Engineering *Escherichia coli* for Effective Expression of Plant Genes

Junhyeong Lee

Department of Bioengineering
School of Nano-Bioscience and Chemical Engineering
Graduate School of UNIST

2015

Engineering *Escherichia coli* for Effective Expression of Plant Genes

Junhyeong Lee

Department of Bioengineering
School of Nano-Bioscience and Chemical Engineering
Graduate School of UNIST

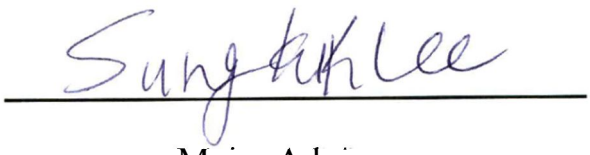
Engineering *Escherichia coli* for Effective Expression of Plant Genes

A thesis
Submitted to the School of Nano-Bioscience and Chemical Engineering and
the Graduate School of UNIST in fulfillment for the degree of
Master of Science

Junhyeong Lee

12. 16. 2014

Approved by



Major Advisor
Sung Kuk Lee, Ph. D.

Engineering *Escherichia coli* for Effective Expression of Plant Genes

Junhyeong Lee

This certifies that the thesis of Junhyeong Lee is approved.

12. 16. 2014

signature

Thesis Supervisor: Sung Kuk Lee, Ph. D.

signature

Robert J. Mitchell, Ph. D.

signature

Cheol-Min Ghim, Ph. D.

Abstract

Various plant-derived useful chemicals have been used for human from ancient to modern time. Because of difference of genetic characters between bacteria and plant, it could cause low plant protein expression or formation of inclusion body composed by insoluble protein precipitates. There are generally two methods for enhancement of heterologous protein expression in *E. coli*. In order to change rare tRNA abundance and chaperone expression level, we conducted chromosomally integrated promoter mutant libraries using recombination with single-strand DNA oligonucleotides. Some chaperone promoter mutants showed enhancement of *Arabidopsis*'s protein expression level. The tRNA and chaperone promoter mutants also showed significant increases in target *Arabidopsis*'s protein expression. We developed *E. coli* mutants for optimal plant protein expression. It is applicable to pharmaceutical production of plant-derived useful chemicals from *E. coli*.

Contents

List of figures

List of tables

Nomenclature

E. coli: *Escherichia coli*

A. thaliana: *Arabidopsis thaliana*

MAGE: Multiplex Automated Genome Engineering

MMR: Methyl-directed Mismatch Repair

OD: Optimal Density

CP: Constitutive Promoter

gDNA: genomic DNA

cDNA: complementary DNA

HTS: High-Throughput Screening

AP: Alkaline Phosphatase

NBT: Nitroblue tetrazolium

BCIP: 5-Bromo-4-chloro-3-indolyl phosphate

IPTG: Isopropyl β-D-1-thiogalactopyranoside

gDS: *Panax ginseng* Dammarenediol Synthase

1. Introduction

Various plant-derived useful chemicals have been used for human from ancient to modern time. [1] For examples, large numbers of herbal remedies, pigments, natural flavoring and natural rubber are extracted from diverse plants.[2, 3] However, the extraction of chemicals from plants takes a long time, and it is relatively low production efficiency and high production cost. [4, 5] Thus, many researchers have developed the rapid, cheap and high productive alternatives for production of plant-derived proteins related in production of useful chemicals.[6, 7] So plant proteins are synthesized in microorganisms rather than direct extraction of desired products from plant.[5] Especially, the methods which are heterologous protein expression in bacterial host have been widely used for academic researches and massive-scale industrial production.[8] However, there are difference of genetic characters between bacteria and plant.[9] It could cause low plant protein expression in bacterial host cell.[9] Sometimes heterologous protein expression causes low growth or cell death, because of formation of inclusion body.[10, 11] The inclusion body was synthesized by truncated or misfolded protein having insoluble character.[12, 13] In order to overcome these problems, various researches of engineering genetic machinary in host organisms have been conducted.

1.1 *Escherichia coli* as an industrial bacterial host for heterologous protein expression

E. coil has been the most well-known bacterial host for the heterologous protein synthesis. Because *E. coli* grows rapidly, it requires cheap substrates for their growth or production of desired molecules.[8] And *E. coli* is well-characterized in genetic information, various vectors, promoters, and many kinds of mutant host strains are available.[8] Furthermore, large-scale production systems have been established for industrial application.[14] Because of these advantages, production time and cost are reduced and various attempts for functional expression of target protein facilitate wide application of *E. coli* strains for heterologous protein expression.[14]

However, there are several disadvantages of heterologous protein expression in *E. coli*. Some foreign protein show toxicity to *E. coli* cell, and this case leads to host cell death.[10] Also, synthesized protein might be no activity because of protein misfolding[15]. Misfolded proteins tend to make truncated protein and formation of inclusion body, it cause low cell growth or death.[12] One of the reason of these problems is difference of codon usage bias between foreign gene and *E. coli* strains.[9, 16] It means that *E. coli* does not have enough translational machinery, especially tRNA pool for foreign gene expression.[6, 16] If these problems are resolved, *E. coli* will be widely used for heterologous protein expression.

1.2 Difference of codon usage bias between *E. coli* and *Arabidopsis thaliana*

A. thaliana is model organism of plant, so it is well-studied.[17] Function of enzymes and genome sequence have been identified. Because of this reason, many researches of *Arabidopsis*-derived heterologous protein expression in *E. coli* were reported.[8] These two organisms show the difference of codon preference of several genes.[18] It means that they use different synonymous codons of encoding same amino acid.[18 19] Why they have different codon usage bias?

The genetic code consists of 61 codons encoding 20 amino acids and 3 stop codons.[20] Therefore, each amino acid is encoded by one codon (Methionine and Tryptophan) and six synonymous codons (Arginine, Leucine and Serine). Codons are recognized by anticodon of complementary tRNAs that have been charged with the amino acid.[21] The degeneracy of the genetic code enables many alternative DNA sequences to encode the same protein sequence.[22] Codon distribution is related to genome GC content and the changes in codon usage bias could be at least partly explained by a mutation–selection equilibrium between the different synonymous codons in each organism.[9] According to arithmetical result of genome analysis of *E. coli* and *A. thaliana*, genome GC content of *E. coli* is 51% and genome GC content of *A. thaliana* is 36%.[23 24] This result means that *Arabidopsis* prefer to AT-rich codon.[25] Although genome GC content is not same with codon composition of several genes, it shows a tendency of AT-rich codon preference.[25] Some researchers have explained that codon usage biases that tend to reduce the diversity of isoacceptor tRNAs reduce the energy and are beneficial to organisms under rapid growth conditions.[26] Whatever the reason, it has become clear which codon biases could have a high impact on the heterologous protein expression in *E. coli*.

Codon usage bias has been believed as the most important factor in *E. coli* for heterologous protein expression.[9, 18] The reason is certainly because codon preference correlate with the abundance of cognate tRNAs available in the cell.[27] This serves to optimize the translation and to balance between codon concentration and isoacceptor tRNA concentration.[28] For example, the $tRNA_4^{Arg}$ which reads the uncommon AGG and AGA codons for Arginine is present at very low levels in *E. coli*.[29] It means that codon usage bias and cognate tRNA concentrations have evolved together.[30] And the selection pressure for this co-evolution could be found in highly expressed genes than genes expressed at low levels.[31] The difference of codon usage bias was represented in Figure 1.

AA	Codon	*E. coli*	*A. thaliana*		AA	Codon	*E. coli*	*A. thaliana*		AA	Codon	*E. coli*	*A. thaliana*		AA	Codon	*E. coli*	*A. thaliana*	
Phe	UUU	0 a	0		Ser	UCU	0	37		Tyr	UAU	0	0		Cys	UGU	0	0	
		2.22 b	2.18	-0.04 c			0.8	2.52	1.72			1.6	1.46	-0.14			0.5	1.05	0.55
	UUC	2	16			UCC	2	1			UAC	3	76			UGC	1	15	
		1.66	2.07	0.41			0.86	1.12	0.26			1.22	1.37	0.15			0.64	0.72	0.08
Leu	UUA	1	6			UCA	1	9		Stop	UAA	0	0		Stop	UGA	1	0	
		1.38	1.27	-0.11			0.7	1.83	1.13			0.2	0.09	-0.11			0.09	0.12	0.03
	UUG	1	10			UCG	1	4			UAG	0	0		Trp	UGG	1	14	
		1.36	2.09	0.73			0.89	0.93	0.04			0.2	0.05	-0.15			1.53	1.25	-0.28
	CUU	0	11		Pro	CCU	0	16		His	CAU	0	0		Arg	CGU	4	9	
		1.1	2.41	1.31			0.7	1.87	1.17			1.29	1.38	0.09			2.1	0.9	-1.20
	CUC	1	1			CCC	1	0			CAC	1	10			CGC	0	0	
		1.11	1.61	0.50			0.55	0.53	-0.02			0.97	0.87	-0.10			2.21	0.38	-1.83
	CUA	1	10			CCA	1	39		Gln	CAA	2	8			CGA	0	6	
		0.39	0.99	0.60			0.84	1.61	0.77			1.54	1.94	0.40			0.35	0.63	0.28
	CUG	4	3			CCG	1	5			CAG	2	9			CGG	1	4	
		5.31	0.98	-4.33			2.33	0.86	-1.47			2.9	1.52	-1.38			0.54	0.49	-0.05
Ile	AUU	0	20		Thr	ACU	0	10		Asn	AAU	0	0		Ser	AGU	0	0	
		3.04	2.15	-0.89			0.89	1.75	0.86			1.76	2.23	0.47			1.87	1.4	-0.47
	AUC	3	0			ACC	2	0			AAC	4	16			AGC	1	13	
		2.52	1.85	-0.67			2.35	1.03	-1.32			2.16	2.09	-0.07			1.6	1.13	-0.47
	AUA	0	5			ACA	1	8		Lys	AAA	6	13		Arg	AGA	1	9	
		0.42	1.26	0.84			0.69	1.57	0.88			3.36	3.08	-0.28			0.2	1.9	1.70
Met	AUG	8	24			ACG	2	6			AAG	0	18			AGG	1	8	
		2.78	2.45	-0.33			1.44	0.77	-0.67			1.03	3.27	2.24			0.11	1.1	0.99
Val	GUU	0	15		Ala	GCU	0	16		Asp	GAU	0	0		Gly	GGU	0	1	
		1.83	2.72	0.89			1.52	2.83	1.31			3.22	3.66	0.44			2.48	2.22	-0.26
	GUC	2	0			GCC	2	0			GAC	3	23			GGC	4	23	
		1.53	1.28	-0.25			2.57	1.03	-1.54			1.91	1.72	-0.19			2.98	0.92	-2.06
	GUA	5	7			GCA	3	10		Glu	GAA	4	12			GGA	1	12	
		1.09	0.99	-0.10			2.01	1.75	-0.26			3.96	3.43	-0.53			0.79	2.42	1.63
	GUG	0	8			GCG	0	7			GAG	0	13			GGG	1	5	
		2.63	1.74	-0.89			3.38	0.9	-2.48			1.79	3.22	1.43			1.1	1.02	-0.08

Figure 1. Comparison of codon usage and number of tRNA in *E. coli* and *A. thaliana*. a, a number of tRNA; b, codon usage bias; c, difference of codon usage between *E. coli* and *A. thailiana*.
*A complete summary of codon usages can be found at http://www.kazusa.or.jp/codon/

1.3 Prevention of inclusion body formation by molecular chaperones

E. coli has many defense mechanisms responding external environmental change. Chaperones are one of the system for synthesis and maintenance of protein in abnormal condition.[32] The major function of chaperones is to prevent both synthesized polypeptide chains and assembled subunits from aggregating into nonfunctional structures.[32] The most chaperones are expressed in heat-shocked circumstance because chaperones have specific promoters corresponding sigma factor 32 (σ32) which is expressed by heat-shock, not sigma factor 70 (σ70) which is expressed constitutively.[33] Because of this region, expression of chaperones is low in normal condition.[33] In heat-shock condition, cytosolic and membrane proteins are damaged and aggregated in the cell and it may cause cell death.[34] Chaperones helps new protein synthesis and damaged protein repair.[35] Especially, DnaKJ and GrpE system help polypeptide elongation in protein synthesis, and GroELS system helps to switch misfolding protein to functional protein by refolding.[36, 37]

1.4 Simultaneous genome editing for introduction of various mutation

In *E. coli* researches, there are two general genome engineering methods such as site-directed genome mutagenesis which uses lambda RED system and KmR cassette from pKD13, and random genome mutagenesis which uses transposon.[38, 39] In this study, lambda RED system and KmR cassette were used to make knockout strains and replacement strains. However this method is only one site mutagenesis in each attempt.[37] So multiple target mutagenesis takes long time and high cost. In order to overcome this problem, simultaneous genome engineering method was developed.

Simultaneous engineering of multiple sites on the genome using short single-stranded oligo nucleotides, it called multiplex automated genomic engineering (MAGE), has currently become a unique genomic engineering tool that could help efficient and accelerated mutagenesis of desired functions during a short time.[40-42] This technique utilizes the lambda RED recombination system (Gamma, Exo, Beta) and short nucleotide oligos in order to introduce desired mutations on *E. coli* genome.[40-42] When DNA replication occurs, MAGE oligos integrated as lagging strand into newly synthetic strand. [40-42] Also, inactivation of the methyl-directed mismatch repair (MMR) system is necessary to avoid correction of engineered regions by native MMR system.[43] Although several modified *E. coli* strain (EcNR2, DY330, and EcHW24) could be used for genome engineering, these strains have several problems such as low cell growth because of the cytotoxic genes from defective prophage.[43] A permanent inactivation of MMR system could cause undesired mutations.[43] In order to avoid these disadvantages of MAGE, our groups developed pRED system which possess all components for oligo-mediated recombination.[43] Also, pRED suicide plasmids could inactivate the host MMR system by insertional inactivation of *mutS*.[43]

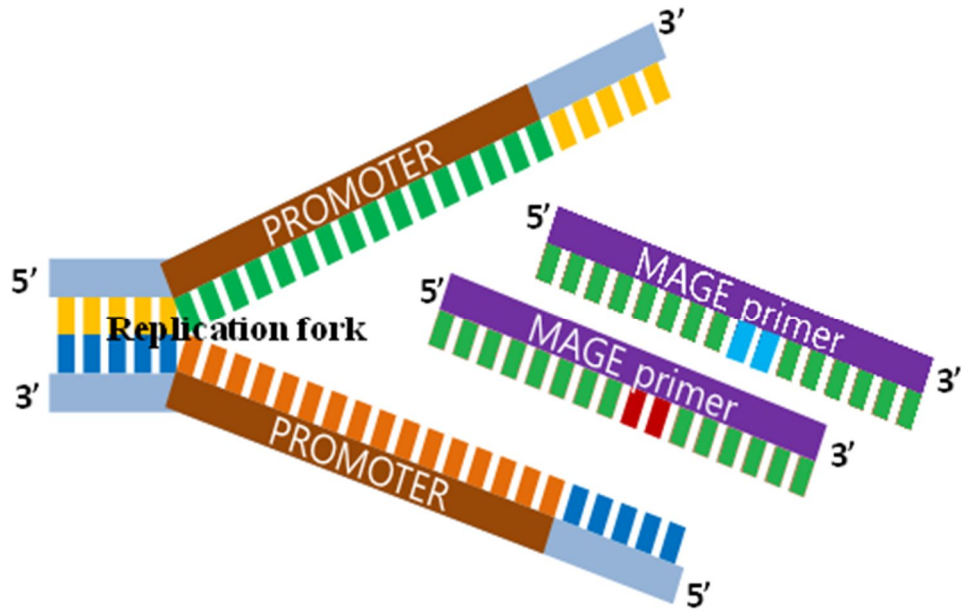


Figure 2. Multiplex Automated Genome Engineering. When DNA replication occurs, MAGE primers integrated into newly synthetic DNA strand as lagging strand.

1.5 Optimal expression of plant genes in *E. coli*

If the effect of different codon usage biases in heterologous protein expression is shown negative result by different tRNA levels, one solution is to be the expansion of the intracellular tRNA pool of the host.[9] For *E. coli*, the important targets to increase the expression of *Arabidopsis*'s genes are the *argU* gene encoding the rare $tRNA_4^{Arg}$ that reads AGG and AGA codons, *ileY* gene encoding the $tRNA_2^{Ile}$ that reads AUA, *leuW* gene encoding the $tRNA_3^{Leu}$ that reads CUA and CUG codons.[44] *E. coli* strains over-expressing these tRNA genes are commercially available from companies such as Stratagene and Novagen.[44-46] Some researchers have shown that expression of proteins whose genes contain rare codons could be dramatically improved when the cognate tRNA is increased in the host.[47, 48]

Aggregation of recombinant heterologous proteins overexpressed in *E. coli* could result either from accumulation of high concentrations of folding intermediates or from inefficient processing by molecular chaperones.[49] By codon usage bias difference, translation speed is low.[50] It causes ribosome pausing and synthesizing polypeptide is separated from ribosome.[50] Finally formation of misfolding or truncated protein occurs.[51] These abnormal proteins are aggregated in the cell, and it causes the formation of inclusion body and cell death.[12] Molecular chaperones contribute new protein synthesis

or refolding of misfolded proteins, in order to make insoluble proteins to soluble proteins.[37] In Kazuyo Nishihara's research, Cryj2 allergen protein from 'Japanese cedar (*Cryptomeria japonica*) pollen' was expressed as insoluble protein.[52] They expressed molecular chaperones containing plasmid.[52] Chaperones helps Cryj2 protein to make it soluble.[52] Above this example, commercial chaperone expressing plasmids have been used to express soluble heterologous protein in *E. coli*.

There are generally two methods for enhancement of heterologous protein expression in *E. coli*. First, expression of rare tRNA extra copies is related to adjust codon usage bias between *E. coli* and *Arabidopsis*'s genes.[30, 48] Second, expression of molecular chaperone is related to convert insoluble protein to soluble protein by helping protein synthesis or repair of misfolding proteins.[37, 52] However existing systems are only plasmid-based tRNA and chaperone expression.[52, 53] In these cases, antibiotics for maintenance of tRNA or chaperone plasmid should be added in culture media. It is disadvantage for large-scale production of heterologous protein expression in *E. coli*, because antibiotic addition and purification after cultivation are dissipation. Thus, we developed plasmid-free optimal plant gene expression *E. coli* strain by engineering promoters of tRNA and chaperone on genome.

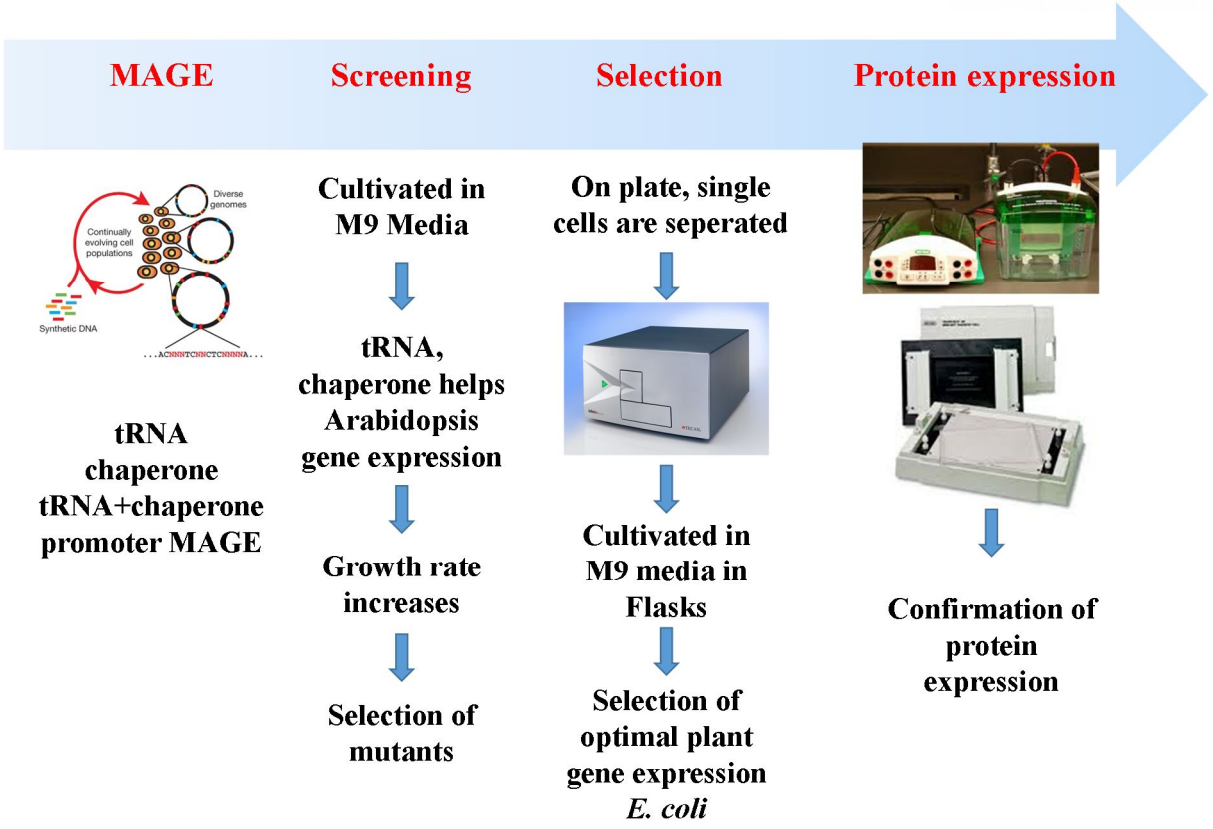


Figure 3. Experiment procedure in this study. MAGE method was used for introduction of mutation on genome. Screening system was constructed to screen MAGE mutants. Selection was conducted by growth. Protein expression was confirmed by SDS-PAGE and western blot.

2. Materials and methods

2.1 Bacterial strains and growth condition

The bacterial strains used in this study are listed in Table 1. *E. coli* MG1655 was used as a parental strain. Luria-Bertani (LB) medium containing suitable antibiotics was used for general cell growth (50 μg/mL of kanamycin, 100 μg/mL ampicillin, or 30 μg/mL chloramphenicol). The reduced-salt LB medium (5g/L NaCl) was used to cultivate *E. coli* strains after MAGE. Strains which was possessing temperature sensitive plasmids and possessing pRED-2 system on genome were grown at 30°C. The others were grown at 37°C. Liquid cultures were incubated in shaking incubator (shaking speed is 200 rpm). Engineered strains were cultured in M9 minimal medium with 0.2% glucose and antibiotics for mutant enrichment.

Strains	Genotype/description	Reference
E. coli MG1655	Wild type, *F⁻ λ⁻ ilvG- rfb-50 rph-1*	24
E. coli BL21(DE3)	*F⁻ ompT gal dcm lon hsdS$_B$(r$_B$- m$_B$-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	54
E. coli DH10B	*F⁻ endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ-*	55
MG-MAGE*	MG1655 which is pRED2 integration to genomic DNA	43
ΔTSA1	MG-MAGE with Δ*trpA* ::*KmR*	This study
ΔADT2	MG-MAGE with Δ*pheA* ::*KmR*	This study
ΔPROC1	MG-MAGE with Δ*proC* ::*KmR*	This study
ΔHIS6	MG-MAGE with Δ*hisA* ::*KmR*	This study
::TSA1	MG-MAGE with Δ*trpA* ::*TSA1-KmR*	This study
::ADT2	MG-MAGE with Δ*pheA* ::*ADT2-KmR*	This study
::PROC1	MG-MAGE with Δ*proC* ::*PROC1-KmR*	This study
::HIS6	MG-MAGE with Δ*hisA* ::*HIS6-KmR*	This study

Table 1. *E. coli* strains used in this study

*MG-MAGE strain was used as wild type

Plasmids	Genotype/description	Reference
pSIM5	pSC101 *ori*. temperature sensitive replication *exo, bet, gam* gene - λ-RED recombinase expression is induced by temperature	56
pKD13	Template of FRT-flanked kanamycin resistant gene. R6K gamma *ori.* Requiring the *pir*$^{+}$ *E. coli*	38
pRED2	R6K gamma *ori.* Requiring the *pir*$^{+}$ *E. coli* *mutS* homologous region with *exo, bet, gam* gene - λ-RED recombinase expression is induced by temperature	43
pYES2.1/V5-His-TOPO	pYES2.1 using the TOPO TA expression kit for cloning of Dammarenediol synthase cDNA	57
pET30a(+)	Kmr, Expression vector	58
pET30a(+)-TSA1	pET30a(+) containing Flag-tagged *A. thaliana* TSA1 gene	This study
pET30a(+)-ADT2	pET30a(+) containing Flag-tagged *A. thaliana* ADT2 gene	This study
pET30a(+)-PROC1	pET30a(+) containing Flag-tagged *A. thaliana* PROC1 gene	This study
pET30a(+)-HIS6	pET30a(+) containing Flag-tagged *A. thaliana* HIS6 gene	This study
pET30a(+)-PSP	pET30a(+) containing Flag-tagged *A. thaliana* PSP gene	This study
pET30a(+)-At2g19940	pET30a(+) containing Flag-tagged *A. thaliana* At2g19940 gene	This study
pET30a(+)-GS1	pET30a(+) containing Flag-tagged *A. thaliana* GS1 gene	This study
pET30a(+)-CGS1	pET30a(+) containing Flag-tagged *A. thaliana* CGS1 gene	This study
pET30a(+)-SAT1	pET30a(+) containing Flag-tagged *A. thaliana* SAT1 gene	This study
pBbB6a	Biobrick BBR1-ori, Amp^{r}, $P_L$lacO-1 promoter, dbl terminator	59
pBbB6a-TSA1	pBbB6a containing FLAG-tagged *A. thaliana* TSA1 gene	This study
pBbB6a-ADT2	pBbB6a containing FLAG-tagged *A. thaliana* ADT2 gene	This study
pBbB6a-PROC1	pBbB6a containing FLAG-tagged *A. thaliana* PROC1 gene	This study
pBbB6a-HIS6	pBbB6a containing FLAG-tagged *A. thaliana* HIS6 gene	This study
pBbB6a-gDS-flag	pBbB6a containing FLAG-tagged *Panax ginseng* Dammarenediol synthase gene	This study

Table 2. Plasmids used in this study

2.2 Strain construction

pSIM5 plasmid harboring the λ-RED recombination system under the control of cI857 (temperature sensitive) was used to for several genes deletion and gene replacement on genome.[56] Chromosomal exchanges were identified by PCR analysis and sequencing. Kanamycin cassette from pKD13 was amplified with primers carrying a 50 bp overhang that is homologous to the target site.[38] For knocking out *E. coli* native genes, the kanamycin cassette PCR product was transformed to MG-MAGE strain. For replacing *E. coli* native genes to *A. thaliana* genes, the kanamycin cassette and the *A. thaliana* gene were linked by SOE-PCR.[60] This SOE-PCR product was transformed to MG-MAGE strain. Cell harboring the plasmid pSIM5 were induced by thermal inactivation of CI at 42°C for 15 minutes. The cells were chilled immediately after induction to arrest the metabolic state of the cell, made electrocompetent, and transformed with the PCR products as described above. SOC medium was added after electroporation and the cells were recovered from electric shock by incubation at 30°C for 2 hours. Transformant colonies carrying the desired modification were directly selected on LB agar plates containing kanamycin. Kanamycin cassette was not deleted because we could know the gene replacement or not. Primers used to construct knockout strains are listed in Table 3 and replacement strains are listed in Table 4.[61] Three synthetic constitutive promoters, CP6, CP12 and CP25 were used for promoter exchange.[62]

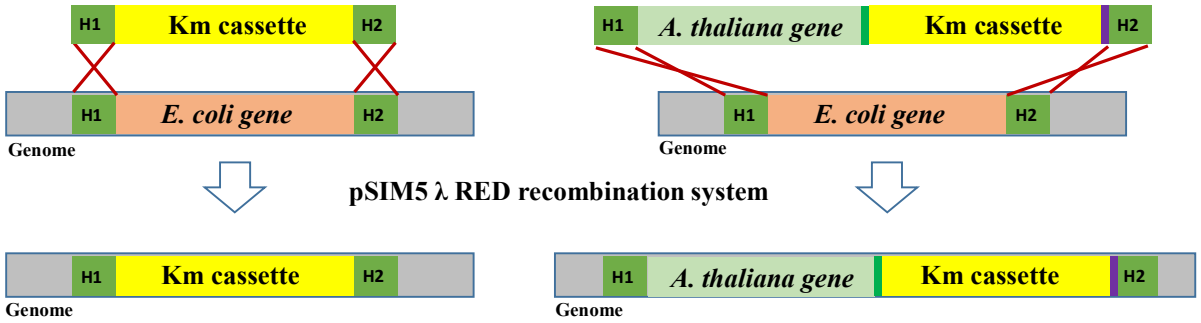


Figure 2. Gene knockout and gene replacement methods. λ RED recombination system was used for exchanging native gene to Km cassette. Arabidopsis and Km cassette fusion genes were synthesized by overlapping PCR method.

Table 3. Primers for gene knockout strain construction

Primers	Template	Primer sequence [a]
trpA KO_F	pKD13	AAAGACATCTTCACCGTTCACGATATTTTGAAAGCACGAGGGGAAATCTGGTGTAGGCTGGAGCTGCTTCG
trpA KO_R		ATTCCGGGGATCCGTCGACCCTTCATTAAAGAAAGTTAAAATGCCGCCAGCGGAACTGGCGGCTGTGGGA
pheA KO_F	pKD13	AGGCCTCCCAAATCGGGGGCCTTTTTTATTGATAACAAAAAGGCAACACTGTGTAGGCTGGAGCTGCTTCG
pheA KO_R		ATTCCGGGGATCCGTCGACCGCCAGTAATAATCCAGTGCCGGATGATTCACATCATCCGGCACCTTTTCA
proC KO_F	pKD13	CTTCTGCCAGCGATTATCAAAACAATGAATTTCACGGCAGGAGTGAGGCAGTGTAGGCTGGAGCTGCTTC
proC KO_R		AGCCGGACGTAACCGCACCGAAGTGGCGGCCTGACGTCCGGCGAAAGTCACTGTCAAACATGAGAATTAA
hisA KO_F	pKD13	CGTTCTGGTGCCGCTGGTGCTAAGTTGCTGAAAAACTTCCTGGAGATGTGGTGTAGGCTGGAGCTGCTTC
hisA KO_R		CGTCGAGACATGGGATTATGCGTTTTGCCAGCATGCGATGGCCTCCTTCACTGTCAAACATGAGAATTAA

[a]All primer sequences are 5′ -> 3′ direction

`Primers	Template	Primer sequence [a]
H1-TSA1 F1	MG1655 gDNA	TGTCTATTACCGATGATGAA
H1-TSA1 R1		GATTTGAAAGCAATCGCCATCAGATTTCCCCTCGTGCTTT
TSA1 F2	*A. thaliana* cDNA	ATGGCGATTGCTTTCAAATC
TSA1-P1 R2		GAAGCAGCTCCAGCCTACACTCAAAGAAGAGCAGATTTAA
P4-H2 TSA1 F3	MG1655 gDNA	TTAATTCTCATGTTTGACAGTCCCACAGCCGCCAGTTCCG
H2 TSA1 R3		GAGGAATAAGTGACTTAGAG
H1-ADT2 F1	MG1655 gDNA	CTGTACTAAAGTCACTTAAG
H1-ADT2 R1		TTCTTCACTTCCATAGCCATAGTGTTGCCTTTTTGTTATC
ADT2 F2	*A. thaliana* cDNA	ATGGCTATGGAAGTGAAGAA
ADT2-P1 R2		GAAGCAGCTCCAGCCTACACTTAGAGCATTGTAGTGTCCA
P4-H2 ADT2 F3	MG1655 gDNA	TTAATTCTCATGTTTGACAGTGAAAAGGTGCCGGATGATG
H2 ADT2 R3		CTGGAGCAGGGCGATAAGCA
H1-PROC1 F1	MG1655 gDNA	CCGCGCGATACAAAATCTCT
H1-PROC1 R1		GGAATCGGAAGAATCTCCATTGCCTCACTCCTGCCGTGAA
PROC1 F2	*A. thaliana* cDNA	ATGGAGATTCTTCCGATTCC
PROC1-P1 R2		GAAGCAGCTCCAGCCTACACTTAGCTCTGTGAGAGCTCGC
P4-H2 PROC1 F3	MG1655 gDNA	TTAATTCTCATGTTTGACAGTGACTTTCGCCGGACGTCAG
H2 PROC1 R3		GGGTGCATGAAGGGCTAAAT
H1-HIS6 F1	MG1655 gDNA	TGTTCACAGCTACGCAATGC
H1- HIS6 R1		TGGGAGCTCAAAGTTCTCATCACATCTCCAGGAAGTTTTT
HIS6 F2	*A. thaliana* cDNA	ATGAGAACTTTGAGCTCCCA
HIS6-P1 R2		GAAGCAGCTCCAGCCTACACTCAATGGAGTGAGTGCTGCT

P4-H2 HIS6 F3	MG1655 gDNA	TTAATTCTCATGTTTGACAGTGAAGGAGGCCATCGCATGC
H2 HIS6 R3		GCTTTTATCTACCACACGGC
P1	pKD13	GTGTAGGCTGGAGCTGCTTC
P4		CTGTCAAACATGAGAATTAA

Table 4. Primers for gene replacement strain construction

[a]All primer sequences are 5′ -> 3′ direction

2.3 Plasmid construction

All DNA manipulations were performed using established protocols.[63] In order to confirm expression condition of *Arabidopsis*'s proteins in *E. coli*, overexpression plasmid backbone pET30a(+) (Novagen) was used with 9 kinds of *Arabidopsis*'s genes. These genes are listed in Table 5. All *Arabidopsis*'s genes were amplified from *A. thaliana* cDNA library. Primers were designed by insertion of restriction enzyme site corresponding with pET30a(+) restriction enzyme site. Forward primers include *NdeI* (CATATG, NEB biolabs) restriction enzyme site. Reverse primers include *XhoI* (CTCGAG, NEB biolabs), *NotI* (GCGGCCGC, NEB biolabs), *HindIII* (AAGCTT, NEB biolabs) restriction enzyme sites. pET vector cloning primers are listed in Table 6. Digested vector and gene were ligated by T4 DNA ligase (NEB biolabs). Ligation mixture was transformed to DH10B competent cell (Invitrogen). In order to confirm the insertion of desire gene to pET30a(+) vector, colony PCR and sequencing (Marcrogen Inc.) were conducted. In order to confirm expression of *Arabidopsis*' protein in MAGE mutant cells, pBbB6a (Biobrick plasmid backbone) was used as expression vector. All insert genes were amplified from pET30(+) containing *Arabidopsis*' gene. Forward primers include *NdeI* restriction enzyme site. Reverse primers include BamHI restriction enzyme site and FLAG-tag binding site. TSA1, ADT2, PROC1 genes amplification used FLAG-*BamHI*-RP primer as universal reverse primer. Because *BamHI* restriction enzyme site located in HIS6, B6a-HIS6_RP including *XhoI* restriction enzyme site was used as reverse primer. *Panax ginseng* Dammarenediol synthase gene[57] was also inserted into pBbB6a vector. Biobrick vector cloning primers are listed in Table 7.

Table 5. Target gene information

***E. coli* gene**	***Arabidopsis* gene**	**Gene product**	**Function**
trpA	TSA1	Tryptophan synthase alpha chain	Tryptophan synthesis
pheA	ADT2	Chorismate mutase / prephenate dehydratase (bifunctional)	Phenylalanine synthesis
proC	PROC1	Pyrroline-5-carboxylate reductase	Proline synthesis
hisA	HIS6	N-(5'-phospho-L-ribosyl-formimino)-5-amino-1-(5'-phosphoribosyl)-4-imidazolecarboxamide isomerase	Histidine synthesis
serB	PSP	Phosphoserine phosphatase	Serine synthesis
argC	At2g19940	N-acetyl-gamma-glutamyl-phosphate reductase	Arginine synthesis
glnA	GS1	Glutamine synthetase	Glutimine synthesis
metB	CGS1	Cystathionine gamma-synthase	Methionine synthesis
cysE	SAT1	Serine acetyltransferase	Cystein synthesis

Primer name	Template	Primer sequence [a]	Restriction site
pET TSA1_F		CAGCATATGGCGATTGCTTTCAAATCCGGCGTCTTC	*NdeI*
pET TSA1_R		CTACTCGAGAAGAAGAGCAGATTTAAGAGAC	*XhoI*
pET ADT2_F		CAGCATATGGCTATGGAAGTGAAGAAGATC	*NdeI*
pET ADT2_R		GTAGCGGCCGCAGCATTGTAGTGTCCACTGGGTAG	*NotI*
pET PROC1_F		CAGCATATGGAGATTCTTCCGATTCCGGCGGAGA	*NdeI*
pET PROC1_R		CTACTCGAGGCTCTGTGAGAGCTCGCGGCTTC	*XhoI*
pET HIS6_F		CAGCATATGAGAACTTTGAGCTCCCAATTATAC	*NdeI*
pET HIS6_R		CTACTCGAGATGGAGTGAGTGCTGCTTGTGATG	*XhoI*
pET PSP_F	*A. thaiana* cDNA	AACTTTAAGAAGGAGATATACATATGACTTTGGGG CATGAAGGCAAC	*NdeI*
pET PSP_R		CTACTCGAGGTCCAATGAGTTTATGAGAGATTC	*XhoI*
pET At2g19940_F		AACTTTAAGAAGGAGATATACATATGGCTAGTAGTT CTGTTAAACCTG	*NdeI*
pET At2g19940_R		CTAAAGCTTCCGCAGAAACGGTGCTGACCCCGGTT ATTTGTTAACTGTTAATTG	*HindIII*
pET GS1_F		AACTTTAAGAAGGAGATATACATATGTCTCTGCTCT CAGATCTCG	*NdeI*
pET GS1_R		TGCGCGGCCGCACCGAGTATGGTCGTCTCAGCGA	*NotI*
pET CGS1_F		CAGCATATGGGTGAAAGATTAGGCCGTGG	*NdeI*
pET CGS1_R		CTACTCGAGGATGGCTTCGAGAGCTTGAAG	*XhoI*
pET SAT1_F		GATCATATGGCAACATGCATCCACACATG	*NdeI*
pET TSA1_R		GATCTCGAGAATCACATAATCAGACCACTCGGT	*XhoI*

Table 6. Primers for plasmid construction of target protein expression

[a]All primer sequences are 5'->3' direction

Primer name	Template	Primer sequence [a]	Restriction site
B6a-TSA1_FP	pET30a(+)-TSA1	CAGCATATGGCGATTGCTTTCAAATCCGGCGTCTTC	*NdeI*
B6a-ADT2_FP	pET30a(+)-ADT2	CAGCATATGGCTATGGAAGTGAAGAAGATC	*NdeI*
B6a-PROC1_FP	pET30a(+)-PROC1	CAGCATATGGAGATTCTTCCGATTCCGGCGGAGA	*NdeI*
FLAG-*BamHI*_RP	FLAG tagged pET30a(+) universal	CGCGGATCCTCACTTGTCATCGTCATCCT	*BamHI*
B6a-HIS6_FP	pET30a(+)-HIS6	CAGCATATGAGAACTTTGAGCTCCCAATTATAC	*NdeI*
B6a-HIS6_RP		CGCCTCGAGTCACTTGTCATCGTCATCCTTGTAATCATGGAGTGAGTGCTGCTTGT	*XhoI*
B6a-gDS-FP	pYES2.1/V5-His-TOPO	TCTCGATCGTTTAAGAAGGAGATATACATATGTGGAAGCTGAAGGTTGC	*NdeI*
B6a-gDS-RP		GCCGCTCGAGCGATCGTTACTTGTCATCGTCATCCTTGTAATCAATTTTGAGCTGCTGGTGCT	*XhoI*

Table 7. Primers for construction of biobrick expression plasmid

[a]All primer sequences are 5′ -> 3′ direction

2.4 Multiplex automated genome engineering (MAGE)

For effective and fast introduction of point and deletion mutations, MAGE method was used. ::TSA1, ::ADT2, ::PROC1, ::HIS6 strains were grown overnight at 30°C in LB supplemented with chloramphenicol and kanamycin. And then, seed was inoculated using 1/100 dilution into 3 mL modified LB - Lennox and grown till the OD_{600} reached to 0.5~0.6 at 30°C. Cells were induced at 42°C water bath for 15 minutes to express λ-recombination system. After induction, cells were cooled on ice for 1 hour to prepare electrocompetent. After cooling, approximately 1 mL ~ 2 mL cells were centrifuged in pre-cooled centrifuge at 4°C and then supernatant were removed. Cell pellet was washed using 4°C chilled autoclaved water, and then spin down and removed supernatant. Using same procedure, repeat at least 3 times for washing the cell pellet. After washing step, the cell pellet was resuspended in 50 μL RNAse free water containing each MAGE primer diluted up to 0.5 μM concentration, respectively. Re-suspended cells were transferred to electrocuvette and transformed via electroporation. Transformed cells were recovered in 3 mL modified LB supplemented with appropriate chloramphenicol and kanamycin in test tube till the OD_{600} reaches 0.5~0.6 at 30°C. The whole procedure was repeated 3 times. After final round, recovery was carried out for over 10 hours.

Table 8. MAGE primers

Target region	MAGE primer sequence [a]
tRNA promoters	
argUp	CGGTAACTGAGTCGTTCCCAACTGGCATATTAAGTGCGCTAATGATGCGTAA CGCCATAGTTGCGCGGGAATCGAGTCAACCTATCTCGT
ileYp	ACCCAAGGTGAATGGGAACGTTTGGCAGTTTATTCTTGACATGTAGTGAGG GGGCTGGTATAATCACATATTTATAAGTATTGTTATTGC
leuWp	CAGTCACTTTCGAGCAATTTTTTGGCAGTTTATTCTTGACATGTAGTGAGGG GGCTGGTATAATCACATACGCAACGCCGATAAGGTATC
argWp	ATAAATGGCGAGGGTTTAAGCAACTTTGGCAGTTTATTCTTGACATGTAGTG AGGGGGCTGGTATAATCATGCCGCATTGTCCTCTTAGT
ileXp	ATTGTTTATAAAAACAGCAGTTTGGCAGTTTATTCTTGACATGTAGTGAGGG GGCTGGTATAATCAAGTTTATTCTTGACATGTAGTGAGGGGGCTGGTATAATC ACATACATACCATGGCCCCTTAGCTCAGT
Chaperone promoters	
DnaKJ-CP12	TCGATACCAATTATTTTACCCATAGCTGTTTCCTGTGTGAACAGTACTCAGGT ATTATATCATTTTGGCCGACTAGTGTCAAGAATAAACTTGTATATGCTCACTC ACTGCGGTTGACT
DnaKJ-CP6	TCGATACCAATTATTTTACCCATAGCTGTTTCCTGTGTGAACAGTACTCAGTT ATTATATCATCCGGAAATATCTGTGTCAAGAATAAACTCCCACATGCTCACTC ACTGCGGTTGACT
DnaKJ-CP25	TCGATACCAATTATTTTACCCATAGCTGTTTCCTGTGTGAACAGTACTATGTG ATTATACCAGCCCCCTCACTACATGTCAAGAATAAACTGCCAAAGCTCACTC ACTGCGGTTGACT
GroESL-CP12	TCATGCAATGGACGAATATTCATAGCTGTTTCCTGTGTGAACAGTACTCAGG TATTATATCATTTTGGCCGACTAGTGTCAAGAATAAACTTGTATATGGATCAG CACAAAATCGGGTG
GroESL-CP6	TCATGCAATGGACGAATATTCATAGCTGTTTCCTGTGTGAACAGTACTCAGT TATTATATCATCCGGAAATATCTGTGTCAAGAATAAACTCCCACATGGATCAG CACAAAATCGGGTG
GroESL-CP25	TCATGCAATGGACGAATATTCATAGCTGTTTCCTGTGTGAACAGTACTATGTG ATTATACCAGCCCCCTCACTACATGTCAAGAATAAACTGCCAAAGGATCAGC ACAAAATCGGGTG
GrpE-CP12	GTTTTCTGTTCTTTACTACTCATAGCTGTTTCCTGTGTGAACAGTACTCAGGT ATTATATCATTTTGGCCGACTAGTGTCAAGAATAAACTTGTATATGGAATTTC TCCGCGTTTTTTT
GrpE-CP6	GTTTTCTGTTCTTTACTACTCATAGCTGTTTCCTGTGTGAACAGTACTCAGTT ATTATATCATCCGGAAATATCTGTGTCAAGAATAAACTCCCACATGGAATTTC TCCGCGTTTTTTT
GrpE-CP25	GTTTTCTGTTCTTTACTACTCATAGCTGTTTCCTGTGTGAACAGTACTATGTG ATTATACCAGCCCCCTCACTACATGTCAAGAATAAACTGCCAAAGGAATTTC TCCGCGTTTTTTT

[a]All primer sequences are 5′ -> 3′ direction

Primers	Sequencing primer sequence [a]
tRNA promoters	
argUp_seq_F	GGGTTCGGAGATAATCGATG
argUp_seq_R	TCATGGCTATCAGCTTGTCG
ileYp_seq_F	TGGAGAGGCAACTGTCAAAA
ileYp_seq_R	AATACATTGCAGTGGCGTGC
leuWp_seq_F	TGTTCACCAGTCGGCGTATA
leuWp_seq_R	GAACCAGGGAATGCCGGTAT
argWp_seq_F	TCTGGTGTTTAACGGTACGC
argWp_seq_R	GCTTTGCTGCTTCAATCTGC
ileXp_seq_F	TGAAAGCCCAGGGTTGATAC
ileXp_seq_R	TGGCGCAGATGCAAATCCCT
Chaperone promoters	
DnaKJ_seq_F	AGGCTGGCGGAAATCGTAAA
DnaKJ_seq_R	GCGACCAATCAGGCGTTTAA
GroESL_seq_F	CAAACACGCCAGTGCCTAAT
GroESL_seq_R	GGCTTCACTTCGCCATTTTC
GrpE_seq_F	CAGCGGTAGAGCATTTCATG
GrpE_seq_R	ACGCAAAATGCCGTCACGTT

Table 9. Sequencing primers for promoters of tRNA and chaperone

[a]All primer sequences are 5′ -> 3′ direction

2.5 High-throughput screening (HTS) for desired mutants after MAGE

In order to obtain cells containing desired mutation after MAGE, screening method should be established because desired mutants are small amount in mutant mixture.[40] After MAGE recovery, seed was inoculated using 1/100 dilution into M9 minimal media (0.2% glucose) supplemented with appropriate chloramphenicol and kanamycin in order to enrich mutant cells in MAGE mutant. Enrichment was repeated 5 times. After enrichment, each MAGE mutant mixture was inoculated into LB medium supplemented with appropriate antibiotics and cultivated overnight. MAGE mutant mixtures were inoculated into M9 minimal medium, and growth of mutant mixtures were analyzed by spectrophotometer (Libra). Mutant mixture diluted until 10^3 cells was plated in M9 plate, in order to separate individual mutants. Single mutant cells were inoculated in 300 uL LB medium with appropriate antibiotics in 96-well 2 mL microplates, respectively. After cultivation, each culture was re-inoculated in 400 uL M9 minimal medium with 0.2% glucose and appropriate antibiotics. Growth of single mutant cells in M9 cultures were analyzed by TECAN after 24 hours. By TECAN growth data, high growing cells were ranked and 4 high ranking strains were cultivated in 30 mL M9 media in culture flasks. Optical density was analyzed by spectrophotometer.

2.6 SDS-PAGE and western blotting of target proteins

SDS-page was applied for separating proteins. Different concentration of acrylamide gels was used (12% or 12.5%). The 12.5% separation gel and 5% stacking gel led to satisfactory separation in the experiment. Electrophoresis was performed at room temperature (25℃) by voltage at maxims 200V and by electrical current at 40mA per gel for one and a half hour. Boil the samples at 95℃ for five minutes, which denatures the samples more and help them running well. All of the samples loaded on the gel had the same concentration.

After SDS-PAGE gels are transferred to nitrocellulose membrane in transfer buffer (500mM Glycine, 50mM Tris-HCl, 0.01% SDS, 20% methanol) buffer at 500 mA for 1h. Membranes are washed in TBS-T (10mM Tris-HCl, 100 mM NaCl, 0.1% Tween20 at pH 7.4) and then blocked with 3% non-fat milk extract in TBS-T for 30min to 1h. Membranes are exposed to primary anti-FLAG antibody (Cell Signaling TECHNOLOGY) in 1% BSA in TBS-T for 1h at room temperature to overnight at 4°C. Membranes are washed with TBS-T three times for ten minutes. Then are incubated with appropriate secondary antibodies (Cell Signaling TECHNOLOGY) coupled to the alkaline phosphatase (AP) for 1h at room temperature to overnight at 4°C. After washing the membranes three times for ten minutes, signals were visualized using mixture of nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) producing response of color formation. (Sigma-Aldrich).

2.7 Heterologous expression of other plant genes in engineered *E. coli* strains

For confirming versatility of our mutant strains, the expression of protein from other plant organism. FLAG-tagged pBbB6a-gDS plasmid was transformed to MAGE mutants showing high *Arabidopsis* protein expression level. In order to compare the enhancement of engineered strains, wild type with pBbB6a-gDS was made, and protein expression of gDS was analyzed by SDS-PAGE and western blotting. Procedures of SDS-PAGE and western blotting were same with 2.6 section.

3. Results

3.1 Analysis of *A. thaliana* protein expression in *E. coli*

When *Arabidopsis*'s genes are heterologously expressed in *E. coli*, there are two fates of protein. First, some proteins are soluble and functional.[49] Second, some cases show insoluble protein expression and formation of inclusion body, or very low protein expression.[49] In order to find target proteins which is not expressed well in *E. coli*, we inserted 9 kinds of *Arabidopsis*'s gene into pET30a(+) vector. This vector uses T7 promoter for expression of desired gene. Because of this reason, large amount of protein expression were observed by SDS-PAGE.

On polyacrylamide gel, 9 *Arabidopsis*'s protein expression were observed. TSA1 and ADT2 proteins showed low expression level on the gel. It means that translation of these two proteins was interrupted. PROC1 and HIS6 proteins showed insoluble expression. These two proteins were 30 KDa and 33 KDa. On the gel, protein bands were located in insoluble fraction. It means that these two proteins form misfolding or truncated structure. It causes formation of inclusion body and cell death. The remainders showed soluble protein expression on the gel. As an outcome of these protein expressions, TSA1, ADT2, PROC1 and HIS6 proteins were selected for our study.

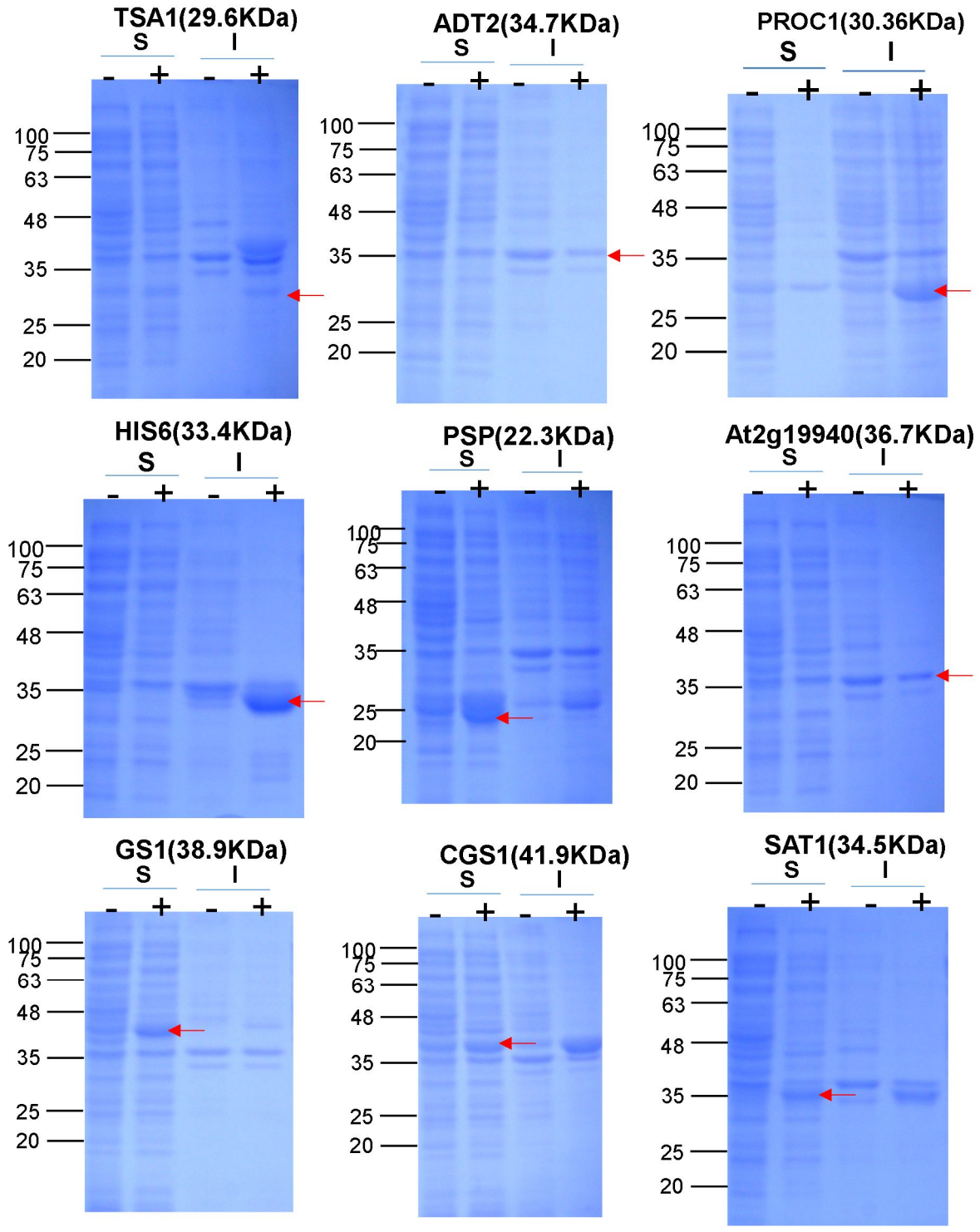


Figure 3. SDS-PAGE of plant gene product expressed in pET30a(+) vector. Red arrow indicated target proteins. S, soluble fraction; I, insoluble fraction

3.2 Replacement of *E. coli* genes with their corresponding *A. thaliana* gene

We selected 4 kinds of *Arabidopsis*' genes by overexpression in *E. coli*. Because target genes encode essential enzymes related in amino acid biosynthesis, knockout strains of *E. coli* genes cannot grow well in M9 minimal media excluded protein source. In the figure, *E. coli* gene knockout strains did not grow well in M9 minimal media. Because essential enzymes for amino acid biosynthesis were deleted on the genome, protein synthesis for cell growth was repressed.

E. coli native genes were replaced with *Arabidopsis*'s genes as same function. If *Arabidopsis*'s protein expression in *E. coli* is well, replacement strains could grow in M9 minimal media without amino acids. However if there are problems of *Arabidopsis*'s protein expression, replacement strains cannot grow well in M9 media. As shown in the figure, TSA1, ADT2 and HIS6 replacement strains showed low growth in M9 media along with knockout strains. It means that these *Arabidopsis*'s proteins were not expressed well. However, PROC1 replacement strain showed higher growth than knockout strain. It means that *E. coli* native proC protein could be substituted for *Arabidopsis*'s PROC1 protein.

If TSA1, ADT2 and HIS6 proteins are expressed well, we expect that replacement strains could grow well in M9 media. This concept could be applied as high-throughput screening of MAGE. By using MAGE, we could introduce various mutation on the genome simultaneously. In this case, these mutation could help soluble and functional expression of *Arabidopsis*'s protein in *E. coli*. It cause increase of cell growth of MAGE mutants in M9 minimal media. We could select desired mutants by this screening method.

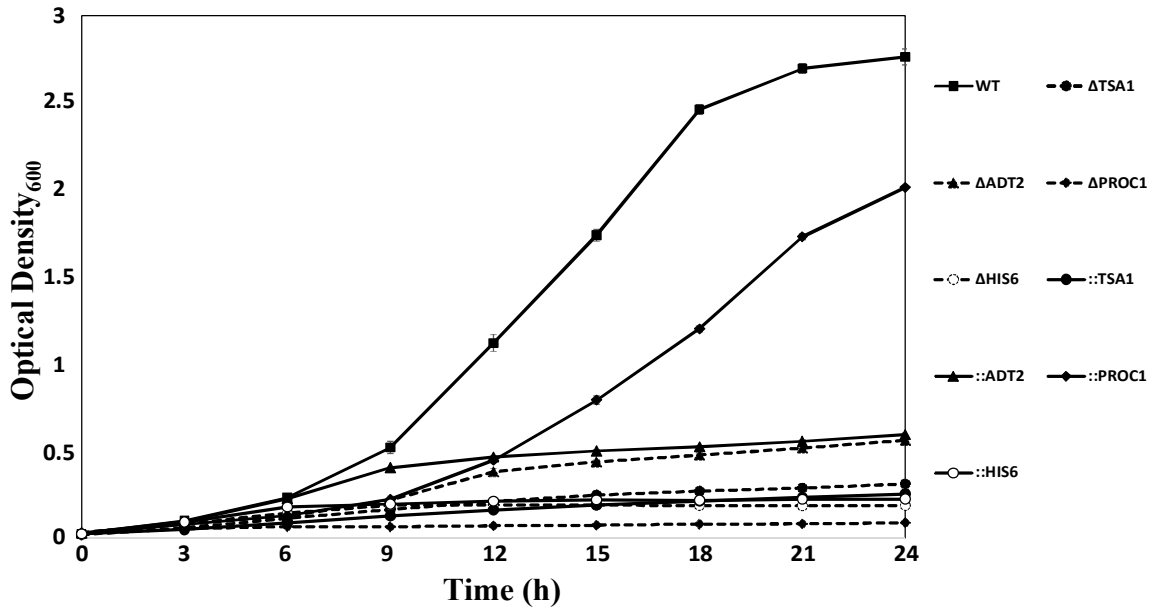


Figure 4. Comparison of cell growth of *E. coli* strains expressing *A. thaliana* genes. Error bars

indicate the standard deviation of experiments performed in triplicate.

3.3 Genome engineering of rare tRNA promoters to enhance tRNA expression

In order to adjust *E. coli* codon usage bias to *Arabidopsis*'s, we engineered rare tRNA promoter by MAGE. *E. coli* representative rare tRNAs are *argU* (AGA, AGG), *ileY* (AUA), *leuW* (CUA). We targeted these rare tRNAs and additional tRNAs such as *ileX* (AUA) and *argW* (AGG) which binds to rare codon. *ileY*, *leuW*, *ileX* and *argW* tRNA promoters were changed to constitutive promoter such as CP25. *argU* tRNA gene is activated by upstream activator sequences (UASs) and have -35 and -10 sequences which are very similar to the sigma70 consensus sequences. However, *argU* tRNA concentration in the cell is very low. Mizuno's group used *argU* promoter from -850 to -43 for reporter protein expression. The results of their research indicate that the limiting sequence could involve nucleotides from +2 to +45, especially +7 and +8 nucleotides. So we changed nucleotides A-7 and C-8 to T-7 and A-8, in order to reduce the extent of the dyad symmetry. These promoter changes were conducted by MAGE. The number of mutation cases are 120. In the figure 7A, TSA1 and ADT2 tRNA MAGE mutant mixture showed slight increase of growth in M9 minimal media. individual mutants of TSA1 tRNA MAGE mutants showed slight increase of growth. Figure 7B indicated that ADT2 tRNA MAGE affected growth increase. AtM4 strain showed higher growth than ::ADT2 strain.

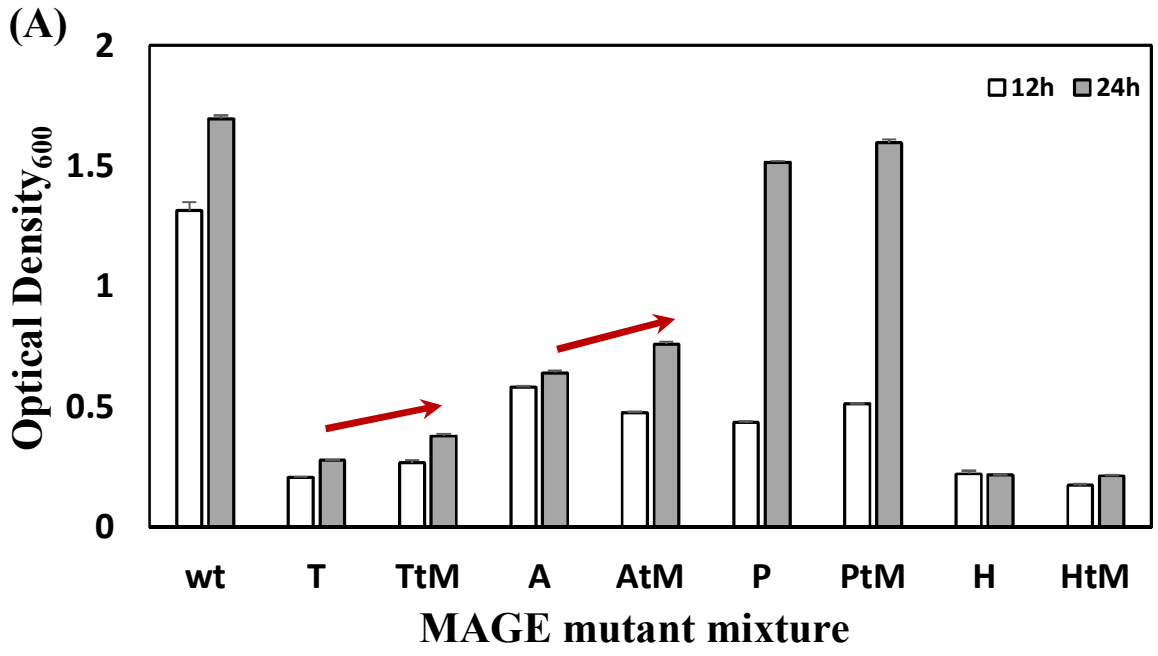


Figure 7A. Cell growth analysis after tRNA promoter engineering. WT, MG-MAGE; T, ::TSA1; TtM, ::TSA1 with tRNA MAGE; A, ::ADT2; AtM, ::ADT2 with tRNA MAGE; P, ::PROC1; PtM, ::PROC1 with tRNA MAGE; H, ::HIS6; HtM, ::HIS6 tRNA MAGE. Error bars indicate the standard deviation of experiments performed in triplicate.

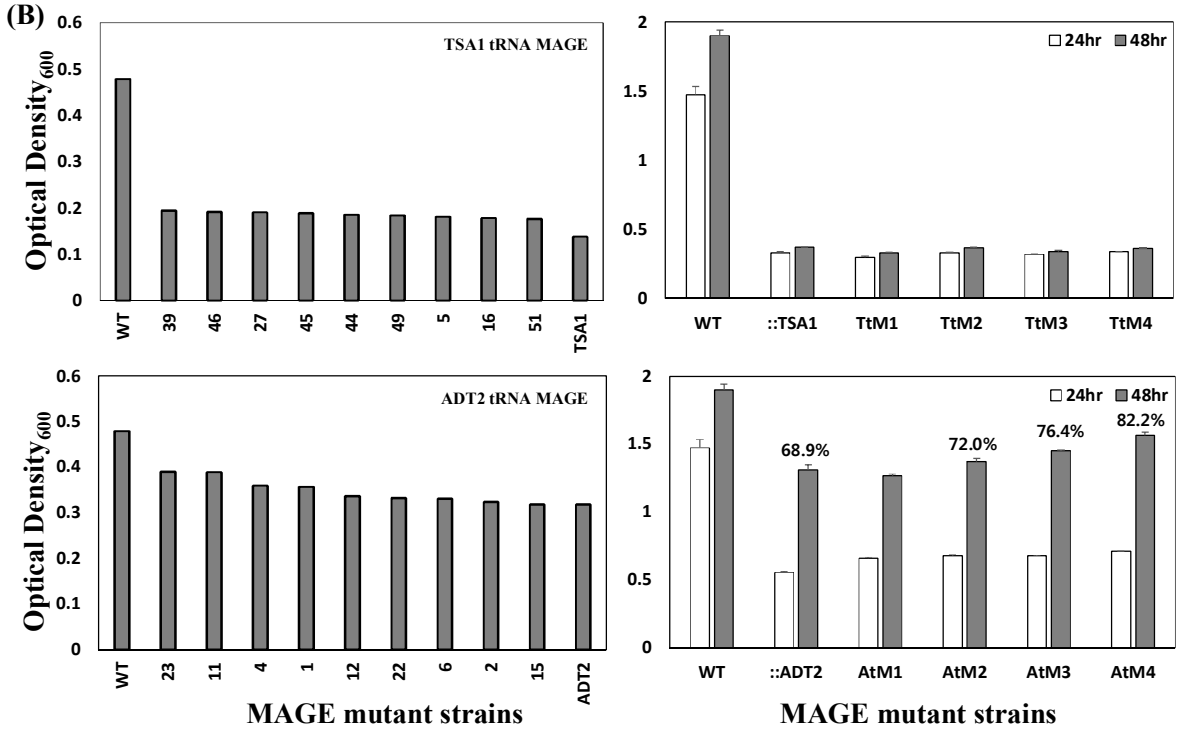


Figure 5B. Comparison of cell growth of individual mutants of tRNA promoter mutants. Individual mutants were ranked by TECAN. (left) 4 high ranked strains were analyzed by OD_{600}. (right) Cell growth percentage was calculated on the basis of wild type strain. Error bars indicate the standard deviation of experiments performed in triplicate.

3.4 Genome engineering of chaperone promoters

For solubility increase of insoluble proteins, molecular chaperone could help protein synthesis or refolding. We engineered the promoters of chaperones such as DnaKJ, GrpE, GroESL. These original promoters correspond to sigma factor 32 of RNA polymerase which is expressed after heat-shock. So we changed original promoters to 3 types of constitutive promoters such as CP12 (weak), CP6 (medium), CP25 (strong). These promoter changes were conducted by MAGE. The number of mutation cases are 216. TSA1 chaperone MAGE mutant mixture showed higher growth than no-MAGE TSA1 strain. When individual mutants were separated, many mutants showed higher growth than no-MAGE TSA1 strain. It means that chaperone expression affects TSA1 protein functionality. By using these fast growing TSA1 chaperone MAGE mutant, TSA1 protein overexpression was conducted by IPTG induction of pBbB6a-TSA1. This result described in section 3.6.

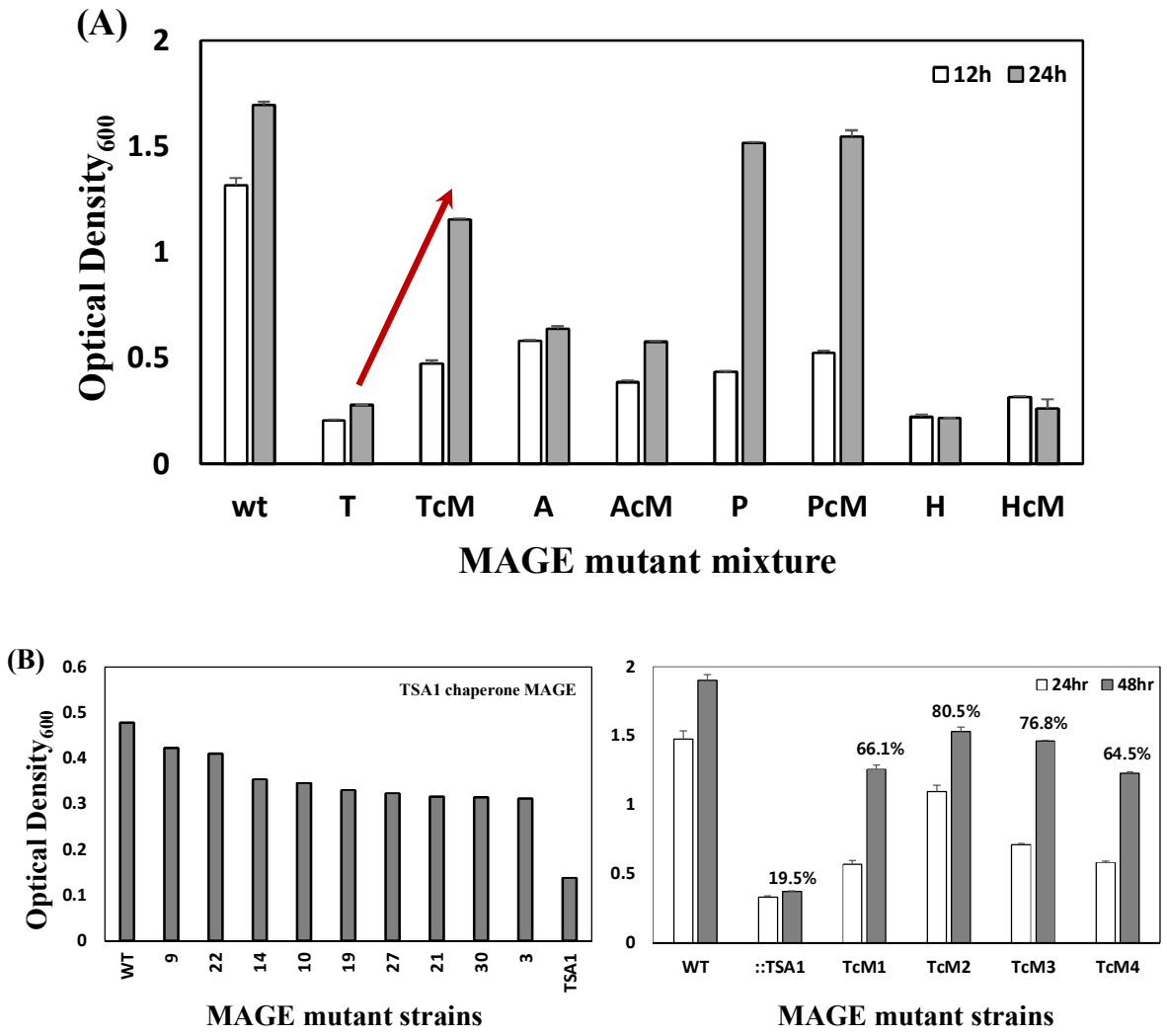


Figure 6. (A) Cell growth analysis after chaperone promoter engineering. WT, MG-MAGE; T, ::TSA1; TcM, ::TSA1 with chaperone MAGE; A, ::ADT2; AcM, ::ADT2 with chaperone MAGE; P, ::PROC1; PcM, ::PROC1 with chaperone MAGE; H, ::HIS6; HcM, ::HIS6 chaperone MAGE. **(B) Comparison of cell growth of individual mutants of chaperone promoter mutants.** Individual mutants were ranked by TECAN. (left) 4 high ranked strains were analyzed by OD_{600}. (right) Cell growth percentage was calculated on the basis of wild type strain. Error bars indicate the standard deviation of experiments performed in triplicate.

3.5 Simultaneous genome engineering of rare tRNA and chaperone promoters

In order to get comprehensive enhancement of *Arabidopsis*'s protein expression, promoter changes of rare tRNA genes (such as *argU*, *ileY*, *leuW*, *ileX* and *argW*) and chaperone genes (such as *dnaK-dnaJ*, *grpE*, *groEL-groES*) were conducted by co-MAGE. The number of mutation cases are 25,920. In MAGE mutant mixture, TSA1 tRNA+chaperone MAGE showed increase of growth and HIS6 tRNA+chaperone MAGE showed increase of growth compared to no-MAGE HIS6 strain.

Growth of MAGE mutant mixtures was analyzed in M9 minimal media. TtcM and HtcM showed growth increase. It means that desired mutants involve in mutant mixture. Individual mutants were separated by plating mixture on M9 agar plate. After selection, growth was compared by TECAN. Ranked cells showed higher growth than unchanged strain. As Figure 9B shown, HtcM4 showed higher growth than ::HIS6 strain.

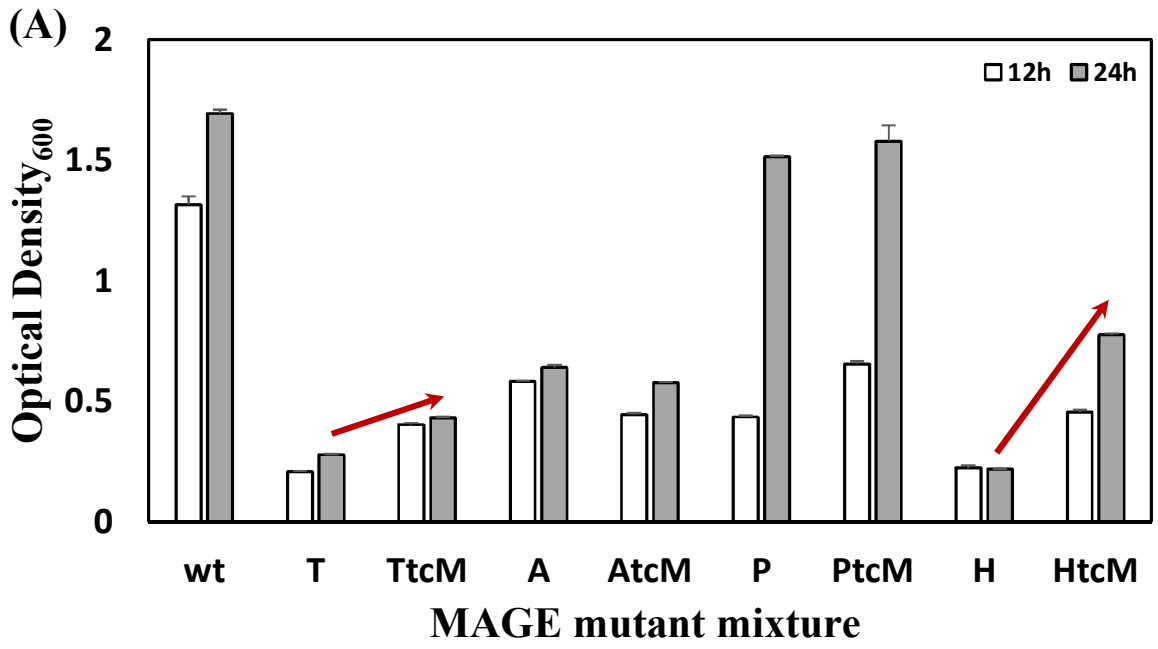


Figure 7A. Cell growth analysis after tRNA and chaperone promoter engineering. WT, MG-MAGE; T, ::TSA1; TtcM, ::TSA1 with tRNA+chaperone co-MAGE; A, ::ADT2; AtcM, ::ADT2 with tRNA+chaperone co-MAGE; P, ::PROC1; PtcM, ::PROC1 with tRNA+chaperone co-MAGE; H, ::HIS6; HtcM, ::HIS6 tRNA+chaperone co-MAGE. Error bars indicate the standard deviation of experiments performed in triplicate.

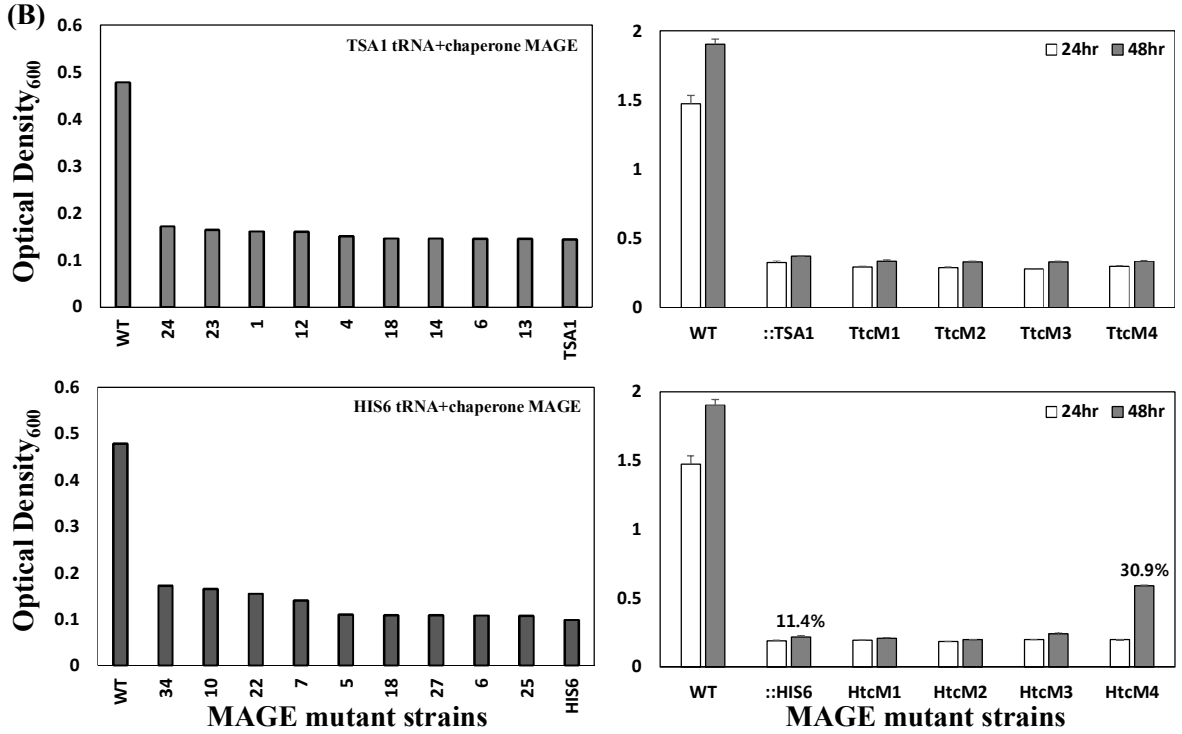


Figure 9B. Comparison of cell growth of individual mutants of tRNA/chaperone promoter mutants. Individual mutants were ranked by TECAN. (left) 4 high ranked strains were analyzed by OD_{600}. (right) Cell growth percentage was calculated on the basis of wild type strain. Error bars indicate the standard deviation of experiments performed in triplicate.

3.6 Protein expression level of wild types and engineered strains

Protein expression level was analyzed by SDS-PAGE and western blotting. After MAGE, some mutants showed increase of growth compared with no-MAGE strain in M9 minimal media. Protein expression confirmation of MG-MAGE strains harboring *Arabidopsis*'s gene were conducted for comparison between wild type strain and engineered strains. In figure 10A, gDS gene was expressed in soluble fraction and HIS6 gene was expressed in insoluble fraction. TSA1 and ADT2 proteins were not observed in western blot results. In order to check protein expression, pBbB6a-TSA1, -ADT2, -HIS6 and -gDS plasmids were transformed to MAGE mutant strains. TSA1 overexpression was conducted in TSA1 chaperone MAGE mutants. ADT2 overexpression was conducted in ADT2 tRNA MAGE mutants. HIS6 overexpression was conducted in HIS6 tRNA+chaperone MAGE mutants. In the results of figure 10B, TSA1 protein expression of TcM2 and TcM3 strains was shown as soluble expression. HIS6 protein expression of HtcM4 was also shown as soluble expression. These results indicated that TcM2, TcM3 and HtcM4 strains have chance of candidate for optimal plant gene expression *E. coli* strain.

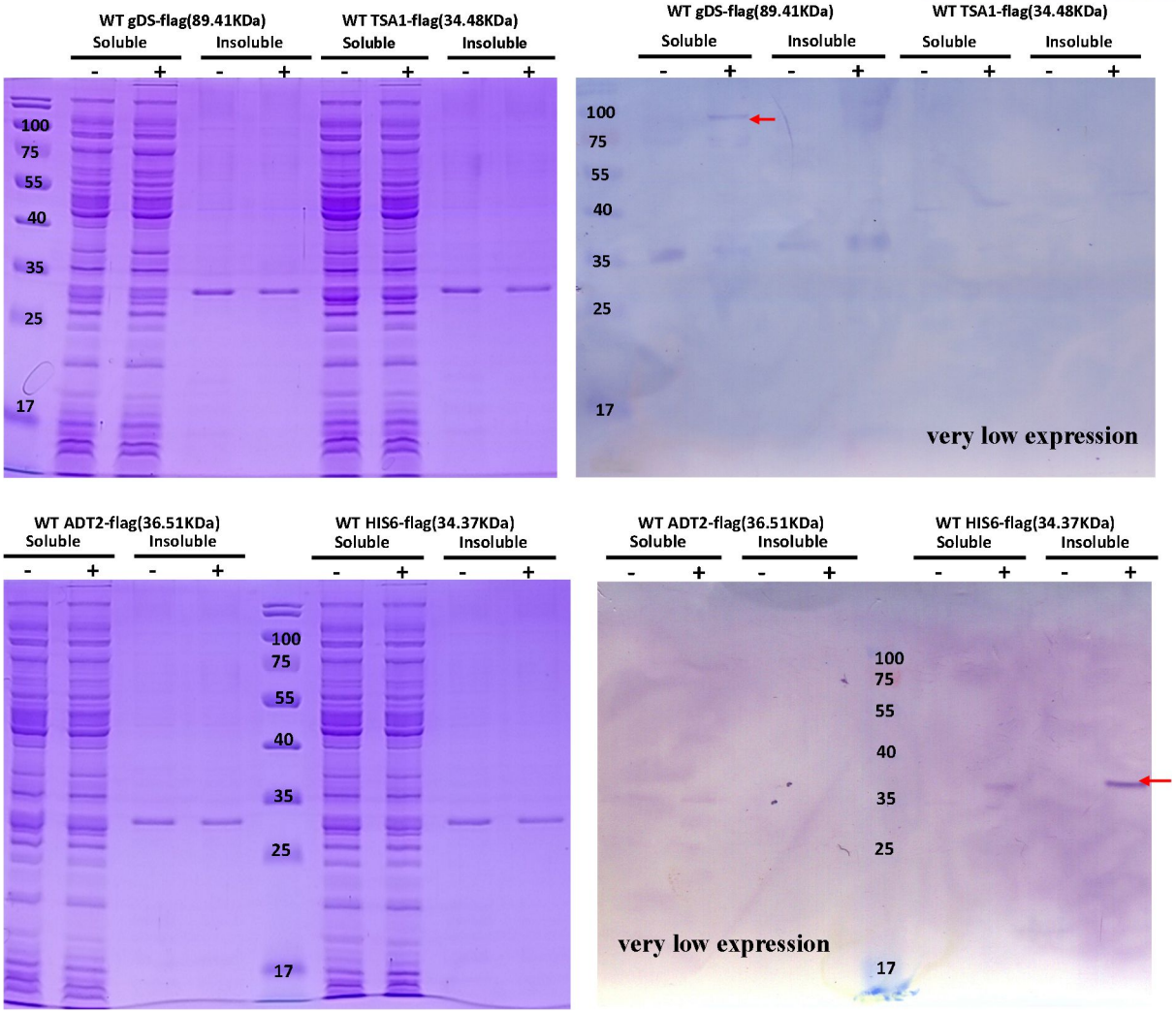


Figure 10A. SDS-PAGE and western blotting of target proteins expressed from wild type strains. Red arrow indicated desired protein. S, soluble fraction; I, insoluble fraction; +, IPTG induced; -, IPTG uninduced.

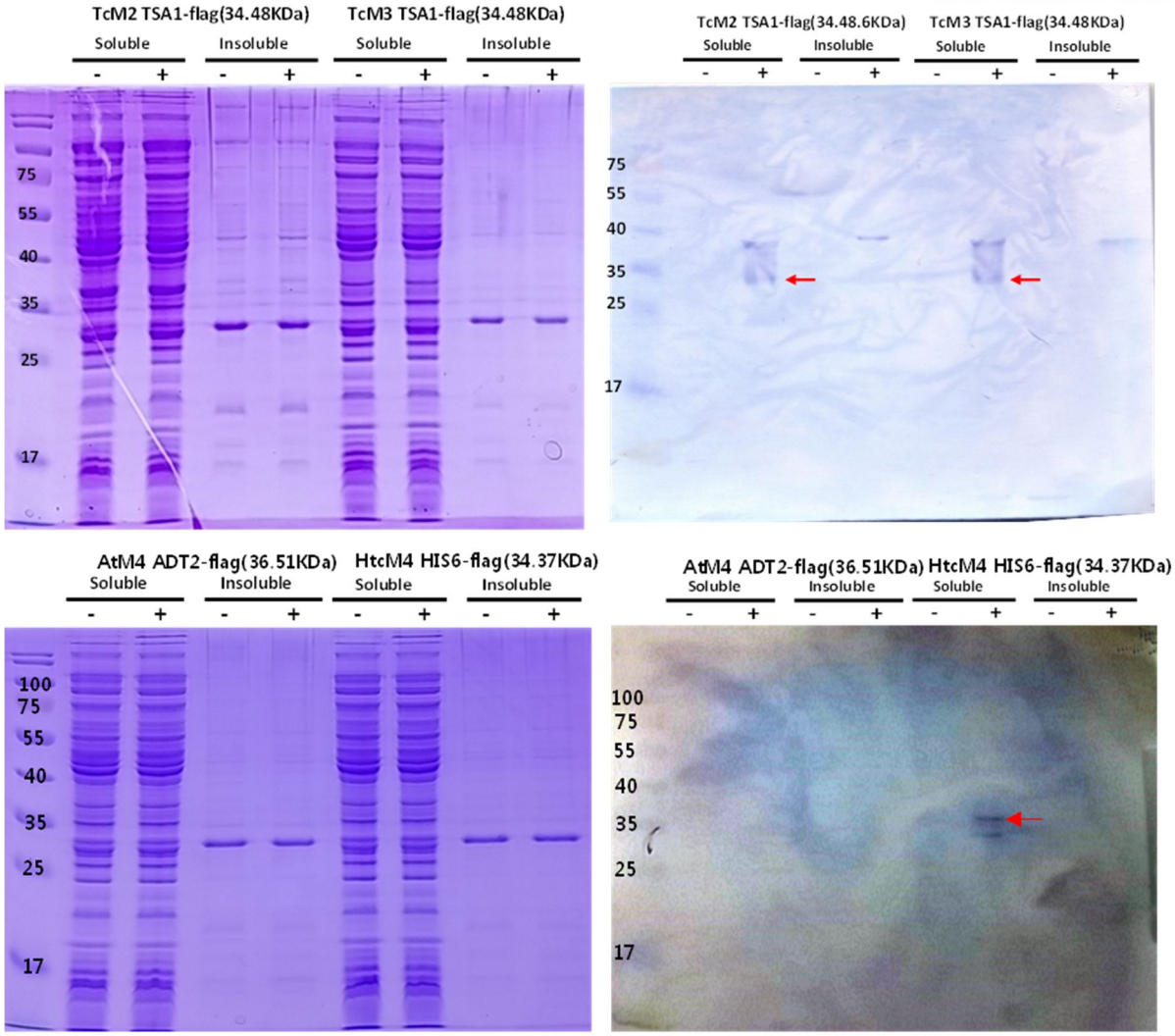


Figure 10B. SDS-PAGE and western blotting of target proteins expressed from engineered strains. Red arrow indicated desired protein. S, soluble fraction; I, insoluble fraction; +, IPTG induced; -, IPTG uninduced.

3.7 Versatility of effective plant gene expression strains

In order to investigate versatility of our engineered strains, we cloned *Panax ginseng* Dammarenediol synthase (gDS) gene into pBbB6a plasmid backbone. gDS is the enzyme related in ginsenoside biosynthesis. Engineered strains harboring pBbB6a-gDS-flag plasmid were cultivated in LB medium. Protein expression was checked by SDS-PAGE and western blot. In western blot result, TcM3 strains showed soluble expression of gDS protein.

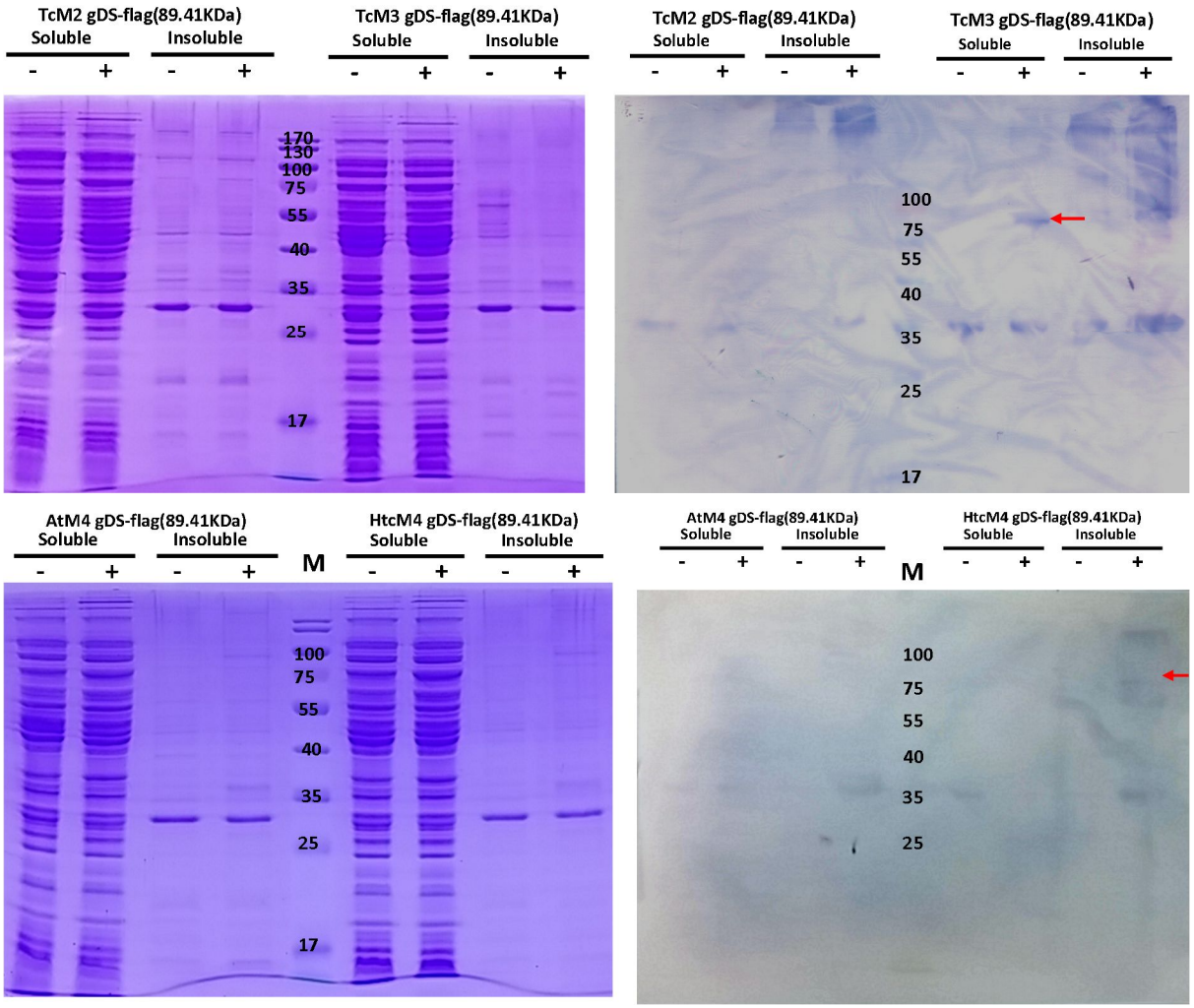


Figure 11. Versatility of engineered strains. *Panax ginseng* Dammarenediol syntheas (gDS) genes expressed in engineered *E. coli*. S, soluble fraction; I, insoluble fraction; +, IPTG induced; -, IPTG uninduced.

4. Discussion

In this work, we have engineered an *Escherichia coli* which enhances heterologous protein expression especially plant-derived protein. Previous studies using plasmid harboring extra copies of rare tRNA genes (*argU, ileY, leuW, ileX* and *argW*) and plasmid harboring additional chaperone genes (*dnaK-dnaJ*, *grpE*, *groES-groEL*) have problems such as addition and purification of antibiotics for these tRNA and chaperone expressing plasmid.[47, 52] However, plasmid-based tRNA and chaperone expression system cannot control combination of tRNAs and chaperones simultaneously. In order to avoid these problems, we constructed *E. coli* strains with engineering tRNA and chaperone promoters on genome. TSA1 tRNA MAGE strains and ADT2 tRNA MAGE strains showed slight increase of growth in M9 minimal media. It means that tRNA promoter engineering by MAGE affect a little of enhancement of *Arabidopsis*'s TSA1 and ADT2 expression. TSA1 chaperone MAGE strains showed considerable increase of growth in M9 minimal media. It means that chaperone expression affects enhancement of *Arabidopsis*'s TSA1 protein soluble and functional. HIS6 tRNA+chaperone MAGE strains showed increase of growth in M9 minimal media. *Arabidopsis*'s HIS6 protein were not expressed well in tRNA MAGE strains and chaperone strains. However, tRNA+chaperone co-MAGE strains expressed soluble and functional HIS6 protein. It means that combination of tRNA and chaperone engineering could help to enhance heterologous protein expression which is not expressed well in wild type *E. coli* strain. In the results of SDS-PAGE and Western blotting, TSA1 chaperone MAGE strains showed enhancement of protein expression on gel. The control ::TSA1 strain showed no expression in western blot result, but TSA1 chaperone MAGE strains showed soluble protein expression. Also, HIS6 tRNA+chaperone MAGE strains showed soluble expression of target protein against the control ::HIS6 strain.

However, all MAGE mutants did not show enhancement of growth and protein expression. The reason of this problem could be 1) insufficient enrichment of MAGE mutant mixture, 2) insufficient MAGE cycle for all coverage, 3) no effect of increase of rare tRNA and chaperone concentration. First, when enrichment of MAGE mutant mixture is not sufficient for predominance of desired mutant. After MAGE, there are a small amount of desired mutants in the mixture. During enrichment cycles, desired mutants grow faster than unchanged strain. So enrichment makes desired mutants to increase its populations. Deficient enrichment leads no difference of growth in M9 minimal media. It could be enhanced by more enrichment. Second, insufficient MAGE cycle could not cover all number of mutation cases.[13, 40] By being more MAGE cycles, MAGE efficiency is increased step by step.[40, 64] Although number of cases are small, mutation efficiency is very low because homologous region length of MAGE primers are small.[13] Because of this reason, a small number of MAGE cycle could

cause no change of target sequence. Third, the concentration of rare tRNAs and chaperones is not sufficient to treat difference of codon usage bias and refolding of misfolded protein. It means that tRNAs and chaperones are expressed from genome. Although these promoter mutation is constitutive promoter, there is no guarantee for capability of necessary tRNA and chaperone concentration. This might be overcome by adding extra copies of tRNA and chaperone on genome, or adding additional strong promoter near original promoter. For *E. coli* strain development of universal optimal plant protein expression, the assistance of tRNA and chaperone is necessary because difference of codon usage bias is major obstacle of heterologous protein expression.[6, 9, 30]

Many bacterial hosts were optimized for heterologous protein production until now.[8] Primarily, all bacteria could be used for heterologous protein production.[8] Many genomics studies offers new information about the bacterial hosts which are frequently or rarely used.[65] Unfrequently used codons could be a reason to alter from an *E. coli* system to another host.[44] However, *E. coli* is still the most generally used host for industrial production of useful chemicals.[8]

Heterologous protein expression involves the analysis of genes and the transfer of the corresponding DNA fragments to hosts other than the original DNA fragments encoding proteins.[10] Protein isolation, directly from plant proteins, could be costly, intractable and taking long time, and heterologous expression provides a convenient way.[66] This method allows large-scale production of plant-derived proteins in *E. coli* to study their biochemical and biophysical characters.

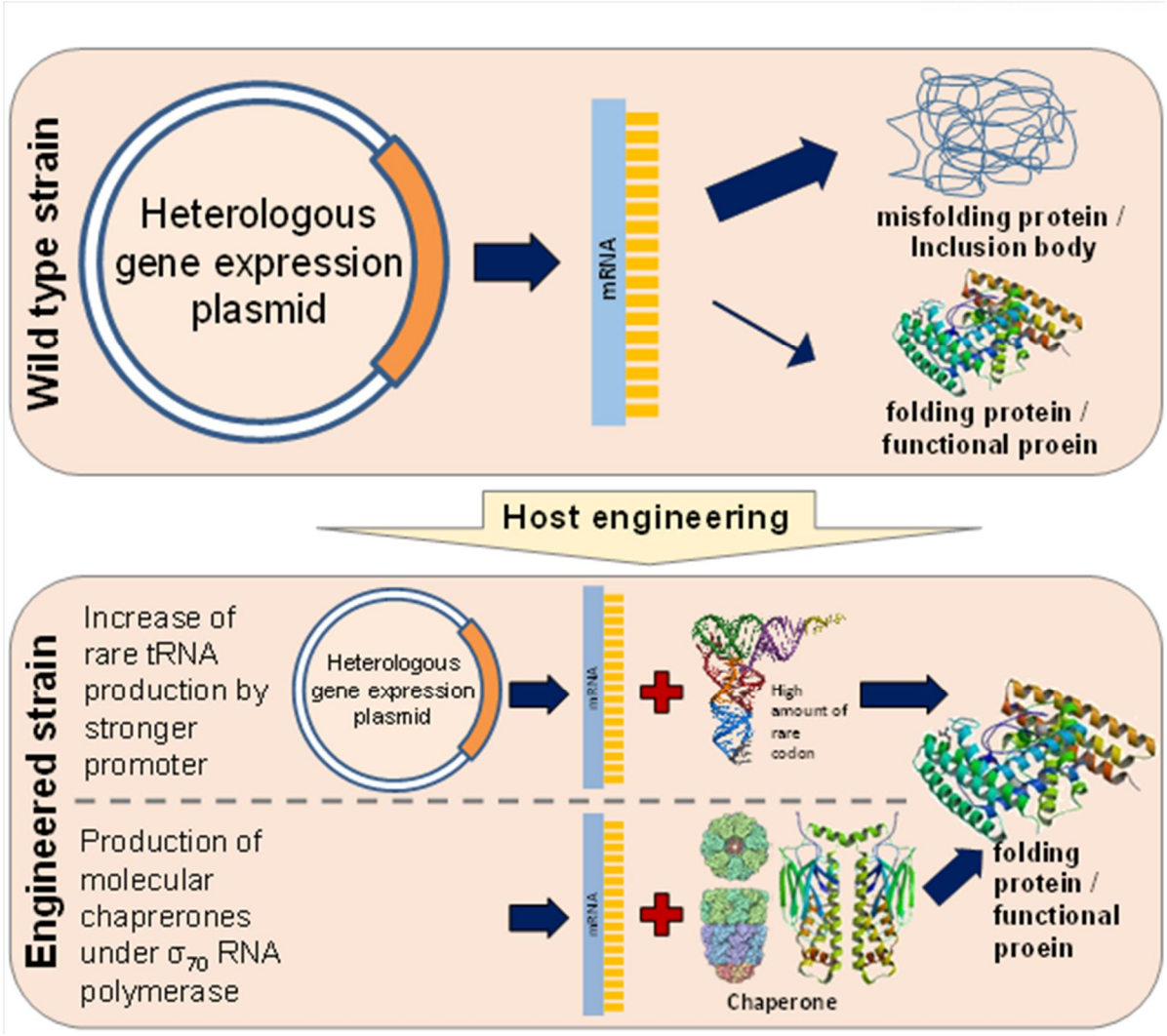


Figure 82. Overview of this study. There are two parts of enhancement of heterologous gene expression in *E. coli* strains.

5. Conclusions

In this study, we constructed optimal plant gene expression *E. coli* strains such as constitutive expression of rare tRNA, chaperone and tRNA+chaperone expressed from genome. By changing native promoters of tRNAs and chaperones to constitutive promoters by MAGE, *Arabidopsis*'s proteins were synthesized to soluble and functional expression. For high-throughput screening of MAGE, we constructed replacement of *E. coli* native genes related in amino acid biosynthesis to corresponding A. thaliana genes. By comparing growth in M9 minimal media, we can know that tRNA and chaperone promoter engineering affects A. thaliana protein expression. Through SDS-PAGE and western blotting data, we confirmed protein expression level and soluble/insoluble expression aspects. As a result, we developed *E. coli* strains of optimal plant protein expression. It is applicable to pharmaceutical production originated various plants and industrial cell factory of plant-derived useful chemicals from *E. coli*. Also, this study could be applied to plant genetic research.

References

1. Balandrin, M.F., Klocke, J.A., Wurtele, E.S. & Bollinger, W.H. Natural plant chemicals: sources of industrial and medicinal materials. *Science* **228**, 1154-1160 (1985).
2. Pryde, E.H., Princen, L. & Mukherjee, K.D. New sources of fats and oils, Vol. 9. The American Oil Chemists Society (1981).
3. Tyler, V., Brady, L. & Robbers, J. Pharmacognosy, 8th edit. *Philadelphia: Lea and Febiger* (1981).
4. Fischer, R., Stoger, E., Schillberg, S., Christou, P. & Twyman, R.M. Plant-based production of biopharmaceuticals. *Current opinion in plant Biology* **7**, 152-158 (2004).
5. Hellwig, S., Drossard, J., Twyman, R.M. & Fischer, R. Plant cell cultures for the production of recombinant proteins. *Nature biotechnology* **22**, 1415-1422 (2004).
6. Rosano, G.L. & Ceccarelli, E.A. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. *Microbial cell factories* **8**, 41 (2009).
7. Rosano, G.L. & Ceccarelli, E.A. Recombinant protein expression in microbial systems. *Frontiers in microbiology* **5** (2014).
8. Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology* **72**, 211-222 (2006).
9. Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. *Trends in biotechnology* **22**, 346-353 (2004).
10. Hannig, G. & Makrides, S.C. Strategies for optimizing heterologous protein expression in Escherichia coli. *Trends in biotechnology* **16**, 54-60 (1998).
11. Kopito, R.R. Aggresomes, inclusion bodies and protein aggregation. *Trends in cell biology* **10**, 524-530 (2000).
12. Vallejo, L.F. & Rinas, U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial Cell Factories* **3**, 11 (2004).
13. Wang, H.H. & Church, G.M. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods Enzymol* **498**, 409-426 (2011).
14. Sørensen, H.P. & Mortensen, K.K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of biotechnology* **115**, 113-128 (2005).
15. Baneyx, F. & Mujacic, M. Recombinant protein folding and misfolding in *Escherichia coli*. *Nature biotechnology* **22**, 1399-1408 (2004).
16. Angov, E., Hillier, C.J., Kincaid, R.L. & Lyon, J.A. Heterologous protein expression is

enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PloS one* **3**, e2189 (2008).

17. Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D. & Koornneef, M. Arabidopsis thaliana: a model plant for genome analysis. *Science* **282**, 662-682 (1998).
18. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic acids research* **28**, 292-292 (2000).
19. Sharp, P.M. et al. Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity. *Nucleic Acids Research* **16**, 8207-8211 (1988).
20. Wang, L., Brock, A., Herberich, B. & Schultz, P.G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498-500 (2001).
21. Crick, F. Codon—anticodon pairing: the wobble hypothesis. *Journal of molecular biology* **19**, 548-555 (1966).
22. Luo, L.F. The degeneracy rule of genetic code. *Origins of Life and evolution of the biosphere* **18**, 65-70 (1988).
23. Initiative, A.G. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796 (2000).
24. Blattner, F.R. et al. The complete genome sequence of *Escherichia coli* K-12. *science* **277**, 1453-1462 (1997).
25. Chiapello, H., Lisacek, F., Caboche, M. & Hénaut, A. Codon usage and gene function are related in sequences of *Arabidopsis thaliana. Gene* **209**, GC1-GC38 (1998).
26. Andersson, G. & Kurland, C. An extreme codon preference strategy: codon reassignment. *Molecular biology and evolution* **8**, 530-544 (1991).
27. Emilsson, V. & Kurland, C.G. Growth rate dependence of transfer RNA abundance in Escherichia coli. *The EMBO journal* **9**, 4359 (1990).
28. Elf, J., Nilsson, D., Tenson, T. & Ehrenberg, M. Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science* **300**, 1718-1722 (2003).
29. Saxena, P. & Walker, J.R. Expression of argU, the *Escherichia coli* gene coding for a rare arginine tRNA. *Journal of bacteriology* **174**, 1956-1964 (1992).
30. Angov, E. Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnology journal* **6**, 650-659 (2011).
31. Lemos, B., Bettencourt, B.R., Meiklejohn, C.D. & Hartl, D.L. Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Molecular biology and evolution* **22**, 1345-1354 (2005).
32. Lilie, H., Schwarz, E. & Rudolph, R. Advances in refolding of proteins produced in *E. coli*.

Current opinion in biotechnology **9**, 497-501 (1998).

33. Morimoto, R.I., Tissières, A. & Georgopoulos, C. The biology of heat shock proteins and molecular chaperones. (Cold Spring Harbor Laboratory Press New York, 1994).
34. Parsell, D. & Lindquist, S. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annual review of genetics* **27**, 437-496 (1993).
35. Ben-Zvi, A.P. & Goloubinoff, P. Review: mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. *Journal of structural biology* **135**, 84-93 (2001).
36. Watanabe, Y.-h., Motohashi, K., Taguchi, H. & Yoshida, M. Heat-inactivated proteins managed by DnaKJ-GrpE-ClpB chaperones are released as a chaperonin-recognizable non-native form. *Journal of Biological Chemistry* **275**, 12388-12392 (2000).
37. Bochkareva, E., Lissin, N., Flynn, G., Rothman, J. & Girshovich, A. Positive cooperativity in the functioning of molecular chaperone GroEL. *Journal of Biological Chemistry* **267**, 6796-6800 (1992).
38. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97**, 6640-6645 (2000).
39. Simon, R., Priefer, U. & Pühler, A. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. *Nature Biotechnology* **1**, 784-791 (1983).
40. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898 (2009).
41. Sawitzke, J.A. et al. Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *Journal of molecular biology* **407**, 45-59 (2011).
42. Wang, H.H. et al. Multiplexed in vivo his-tagging of enzyme pathways for in vitro single-pot multienzyme catalysis. *ACS synthetic biology* **1**, 43-52 (2012).
43. Ryu, Y.S. et al. A Simple and Effective Method for Construction of Escherichia coli Strains Proficient for Genome Engineering. *PloS one* **9**, e94266 (2014).
44. Kane, J.F. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current opinion in biotechnology* **6**, 494-500 (1995).
45. Bonekamp, F., Andersen, H.D., Christensen, T. & Jensen, K.F. Codon-defined ribosomal pausing in Escherichia coli detected by using the *pyrE* attenuator to probe the coupling between transcription and translation. *Nucleic acids research* **13**, 4113-4123 (1985).
46. Deana, A., Ehrlich, R. & Reiss, C. Silent mutations in the *Escherichia coli* ompA leader peptide region strongly affect transcription and translation in vivo. *Nucleic acids research* **26**, 4778-4782 (1998).

47. Carstens, C.-P. in *E. coli* Gene Expression Protocols 225-233 (Springer, 2003).
48. Mirzahoseini, H., Mafakheri, S., Mohammadi, N.S., Enayati, S. & Mortazavidehkordi, N. Heterologous proteins production in *Escherichia coli*: an investigation on the effect of codon usage and expression host optimization. *Cell Journal(Yakhteh)* **12**, 453-458 (2011).
49. Sørensen, H.P. & Mortensen, K.K. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial cell factories* **4**, 1 (2005).
50. Sørensen, M.A., Kurland, C. & Pedersen, S. Codon usage determines translation rate in *Escherichia coli*. *Journal of molecular biology* **207**, 365-377 (1989).
51. Dobson, C.M. Protein folding and misfolding. *Nature* **426**, 884-890 (2003).
52. Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H. & Yura, T. Chaperone Coexpression Plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese Cedar Pollen, Cryj2, in *Escherichia coli*. *Applied and environmental microbiology* **64**, 1694-1699 (1998).
53. Smith, D.W. Problems of translating heterologous genes in expression systems: the role of tRNA. *Biotechnology progress* **12**, 417-422 (1996).
54. Miroux, B. & Walker, J.E. Over-production of Proteins in< i> Escherichia coli</i>: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *Journal of molecular biology* **260**, 289-298 (1996).
55. Casadaban, M.J. & Cohen, S.N. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *Journal of molecular biology* **138**, 179-207 (1980).
56. Datta, S., Costantino, N. & Court, D.L. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109-115 (2006).
57. Kim, O.T. et al. Upregulation of ginsenoside and gene expression related to triterpene biosynthesis in ginseng hairy root cultures elicited by methyl jasmonate. *Plant Cell, Tissue and Organ Culture (PCTOC)* **98**, 25-33 (2009).
58. Marshak, D.R. et al. Strategies for protein purification and characterization: a laboratory course manual. (Cold Spring Harbor Laboratory, 1996).
59. Lee, T.S. et al. BglBrick vectors and datasheets: a synthetic biology platform for gene expression. *Journal of biological engineering* **5**, 1-14 (2011).
60. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. & Pease, L.R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61-68 (1989).
61. Warrens, A.N., Jones, M.D. & Lechler, R.I. Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene* **186**, 29-35 (1997).
62. Jensen, P.R. & Hammer, K. The sequence of spacers between the consensus sequences

modulates the strength of prokaryotic promoters. *Applied and environmental microbiology* **64**, 82-87 (1998).

63. Sambrook, J., Russell, D.W. & Russell, D.W. Molecular cloning: a laboratory manual (3-volume set). (Cold spring harbor laboratory press Cold Spring Harbor, New York:, 2001).
64. Wang, H.H. et al. Genome-scale promoter engineering by coselection MAGE. *Nature methods* **9**, 591-593 (2012).
65. Wenzel, S.C. & Müller, R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Current opinion in biotechnology* **16**, 594-606 (2005).
66. Peti, W. & Page, R. Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein expression and purification* **51**, 1-10 (2007).