Study on the influence of sequence of green fluorescent protein gene on expression efficiency in *Escherichia coli*

(大腸菌内での緑色蛍光タンパク質遺伝子の塩基配列が発現効率に及ぼす影響に関する研究**)**

Xiaoliang Hao

Saitama Institute of Technology

September, 2015

Contents

Abstract

This dissertation describes the influence of sequence of green fluorescent protein (GFP) gene on expression efficiency in *Escherichia coli* (*E. coli*).

Gene expression in *E. coli* has been the most popular means of producing recombinant proteins. *E. coli* is a very useful host that offers efficient culturing efficiency, easy manipulation and low cost media. People have produced a wide variety of different types of proteins in *E. coli*. Although many proteins are useful for the humans, these proteins can only rarely be obtained in sufficiently large quantities from their natural sources. Therefore, such proteins are often produced in *E. coli* via genetic engineering. The expression efficiency of genes in *E. coli* is very important in genetic engineering, and the codon composition of coding sequences plays an important role in the regulation of gene expression. To master the relationship between the gene sequence and gene expression is helpful for understanding principles of gene expression and effectively controlling the production of protein. It is hoped that the results would be helpful for the further study of gene expression. The dissertation composes of the following 5 chapters:

Chapter 1 General introduction

This chapter introduced background and significance of this research. It included the definition, principles of genetic engineering, the definition of *E. coli*, also included the definition, characteristics, application of GFP, gene expression, etc.

Chapter 2 Influence of last sense codon on expression efficiency of GFP gene in *E. coli*

In this chapter, the relationship between changes in the last sense codon (i.e., the codon preceding the stop codon) and the efficiency of gene expression was studied. pKK223-3 was used as the expression vector, 64 kinds of last sense codons were inserted at the 5’ end of the stop codon of the GFP gene and the modified GFP genes

were acquired. GFP-fusion fluorescence intensity was an excellent indicator of over-expression potential. Because fluorescence was one of the most convenient ways to follow a gene expression, so the fluorescence intensity was used to analyze the expression efficiency of GFP gene. Because the fluorescence intensity was easy to be measured and quantified, it was used to compare the expression efficiency of GFP gene. Relative fluorescent intensity (RFI) was normalized with the fluorescence intensity of the wild type GFP, and had been applied in analyzing the expression efficiency. By research, the strongest RFI (2.1) was observed with a CCG last sense codon, while the lowest RFI (0.33) was observed with a CTA last sense codon. The RFIs of the GFP genes with the other 62 kinds of last sense codons were 0.7-1.5.

For the last sense codons which belonged to the same amino acids, the GFP genes with the last sense codons showed the nearly same RFIs. For example, for phenylalanine (coded by TTT and TTC), the RFIs were close, indicating similar in the quantities of GFP produced, but not variation in the amino acid sequence. Similar tendency was seen for the other amino acids. However, for some last sense codons, the differences were clear; for example, the RFIs were different among the six last sense codons for leucine (TTA, TTG, TCT, TCC, TCA, and TCG).

Our experiments revealed that the last sense codon influenced the expression efficiency of the GFP gene in *E. coli*; however, clear rule between the last sense codon and the expression efficiency of the GFP gene had not been found.

Chapter 3 Influence of the combination of last sense codon and stop codon on expression efficiency of GFP gene in *E. coli*

In this chapter, when the expression vector was pKK223-3, 18 kinds of last sense codons were chosen, stop codon was changed from UAA to UAG and UGA, and the influence of the combination of last sense codon and stop codon on expression efficiency of GFP gene in *E. coli* by the expression vector pKK223-3 was researched. Later, the expression vector was changed from pKK223-3 to pET21(+). 12 kinds of last sense codons were inserted into GFP genes, the influences of the combination of last sense codon and stop codon on expression efficiency of GFP gene were compared

by the expression vector pET21(+**)**.

Chapter 4 Research on the relationship between mutant gene and protein synthesis efficiency of GFP in *E. coli*

In our previous work, 20 mutant GFP genes had been acquired. In order to know the relationship between mutant gene and protein synthesis efficiency easily, the mutant GFP gene with a single point mutation was chosen. Next, one special mutant GFP(ACC) was found, whose inserted last sense codon was ACC, had three mutant sites, at 56, 357 and 443 position of the GFP gene. The RFI of *E. coli* having this vector indicated 0.1. Meantime, other four mutant GFP genes with one mutant site were acquired. Finally, the influence of six mutant GFP genes on expression efficiency in *E. coli* was studied, and the relationship between the mutant gene and expression efficiency was discussed. It was found that some mutant genes of GFPs had influenced on the expression efficiency of GFP gene and the mutant A443G showed the lowest RFI (0.04), and mutant A88G showed the lower RFI (0.16), mutant A402G showed the stronger RFI (1.2).

Chapter 5 Conclusions and recommendations

In this chapter, the results in the preceding chapters were summarized, the conclusions were given. The gene sequence had close relationship with the expression efficiency of the GFP gene in *E. coli*. Some GFP genes with the last sense codons had stronger fluorescence intensities and got the more expression quantities of proteins. The expression vector (pKK223-3 and pET21(+)) had some same points, but showed great differences on genetic expression. The point mutation of GFP gene could greatly influence the gene expression.

Finally unresolved problems were posed in this research, such as no clear rules were complied with and the influences of two or more mutant sites which had affected on the expression efficiency were not researched.

Chapter 1 General introduction

1.1 Background and significance

Genetic engineering was introduced around 1970 as a strongly potent strategy for genetic research at the level of DNA molecules, the carriers of genetic information. This strategy consists principally of introducing nucleotide sequence alterations into DNA molecules, such as by site-directed mutagenesis and by splicing DNA segments from different locations in the genome or from different kinds of organisms (recombinant DNA molecules) [1].

Since the 1990s, genetic modification has been the focus of debate, owing to its increased use and effect on our lives. There is evidence that genetic modification, including genetic engineering and biotechnology, has appeared in a number of areas, ranging from agriculture, chemical engineering, industry and the food industry, to medicine, molecular biology, environmental protection and human health. However, this increasing appearance of genetic modification in various areas has resulted in growing interest, concerns, ethical and social issues not only in scientific communities but in public ones as well; so, genetic modification is recognized as a socioscientific issue [2].

Protein is an essential nutrient which helps forming structural component of body tissues and is used within many biological processes, for example protein is used to make enzymes, antibodies to help us fight infection as well as DNA the building blocks to life. It’s also needed to make up muscle tissue which in turn helps to keep our bodies active, strong, and healthy. Biomedical and biotechnological research relies on processes leading to the successful expression and production of key biological products. Strong-quality proteins are required for many purposes, including protein structural and functional studies. Gene expression is the culmination of multistep processes involving regulation at the level of transcription, mRNA turnover, translation, and post-translational modifications leading to the formation of a stable product [3-4].

As the important proteins, the rapid rise of fluorescent proteins in biomedical research has often been termed a revolution. Green fluorescent protein (GFP) was

elegantly used to strong-light sensory neurons in the nematode [5-6]. Since then, the race has been on to produce new and improved versions that are brighter, cover a broad spectral range, and also exhibit enhanced photostability, reduced oligomerization, pH insensitivity, and faster maturation rates. GFP has attracted widespread interest and is considered to have several advantages over other visual marker genes. First, the fluorescence emission of GFP does not require a cofactor or a substrate [7]. The second advantage of GFP is that it is relatively small and can tolerate both N- and C-terminal protein fusions, lending itself to studies of protein localization and intracellular protein trafficking [8-10]. The third advantage of GFP is that GFP mutants with shifted wave-lengths of absorption and emission have been isolated, which permits simultaneous use and detection of multiple reporter genes [11-13].

Gene expression in *Escherichia coli* (*E. coli*) has been the most popular means of producing recombinant proteins. *E. coli* is a very useful host that offers efficient culturing efficiency, easy manipulation and low cost media. People have produced a wide variety of different types of proteins by *E. coli* [14-21]. Although many proteins are useful for the humans, these proteins can only rarely be obtained in sufficiently large quantities from their natural sources. Therefore, such proteins are often produced in *E. coli* via genetic engineering.

The expression efficiency of genes in *E. coli* is very important in genetic engineering, and the codon composition of coding sequences plays an important role in the regulation of gene expression. To master the relationship between the gene sequence and gene expression is helpful for understanding principles of gene expression and effectively controlling the production of protein. The researches about codon, protein, and gene expression have attracted wide attention of people and made great achievements, for example some models had been constructed [22-24], some new methods had been detected and applied to research [25-27].

Advanced recombinant DNA technologies have then allowed the transfer of genes to a suitable host for rapid propagation and or strong level protein production. The development of small size and automatizable chemical or biological assays allow the

screening of large numbers of samples for selection of desired functions. There is a significant association between codon bias and translation efficiency across all endogenous genes in *E. coli* [28]. It has long been known that codon preference is related to gene expression [29-32]. Eighteen of the 20 amino acids are each encoded by two or more synonymous codons in the standard genetic code, yet the synonymous codons are often used unequally in a genome. Such codon usage bias has been extensively documented in all three domains of life [30, 31, 33]. Within a genome, strongly expressed genes tend to have stronger codon usage bias than lowly expressed ones and the codons preferentially used in strongly expressed genes of a species are referred to as preferred codons. Codon-usage bias has been observed in almost all genomes and is thought to result from selection for efficient and accurate translation of strongly expressed genes [34-37]. Codon usage is also implicated in the control of transcription, splicing and RNA structure [38-40].

1.2 Genetic engineering

Genetic engineering also can be the rearrangement of gene location or the removal of genes. The “altered” organism then makes new substances or performs new functions based on its new DNA. The chart of Genetic engineering was shown in Fig. 1.1. For example, the protein insulin, used in the treatment of diabetes, now can be produced in large quantities by genetically engineered bacteria and yeasts. Insulin was formerly extracted from pigs or cows [41].

Genetic engineering has already supplied us with products that alleviate illness, clean up the environment, and increase crop yields, among other practical benefits to humanity and the ecosystem. For example, the first genetically engineered life form to be granted patent protection was developed by Ananda Chakrabarty, who genetically engineered a common bacterium into *Burkholderia cepacia*, a variant that digests petroleum products. He obtained a patent for his new life form, and helped establish the Supreme Court precedent that, to this day, enables inventors to patent genetically engineered life forms [42]. The bacterium cleans up oil spills and has proven to be both safe and useful. Since this precedent, tens of thousands of patents have issued for

genetically engineered life forms. Genetic engineering has also helped create thousands of organisms and processes useful in medicine, research, and manufacturing. Genetically engineered bacteria churn out insulin for treating human diabetes, production of which would be substantially more expensive without the use of genetic engineering [14].

1.3 *Escherichia coli* (*E. coli*)

E. coli is a gram-negative bacillus, which is made up of species present in the human and other animal intestine. When eliminated in the environment together with feces it contaminates water, soil and food. *E. coli* is a convenient host for gene expression and one of the organisms of choice for the production of recombinant proteins, it has become the most popular expression platform. Its advantages include strong levels of heterologous gene expression and scalability of experiments, low cost, fast growth, a lack of posttranslational modification and an ablity to express labeled proteins and is one of the organisms of choice for the production of recombinant proteins. Its use as a cell factory is well-established and it has become the most popular expression platform [16-21, 43].

1.4 Green fluorescent protein (GFP)

1.4.1 Definition of GFP

GFP was discovered by Shimomura *et al* (1962) as a companion protein to aequorin, the famous chemiluminescent protein from *Aequorea victoria*. GFP has been widely used as a reporter in the determination of gene expression and protein localization [44-47]. The green light emitted by *Aequorea victoria* is produced by two proteins that work together: *Aequorin* and green fluorescent protein (GFP) [48]. Fig. 1.2 showed the figure of jellyfish *Aequorea*, Fig. 1.3 showed Top view of a specimen of *Aequorea* when stimulated in a darkroom [49]. In just recent years, GFP has valued from obscurity to become one of the most widely studied and exploited proteins in biochemistry and cell biology [50].

1.4.2 Characteristics of GFP

GFP is composed of 238 amino acids and hence a rather small protein with a molecular weight of roughly 27,000 [51]. Excitation at 396 nm results in an emission maximum at 508 nm [52]. The expression of GFP gene in *E.coli* was shown in Fig. 1.4.

GFP has been used extensively throughout the biological sciences. GFP has been linked to proteins and promoters from archea to zebrafish and mice. It can quantify cells, components of cells, or reactions within subcellular spaces. Reviews detailing GFP applications are prevalent in the literature and are summarized in a comprehensive review of the chemistry and functional properties of GFP [48].

1.4.3 Application of GFP

The use of GFP as a biosensor for genotoxic compounds has also been gaining momentum over the past few years. The most common use of GFP is its remains as a reporter of the protein. GFP imparts stability to its fusion partners and allows for facile estimates of protein locale and quantity. GFP is used to accurately monitor bioprocesses, both at the laboratory scale and on an industrial scale. Poppenborg and coworkers provided one of the first examples of bioprocess monitoring using GFP. GFP fluorescence has been used to investigate a remarkable array of properties and behaviors. The main reasons for this are that the chromophore of GFP is produced through an internal posttranslational autocatalytic cyclization that does not require any cofactors or substrates, fusion of GFP to a protein rarely affects the proteins activity or mobility, and GFP is nontoxic in most cases [48, 53].

The biggest advantage of GFP is that it is heritable, since it is able to be transformed with the use of DNA encoding GFP. Additionally, visualizing GFP is noninvasive; so just by shining light on the protein, it is able to detect it. Furthermore, GFP is a relatively small and inert molecule, that doesn't seem to interfere with any biological processes of interest. Moreover if used with a monomer it is able to diffuse readily throughout cells [54]. This research used above advantages of GFP to study the influence of gene sequence on expression efficiency of the gene.

References

1. Arber, W. Genetic engineering compared to natural genetic variations. *New Biotechnology*., **27**, 517-521 (2010).

2. Sadler, T.D., Zeidler, D. L. Patterns of informal reasoning in the context of socioscientific decision making. *Journal of Research in Science Teaching*., **42**, 112-138 (2005).

3. Ordu, E., Karaguler, N. G. *Protein Engineering Applications on Industrially Important Enzymes: Candida methylica FDH as a Case Study*. Protein Engineering. Chapter 4, 75-98 (2015).

4. Angov, E. Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnol. J*., **6**, 650-659 (2011).

5. Shaner, NC., Patterson, GH., Davidson, MW. Advances in fluorescent protein technology. *Journal of Cell Science*., **120 (24)**, 4247-4260 (2007).

6. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science*., **263**, 802-805 (1994).

7. Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G., Ward, W. W. Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein. Ward. *Biochemistry*., **32 (5)**, 1212-1218 (1993).

8. Wang, S., Hazelrigg, T. Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis. *Nature*., **369(6479)**, 400-403 (1994).

9. Davis, I., Girdham, CH., O'Farrell, PH. A nuclear GFP that marks nuclei in living Drosophila embryos; maternal supply overcomes a delay in the appearance of zygotic fluorescence. *Dev Biol*., **170(2)**, 726-729 (1995).

10. Kaether, C., Gerdes, H. H. Visualization of protein transport along the secretory pathway using green fluorescent protein. *FEBS Lett*., **369(2-3)**, 267-271 (1995).

11. Heim, R., Prasher, D. C., Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci*., **91(26)**, 12501-12504 (1994).

12. Heim, R., Cubitt, AB., Tsien, RY. Improved green fluorescence. *Nature*., **373**, 663-664 (1995).

13. Delagrave, S., Hawtin, RE., Silva, CM., Yang, MM. Red-shifted excitation mutants of the green fluorescent protein. *Bio Technology*., **13**, 151-154 (1995).

14. Tolia, N. H., Leemor, J. T. Strategies for protein coexpression in *Escherichia coli*. *Nature methods*., **3**, 55-64 (2006).

15. Rosano, G. L., Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*, advances and challenges. *Frontiers in microbiology*., **5**, 1-17 (2014).

16. Hannig, G., Makrides, SC. Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Tendency in Biotech*., **16**, 54-60 (1998).

17. Baneyx, F. Recombinant protein expression in *Escherichia coli*. *Curr Opin Biotechnol*., **10**, 411-421 (1999).

18. Panda, AK. Bioprocessing of therapeutic proteins from the inclusion bodies of *Escherichia coli*. *Adv Biochem Eng Biotechnol*., **85**, 43-93 (2003).

19. Palomares, LA., Estrada-Mondaca, S., Ramirez, OT. Production of recombinant proteins, challenges and solutions: *Methods Mol Biol*., **2**, 15-52 (2004).

20. Mergulhao, FJ., Summers, DK., Monteiro, GA. Recombinant protein secretion in *Escherichia coli*. *Biotechnol Adv*., **23(3)**, 177-202 (2005).

21. Yin, J., Li, G., Ren, X., Herrler, G. Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *J Bacteriol*., **127**, 335-347 (2007).

22. Gilchrist, M. A., Wagner, A. A model of protein translation including codon bias, nonsense errors, and ribosome recycling. *Journal of Theoretical Biology*., **239**, 417-434 (2006).

23. Kosiol, C., Holmes, I., Goldman, N. An Empirical Codon Model for Protein Sequence Evolution. *Mol. Biol. Evol*., **24(7)**, 1464-1479 (2007).

24. Jukes, T. H., Cantor, C. R. Evolution of protein molecules: Mammalian protein metabolism. *Academic Press. New York*., **3**, 21-132 (1969).

25. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol*., **16**,

111-120 (1980).

26. Bielawski, J. P., Yang, Z. H. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *Journal of Molecular Evolution*., **59**, 121-132 (2004).

27. Suzek, B. E., Ermolaeva, M. D., Schreiber, M., Salzberg, S. L. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics*., **17**, 1123-1130 (2001).

28. Tuller, T., Waldman, Y. Y., Kupie, M., Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *Current Issue*., **107**, 3645-3650 (2009).

29. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., Mercier, R. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res*., **9**, 43-74 (1981).

30. Ikemura, T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*., **2**, 13-34 (1985).

31. Sharp, PM., Cowe, E., Higgins, DG., Shields, DC., Wolfe, KH., Wright, F. Codon usage patterns in *Escherichia coli*, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogasterand Homo sapiens; a review of the considerable within species diversity. *Nucleic Acids Res*., **16**, 8207-8211 (1988).

32. Médigue, C., Rouxel, T., Vigier, P., Hénaut, A., Danchin, A. Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J Mol Biol*., **222**, 851-856 (1991).

33. Hershberg, R., Petrov, DA. General rules for optimal codon choice. *Plos Genet*., **5(7)**, 1-10 (2009).

34. Ikemura, T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol*., **151**, 389-409 (1981).

35. Goodman, D. B., Church, G. M., Kosuri, S. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science*., **342 (6157)**, 475-479 (2013).

36. Plotkin, J. B., Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*., **12**, 32-42 (2011).

37. Drummond, D. A., Wilke, C. O. Mistranslation induced protein missing as a dominant constraint on coding sequence evolution. *Cell*., **134**, 341-352 (2008).

38. Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., Pilpel, Y. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*., **141,** 344-354 (2010).

39. Gu, W., Zhou, T., Wilke, C. O. A universal tendency of reduced mRNA stability near the translation initiation site in prokaryotes and eukaryotes. *Plos Comput Biol*., **6(2)**, 1-8 (2010).

40. Cannarozzi, G., Schraudolph, NN., Faty, M., Von, R. P., Friberg, MT., Roth, AC., Gonnet, P., Gonnet, G., Barral, Y. A role for codon order in translation dynamics. *Cell*., **141**, 355-367 (2010).

41. Pecarski, D., Jugovic, Z., Stevanovic, A. Genetic engineering. 15th International Research / Expert Conference. "Tendency in the Development of Machinery and Associated Technology"TMT 2011, Prague, Czech Republic, 12-18 (2011).

42. Singleton, P. Bacteria in Biology, Biotechnology and Medicine (5th ed.). *Wiley*., 444-454 (1999).

43. Koepsell, D. The ethics of genetics of genetic engineering. *A position paper from the center for inquiry office of public policy*., 1-22 (2007).

44. Shimomura, O., Johnson, F. H., Saiga, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea. J. Cell Comp. Physiol*., **59**, 223-239 (1962).

45. Johnson, FH., Shimomura, O., Saiga, Y., Gershman, LC., Reynolds, GT., Waters, JR. Quantum efficiency of *Cypridina* luminescence, with a note on that of *Aequorea*. *J. Cell. Comp. Physiol.*, **60**, 85-103 (1962).

46. Morin, JG., Hastings, JW. Energy transfer in a bioluminescent system. *J. Cell. Physiol.*, **77**, 313-318 (1971).

47. Morise, H., Shimomura, O., Johnson, FH., Winant, J. Intermolecular energy transfer in the bioluminescent system of Aequorea. *Biochemistry.*, **13**, 2656-62 (1974).

48. Zimmer, M. Green Fluorescent Protein (GFP), Applications, Structure, and Related Photophysical Behavior. *Chem. Rev*., **102**, 759-781 (2002).

49. Atsushi, M. GFP and bioimaging-From the basic detection and expression of fluorescent protein to visualize the biological function (experimental course of post-genome era). *Yodosha press*., 16 (2000).

50. Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.*, **67**, 509-544 (1998).

51. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., Cormier, M. J. Primary structure of the Aequorea victoria green fluorescent protein. *Gene (Amst.)*., **111**, 229-233 (1992).

52. March, J. C., Rao, G., Bentley, W. E. Biotechnological applications of green fluorescent protein. *Appl Microbiol Biotechnol*., **62**, 303-315 (2003).

53. Poppenborg, L., Friehs, K., Flaschel, E. The green fluorescent protein is a versatile reporter for bioprocess monitoring. *J Biotechnol*., **58**, 79-88 (1997).

54. Chalfie, M. GFP: Lighting up life. *Proceedings of the National Academy of Sciences of the United States of America*., **106** (25), 10073-10080 (2009).

Figures

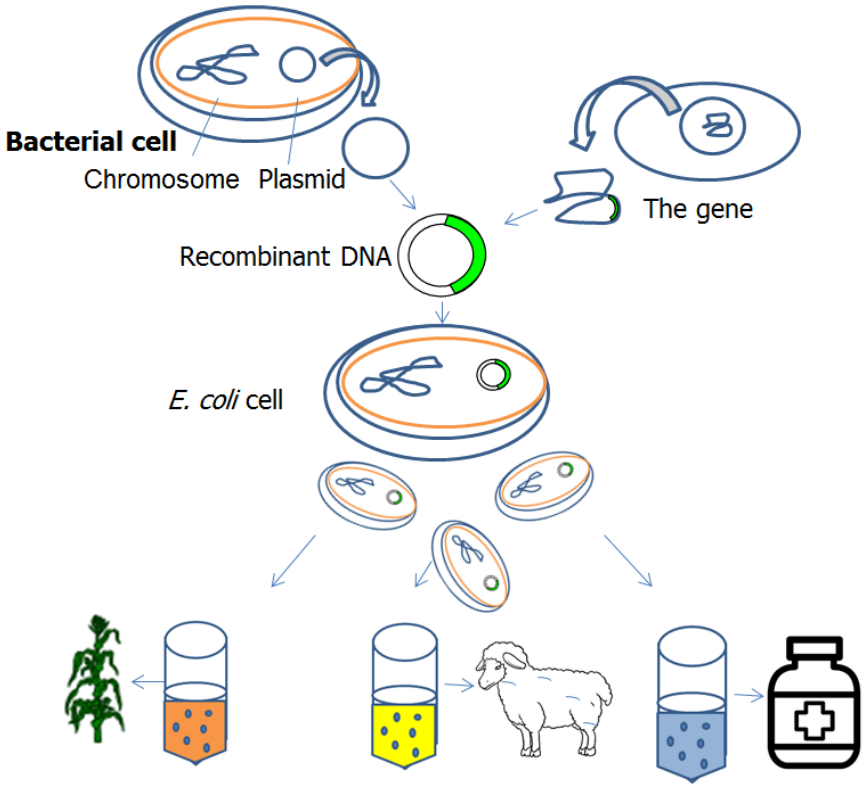


Figure 1.1 The chart of genetic engineering

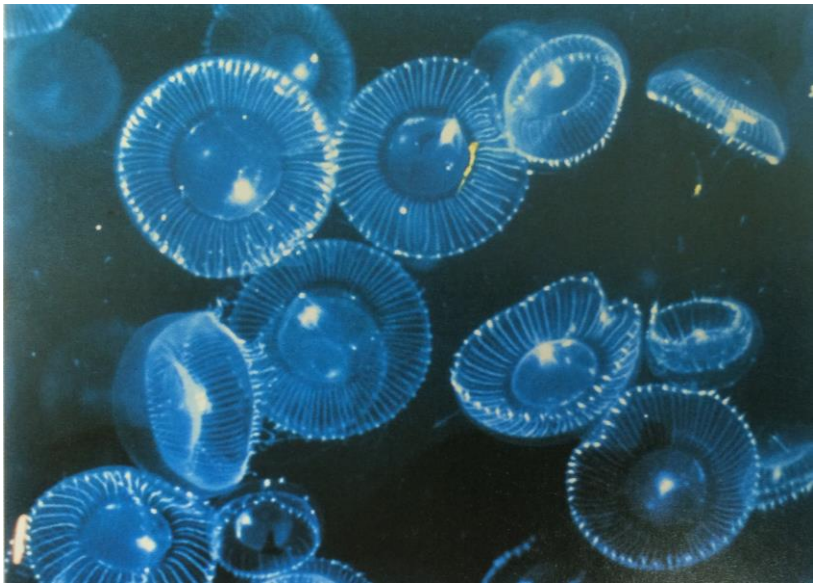


Figure 1.2 The jellyfish *Aequorea* (ref. 49)

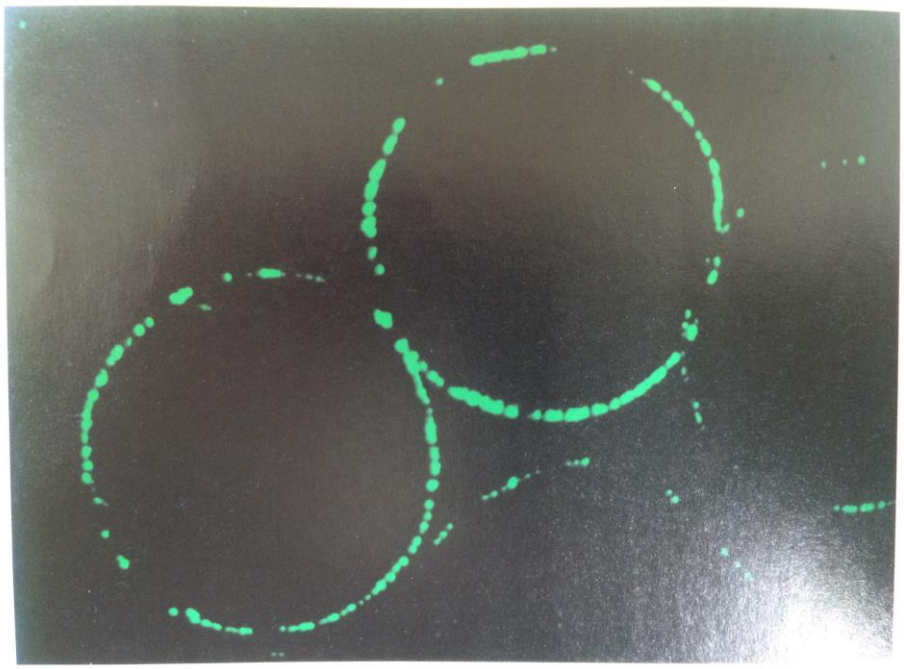


Figure 1.3 Top view of a specimen of *Aequorea* when stimulated in a darkroom (ref. 49)

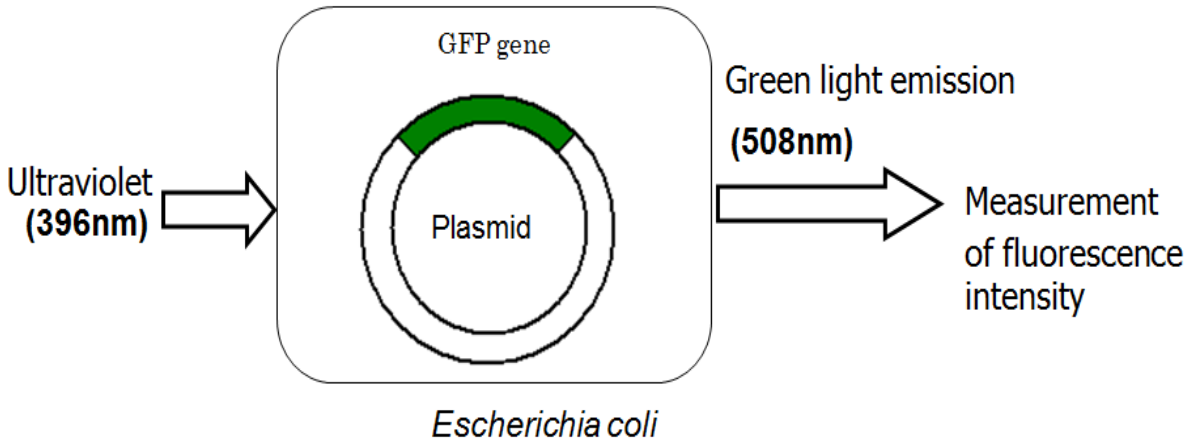


Figure 1.4 The gene expression and fluoresence of GFP in *E. coli*

Chapter 2 Influence of last sense codon on expression efficiency of GFP gene in *E. coli*

2.1 Introduction

Protein synthesis is a complex, continuous process comprising the main stages of initiation, elongation, and termination. The initiation phase is believed mainly to control translation in bacterial cells [1]. A well-characterized element that influences the initiation step is the Shine-Dalgarno (SD) sequence, purine-rich nucleotides in the upstream of an open reading frame [2]. The SD sequence interacts with the complementary sequence in the 16S rRNA by base-pairing, which promotes the binding of mRNA to the 30S ribosomal subunit to form a pre-initiation complex. The importance of the base pairing between the SD sequence and 16S rRNA during the translational initiation step has been established by a series of biochemical experiments [2-3]. It is very important how to overexpress a gene in genetic engineering. It is known that the sequences around the initiation codon, such as SD sequence [2] or second codons, influence on the translation efficiency in *E. coli*. About the second codon, it was seen that the two most frequent second codons, AAA and AAU, enhanced the translation efficiency compared with the wild type, whereas the effects of lower frequency codons were not significant [4]. From the researches above, it is known that the sequences around the initiation codon influence gene expression in *E. coli*. However, the influence of the sequences around the stop codon has not been clarified. So in our lab, 64 kinds of last sense codons (i.e., the codon preceding the stop codon) at the 5' end of the stop codon had been inserted into the GFP gene (Fig. 2.1). GFP-fusion fluorescence intensity was an excellent indicator of overexpression potential [5]. Because fluorescence was one of the most convenient way to follow a gene expression [6], so the fluorescence intensity was used to analyze the expression efficiency of proteins. Because the fluorescence intensity was easy to be measured and quantified, we used relative fluorescence intensity (RFI) to compare the expression efficiency [7-8]. RFI was normalized with that of the wild type GFP gene, and had

been applied in analyzing the expression efficiency. In our previous work, it was found that the mutant GFP genes with the last sense codon ACC and AGC had the lowest RFIs (0.1 below). So the test experiments were done by sequence, it was found that there were mutant nucleotides on the GFP chains. After 64 kinds of GFP genes with the last sense codons were tested, 20 kinds of mutant GFP genes were found. Next, 20 kinds of last sense codons were inserted into the GFP genes separately by polymerase chain reaction (PCR), constructed the expression vector, and researched the relationship between the last sense codon and the expression efficiency. This text aimed to research the influence between gene sequence and expression efficiency of protein in *E. coli* from the perspective of genetic engineering, probe how to gain more and more useful proteins to serve people.

2.2 Experimental

2.2.1 Bacterial strains and plasmids

E. coli JM109 was used as the host strain, and the pKK223-3 was used as the expression vector. Primers for PCR were purchased from Sigma-Aldrich Japan (Tokyo, Japan). The reagents and equipments were written in appendix.

2.2.2 Insertion of last sense codon to GFP gene

In our previous work, 20 kinds of mutant GFP genes were found (Tab. 2.1). So in this chapter, 20 kinds of primers were inserted into GFP genes by PCR and the modified GFP genes were acquired (Tab. 2.2). The part with the underline was the stop codon (UAA), italic letters were the 20 kinds of last sense codons, the shaded parts indicated digested sites of restriction enzyme corresponding to the primer description. For example, when italic letter was “TAA”, according to the rule of complementarity principle, from 3’ side of DNA gene, it was understood that the last sense codon “TTA” was inserted.

Fig. 2.2 was the flow diagram of experiment in this chapter. The last sense codon had been inserted into the GFP gene by PCR. Later the PCR product was digested by

the restriction enzymes (*Eco*RI and *Hin*dIII). The plasmid pKK223-3 had been digested by the restriction enzymes (*Eco*RI and *Hin*dIII). The GFP gene and the pKK223-3 were ligated by T4 DNA ligase and the modified GFP gene which the last sense codon had been inserted was acquired. After *E. coli* was transformed by the expression vector containing the modified GFP gene, the colony was acquired. The expression vectors were isolated from *E. coli*, and the sequence of the modified GFP gene was determined.

The principle of PCR was shown in Fig. 2.3. There are three major steps involved in the PCR technique: denaturation, annealing, and extension. In step one; the DNA is denatured at 94 °C. In step two, primers anneal to the DNA template strands to prime extension at 48 °C. In step three, extension occurs at the end of the annealed primers to create a complementary copy strand of DNA by *Tag* polymerase at 68 °C. This process results in the duplication and modification of the original DNA. The beauty of the PCR cycle and process is that it is very fast compared to other techniques and each cycle doubles the number of copies of the desired DNA strand. After 25 cycles, plenty of copies of DNA were acquired [9].

Later the PCR products were transferred into 1.5 mL microtube, 100 µL of PCI was added into the solution, centrifuged (4 °C, 12000 rpm, 7 min), the lower solution was discarded and the supernatant was transferred into the new microtube, 100 µL of CIA was added into the solution, centrifuged (4 °C, 12000 rpm, 1 min), the lower solution was discarded and the supernatant was transferred into the new microtube, centrifuged (4 °C, 12000 rpm, 1 min) again, the lower solution was discarded with the pipette, 10 µL of 3 M Sodium acetate (pH 5.2) and 250 µL of 99.5 % ethanol were added into the solution, cooled at - 85 °C for 60 min, centrifuged (4 °C, 15000 rpm, 20 min), the supernatant was discarded and 100 µL of 70 % of ethanol was added into the below precipitation, centrifuged (4 °C, 15000 rpm, 10 min), the supernatant was discarded and the precipitation was dried under reduced pressure for 10 min, 50 µL of TE buffer was added, and kept at 4 °C.

2.2.3 Digestion with the restriction enzymes *Eco*RI and *Hin*dIII

41 µL of sterile water, 10 µL of 10 X H Buffer, 2 µL of *Eco*RI were added into the 1.5 mL microtube with the PCR product. The microtube was incubated at 37 °C for 60 min. The following procedure was the same with that of aftertreatment of PCR procedure.

After the solution was digested with *Eco*RI, the 25 µL of pKK223-3 / *Eco*RI solution was digested with *Hin*dIII by adding of 63 µL of sterile water, 10 µL of 10 X M Buffer and 2 µL of *Hin*dIII solution. The solution was incubated at 37 °C for 60 min. The following procedures were the same with that of aftertreatment of PCR procedure.

2.2.4 Dephosphorylation with calf intestine alkaline phosphatase (CIAP)

In order to make the phosphate group attached to the 5' end of the plasmid remains connecting when used the restriction enzyme, the probability of self-ligated with ligation reaction is strong. Calf intestine alkaline phosphatase (CIAP) was used to remove the phosphate group at the 5' end by a CIAP treatment to prevent self-ligation.

The composition of the CIAP reaction contained 25 µL of the solution after digestion by *Eco*RI and *Hin*dIII, 19 µL of sterile water, 5 µL of 10 X Alkaline Phosphatase Buffer and 1 µL of CIAP. The mixture was incubated at 37 °C for 30 min. The following procedures were the same with that of aftertreatment of PCR procedure.

2.2.5 Purification of pKK223-3 digested with *Eco*RI and *Hin*dIII, and PCR products of GFP gene digested with *Eco*RI and *Hin*dIII

Wizard SV Gel and PCR Clean-Up System was used to purify pKK223-3 digested with *Eco*RI and *Hin*dIII, and PCR product of GFP gene digested with *Eco*RI and *Hin*dIII, which allowed rapid and efficient purification of DNA.

2.2.6 Ligation and transformation

Ligation is the process of joining two pieces of linear DNA into a single piece through the use of an enzyme called DNA ligase. The rate (mol) between the plasmid (expression vector) and PCR-product is 1:4. Plasmid vector (200 fmol), PCR-Product

(50 fmol) and ligation Mix were added into the 0.5 mL microtube, and incubated at 16 °C for overnight.

200 μL of competent cell of *E. coli* JM109 was placed on the ice for 30 min before the transformation precedure. 10 μL of ligated DNA was added into the 0.5 ml microtube of competent cell, slowly stirred with the pipette, and the microtube was placed on the ice for 30 min again. Later, incubated at 42 °C for 1 min, the microtube was placed on the ice for 2 min. 800 μL of SOC was added into the microtube, shaked in the incubator (37 °C, 200 rpm, 60 min). After incubator, the microtube was centrifuged (4 °C, 6000 rpm, 3 min), 900 μL of supernatant was discarded, the residual liquid was suspended, the coloney was incubated at 37 °C for 18 h, and kept at 4 °C.

2.2.7 Sequence of GFP gene in the expression vector pKK223-3

In order to analyze the sequence of inserted last sense codon, the sequence experiment was carried out using Genetic Analyzer PRISM 310 (Thermo Fischer Scientific). The composition of the reaction solution for sequence contained the sterile water, 5 X sequence buffer, primers and plasmid DNA. The concentration of the plasmid DNA was adjusted to be between 100 and 200 ng. The HS premix of Big Dye Terminator v1.1 Cycle Sequencing Kit was added into the solution. After the sequencing reaction, 16 μL of sterile water and 64 μL of 95 % ethanol were added into the solution, standed for 15 min, the supernatant was thrown away after the centrifugation (20 min, 16000 rpm, 25 °C) and 250 μL of 70 % ethanol was added into the solution and centrifuged (10 min, 16000 rpm, 25 °C) once again, next the precipitation was dried at 90 °C for 1 min, later 25 μL of Hi-Di formamide was added into each sample tube and heated at 95 °C for 2 min, finally the sample tube was kept on the ice for 3 min, prepared for the sequence equipment to test the gene arrangement.

2.2.8 Measurement of fluorescence intensity of GFP in *E. coli*

After all the 64 kinds of GFP genes with different the last sense codons were acquired, the distinctions of fluorescence intensities among different last sense codons were compared. So the cells were cultivated in LB medium supplemented with 0.1 mg

/ mL ampicillin (amp) and 40 μM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37 °C for 18 h. The absorbance at 600 nm of the *E. coli* culture diluted with LB medium was measured, and the fluorescence intensity at 508 nm excited at 396 nm of the same culture by the fluorescence microplate reader (Nihon Molecular Devise) was analyzed. The expression efficiency of GFP gene was compared as the value of the fluorescence intensity / the absorbance at 600 nm. Fluorescence spectroscopy could be used to measure the concentration of a compound because the fluorescence intensity was increased in proportion to the absorbance at 600 nm, so the fluorescence intensity could be used to compare the expression efficiency of GFP genes with different last sense codons (Fig. 2.4). The calculated method of RFI was that the value of fluorescence intensity of wild type was divided by the value of fluorescence intensity of modified GFP gene.

$$\mathrm{RFI} = \frac{\text{Value of fluorescence intensity of modified GFP gene}}{\text{Value of fluorescence intensity of wild type}}$$

2.2.9 SDS-PAGE (SDS-Polyacrylamide Gel Electrophoresis) of soluble proteins extracted from *E. coli*

After the fluorescence intensities of acquired modified GFP genes were known, we wanted to compare the expression quantities of proteins, and analyzed the influences of the modified GFP genes on expression efficiency, therefore sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) experiment was done. First, 20 mL of LB medium, 20 μL of 100 mg / mL amp and 50 μL of inoculated *E. coli* on the colony from glycerol stock were added into each steriled 100 mL erlenmeyer flask. The erlenmeyer flasks were cultured in the culture-shaker (37 °C, 200 rpm). After three hours, 40 μL of IPTG (20 mM) was added into the erlenmeyer flasks. After the solutions were cultured for 15 hours, the solutions were transferred into 50 mL of falcon tube from erlenmeyer flasks, centrifuged (4 °C, 6000 rpm, 10 min). Next, the supernatant was discarded, the falcon tube was freezed at - 85 °C for 3 h. 700 μL of TE

was added into the falcon tube to suspend the precipitate. The weight of new 1.5 mL microtube was measured and the solution was transferred into the new microtube. After be centrifuged (4 °C, 5000 rpm, 10 min), the supernatant was discarded, at this time the weight of 1.5 mL microtube was measured again, the weight of the precipitate of *E. coli* was calculated, and the volumn of TE was calculated to 250 μL against 0.1 mg of *E. coli*. According to the proportion, TE, and Lisozyme were added into the 1.5 mL microtube. The microtube was placed at the room temperature for 60 min. 4 % Sodium Deoxycholate and DNase I were added into the microtube, the microtube was incubated at 37 °C for 60 min, centrifuged (4 °C, 13000 rpm, 10 min), and the soluble fraction was analyzed by SDS-PAGE.

2.3 Results and discussions

2.3.1 Influence of the last sense codons on expression efficiency

Fig. 2.5 was one part of the GFP gene sequence's figure, three bases which formed complementary base pairs, the part with the underline was the last sense codon, the part with the square was the stop codon. According to the rule of complementarity principle of DNA (Fig. 2.6), nucleotide "A" combined with nucleotide "T", nucleotide "C" combined with nucleotide "G", from 5' side of DNA, it was understood that the the last sense codon TCG had been inserted into the DNA chain, no mutation had been occurred in the base sequence.

As shown in Fig. 2.7, the RFIs of 64 GFP genes were compared with the last sense codons. It was shown that the strongest RFI (2.1) was observed with the insertion of a CCG last sense codon, while the lowest RFI (0.33) was observed with that of a CTA last sense codon. The RFIs of GFP(CCG), GFP(GGA), GFP(TAA), GFP(TTT), GFP(TTG), GFP(CTA) were also observed 2.1, 1.49, 0.94, 1.01, 0.64, 0.33, respectively. The RFIs of most of GFP genes with the insertion of last sense codons were 0.7-1.5.

As shown in Tab. 2.3, for the last sense codons which belonged to the same amino acids, the GFP genes with these last sense codons had the similar RFIs. For example, for phenylalanine (coded by TTT and TTC), the RFIs of GFP(TTT) and GFP(TTC)

were similar, indicating differences in the quantities of GFP produced, but not variation in the amino acid sequence. Similar tendency was seen for the other amino acids. However, for most of the amino acids, the differences were clear; for example, the RFIs were different among the six last sense codons for leucine (TTA, TTG, TCT, TCC, TCA, and TCG).

Codon usage frequency (CUF) has always been considered in the research and production of gene expression. When our research results were compared with the usage frequency of the codon commodities, it was found that some rules were the same, for example, as shown in Tab. 2.4, GFP(CCG) had the strongest RFI, and the CUF was bigger; GFP(CTA) had the lowest RFI, and the CUF was smaller. However, the CUF and RFI had not always the same tendency. For example, GFP(CTG) had the lower RFI, but the CUF was bigger, similarly, GFP(GGA) had the stronger RFI, but the CUF was smaller. It was seemed that there were not clear rules between RFI and CUF.

2.3.2 SDS-PAGE of soluble proteins extracted from *E. coli*

To confirm that the RFI was related to the expression efficiency of the GFP gene, SDS-PAGE of soluble proteins extracted from *E. coli* was done. In our experiments, GFP(CCG) (2.1), GFP(GGA) (1.49), GFP(TAA) (0.94), GFP(TTT) (1.01), GFP(TTG) (0.64), GFP(CTA) (0.33) were chosen and the expression quantities were compared according to the thickness of band around 27 kDa. As shown in Fig. 2.7, GFP(CCG), GFP(GGA) had the strongest RFIs, and the bands around 27 kDa showed the thickest; GFP(TTG), GFP(CTA) had the lowest RFIs, and the bands around 27 kDa showed the thinnest; GFP(TAA), GFP(TTT) had the similar RFIs with the wild type, the bands around 27 kDa showed the medium thicknesses. From the photograph of the extracts of *E. coli* having the same last sense codon as (a) on transilluminator, it was shown that GFP(CCG), GFP(GGA) had the strongest fluorescences, GFP(TTG), GFP(CTA) had the lowest fluorescences, the results of SDS-PAGE and transilluminator had the same tendency with the ones of RFIs (Fig. 2.8).

It was confirmed that the RFI was associated with the quantity of synthesized

protein by SDS-PAGE. Since GFP(CCG) had the strongest quantity of synthesized protein, it could be used as the last sense codon in genetic engineering to produce useful proteins. Our findings would be useful for further studies of gene expression and may be applicable to genetic engineering.

2.4 Conclusions

In this chapter, the influence of last sense codon on expression efficiency of GFP gene in *E. coli* was observed. Insertion of some of last sense codons could increase the expression efficiency of GFP gene, this fact would be useful for the gene expression. When we compared the expression efficiency of GFP gene with the CUF of the inserted last sense codon, our results were not explained only by the CUF. The result of SDS-PAGE confirmed that the fluorescence intensity had the same tendency with the expression quantity of GFP.

References

1. Kozak, M. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev*., **47**, 1-45 (1983).
2. Shine, J., Dalgarno, L. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA*., **71**, 1342-1346 (1974).
3. Steitz, J. A., Jakes, K. How ribosomes select initiator regions in mRNA: Base pair formation between the 3'-teminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli. Proc. Natl. Acad. Sci. USA*., **72**, 4734-4738 (1975).
4. Sato, T., Terabe, M., Watanabe, H., Gojobori, T., Hori-Takemoto, C., Miura, K. I. Codon and Bases after the Initiation Codon of the Open Reading Frames in the *Escherichia coli* Genome and Their Influence on the Translation Efficiency. *J. Biochem*., **129**, 851-860 (2001).
5. Hammon, J., Palanivelu, DV., Chen, J., Patel, C., Minor, DL. A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms. *Protein Sci*., **18**, 121-133 (2009).
6. Waldo, GS., Standish, BM., Berendzen, J., Terwilliger, TC. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol*., **17**, 691-695 (1999).
7. Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H., Sheen, J. Engineered GFP as a vital reporter in plants. *Curr. Biol.*, **3**, 325-330 (1996).
8. Yang, F., Moss, G., Maheswaran, G. The molecular structure of green fluorescent protein. *Nat. Biotechnol*., **14**, 1246-1251 (1996).
9. Joshi, M., Deshpande, J. D. Polymerase chain reaction: methods, principles and application. *International Journal of Biomedical Research*., **1(5)**, 81-97 (2010).

Figures and tables

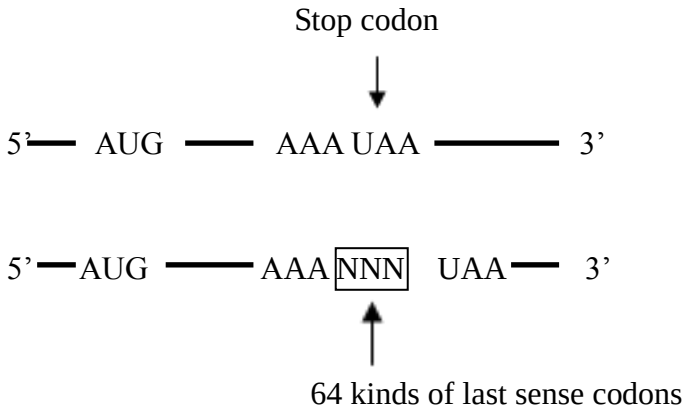


Figure 2.1 The insertion of 64 kinds of last sense codons

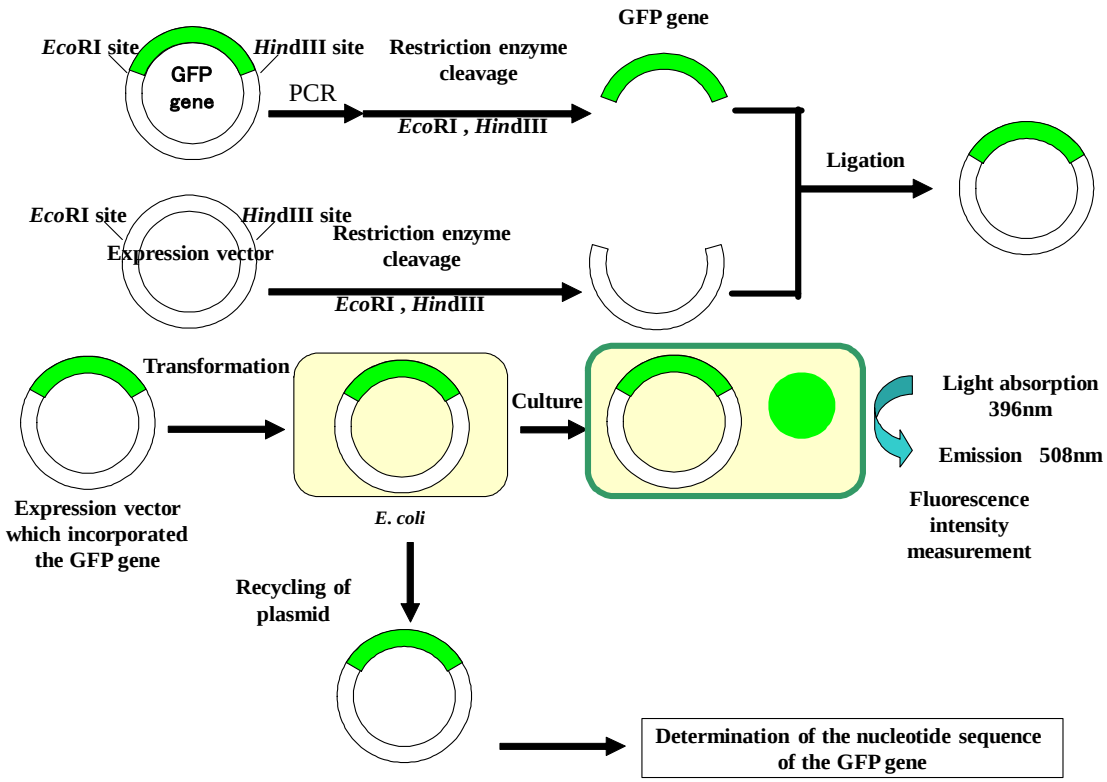


Figure 2.2 Flow diagram of experiment

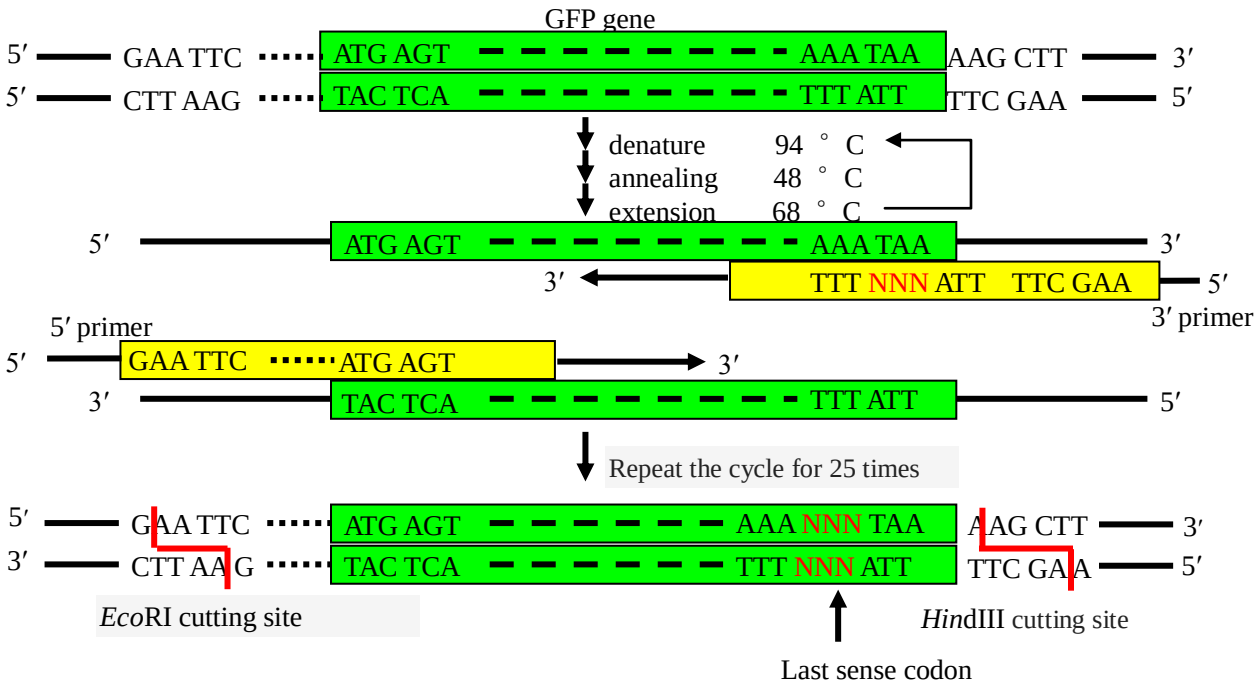


Figure 2.3 The principle of PCR

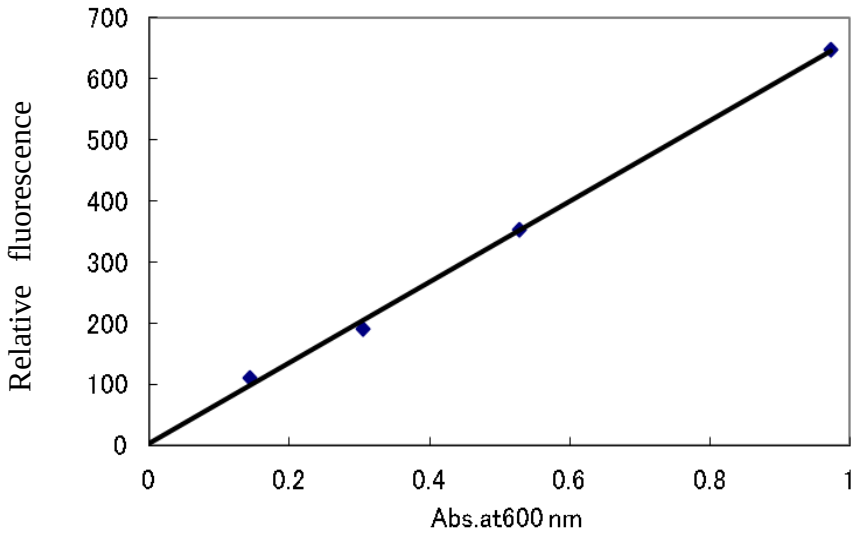


Figure 2.4 GFP fluorescence was proportional to GFP mRNA yield.

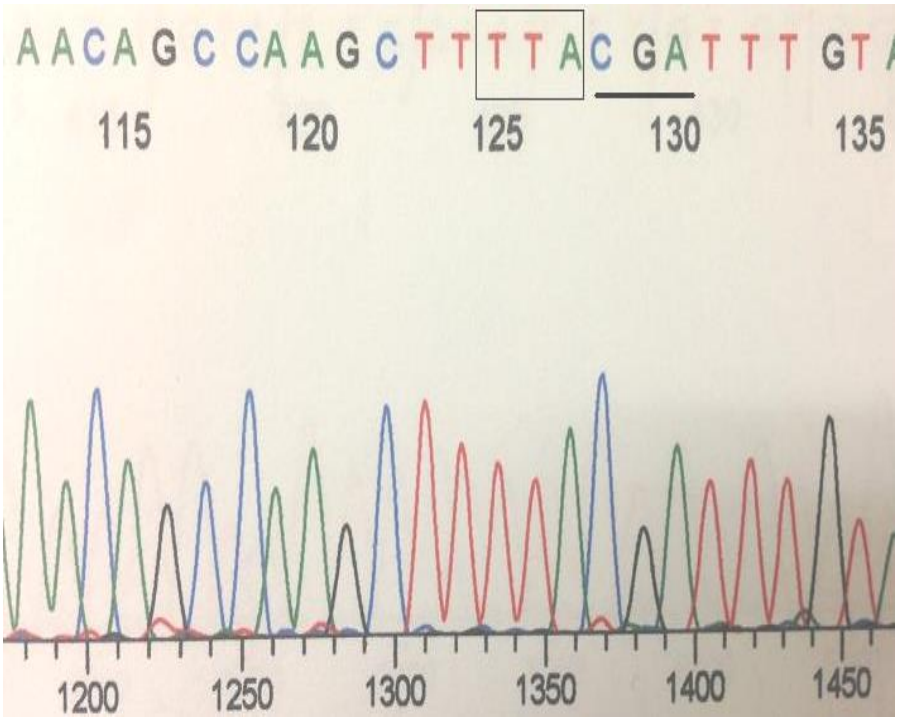


Figure 2.5 One part of sequence analysis

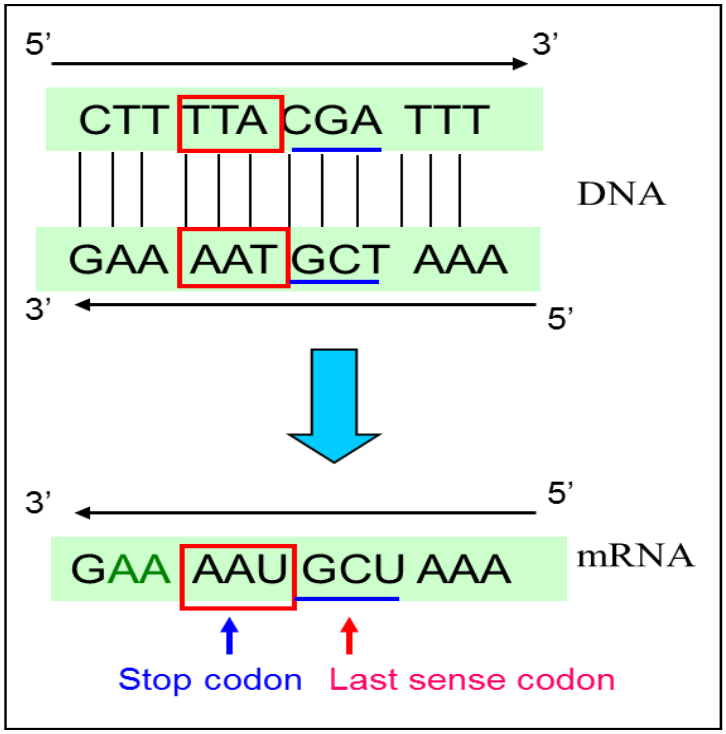
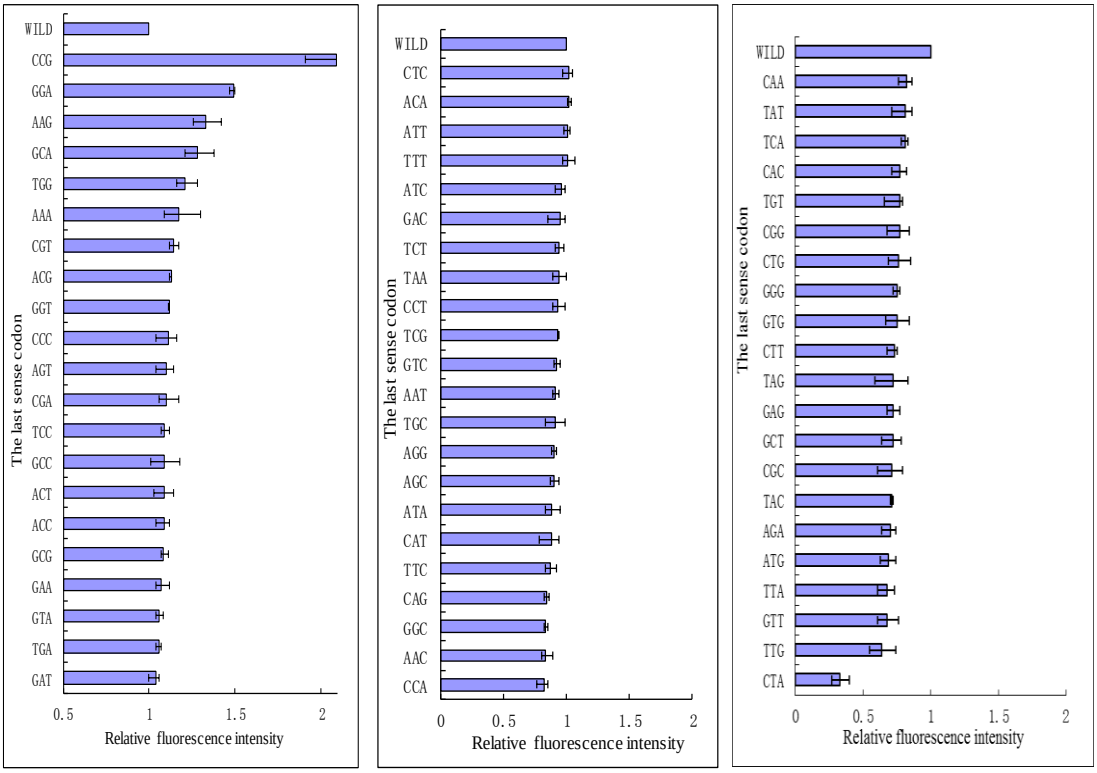


Figure 2.6 Determination of the last sense codon from DNA sequence



(a) (b) (c)

Figure 2.7 RFIs of GFP expressed by genes with insertions of last sense codons in *E. coli*

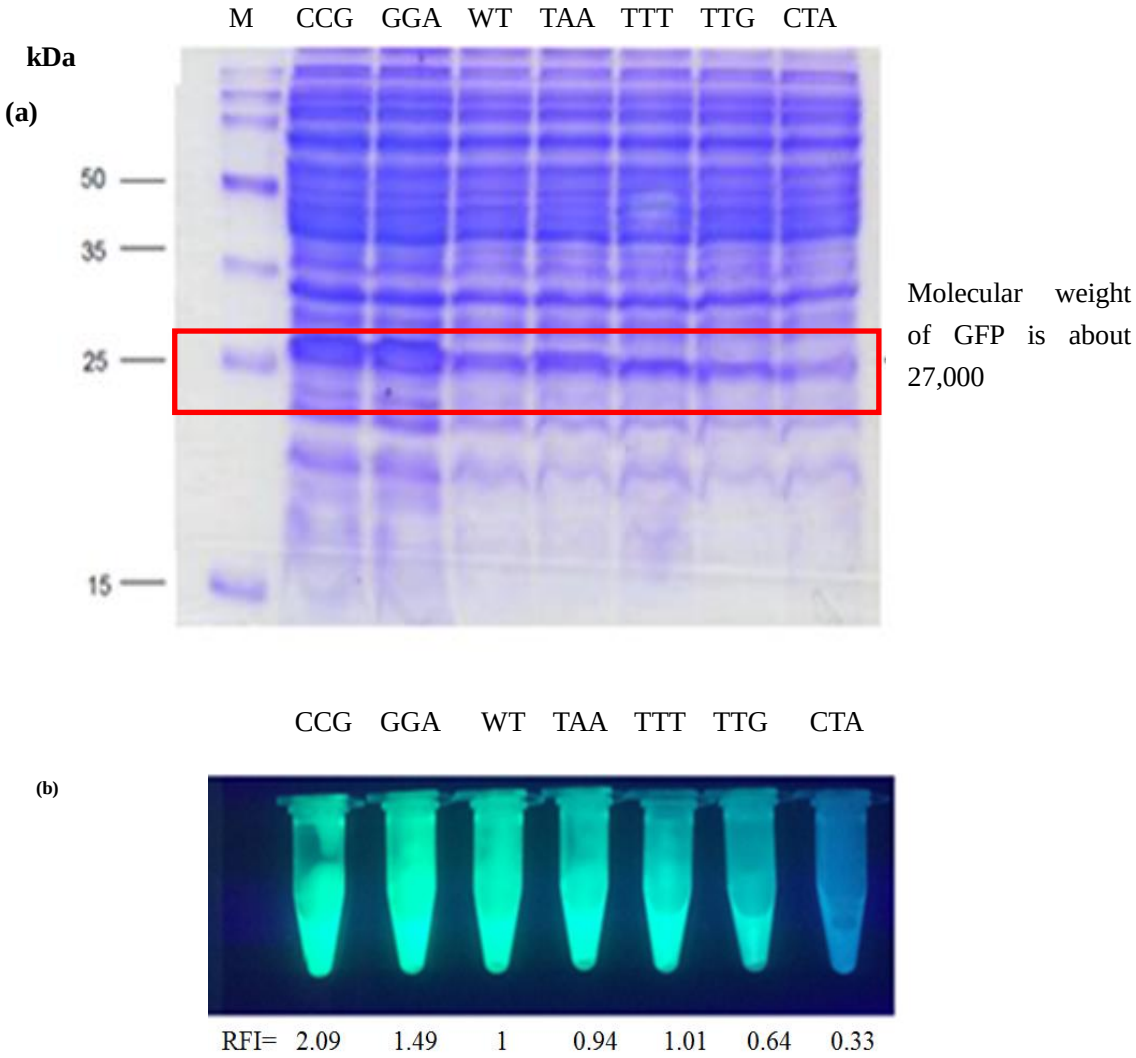


Figure 2.8 (a) Results of SDS-PAGE of the extracts of *E. coli* transformed by the expression vectors with the different last sense codons in the GFP genes. M: Size marker; WT: wild type. Arrow shows the molecular weight of GFP (27,000) (b) Photograph of the extracts of *E. coli* having the same last sense codon as (a) on transilluminator.

Table 2.1 20 kinds of mutant GFP genes

		Second							
		T		C		A		G	
First	T	Phe(F)	TTT	Ser(S)	TCT	Tyr(Y)	TAT	Cys(C)	TGT
			TTC		TCC		TAC		TGC
		Leu(L)	TTA		TCA	STOP	TAA	STOP	TGA
			UUG		TCG		TAG	Trp(W)	TGG
	C	Leu(L)	CTT	Pro(P)	CCT	His(H)	CAT	Arg(R)	CGT
			CTC		CCC		CAC		CGC
			CTA		CCA	Gln(Q)	CAA		CGA
			CTG		CCG		CAG		CGG
	A	Ile(I)	ATT	Thr(T)	ACT	Asn(N)	AAU	Ser(S)	AGT
			ATC		ACC		AAC		AGC
			ATA		ACA	Lys(K)	AAA	Arg(R)	AGA
		Met(M)	ATG		ACG		AAG		AGG
	G	Val(V)	GTT	Ala(A)	GCT	Asp(D)	GAT	Gly(G)	GGT
			GTC		GCC		GAC		GGC
			GTA		GCA	Glu(E)	GAA		GGA
			GTG		GCG		GAG		GGG

Table 2.2 The primers of 20 kinds of last sense codons

Primer	Sequence	Last sense codon
3’	5’－CTCTAAGCTTTTA*TAA*TTTGTATAGTTCATCCATG－3’	TTA
3’	5’－CTCTAAGCTTTTA*AAG*TTTGTATAGTTCATCCATG－3’	CTT
3’	5’－CTCTAAGCTTTTA*TAG*TTTGTATAGTTCATCCATG－3’	CTA
3’	5’－CTCTAAGCTTTTA*TAC*TTTGTATAGTTCATCCATG－3’	GTA
3’	5’－CTCTAAGCTTTTA*CAC*TTTGTATAGTTCATCCATG－3’	GTG
3’	5’－CTCTAAGCTTTTA*AGA*TTTGTATAGTTCATCCATG－3’	TCT
3’	5’－CTCTAAGCTTTTA*CGA*TTTGTATAGTTCATCCATG－3’	TCG
3’	5’－CTCTAAGCTTTTA*AGT*TTTGTATAGTTCATCCATG－3’	ACT
3’	5’－CTCTAAGCTTTTA*GGT*TTTGTATAGTTCATCCATG－3’	ACC
3’	5’－CTCTAAGCTTTTA*TGT*TTTGTATAGTTCATCCATG－3’	ACA
3’	5’－CTCTAAGCTTTTA*CGT*TTTGTATAGTTCATCCATG－3’	ACG
3’	5’－CTCTAAGCTTTTA*GTA*TTTGTATAGTTCATCCATG－3’	TAC
3’	5’－CTCTAAGCTTTTA*TTA*TTTGTATAGTTCATCCATG－3’	TAA
3’	5’－CTCTAAGCTTTTA*TTG*TTTGTATAGTTCATCCATG－3’	CAA
3’	5’－CTCTAAGCTTTTA*CTG*TTTGTATAGTTCATCCATG－3’	CAG
3’	5’－CTCTAAGCTTTTA*GTT*TTTGTATAGTTCATCCATG－3’	AAC
3’	5’－CTCTAAGCTTTTA*CTT*TTTGTATAGTTCATCCATG－3’	AAG
3’	5’－CTCTAAGCTTTTA*TTC*TTTGTATAGTTCATCCATG－3’	GAA
3’	5’－CTCTAAGCTTTTA*TCA*TTTGTATAGTTCATCCATG－3’	TGA
3’	5’－CTCTAAGCTTTTA*GCT*TTTGTATAGTTCATCCATG－3’	AGC

Table 2.3 Relationship between the last sense codon and the RFI

(normalized with wild type GFP gene)

Amino acid	Codon	RFI	Amino acid	Codon	RFI	Amino acid	Codon	RFI	Amino acid	Codon	RFI
Phe	TTT	±	Ser	TCT	±	Tyr	TAT	±	Cys	TGT	−
	TTC	±		TCC	±		TAC	−		TGC	±
Leu	TTA	−		TCA	±	STOP	TAA	±	STOP	TGA	±
	TTG	−		TCG	±		TAG	−	Trp	TGG	+
Leu	CTT	−	Pro	CCT	±	His	CAT	±	Arg	CGT	±
	CTC	±		CCC	±		CAC	−		CGC	−
	CTA	−−		CCA	±	Gln	CAA	±		CGA	±
	CTG	−		CCG	++		CAG	±		CGG	−
Ile	ATT	±	Thr	ACT	±	Asn	AAT	±	Ser	AGT	±
	ATC	±		ACC	±		AAC	±		AGC	±
	ATA	±		ACA	±	Lys	AAA	±	Arg	AGA	−
Met	ATG	−		ACG	±		AAG	+		AGG	±
Val	GTT	−	Ala	GCT	−	Asp	GAT	±	Gly	GGT	±
	GTC	±		GCC	±		GAC	±		GGC	±
	GTA	±		GCA	+	Glu	GAA	±		GGA	+
	GTG	−		GCG	±		GAG	−		GGG	−

RFI : 1.5 above ++ From 1.2 to 1.5 + From 0.8 to 1.2 ±

From 0.5 to 0.8− 0.5 below −−

Table 2.4 CUFs of last sense codon

		Second															
		T				C				A				G			
First	T	Phe (F)	TTT	C	19.7	Ser (S)	TCT	C	5.7	Tyr (Y)	TAT	C	16.8	Cys (C)	TGT	D	5.9
			TTC	C	15.0		TCC	C	5.5		TAC	D	14.6		TGC	C	8.0
		Leu (L)	TTA	D	15.2		TCA	C	7.8	STOP	TAA	C	1.8	STOP	TGA	C	1.0
			TTG	D	11.9		TCG	C	8.0		TAG	D	0.0	Trp (W)	TGG	B	10.7
	C	Leu (L)	CTT	D	11.9	Pro (P)	CCT	C	8.4	His (H)	CAT	C	15.8	Arg (R)	CGT	C	21.1
			CTC	C	10.5		CCC	C	6.4		CAC	D	13.1		CGC	D	26.0
			CTA	E	5.3		CCA	C	6.6	Gln (Q)	CAA	C	12.1		CGA	C	4.3
			CTG	D	46.9		CCG	A	26.7		CAG	C	27.7		CGG	D	4.1
	A	Ile (I)	ATT	C	30.5	Thr (T)	ACT	C	8.0	Asn (N)	AAT	C	21.9	Ser (S)	AGT	C	7.2
			ATC	C	18.2		ACC	C	22.8		AAC	C	24.4		AGC	C	16.6
			ATA	C	3.7		ACA	C	6.4	Lys (K)	AAA	C	33.2	Arg (R)	AGA	D	1.4
		Met (M)	ATG	D	24.8		ACG	C	11.5		AAG	B	12.1		AGG	C	1.6
	G	Val (V)	GTT	D	16.8	Ala (A)	GCT	D	10.7	Asp (D)	GAT	C	37.9	Gly (G)	GGT	C	21.3
			GTC	D	11.7		GCC	C	31.6		GAC	C	20.5		GGC	C	33.4
			GTA	C	11.5		GCA	B	21.1	Glu (E)	GAA	C	43.7		GGA	B	9.2
			GTG	D	26.4		GCG	C	38.5		GAG	D	18.4		GGG	D	8.6

RFI: A: 1.5 above B: From 1.2 to 1.5 C：From 0.8 to 1.2

D：From 0.5 to 0.8 E: 0.5 below

Chapter 3 Influence of the combination of last sense codon and stop codon on expression efficiency of GFP gene in *E. coli*

3.1 Introduction

In the former chapter, the influence of the last sense codon on expression efficiency of GFP gene in *E. coli* when the stop codon UAA had been researched, if the stop codon was changed, the influence of last sense codon was whether the same or not, in order to confirm this point, the stop codon was changed from UAA to UAG and UGA. 18 kinds of last sense codons were inserted into GFP gene and the amount of produced GFP in *E. coli* was measured by fluorescence intensity.

From Tab. 3.1, according to the former research results, GFP(CCG) and GFP(GGA) showed the strongest RFIs (2.1, 1.49), GFP(CTA) and GFP(TTG) showed the lower RFIs (0.33, 0.64), GFP(ATA), GFP(CCC), GFP(CCA), GFP(ACA), GFP(AGT) and GFP(AGG) showed the medium RFIs (0.88, 1.11, 0.82, 1.02, 1.1 and 0.9, respectively), TAA, TAG and TGA were stop codons, according to the reasons above, 18 kinds of last sense codons were chosen from 64 kinds of last sense codons.

One expression system was developed in 1986 by W. F. Studier and B. A. Moffatt, who created a RNA polymerase expression system which was strongly selective for bacteriophage T7 RNA polymerase. The initial system involved two different methods of maintaining T7 RNA polymerase into the cell - in one method, a lambda bacteriophage was used to insert the gene which codes for T7 RNA polymerase, and in the other, the gene for T7 RNA polymerase was inserted into the host chromosome [1]. This expression system has become known as the pET expression System, which is based on the T7 phage RNA polymerase promoter and is now widely used because of its ability to mass-produce proteins, the specificity involved in the T7 promoter which only binds T7 RNA polymerase, and also the design of the system which allows for the easy manipulation of how much of the desired protein is expressed and when that expression occurs [2].

Cell-free protein synthesis harnesses the synthetic power of biology, programming

the ribosomal translational machinery of the cell to create macromolecular products. Like PCR, which uses cellular replication machinery to create a DNA amplifier, cell-free protein synthesis is emerging as a transformative technology with broad applications in protein engineering, biopharmaceutical development, and post-genomic research [3].

In vitro synthesis of proteins in cell-free extracts is an important tool for molecular biologists and has a variety of applications, including the rapid identification of gene products, protein folding studies, incorporation of modified or unnatural amino acids for functional studies, and investigation of protein-protein and antibody epitope mapping [4].

Different expression vectors have different elements for gene expression. For the expression vectors (pKK223-3 and pET21(+)), the two expression vectors were commonly used for gene expression. Each of them was reported to be strongly efficient, robust and costeffective. The expression vector pKK223-3 uses *E. coli* RNA polymerase. Translation system *in vitro* used T7 RNA polymerase. *In vitro* translation system was shown in Fig. 3.1. In order to analyze the influence between genetic codon and expression efficiency, the translation system was changed from *E. coli* to *in vitro*. In this case, the expression vector pKK223-3 could not be expressed, so the expression vector was changed from pKK223-3 to pET21(+), also competent cell was changed from *E. coli* JM109 strain to *E. coli* BL21 strain (Fig. 3.2).

It was seen that in Tab. 3.1, firstly 18 kinds of last sense codons were chose to research the influence of the combination of last sense codon and stop codon on expression efficiency of GFP gene in *E. coli* by the expression vector pKK223-3. Later the expression vector was changed from pKK223-3 to pET21(+) and 12 kinds of last sense codons were chosen to research the influence of the combination of last sense codon and stop codon on expression efficiency of GFP gene in *E. coli* by the expression vector pET21(+) (Tab. 3.2).

3.2 Experimental

When the expression vector was pKK223-3, 18 kinds of last sense codons were chosen (Tab. 3.1), and were inserted into GFP genes by PCR (Fig. 3.3). When the expression vector was pET21(+), 12 kinds of last sense codons were chosen (Tab. 3.2). All the primers were shown in Tab. 3.3. The part with the underline was the stop codon. The shaded parts indicated restriction enzyme digested sites corresponding to the primer description.

3.3 Results and discussions

3.3.1 Comparison of three kinds of stop codons without insertion of last sense codon by the expression vector pKK223-3

The stop codon had three kinds (UAA, UAG, and UGA), the wild types of three kinds of stop codons under the same conditions were determined the fluorescence intensities, it was found that the sequence of fluorescence intensity for three kinds of GFP genes of stop codons was wild(UGA) > wild(UAA) > wild(UAG). The result of comparison was seen in Fig. 3.4.

3.3.2 Comparison of the RFI by the stop codon was changed by the expression vector pKK223-3

As the results (Fig. 3.5), for most of the last sense codons, with one kind of last sense codon, when the stop codon was replaced, the change of fluorescence intensity was small. But some specific phenomena were seen, for example, for the last sense codon CCG, the RFI of GFP(CCG) was 2.1 when the stop codon was UAA, but in comparison, the RFI was 1.1 when the stop codon was changed from UAA to UAG. For last sense codon TAG, the RFI of GFP(TAG) with the stop codon UAG was stronger than that with the stop codons UAA and UGA.

Comparison to wild type GFP, some GFP genes had the stronger or lower RFIs when the three kinds of stop codons. For example, when the expression vector was pKK223-3, GFP(CCG), GFP(GGA) always had the stronger fluorescence intensities and GFP(GTT), GFP(TTG), GFP(CTA) always had the lower fluorescence intensities when the three stop codons (UAA, UAG, and UGA) comparison to wild type GFP.

For the last sense codons which belonged to the same amino acid, the GFP genes with these last sense codons had the nearly same RFIs, for example, the last sense codons GTT and GTG belonged to Valine, the last sense codon ACC and ACA belonged to Threonine, the last sense codon AGT and AGC belonged to Serine. Their fluorescence intensities of GFP genes with these last sense codons were separately the same. There were also special cases, for example, the last sense codon CCC, CCA and CCG belonged to Proline, GFP(CCC) and GFP(CCA) had the near fluorescence intensities, but for the last sense codon CCG, when the stop codon was UAA, the RFI of GFP(CCG) was particularly strong (2.1).

The recognition of stop codons by release factors 1 (RF-1) and 2 (RF-2) leads to peptide chain termination during translation. Stop codons (UAG and UAA) are recognized by RF-1 while stop codons (UAA and UGA) are recognized by RF-2. The difference of the influence on expression efficiency among stop codons might be caused by the change of the release factors, this needed further verification.

3.3.3 Comparison of the RFI by the stop codon was changed by the expression vector pET21(+)

When the expression vector was pET21(+), 12 kinds of last sense codons (TGA, ACA, AGC, TAG, CCA, TAA, GTG, CCG, AAA, ACC, AGT, GTT) was inserted into GFP for the stop codon UAG, respectively. From the result of our experiment (Fig. 3.6), for the stop codon UAG, comparison to wild type GFP, GFP(TGA) had the strongest RFI (1.38), GFP(ACA) and GFP(ACC) had the medium RFIs (1.04, 0.94), GFP(CCG) had the lower RFI (0.85), other GFP genes with the last sense codons had the near fluorescence intensities by the expression vector pET21(+). As shown in Fig. 3.7, the fluorescence intensities of three kinds of stop codons were near, GFP(TGA) had the strongest RFI. Other GFP genes with the last sense codons had the near RFIs.

3.3.4 Comparison of expression vectors (pKK223-3 and pET21(+))

The expression vectors (pKK223-3 and pET21(+)) had some same points on

expression efficiency, especially when the stop codon was UAG or UGA (Fig. 3.8, Fig. 3.9). When the stop codon was UAA, there were great differences on the expression efficiency. For example, for the last sense codon CCG, the RFI of GFP(CCG) was 0.89 by the expression vector pET21(+), the RFI of GFP(CCG) was 2.1 by the expression vector pKK223-3, the clear rule was not found (Fig. 3.10).

3.3.5 SDS-PAGE of soluble proteins extracted from *E. coli*

To confirm that the RFI was related to the expression efficiency of GFP gene, SDS-PAGE of soluble proteins extracted from *E. coli* was done. The result of SDS-PAGE was seen in Fig. 3.11 by the expression vector pKK223-3. Three kinds of last sense codons CCG, AGT, CTA were selected, and the RFIs by three kinds of stop codons were compared. As shown in Fig. 3.6, the results of RFIs were: CCG(UAA) 2.1, CCG(UAG) 1.1, CCG(UGA) 1.1, AGT(UAA) 1.1, AGT(UAG) 0.92, AGT(UGA) 0.91, CTA(UAA) 0.33, CTA(UAG) 0.73, and CTA(UGA) 0.71. Molecular weight of GFP was about 27,000, from the band of GFP, it was shown that the band of CCG(UAA) was thickest, the bands of CTA(UAA), CTA(UAG), CTA(UGA) were thinner, the thicknesses of other bands were nearly the same. The result of RFIs had the same tendency with that of SDS-PAGE.

According to the result of the RFIs, several last sense codons were selected to do the SDS-PAGE. For the expression vector pET21(+), when the stop codon was UAG, among 12 kinds of last sense codons, wild type GFP (RFI = 1), GFP(TGA) (1.38), GFP(ACA) (1.04), GFP(ACC) (0.94), GFP(CCG) (0.85) and the expression vector pET21(+) were selected to do the SDS-PAGE. The molecular weight of GFP was about 27,000, as shown in Fig. 3.12, the bands of GFP(TGA) and GFP(ACA) were thickest and the bands of GFP(ACC) and GFP(CCG) were thinner than the bands of wild type, GFP(TGA), and GFP(ACA). The results of SDS-PAGE had the same tendency with the ones of RFIs. Because GFP would give out green light under ultraviolet light, the samples were laid up under ultraviolet light. It was shown that the result of luminescenties (Fig. 3.13) stayed the same tendency with that of the result of SDS-PAGE (Fig. 3.12).

3.4 Conclusions

In this chapter, the influence of the combination of last sense codon and stop codon on expression efficiency of GFP gene in *E. coli* was researched. Besides the last sense codon, the stop codon had the effects on the expression of genes. The differences of two expression vector (pKK223-3 and pET21(+)) had effects on the gene expression. When the stop codon was changed, for most of GFP genes which were inserted last sense codons, the fluorescence intensities were not changed. Some GFP genes always had the stronger expression efficiencies, such as GFP(CCG), GFP(GGA), when these last sense codons were inserted, the quantities of proteins would be increased. It was confirmed that the combination of stop codon and the last sense codon (i.e., the codon preceding the stop codon) had influenced the expression efficiency of GFP gene.

References

1. Moffatt, BA., Studier, FW. Use of Bacteriophage T7 RNA Polymerase to Direct Selective Strong-Level Expression of Cloned Genes. *Molecular Biology*., **189(1)**, 113-130 (1986).
2. Unger, TF. Show Me The Money: Prokaryotic Expression Vectors And Purification Systems. *Scientist- Philadelphia*., **11**, 20-22 (1997).
3. Whittaker, J. W. Cell-free protein synthesis: the state of the art. *Biotechnol Lett*., **35(2)**, 143-152 (2013).
4. Caschera, F., Bedau, M. A., Buchanan, A., Cawse, J., Lucrezia, D., Gazzola, G., Hanczyc, M. M., Packard, N. H. Coping With Complexity: Machine Learning Optimization of Cell-Free Protein Synthesis. *Biotechnol Bioeng*., **108(9)**., 2218-28 (2011).

Figures and tables

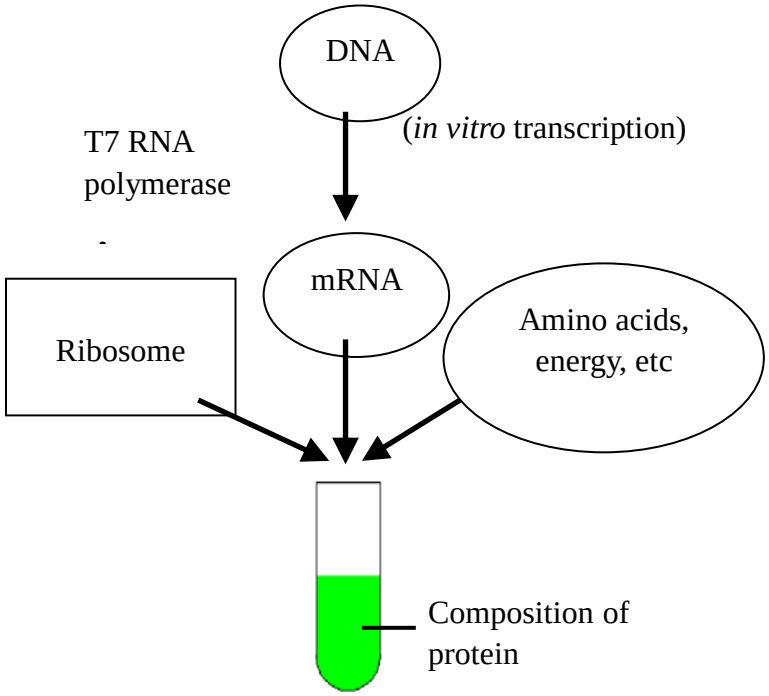


Figure 3.1 *In vitro* translation system

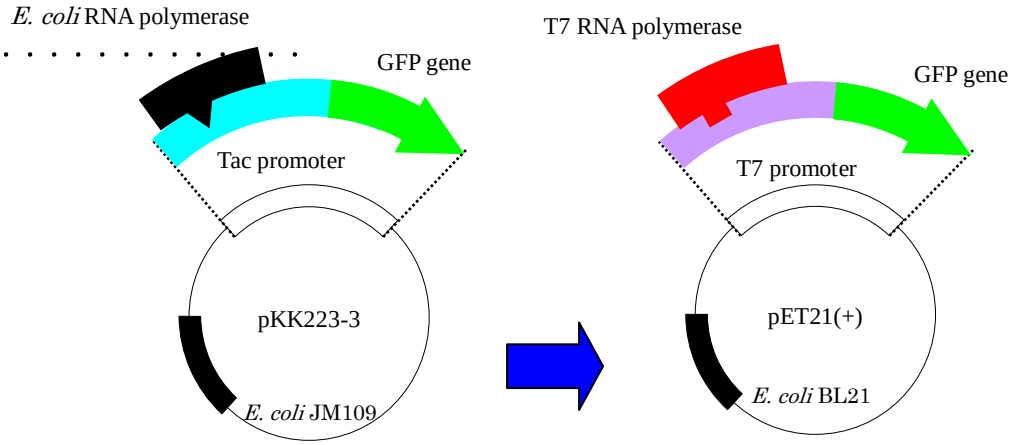


Figure 3.2 Change of expression vector (From pKK223-3 to pET21(+))

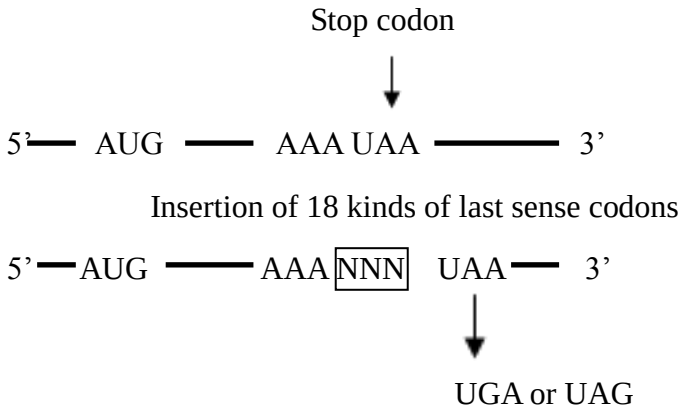


Figure 3.3 The insertion of last sense codons and change of stop codon in GFP gene

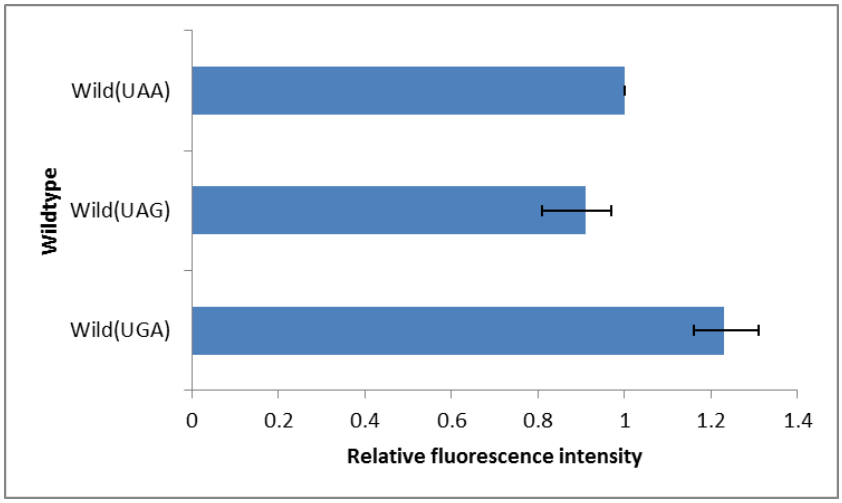


Figure 3.4 Comparison of the RFIs by three kinds of stop codons without insertion of last sense codon by the expression vector pKK223-3

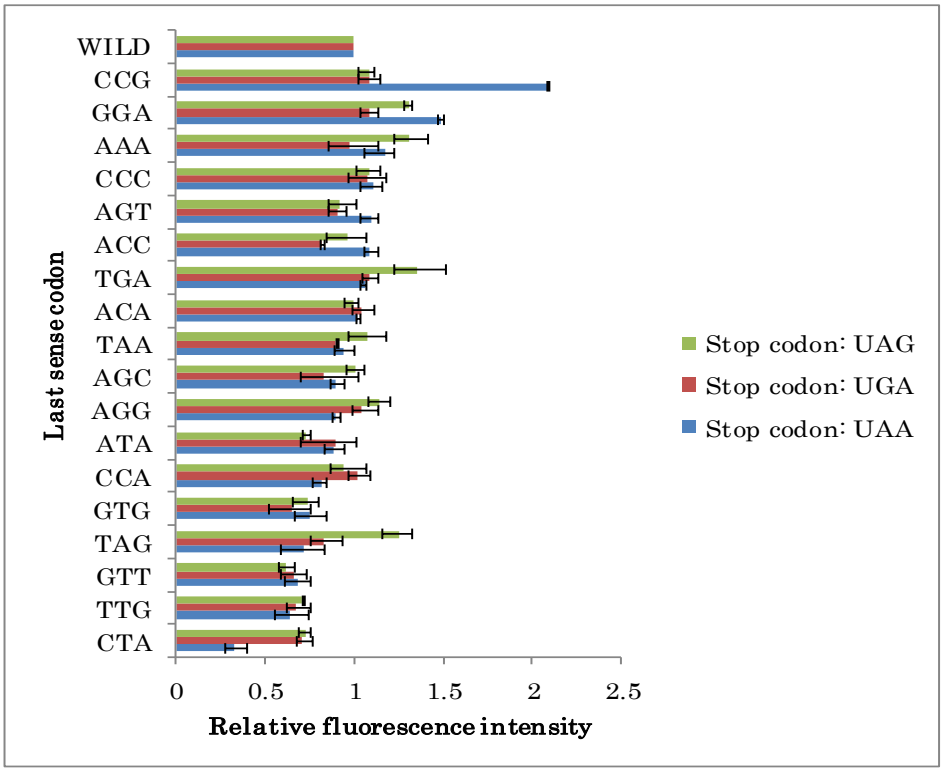


Figure 3.5 RFIs of GFP expressed by genes with insertion of last sense codons by the expression vector pKK223-3

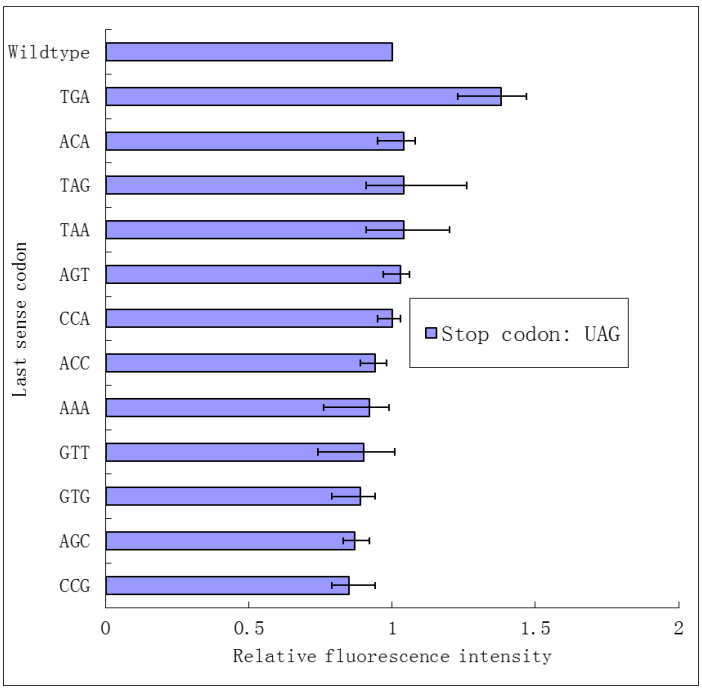


Figure 3.6 Comparison of the RFIs under the stop codon UAG

by the expression vector pET21(+)

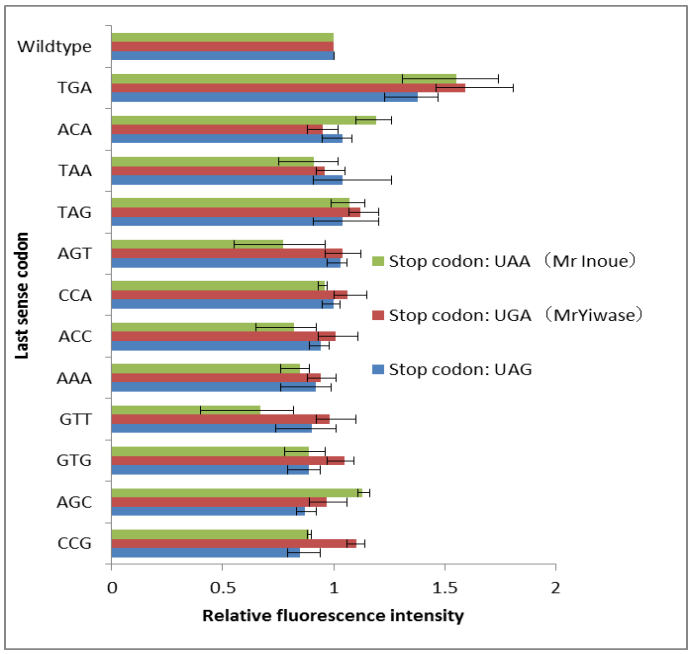


Figure 3.7 Comparison of the RFIs under three kinds of stop codons by the expression vector pET21(+)

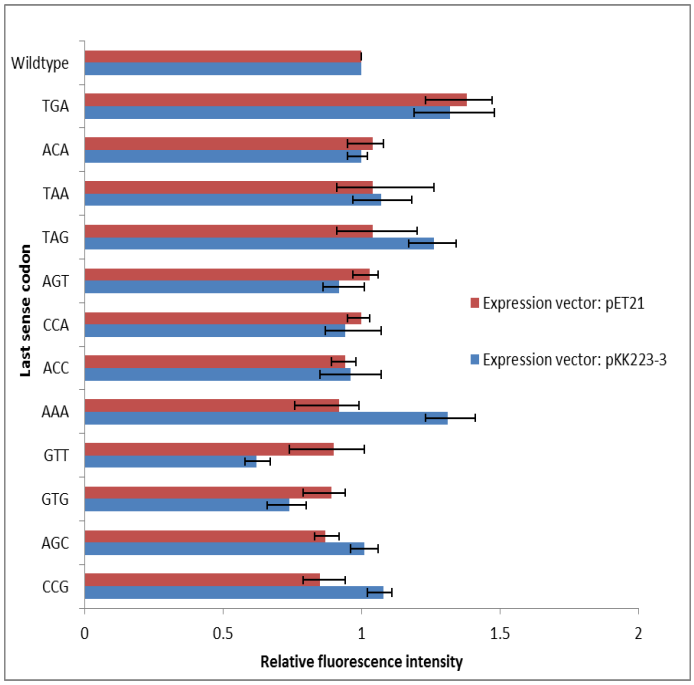


Figure 3.8 Comparison of the RFIs of the expression vectors (pKK223-3 and pET21) by the stop codon, UAG

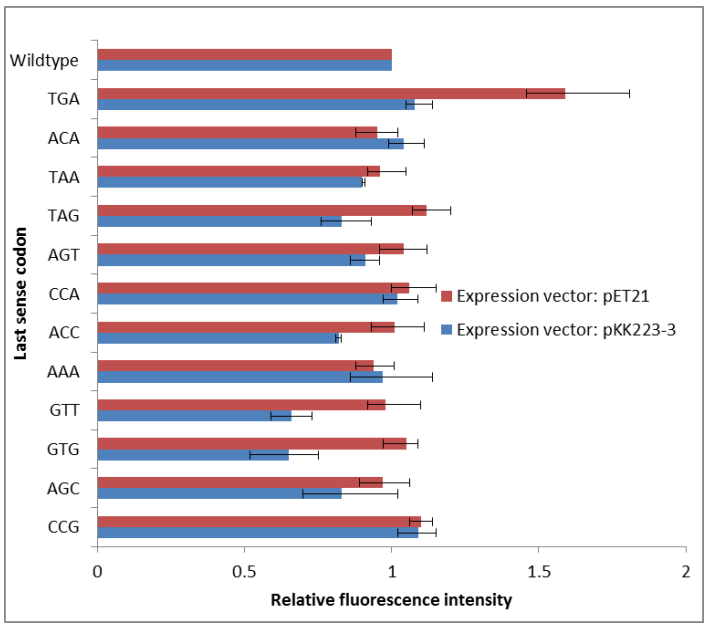


Figure 3.9 Comparison of the RFIs of the expression vectors (pKK223-3 and pET21) by the stop codon, UGA

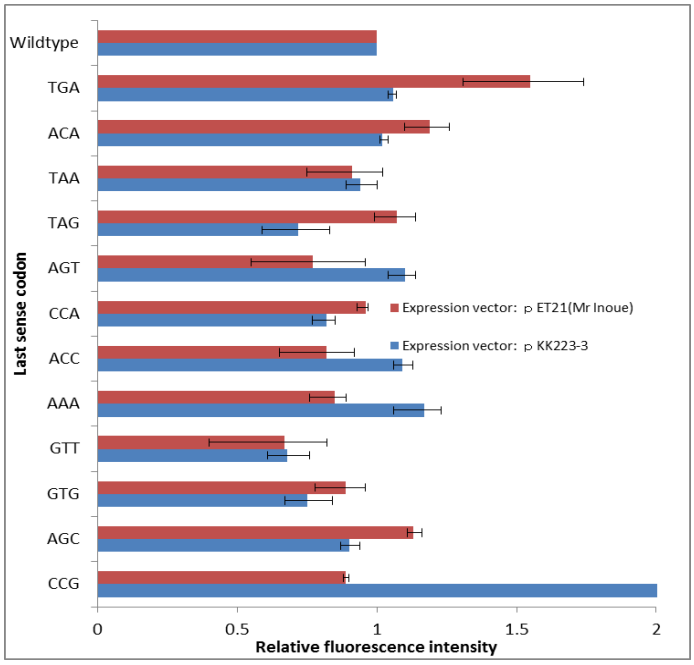


Figure 3.10 Comparison of the RFIs of the expression vectors (pKK223-3 and pET21) by the stop codon, UAA

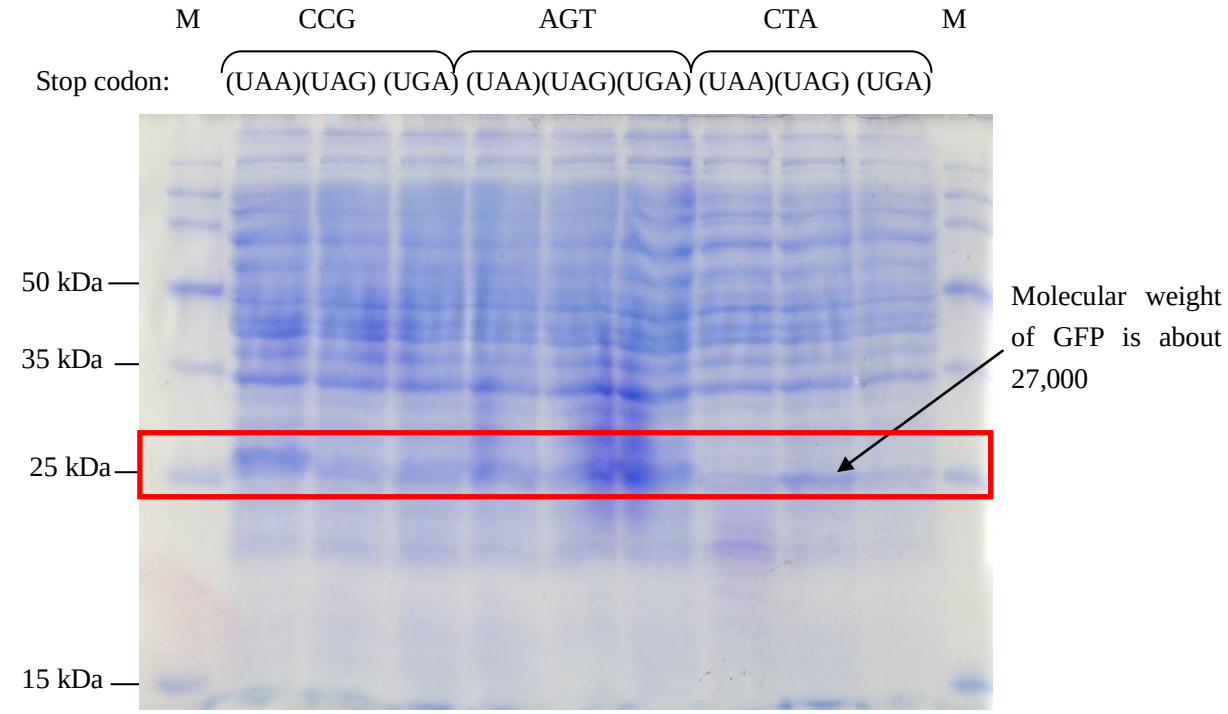


Figure 3.11 SDS-PAGE of soluble proteins extracted from *E. coli* transformed by having three kinds of last sense codons in GFP genes by the expression vector pKK223-3

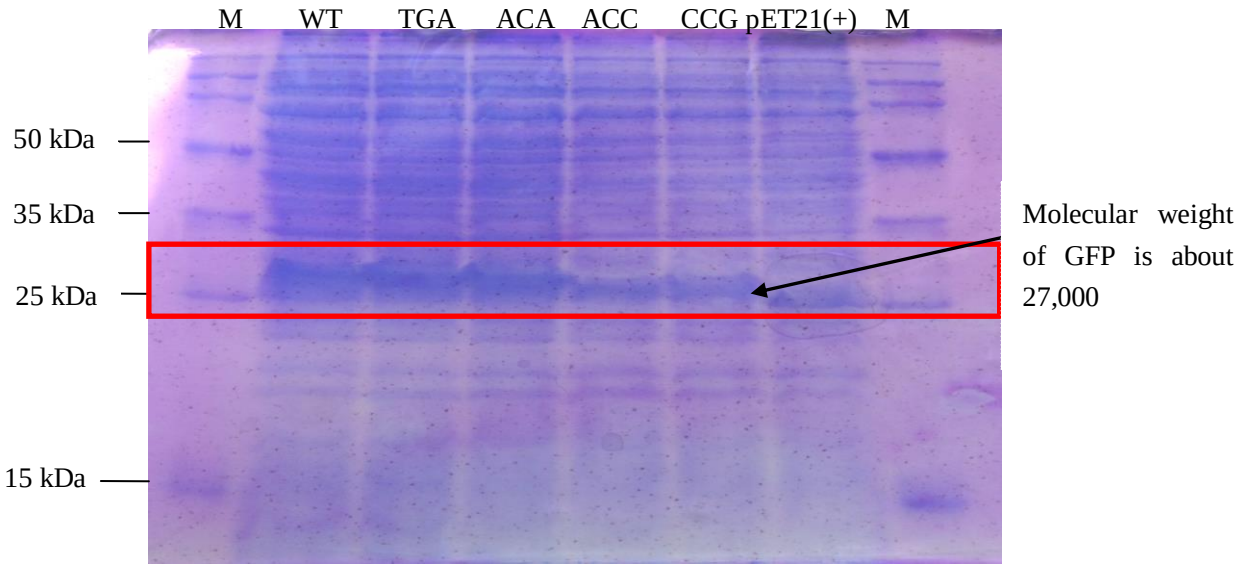


Figure 3.12 SDS-PAGE of soluble proteins extracted from *E. coli* transformed by having four kinds of last sense codons in GFP genes by the expression vector pET21(+)

WT TGA ACA ACC CCG pET21(+)

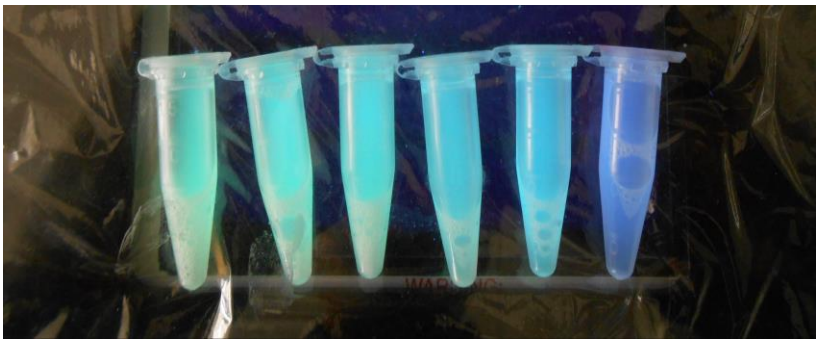


Figure 3.13 Comparison of fluorescence of GFP under ultraviolet light

Table 3.1 18 kinds of last sense codons

		Second											
		T			C			A			G		
First	T	Phe (F)	TTT		Ser (S)	TCT		Tyr (Y)	TAT		Cys (C)	TGT	
			TTC			TCC			TAC			TGC	
		Leu (L)	TTA			TCA		STOP	TAA	C	STOP	TGA	C
			TTG	D		TCG			TAG	D	Trp (W)	TGG	
	C	Leu (L)	CTT		Pro (P)	CCT		His (H)	CAT		Arg (R)	CGT	
			CTC			CCC	C		CAC			CGC	
			CTA	E		CCA	C	Gln (Q)	CAA			CGA	
			CTG			CCG	A		CAG			CGG	
	A	Ile (I)	ATT		Thr (T)	ACT		Asn (N)	AAT		Ser (S)	AGT	C
			ATC			ACC (M)	E		AAC			AGC (M)	E
			ATA	C		ACA	C	Lys (K)	AAA		Arg (R)	AGA	
		Met (M)	ATG	D		ACG			AAG			AGG	C
	G	Val (V)	GTT	D	Ala (A)	GCT		Asp (D)	GAT		Gly (G)	GGT	
			GTC			GCC			GAC			GGC	
			GTA			GCA		Glu (E)	GAA			GGA	B
			GTG	D		GCG			GAG			GGG	

RFI : A: 1.5 above B: From 1.2 to 1.5 C：From 0.8 to 1.2

D：From 0.5 to 0.8 E: 0.5 below

Table 3.2 Comparison of 12 kinds of last sense codons

		Second							
		T		C		A		G	
First	T	Phe(F)	TTT	Ser(S)	TCT	Tyr(Y)	TAT	Cys(C)	TGT
			TTC		TCC		TAC		TGC
		Leu(L)	TTA		TCA	STOP	TAA	STOP	TGA
			TTG		TCG		TAG	Trp(W)	TGG
	C	Leu(L)	CTT	Pro(P)	CCT	His(H)	CAT	Arg(R)	CGT
			CTC		CCC		CAC		CGC
			CTA		CCA	Gln(Q)	CAA		CGA
			CTG		CCG		CAG		CGG
	A	Ile(I)	ATT	Thr(T)	ACT	Asn(N)	AAT	Ser(S)	AGT
			ATC		ACC		AAC		AGC
			ATA		ACA	Lys(K)	AAA	Arg(R)	AGA
		Met(M)	ATG		ACG		AAG		AGG
	G	Val(V)	GTT	Ala(A)	GCT	Asp(D)	GAT	Gly(G)	GGT
			GTC		GCC		GAC		GGC
			GTA		GCA	Glu(E)	GAA		GGA
			GTG		GCG		GAG		GGG

Table 3.3 Primer sequences

NNN: stop codon (UAG, UGA)

Primer	Sequence	Last sense codon
3’	5’－CTCTAAGCTTNNN***CGG***TTTGTATAGTTCATCCATG－3’	CCG
3’	5’－CTCTAAGCTTNNN***TCC***TTTGTATAGTTCATCCATG－3’	GGA
3’	5’－CTCTAAGCTTNNN***TTT***TTTGTATAGTTCATCCATG－3’	AAA
3’	5’－CTCTAAGCTTNNN***GGG***TTTGTATAGTTCATCCATG－3’	CCC
3’	5’－CTCTAAGCTTNNN***ACT***TTTGTATAGTTCATCCATG－3’	AGT
3’	5’－CTCTAAGCTTNNN***GGT***TTTGTATAGTTCATCCATG－3’	ACC
3’	5’－CTCTAAGCTTNNN***TCA***TTTGTATAGTTCATCCATG－3’	TGA
3’	5’－CTCTAAGCTTNNN***TGT***TTTGTATAGTTCATCCATG－3’	ACA
3’	5’－CTCTAAGCTTNNN***TTA***TTTGTATAGTTCATCCATG－3’	TAA
3’	5’－CTCTAAGCTTNNN***CCT***TTTGTATAGTTCATCCATG－3’	AGG
3’	5’－CTCTAAGCTTNNN***GCT***TTTGTATAGTTCATCCATG－3’	AGC
3’	5’－CTCTAAGCTTNNN***TAT***TTTGTATAGTTCATCCATG－3’	ATA
3’	5’－CTCTAAGCTTNNN***TGG***TTTGTATAGTTCATCCATG－3’	CCA
3’	5’－CTCTAAGCTTNNN***CAC***TTTGTATAGTTCATCCATG－3’	GTG
3’	5’－CTCTAAGCTTNNN***CTA***TTTGTATAGTTCATCCATG－3’	TAG
3’	5’－CTCTAAGCTTNNN***AAC***TTTGTATAGTTCATCCATG－3’	GTT
3’	5’－CTCTAAGCTTNNN***CAA***TTTGTATAGTTCATCCATG－3’	TTG
3’	5’－CTCTAAGCTTNNN***TAG***TTTGTATAGTTCATCCATG－3’	CTA

Chapter 4 Study on the relationship between mutation and expression efficiency of GFP gene in *E. coli*

4.1 Introduction

The research about GFP and its some mutants has always been concerned. People had found mutant was a good way, seven strategies were used to generate mutations in GFP [1]. A library had been constructed in *E. coli* of mutant GFP genes [2]. A new method had been developed to determine the mutagenic efficacy of a suspected mutagen by employing GFP as a direct biosensor for mutation detection [3]. A new strategy of molecular evolution aimed at generating a red-emitting mutant of GFP had been used and had succeeded in producing the first GFP mutant that substantially matures to the red-emitting state with excitation and emission maxima at 555 and 585 nm, respectively [4].

Extensive efforts had been applied to create GFP variants. As a result, blue and cyan mutants with Tyr66 substituted with Phe, His, or Trp were generated. Also, it was demonstrated that certain mutations can strongly influence a proportion of neutral / anionic chromophore [3]-[5]. Two plasmid vectors had been constructed for the expression of the synthetic SGFP-TYG gene in ascomycete species, indicating that SGFP-TYG can be used as a strongly effective vital marker in ascomycetes [6]. The most widely studied of the folding mutants is F99S/M153T/V163A mutant with a modified codon usage for better expression in E. coli, which was found by forming random mutations with PCR [7].

Relative to the wild type of GFP, Emerald GFP was originally reported to have 5 amino acid changes (S65T, S72A, N149K, M163T, and I167T) [8]. Point mutations have been inserted, which resulted in much faster chromophore formation and up to a 100-fold increase in the fluorescence of the produced protein [5], [9], [10].

To research the properties of the mutations could be helpful for understanding the mechanism of mutant sites and finding the phenomenon in my favor. Compared to the newly discovered mutants, the previously characterized mutants GFP-S65T and GFP-Y66H had been employed [11]. Many properties relevant to quantitative imaging in living cells of wildtype, AlphaGFP (F99S/M153T/V163A), S65T, EGFP (F64L / S65T) , EBFP (F64L /S65T / Y66H / Y145F) had been investigated. The five GFP variants had been used extensively or are potentially useful [12].

In our previous work, we acquired the 20 mutant recombinant GFP genes. As shown in Tab. 4.1, for last sense codon TAC, the difference of RFI between the normal GFP gene and the mutant GFP gene was not big; for the last sense codon CTT, the nucleotide "G" on the GFP gene chain was missed; for the last sense codon CAA, the mutant had two mutant sites, the influence of two mutant sites on expression efficiency was difficult for us. It was shown that one special mutant GFP(ACC) was found, whose inserted last sense codon was ACC, had three mutant sites, at 56, 357 and 443 of the GFP gene. The RFI of *E. coli* having this vector indicated 0.1. By designing primers, mutant GFP(ACC) could be separated and the GFP gene with one mutant site could be acquired, for the reasons above, GFP(TCG), GFP(ACA), GFP(CTA), GFP(ACT), and GFP(ACC) were chosen to research the relationship between mutant gene and protein synthesis efficiency.

4.2 Experimental

pKK223-3 was used as the expression vector in this study. In order to research the relationship between mutant gene and the expression efficiency, PCR was used to remove these last sense codons of the modified GFP genes.

It was shown in Fig. 4.1 that mutant GFP(ACC) had three mutant sites by determining sequence. It had been known that GFP gene had 717 bases. The site of "A"

on the 5’ side of GFP gene was named “1”, according to this way, three mutant sites of GFP(ACC) was found as the sites of 56, 357, and 443. In order to separate the three mutant sites, several enzymes and primers were designed (Fig. 4.2). “CACGTG” was the recognition site of *Pma*CI, “GTTAAC” was the recognition site of *Hpa*I. The digested sites of *Eco*RI and *Hin*dIII were located in the 5’ and 3’ side of GFP gene.

The constructions of T357C and A443G were different from the other mutants. The restriction enzymes (*Pma*CI and *Hin*dIII) were used to digest the wild type of GFP and created T357C. Later the restriction enzymes (*Eco*RI and *Hpa*I) were used to digest the wild type GFP and created A443G. The used primers were seen below:

(1) T357C (*Pma*CI and *Hin*dIII)

GFP *Pma*CI Primer (5’Primer)

5’-AAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTCGTTAATAGA ATCGAGTAAA-3’ (underline shows the mutation site, from “T” to “C”)

GFP *Hin*dIII New Primer (3’Primer)

5’-CTCTAAGCTTTTATTTGTATAGTTCATCCA-3’

The 5’Primer *Pma*CI had one point mutation “C”, according to the principle of PCR, when the 5’Primer was connected with the DNA molecule, the point mutation “C” was inserted into the GFP gene. After the expression vector was ligazed with the GFP gene, the modified GFP with one mutant site was acquired.

(2) A443G (*Eco*RI and *Hpa*I)

GFP *Eco*RI Primer (5’Primer)

5’-CCCGAATTCTTTAACTTTAGGAAACACAATTCATGAGTAAAGGAGAAG AACTT-3’

GFP *Hpa*I Primer (3’Primer)

5’-TGAAGTTAACTTTGATTCCATTCTTTTGTTTGTCTGCCATGATGTATACA TTGCGTGAGTTATAGTTGTATTCC-3’ (underline shows the mutation site, from “T”

to “C”)

Apart from T357C and A443G, there were other four mutant GFP genes (A402G, A675C, A462G, A88G). The four mutant GFP genes had the last sense codons. From the former researches, it could learn that the last sense codon had influenced the expression efficiency. When the last sense codon was existed, the relationship between mutant gene and protein synthesis efficiency was difficult, so it was necessary to remove the last sense codons. So the primers (*Eco*RI and *Hin*dIII) were designed to remove the last sense codons of the four mutant GFP genes. The primers were:

GFP *Eco*RI Primer (5’Primer):

5’-CCCGAATTCTTTAACTTTAGGAAACACAATTCATGAGTAAAGGAGAAG AACTT-3’

GFP *Hin*dIII New Primer (3’Primer):

5’-CTCTAAGCTTTTATTTGTATAGTTCATCCA-3’

4.3 Results and discussions

4.3.1 Construction of mutant genes

By the study, due to the experimental conditions and operational constraints, it was failure to acquire the first mutant GFP(ACC-M1), finally the second mutant (T357C) and the third mutant (A443G) were acquired.

As shown in Fig. 4.3, the enzymes *Pma*CI and *Hin*dIII were used to digest wild type GFP to acquire the expression vector which had the *Pma*CI and *Hin*dIII digested sites, the primer *Pma*CI which had one mutant site and primer *Hin*dIII were used to acquire the GFP gene of T357C by PCR. As shown in Fig. 4.4, the 5’ primer *Pma*CI had one mutant, during PCR, 5’ primer was ligated with the DNA molecule, after PCR, the GFP gene fragment had one mutant site. By ligation, the expression vector and the GFP gene fragment was ligated, T357C was acquired, without last sense codon, but with only one

mutant site. Similarly, the enzymes *Eco*RI and *Hpa*I were used to digest wild type GFP to acquire the expression vector which had the *Eco*RI and *Hpa*I digested sites, the primer *Hpa*I which had one mutant site and primer *Eco*RI were used to acquire the GFP gene fragment of A443G by PCR. After the expression vector and GFP gene fragment were ligated, A443G was acquired (Fig. 4.5).

As shown in Fig. 4.6, the enzymes *Eco*RI and *Hin*dIII were used to digest the pKK223-3. PCR procedure of A402G was shown in Fig. 4.7. The *Eco*RI and *Hin*dIII were used to acquire the common GFP gene fragment by PCR, it was shown that the GFP gene had one point mutation "G", and last sense codon was "TCG". 3' *Hin*dIII new primer was used, because the primer had not complementary codon with the last sense codon TCG, after PCR, the last sense codon was removed from the GFP gene, after the expression vector and the GFP gene were ligated, the mutant A402G was acquired.

As shown in Tab. 4.2, each mutant GFP gene had only one mutant site. For now, getting the mutant with a single mutant site was more easier to compare the influence of mutant site on expression efficiency of the genes than getting the mutant with two and more mutant sites. For A402G, the codon was changed from GGA to GGG. The amino acids on this position were both Glycine and were not substituted. The representations of other mutant GFP genes were similar with A402G, but for A88G and A443G, the amino acid was substituted from Serine to Glycine in A88G, from Histidine to Arginine in A443G, respectively.

As the results, six kinds of GFP mutants with only one mutant site were acquired and the fact that some mutant sites had affected on the RFI. From the results of the RFIs (Fig. 4.8), it was shown that the RFI of A402G was 1.2, A675C was 1.1, T357C was 1.0, A462G was 1.0, A88G was 0.16, A443G was 0.04.

4.3.2 SDS-PAGE of soluble proteins extracted from *E. coli*

Next six kinds of GFP genes were done the SDS-PAGE experiment. When the bands around 27,000 were compared, it was shown that the band of A402G was thickest, the bands of A88G and A443G were thinner than that of wild type. The results of SDS-PAGE were the same tendency with the results of fluorescence intensities (Fig. 4.8). SDS-PAGE confirmed that the RFI was associated with the quantity of synthesized protein. Based on the above results, it was learned that there were some differences in the expression efficiency among six kinds of GFP mutants. For the A402G, even the amino acid was not substituted, when the nucleotide was changed from "A" to "G" on the gene site 402, the RFI was increased to 1.2, it meaned that the expression quantity of the protein could be increased by the point mutation.

It was shown that though for A88G and A443G the amino acids were changed in Tab. 4.2. It was concluded that the change of RFIs was not based on the change of amino acids. The consequences of base substitution mutations in protein coding regions of a gene depended on the substitution and its location. It might be that when the nucleotide was mutated, inhibition might be produced, the inhibition might make the process of protein synthesis be tended to stop. This is only our speculation, this needed further verification.

4.4 Conclusions

In this chapter, the relationship between point mutaion and expression efficiency of GFP gene in *E. coli* was researched. Some mutations had effects on the expression efficiency of GFP genes. The RFI of mutant A402G was increased to 1.2. On the other hand, the RFI of A443G and A88G were decreased to 0.04 and 0.16, respectively.

References

1. Zacharias, DA., Tsien, RY. *Molecular biology and mutation of green fluorescent protein*, Green fluorescent protein: Properties, Applications, and Protocols, Second Edition., 83-118 (2006).

2. Cormack, BP., Valdivia, R., Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene.*, **173**, 371-374 (1996).

3. Tak, YK., Naoghare, PK., Lee, KH., Park, SS., Song, JM. Green fluorescent protein (GFP) as a direct biosensor for mutation detection: Elimination of false-negative errors in target gene expression. *Analytical Biochemistry*., **380**, 91-98 (2008).

4. Mishin, AS., Subach, FV., Yampolsky, IV., King, W., Lukyanov, KA., Verkhusha, VV. The first mutant of the aequorea victoria green fluorescent protein that forms a red chromophore. *Biochemistry*., **47**, 4666-4673 (2008).

5. Heim, R., Tsien, RY. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.*, **6**, 178-182 (1996).

6. Heim, R., Prasher, DC., Tsien, RY. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 12501-12504 (1994).

7. Crameri, A., Whitehorn, E. A., Tate, E., Stemmer, W. P. C. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol*., **14**, 315-319 (1996).

8. Maor, R., Puyesky, M., Horwitz, BA., Sharon, A. Use of green fluorescent protein (GFP) for studying development and fungal-plant interaction in Cochliobolus heterostrophus. *Mycol. Res.*, **102**, 491-496 (1998).

9. Tsien, RY. The green fluorescent protein. *Biochem.*, **67**, 509-544 (1998).

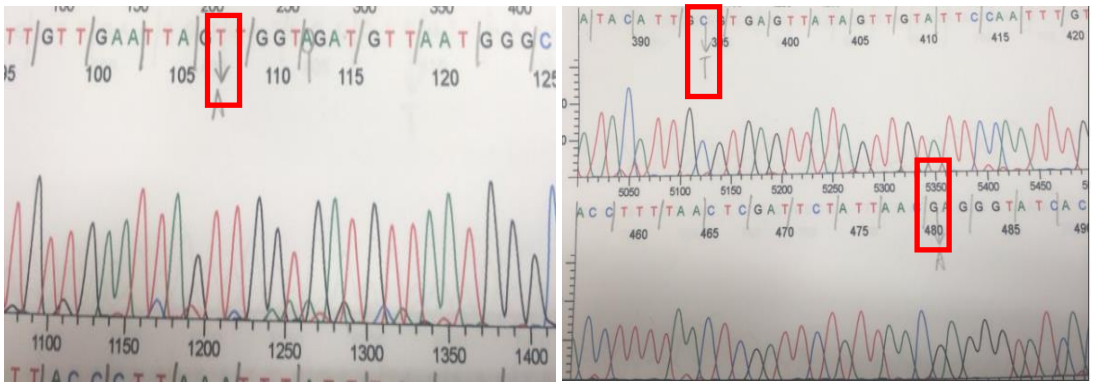
10. Heim, R., Cubitt, AB., Tsien, RY. Improved green fluorescence. *Nature*., **373**, 663-664 (1995).

11. Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H., Sheen, J. Engineered GFP as a vital reporter in plants. *Current Biology*., **6**, 325-330 (1996).

12. Von Arnim, AG., Deng, XW., Stacey, MG. Cloning vectors for the expression of green fluorescent protein fusion proteins in transgenic plants. *Gene*., **221**, 35-43 (1998).

13. Patterson, GH., Knobel, SM., Sharif, WD., Kain, SR., Piston, DW. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys.*, **73**, 2782-2790 (1997).

Figures and tables



(a) (b)

Figure 4.1 The sequence result of GFP(ACC)

5’ *Eco*RI site

ATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAG$\underline{A_1}$TGGT
GATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGA
AAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTG
TCACTACTTTCTCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACGGCAT
GACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTCAAAG
ATGACGGGAACTACAAGA***CACGTG***CTGAAGTCAAGTTTGAAGGTGATACCCT$\underline{T_2}$GTTAATA
GAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGG
AATACAACTATAACTCAC$\underline{A_3}$CAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCA
AA***GTTAAC***TTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATT
ATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTC
CACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGA
GTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAA *Hin*dIII site 3’

1- Mutant site of GFP(ACC-M1) 2- Mutant site of T357C

3- Mutant site of A443G

CACGTG: Recognition site of *Pma*CI GTTAAC: Recognition site of *Hpa*I

Figure 4.2 The mutant sites of GFP (ACC)

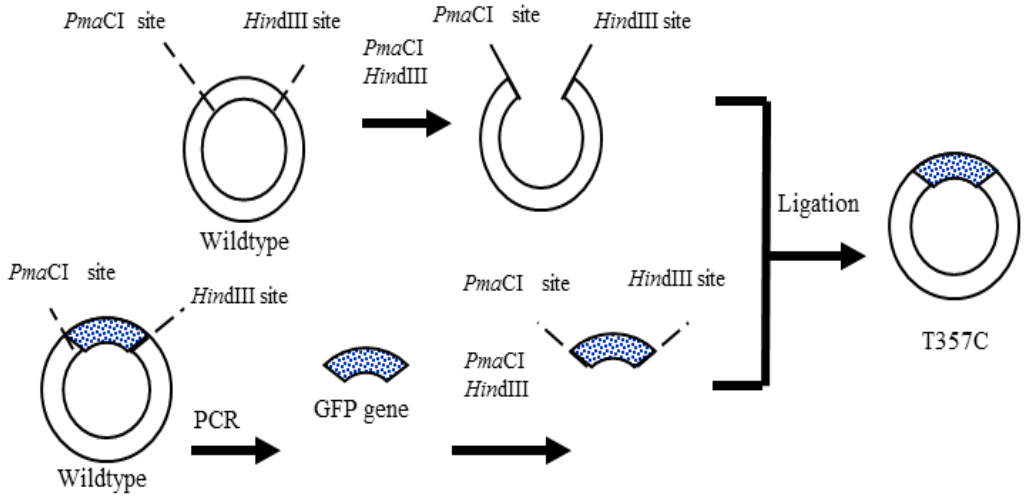


Figure 4.3 The construction of T357C

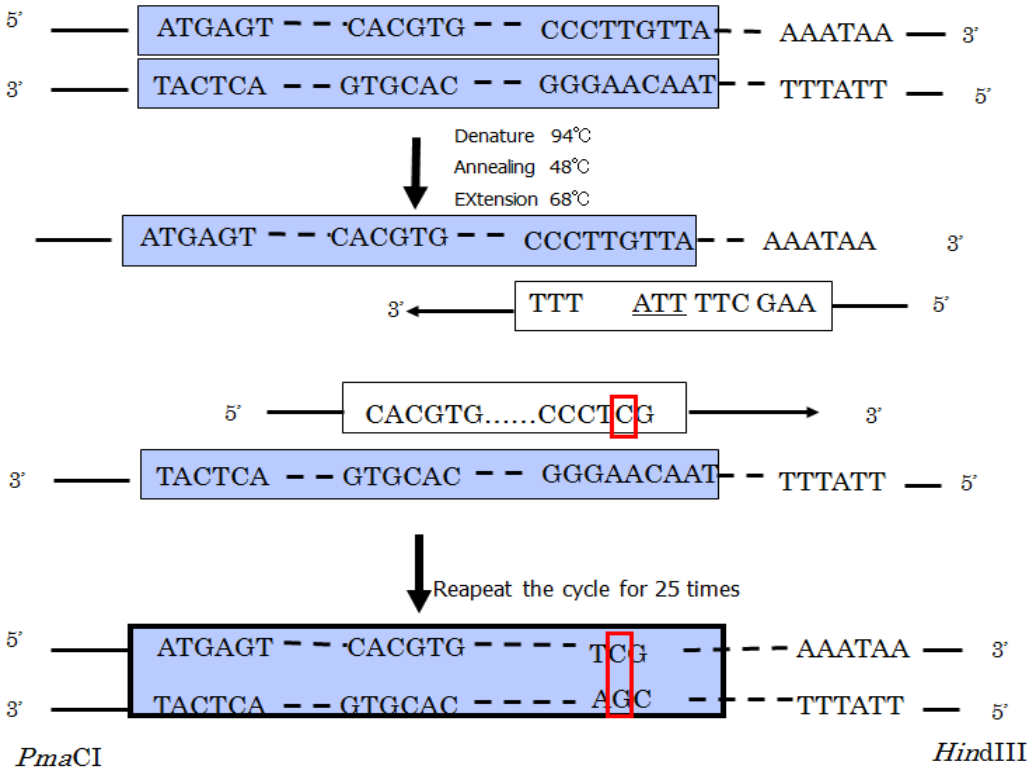


Figure 4.4 The PCR of T357C

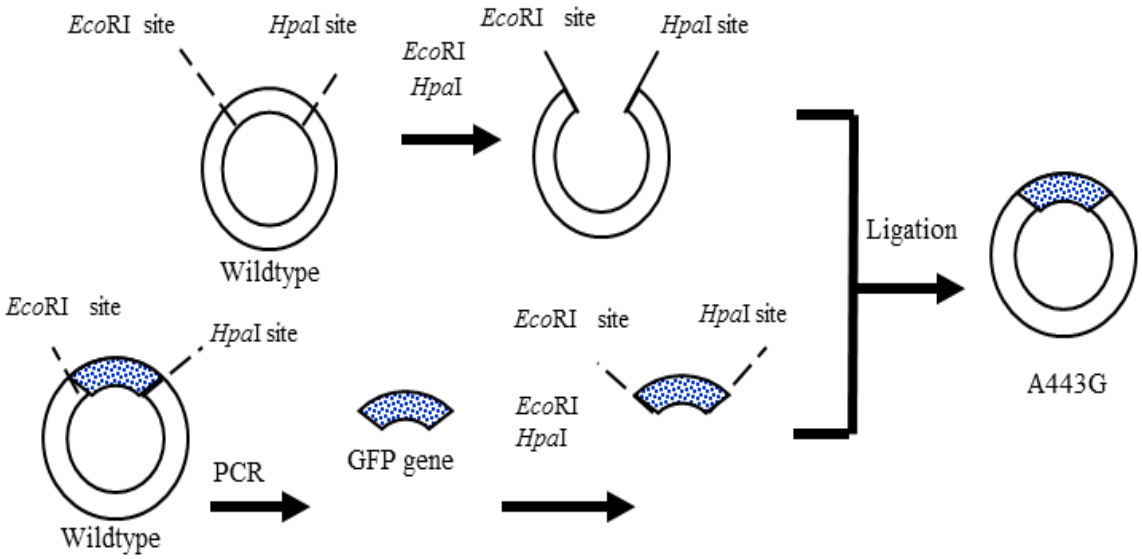


Figure 4.5 The construction of A443G

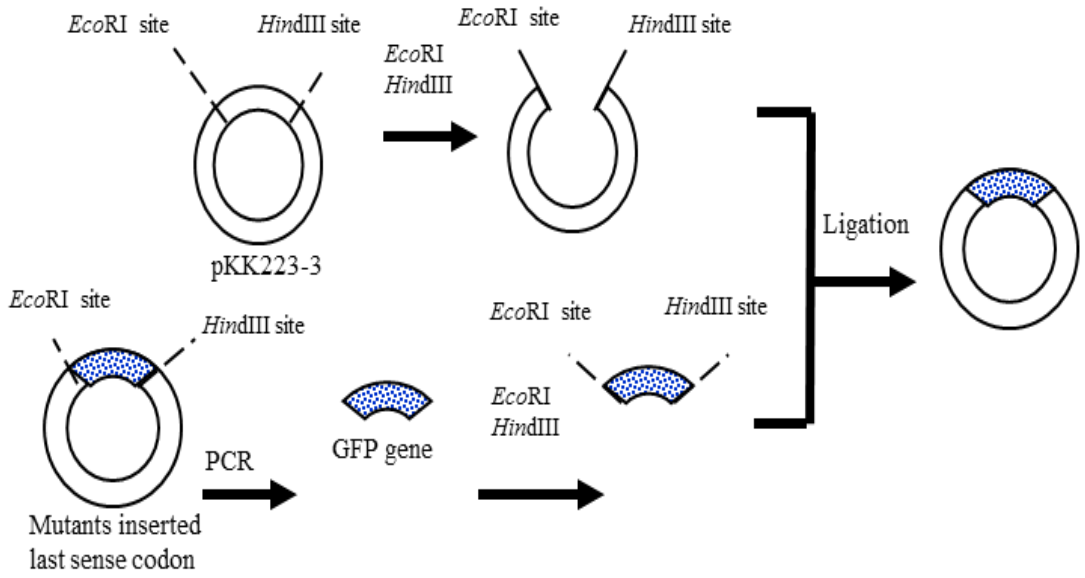


Figure 4.6 The constructions of A402G, A675C, A462G, A88G

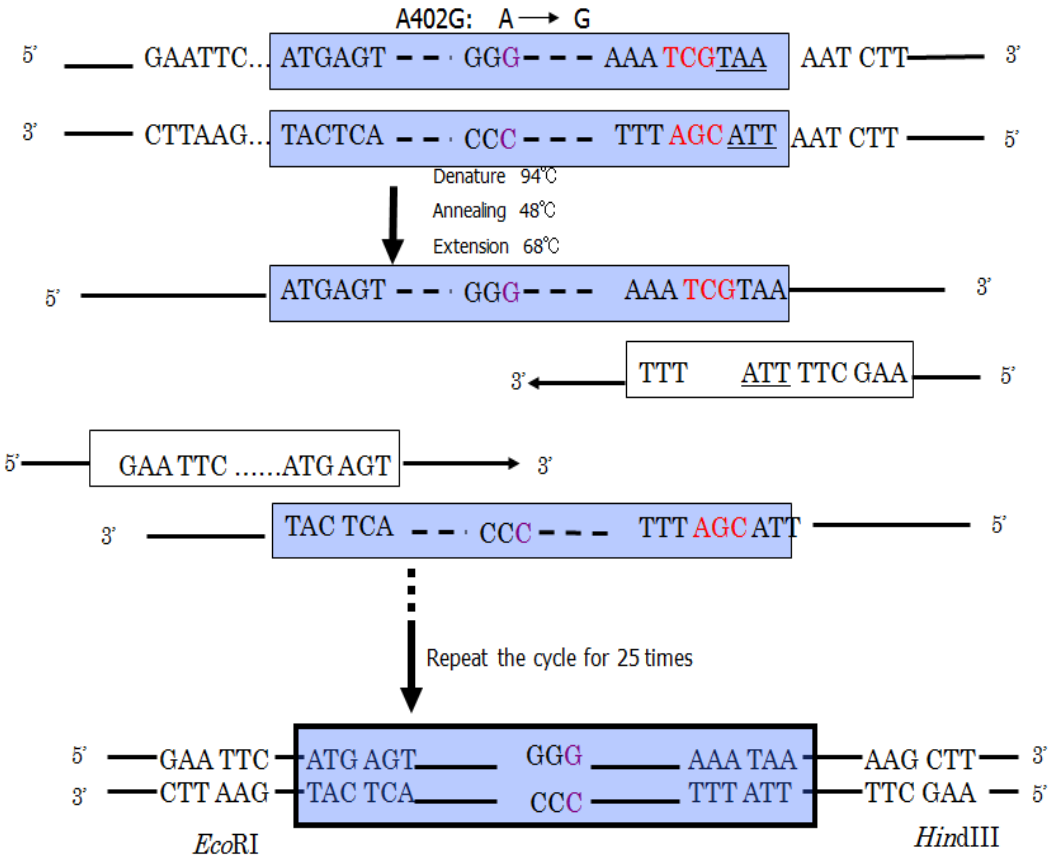


Figure 4.7 The PCR of A402G

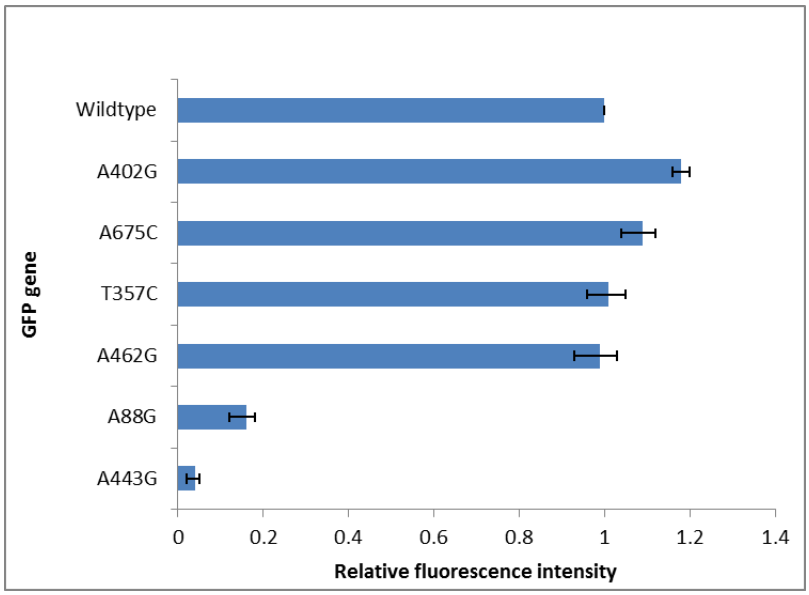


Figure 4.8 Comparison of the RFIs of mutants in *E. coli*

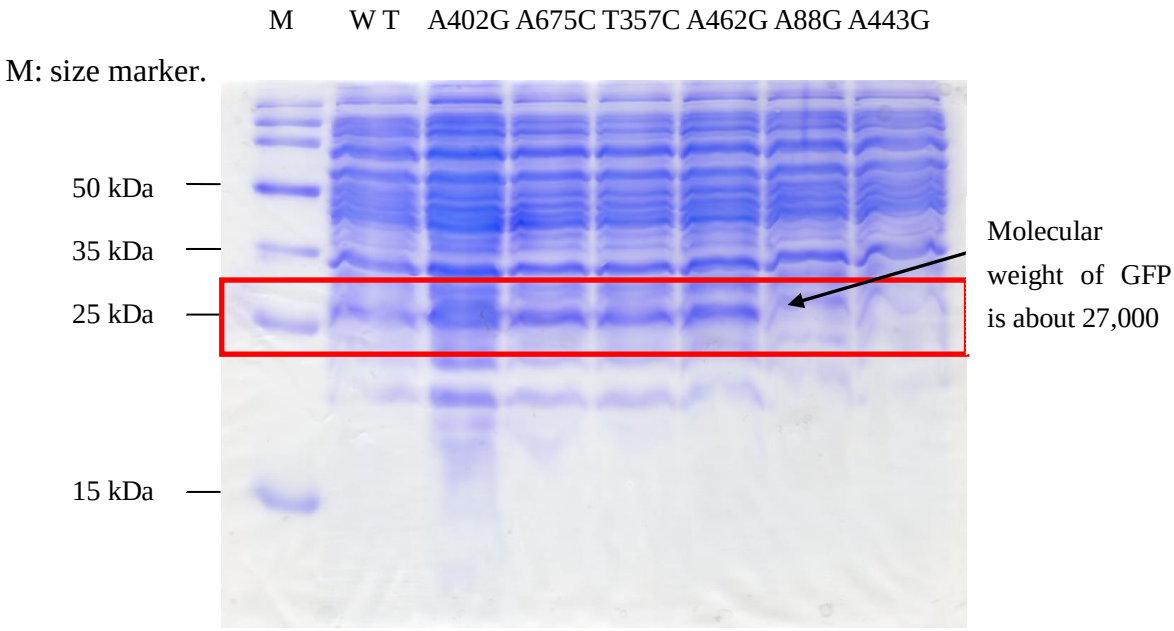


Figure 4.9 SDS-PAGE of soluble proteins extracted from *E. coli* in GFP genes

Table 4.1 20 kinds of mutants

No	Last sense codon	Mutant site	Mutant codon	RFI(Mutants)	RFI (Right GFP gene)	Change of amino acids
1	TAC	150	ACU-ACA	1.01	0.91	Thr～Thr
2	CTT（Codon G missing）	395	CUA-UUA	0.01	0.73	Leu～Leu
3	GAA	326	CGU-CAU	0.71	1.07	Arg～His
4	CAA		TAT-TAC	0.62	0.82	
			AAG-AAA			
5	AAG	Primer overlapping				
6	TCG	402	GGA-GGG	1.21	0.93	Gly～Gly
7	ACA	675	ACA-ACC	1.15	1	Thr～Thr
8	ACG	Primer overlapping		1.22	1.13	
9	TGA	30	ACA-GCA	1.48	1.06	Thr～Ala
10	CTA	462	GCA-GCG	0.99	0.33	Ala～Ala
11	ACT	88	AGU-GGU	0.14	1.09	Ser～Gly
12	TAA（Codon T missing）	630	GUA-GAC	1.31	0.94	Asn～Asn
13	AGC	M1:407	ACT-ATT	0.13	0.9	M1 Thr～Lle
		M2:493	CTC-TTC			M2 Leu～Phe
14	ACC	M1:56	GTT-GAT	0.05	1.09	M1Val～Asp
		M2:357	CTC-CTT			M2Leu～Leu
		M3:443	CGC-CAC			M3Arg～His
15	TCT	656	GUC-GCC	0.11	0.94	Val～Ala
		447	AAT-AAC			
16	GTA	632	CCC-CAC	0.23	1.06	Val～His
17	AAC	584	UUA-UCA	0.38	0.83	Leu～Ser
18	TTA	158	CUA-CCA	0.07	0.68	Leu～Pro
		223	CCA-UCA			
19	CAG	645	UUU-UCU	0.24	0.84	Phe～Ser
20	GTG	50	GGA-GAA	0.13	0.75	Gly～Glu

Table 4.2 The mutant GFP genes in the research

No	Mutant	Change of codon	Amino acids substitution
1	A402G	GGA-GGG	Nothing
2	A675C	ACA-ACC	Nothing
3	T357C	CTT-CTC	Nothing
4	A462G	GCA-GCG	Nothing
5	A88G	AGT-GGT	S30G
6	A443G	CAC-CGC	H148R

Chapter 5 Conclusions and unresolved problems

5.1 Conclusions

From the preceding chapters, it was found that gene sequence had close relationship with the expression efficiency of the GFP gene in *E. coli*. In this study, the influence of 64 kinds of last sense codons on the expression efficiency, the influence of the combination of last sense codon and stop codon on expression efficiency, the relationship between mutant gene and protein synthesis efficiency were researched (Fig. 5.1).

Based on the above researches, some conclusions were given:

(1) Gene sequence had close relationship with the expression efficiency of the GFP gene in *E. coli*. For the last sense codon, it was shown in Fig. 2.7 that by the stop codon UAA and the expression vector pKK223-3, the RFI of GFP(CCG) was 2.1, which could greatly increase the expression efficiency of GFP gene. On the contrary, the RFI of GFP(CTA) was 0.33, which could suppress the gene expression.

(2) For the same expression vector, even the stop codons were changed, some GFP genes with the last sense codons had the stronger RFIs and got the more expression quantities of proteins; some GFP genes with the last sense codons had the lower RFIs. For example, when the expression vector was pKK223-3, GFP(CCG), GFP(GGA) always had the stronger RFIs, the last sense codons GFP(GTT), GFP(TTG), GFP(CTA) always had the lower RFIs (Fig. 3.5); when the expression vector was pET21(+), GFP(TGA) always had the strongest RFI, other last sense codons had the similar RFIs (Fig. 3.7).

(3) The expression vector (pKK223-3 and pET21(+)) had the different expression efficiency. By the results, when the stop codon was UAG or UGA, for the two expression vectors, even most of GFP genes with the last sense codons had the nearly

same RFIs (Fig. 3.8, Fig. 3.9), when the stop codon was UAA, some differences were observed. For example, as shown in Fig. 3.10, when the expression vector was pKK223-3, GFP(CCG) always had the strongest RFI (2.1); but when the expression vector was pET21(+), the RFI of GFP(CCG) was 0.89.

(4) The point mutation of GFP gene could greatly influence the gene expression. The RFI of A402G was increased to 1.2. On the other hand, the RFIs of A443G and A88G were decreased to 0.04 and 0.16, respectively.

5.2 Unresolved problems

By this study, the conclusions about the influences of GFP gene on expression efficiency in *E. coli* were given. However, there were some problems which could not be resolved. Those were:

(1) Although it was confirmed that there were close relationships between gene sequence and expression efficiency of gene, there were not clear rules to comply with.

(2) Mutant gene had affected on the expression of gene. In our research, six mutant GFP genes with only one mutant site were acquired. The influences of other GFP's mutant site on expression efficiency and the influences of two or more mutant sites had affected on the expression efficiency had not been researched.

Figure

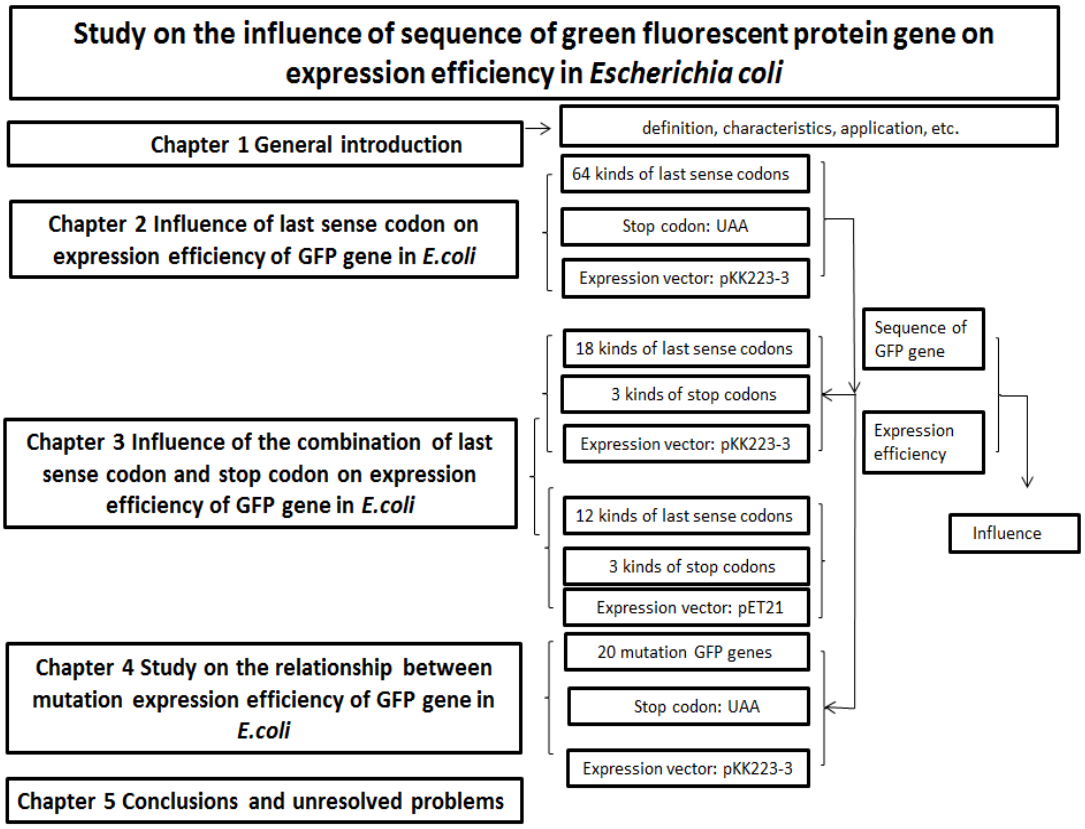


Figure 5.1 The flow chart of this study

Appendix: Abbreviation

Ampicillin (Amp)

Base pair (bp)

Chloroform- Isoamylalchohol (CIA)

Calf intestin alkaline phosphatase (CIAP)

Codon usage frequency (CUF)

Escherichia coli (*E. coli*)

Green fluorescent protein (GFP)

Isopropyl-β-D-thiogalactopyranoside (IPTG)

Molecular weight (MW)

Polymerase chain reaction (PCR)

Phenol-Chloroform Isoamyl Alcohol (PCI)

Relative fluorescence intensity (RFI)

Revolution per minute (rpm)

SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Shine-Dalgarno (SD)

Ultraviolet (UV)

Wild type (WT)

Appendix：Reagents and equipments

Reagents

○Wako Pure Chemical Industries, Ltd

Strong grade Ethanol 95 %

Ethanol 99.5 %

Chloroform

Sodium acetate trihydrate

Sodium chloride

Sodium hydroxide (granulous)

Sodium dodecyl sulfate

Tris (hydroxymethyl) aminomethane

Agar (for the use of bacterial medium)

Bacto Tryptone

Bacto Yeast Extract

○Takara Bio Inc.

DNA Ligation kit

TaKaRa Ex Taq HS

*Eco*RI

*Hin*dIII

λ*Hin*dIII

Agroase (powder)

E. coli JM109 Competent Cells

○Novagen

E. coli BL 21 Competent Cells

○SIGMA

Ampicillin

Ribonuclease A (RNaseA)

isopropyl-β-D-thiogalactopyranoside (IPTG)

○Nacalai Tesque, Inc.

0.5 M EDTA (autoclaved)

Phenol-Chloroform Isoamyl alcohol (25:24:1) Mixed, pH7.9

○Promega

Wizard SV Gel and PCR Clean-Up System

○Thermo Fisher Scientific (Applied Biosystems)

BigDye Terminator v 1.1 Cycle Sequencing Kit

Equipments

○Biometra

T3 Thermocycler (PCR)

○TOMY

High Speed Refrigerated Micro Centrifuge MX-107 and MX-150

○Thermo Fisher Scientific (Applied Biosystems)

ABI PRISM 310 Genetic Analyzer (DNA sequencer)

○TAITEC

BIO SHAKER BR-15LF (Shaker Incubator)

○SANYO

Incubator MIR-153

○Elma

Spectrophotometer AE-45

○Nihon Molecular Device

SPECTRA Max Gemini (Fluorescence microplate reader)

List of publications

[1] Xiaoliang Hao, Shota Inoue, Masahide Ishikawa. Influence of insertion of the last sense codon on expression efficiency of green fluorescent protein gene in *Escherichia coli*. *Journal of Materials Science and Chemical Engineering*., vol.3, pp13-18, 2015.

[2] Xiaoliang Hao. Influence of some point mutations of green fluorescent protein gene on expression efficiency in *Escherichia coli*. *International Journal of Chemical Engineering and Applications*., in press.

Acknowledgements

I would like to express my gratitude to all those who helped me during the writting of this thesis. My deepest gratitude goes first and foremost to my advisor, Prof. Ishikawa Masahide. Thanks for three and half years of valuable guidance, enthusiastic supervision and patient support throughout my doctoral study. Teacher's passion and attitude on research inspire me greatly. From him, I experience the true meaning of research and give me power to keep on the research in all of my life. It is unforgettable memory in my gratitude to the past and present members of Prof. Ishikawa Masahide's laboratory, I study too much knowledge and acquire many helps from them.

I woluld also like to thank Prof. Tokio Hagiwara, Prof. Takashi Kumazawa, Prof. Yasushi Hasebe, for their insightful comments and kindly advice for my work. Many thanks go to all the professors of Department of life Science and Green Chemistry in SIT for their kindess help and encouragement on my research.

Especially, I would like to express my sincere gratitude to Prof. Dong-Ying Ju for his infinite kindess and endless help no matter I was in China or came to Japan. When I meet troubles, Prof. Dong-Ying Ju always takes care of my studies and life, just like my father, I will remember teacher's help for ever.

Whatever thanks could not express my appreciation exactly. During three and half years, I acquired lots of help from the friendly Janpanese, and for other Chinese students, I always feel warm because of the existence of them. Thanks a lot.

My deepest grateful to my loving wife, parents and my lovely daughter. Without their continuously support and encourage, without me today.

Thank Saitama Institute of Technology, I am so proud of be your student, thank you very much.

Finally, it is my sincere hope that thank everyone who ever gave me help.