant protein. *J Biol Chem* **281**, 37559-37565.
- Suzuki, M., Zhang, J., Liu, M., Woychik, N.A. and Inouye, M. (2005) Single protein production in living cells facilitated by an mRNA interferase. *Mol Cell* **18**, 253-261.
- Takeda, M., Hallenga, K., Shigezane, M., Waelchli, M., Löhr, F., Markley, J.L. and Kainosho, M. (2011) Construction and performance of an NMR tube with a sample cavity formed within magnetic susceptibility-matched glass. *J Magn Reson* **209**, 167-173.
- Takeda, M. and Kainosho, M. (2012) Cell-free protein synthesis using *E. coli* cell extract for NMR studies. *Adv Exp Med Biol* **992**, 167-177.
- Takeda, M., Sugimori, N., Torizawa, T., Terauchi, T., Ono, A.M., Yagi, H., Yamaguchi, Y., Kato, K., Ikeya, T., Jee, J., Güntert, P., Aceti, D.J., Markley, J.L. and Kainosho, M. (2008) Structure of the putative 32 kDa myosinase-binding protein from Arabidopsis (At3g16450.1) determined by SAIL-NMR. *FEBS J* **275**, 5873-5884.
- Tugarinov, V. and Kay, L.E. (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* **125**, 13868-13878.
- Tugarinov, V., Muhandiram, R., Ayed, A. and Kay, L.E. (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *J Am Chem Soc* **124**, 10025-10035.
- Vaiphei, S.T., Mao, L., Shimazu, T., Park, J.H. and Inouye, M. (2010) Use of amino acids as inducers for high-level protein expression in the single-protein production system. *Appl Environ Microbiol* **76**, 6063-6068.
- Vaiphei, S.T., Tang, Y., Montelione, G.T. and Inouye, M. (2011) The use of the condensed single protein production system for isotope-labeled outer membrane proteins, OmpA and OmpX in *E. coli*. *Mol Biotechnol* **47**, 205-210.
- Velyvis, A., Ruschak, A.M. and Kay, L.E. (2012) An economical method for production of (2)H, (13)CH₃-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. *PLoS One* **7**, e43725.
- Verstraeten, N., Knapen, W.J., Kint, C.I., Liebens, V., Van den Bergh, B., Dewachter, L., Michiels, J.E., Fu, Q., David, C.C., Fierro, A.C., Marchal, K., Beirlant, J., Versées, W., Hofkens, J., Jansen, M., Fauvart, M. and Michiels, J. (2015) O₂ and Membrane Depolarization Are Part of a

Microbial Bet-Hedging Strategy that Leads to Antibiotic Tolerance. *Mol Cell* **59**, 9-21.

- Vogel, J., Argaman, L., Wagner, E.G. and Altuvia, S. (2004) The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr Biol* **14**, 2271-2276.
-
- Wang, G. (2015) Improved methods for classification, prediction, and design of antimicrobial peptides. *Methods Mol Biol* **1268**, 43-66.
- Wang, L., Brock, A. and Schultz, P.G. (2002) Adding L-3-(2-Naphthyl)alanine to the genetic code of *E. coli*. *J Am Chem Soc* **124**, 1836-1837.
- Wang, Y., Chen, J., Zheng, X., Yang, X., Ma, P., Cai, Y., Zhang, B. and Chen, Y. (2014) Design of novel analogues of short antimicrobial peptide anoplin with improved antimicrobial activity. *J Pept Sci* **20**, 945-951.
- Wenk, M., Baumgartner, R., Holak, T.A., Huber, R., Jaenicke, R. and Mayr, E.M. (1999) The domains of protein S from *Myxococcus xanthus*: structure, stability and interactions. *J Mol Biol* **286**, 1533-1545.
- Wenk, M. and Mayr, E.M. (1998) *Myxococcus xanthus* spore coat protein S, a stress-induced member of the betagamma-crystallin superfamily, gains stability from binding of calcium ions. *Eur J Biochem* **255**, 604-610.
- Worst, E.G., Exner, M.P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N. and Ott, A. (2015) Cell-free expression with the toxic amino acid canavanine. *Bioorg Med Chem Lett* **25**, 3658-3660.
- Xi, D., Wang, X., Teng, D. and Mao, R. (2014) Mechanism of action of the tri-hybrid antimicrobial peptide LHP7 from lactoferricin, HP and plectasin on *Staphylococcus aureus*. *Biometals* **27**, 957-968.
- Yamaguchi, Y. and Inouye, M. (2009) mRNA interferases, sequence-specific endoribonucleases from the toxin-antitoxin systems. *Prog Mol Biol Transl Sci* **85**, 467-500.
- Yamaguchi, Y., Nariya, H., Park, J.H. and Inouye, M. (2012) Inhibition of specific gene expressions by protein-mediated mRNA interference. *Nat Commun* **3**, 607.
- Yamaguchi, Y., Park, J.H. and Inouye, M. (2009) MqsR, a crucial regulator for quorum sensing and biofilm formation, is a GCU-specific mRNA interferase in *Escherichia coli*. *J Biol Chem* **284**, 28746-28753.
- Yang, S.T., Lim, S.I., Kiessling, V., Kwon, I. and Tamm, L.K. (2016) Site-specific fluorescent labeling to visualize membrane translocation of a myristoyl switch protein. *Sci Rep* **6**, 32866.

- Yi, T., Sun, S., Huang, Y. and Chen, Y. (2015) Prokaryotic expression and mechanism of action of α -helical antimicrobial peptide A20L using fusion tags. *BMC Biotechnol* **15**, 69.
- Yoshizumi, S., Zhang, Y., Yamaguchi, Y., Chen, L., Kreiswirth, B.N. and Inouye, M. (2009) *Staphylococcus aureus* YoeB homologues inhibit translation initiation. *J Bacteriol* **191**, 5868-5872.
- Zhang, C., He, X., Gu, Y., Zhou, H., Cao, J. and Gao, Q. (2014) Recombinant scorpine produced using SUMO fusion partner in *Escherichia coli* has the activities against clinically isolated bacteria and inhibits the *Plasmodium falciparum* parasitemia in vitro. *PLoS One* **9**, e103456.
- Zhang, J., Movahedi, A., Wei, Z., Sang, M., Wu, X., Wang, M., Wei, H., Pan, H., Yin, T. and Zhuge, Q. (2016) High-level SUMO-mediated fusion expression of ABP-dHC-cecropin A from multiple joined genes in *Escherichia coli*. *Anal Biochem* **509**, 15-23.
- Zhang, J., Zhang, Y. and Inouye, M. (2003a) Characterization of the interactions within the *mazEF* addiction module of *Escherichia coli*. *J Biol Chem* **278**, 32300-32306.
- Zhang, L., Li, X., Wei, D., Wang, J., Shan, A. and Li, Z. (2015a) Expression of plectasin in *Bacillus subtilis* using SUMO technology by a maltose-inducible vector. *J Ind Microbiol Biotechnol* **42**, 1369-1376.
- Zhang, Y. and Inouye, M. (2009) The inhibitory mechanism of protein synthesis by YoeB, an *Escherichia coli* toxin. *J Biol Chem* **284**, 6627-6638.
- Zhang, Y., Teng, D., Wang, X., Mao, R., Cao, X., Hu, X., Zong, L. and Wang, J. (2015b) In vitro and in vivo characterization of a new recombinant antimicrobial peptide, MP1102, against methicillin-resistant *Staphylococcus aureus*. *Appl Microbiol Biotechnol* **99**, 6255-6266.
- Zhang, Y., Zhang, J., Hara, H., Kato, I. and Inouye, M. (2005a) Insights into the mRNA cleavage mechanism by MazF, an mRNA interferase. *J Biol Chem* **280**, 3143-3150.
- Zhang, Y., Zhang, J., Hoeflich, K.P., Ikura, M., Qing, G. and Inouye, M. (2003b) MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in *Escherichia coli*. *Mol Cell* **12**, 913-923.
- Zhang, Y., Zhu, L., Zhang, J. and Inouye, M. (2005b) Characterization of ChpBK, an mRNA interferase from *Escherichia coli*. *J Biol Chem* **280**, 26080-26088.

- Zhu, L., Inoue, K., Yoshizumi, S., Kobayashi, H., Zhang, Y., Ouyang, M., Kato, F., Sugai, M. and Inouye, M. (2009) *Staphylococcus aureus* MazF specifically cleaves a pentad sequence, UACAU, which is unusually abundant in the mRNA for pathogenic adhesive factor SraP. *J Bacteriol* **191**, 3248-3255.