

Influence of the combination of last sense codon and stop codon on expression efficiency of green fluorescent protein gene in *Escherichia coli* when the expression vector pKK223-3

Xing Zhang, Aiping Fei, Yong Wang, Zhigang Fang, Yingxue Teng, Xiaoliang Hao*

University of Science and Technology Liaoning, Anshan China, 114000

Abstract. In this paper, the influence of the combination of last sense codon and stop codon on expression efficiency was studied. By our study, for the last sense codon CCG, the RFI of GFP(CCG) was 2.1 fold when the stop codon was UAA, but in comparison, the RFI was 1.1 fold 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.

1 INTRODUCTION

Green fluorescent protein (GFP) was discovered as a companion protein to aequorin, the famous chemiluminescent protein from *Aequorea* jellyfish^[1]. GFP is composed of 238 amino acids and hence a rather small protein with a molecular weight of roughly 27 kDa^[2]. Excitation at 396 nm results in an emission maximum at 508 nm^[3]. The discovery of GFP led to a new revolution in molecular biology, whose different mutants had been engineered over the last few years^[4-5]. Due to its favourable features, GFP rapidly became a popular tool in various applications in biology research. During the last decade, it has been introduced into a wide range of organisms, including bacteria, yeasts^[6-7].

Meantime, *Escherichia coli* (*E.coli*) is a convenient host for protein expression and one of the organisms of choice for the production of recombinant proteins in high quantities and low production costs, it has become the most popular expression platform. *E.coli* derived gene has become one of the most widely used. Its popularity is due to its lack of interference with plant metabolism, the ease and sensitivity of the assay and its stability both in vivo and in vitro^[8-17].

As we all know, gene contains a specific sequence of nucleotides which gives the instructions for the specific sequence of amino acids, if the gene is mutated, may change the formation of amino acids, and then influence the structure and function of the protein^[18]. To master the relationship between the gene sequence and protein expression is helpful for understanding principles of gene expression and effectively controlling the production of protein. GFP and its some mutants have many useful applications, such as real-time detection, no disruption or toxicity to the host cells, no requirement for cofactors,

and feasibility of fusion with the target proteins, have become one of the most used fluorescent probes in cell biology and molecular biology^[19-21].

Fluorescent proteins have proven to be excellent tools for live cell imaging. A mutation which results in an exchange of the amino acid serine at position 65 to threonine dramatically increased the intensity of green fluorescence^[22].

Some texts about the influence of the codon on the protein expression were greatly researched. Codon usage bias was considered to influence the elongation rate because each codon was decoded at a different rate^[23-24].

The C-terminal region contains all the information required for efficient translocation and can therefore be used as a signal sequence for recombinant protein targeting^[25]. In addition, the sequence context motif surrounding the ATG initiation codon (ATG-context) is an important factor that increases protein production^[26-32]. Introduction of the rare codons GCC, CGG, and ACC for alanine, arginine, and threonine reduced GFP production 2.1, 3.3, and 1.7 fold in comparison to the favored codons GCU, CGU, and ACA, respectively^[33].

During our former researches, my collaborators and I inserted 64 kinds of last sense codons at the 5' end of the stop codon of the GFP gene and studied on influence of last sense codon on expression in *E.coli*^[34]. In the process, in order to study the influence of the combination of last sense codon and stop codon on expression efficiency of GFP gene in *E.coli* when the expression vector pKK223-3, I changed the stop codon from UAA to UAG and UGA, selected 18 kinds of last sense codons for each stop codon, and studied the influence of the combination of last sense codon and stop codon on expression efficiency of GFP gene in *E.coli* when the expression vector pKK223-3.

* Corresponding author: Ha Xiaoliang1980@163.com

2 MATERIALS AND METHODS

2.1 Constructions of Plasmid and Cloning Vectors

In order to research the combination of last sense codon and stop codon on expression efficiency of GFP gene, 18 kinds of GFP genes with special features were selected, such as had the higher or lower fluorescence intensities, medium fluorescence intensities, the same name with the stop codon, etc. NNN was the insertion of 18 kinds of last sense codons. The stop codon was changed from UAA to UGA and UAG (Figure 1).

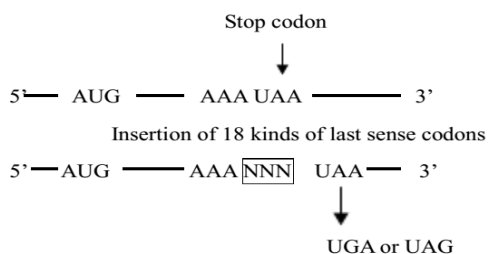


Figure 1. The insertion of last sense codons and change of stop codon in GFP gene.

pKK223-3 was used for the expression vector in this study. In order to research the relationship between the last sense codon and the protein expression efficiency, polymerase chain reaction (PCR) was used for removing these last sense codons of the modified GFP genes. PCR was performed at 94 °C for 10 min, followed by 25 cycles of 94 °C for 45 s, 48 °C for 1 min and 68 °C for 45 s. The principle of PCR was seen in Figure 2. After PCR, the plasmids were constructed by introduction of the modified GFP genes into expression vector. *E. coli* JM109 was transformed with the ligation mixture, the resulting clones were analyzed by sequencing. Primers (PCR) were purchased from Sigma-Aldrich (Tokyo, Japan). Restriction enzymes (EcoRI, HindIII), a PCR amplification kit and T4 DNA ligase had been purchased from TaKaRa (Otsu, Japan). We inserted the last sense codons by PCR, DNA fragments were acquired by enzyme digestion, and recombined by transformation and ligation. The flow diagram of experiment was seen in Figure 3.

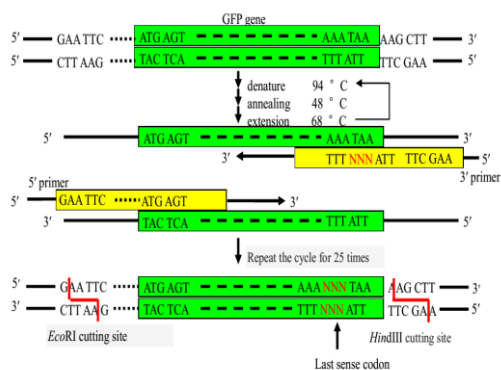
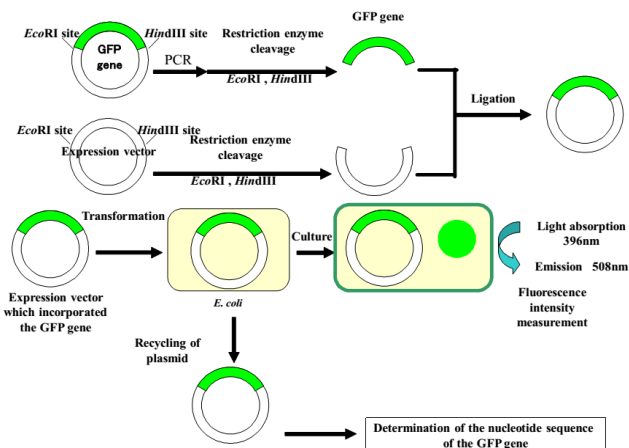


Figure 2. The principle of PCR.



When the expression vector was pKK223-3, 18 kinds of last sense codons were chosen and were inserted into GFP genes by PCR (Figure 2). The used primers of 5' site was the same :

GFP EcoRI Primer (5'Primer)
 5'-CCCGAATTCTTTAACTTTAGGAAACACAATT
 CATGAGTAAAGGAGAAGAACTT-3'

The used primers of 3' site was shown in Table 1. The NNN part was the stop codon (UAA, UGA, UAG). The shaded parts indicated restriction enzyme digested sites corresponding to the primer description.

2.2 Measurement of GFP Relative Fluorescence Intensity in *E. coli*

GFP fusion fluorescence intensity was an excellent indicator of over-expression potential^[35]. Because fluorescence was one of the most convenient ways to follow a protein expression and purification procedure^[36], so the fluorescence intensity was used to analyze the expression efficiency of proteins. The cells were cultivated in LB medium supplemented with 0.1 mg/mL ampicillin and 40 μM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37 °C for 18 h^[37]. The culture of *E. coli* was measured its absorbance at 600 nm, and the fluorescence intensity at 508 nm was excited at 396 nm of the same culture and measured 3 times by the Gemini fluorescence microplate reader (Nihon Molecular Devise). The expression efficiency of GFP gene was compared as the value of the fluorescence intensity / the absorbance at 600 nm. The calculation formula of RFI was shown by formula (1).

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

(1)

2.3 SDS-PAGE Analysis of Extracts of *E. coli* Having last sense codon of GFP Gene

After the fluorescence intensities of acquired GFP variants were known, the GFP genes were expressed in the proteins, the quantities of proteins were compared, and the influences of the last sense codon and stop codon on protein expression were analyzed. In order to be easily compared, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) experiment was done. *E.coli* JM109 was transformed by the expression vector including GFP gene, and cultured in the LB medium of 20 mL under the presence of 40 μM IPTG. DNase was added into the solution and the solution was incubated at 37 °C for 1 h to remove the remained DNA of the solution. The insoluble parts were separated from the solution by centrifugation (4°C, 13,000 rpm, 10 min) and soluble proteins were analyzed by SDS-PAGE.

3 RESULTS AND DISCUSSION

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 Figure 4.

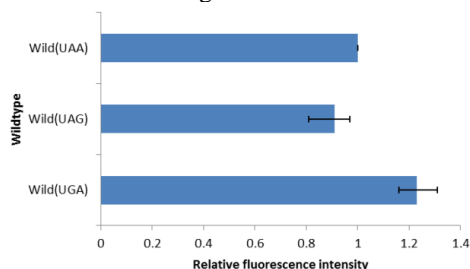


Figure 4. Comparison of the RFIs by three kinds of stop codons without insertion of last sense codon by the expression vector pKK223-3.

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

From Figure 5, it was seen that the last sense codon TCG had been inserted into GFP gene, the stop codon UAA was marked in square.

As the results (Figure 4), 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 fold when the stop codon was UAA, but in comparison, the RFI was 1.1 fold 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 fold).

The recognition of stop codons by release factors 1 (RF-1) and 2 (RF-2) might lead 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.

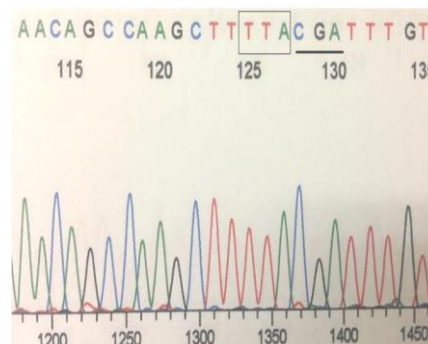


Figure 5. The insertion of last sense codon TCG.

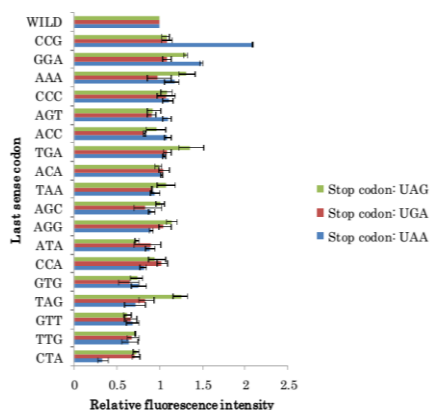


Figure 6. RFIs of GFP expressed by genes with insertion of last sense codons by the expression vector pKK223-3.

3.3 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 Figure 7 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 Figure 6, the results of RFIs were: CCG(UAA) 2.1 fold, CCG(UAG) 1.1 fold, CCG(UGA) 1.1 fold, AGT(UAA) 1.1 fold, AGT(UAG) 0.92 fold, AGT(UGA) 0.91 fold, CTA(UAA) 0.33 fold, CTA(UAG) 0.73 fold, and CTA(UGA) 0.71 fold. 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 (Figure 7).

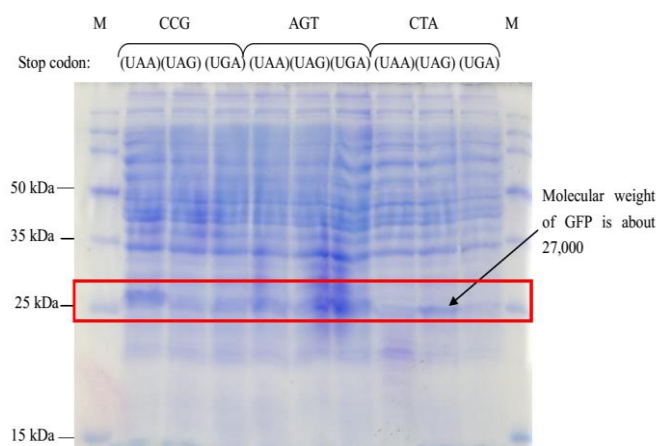


Figure7. SDS-PAGE of soluble proteins extracted from *E.coli* in GFP gene M: size marker. MW of GFP: ~ 27kDa

Conclusions

In our study, 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. 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 had influenced the expression efficiency of GFP gene. By insertion of last sense codon, we changed some fluorescence intensities of GFP variants, that meant that this way could control the increase and decrease of the protein expression efficiency. In the future research, it is expected to insert the codon from different site of GFP gene, compare the different influence of GFP variants by insertion the codon,

especially some GFP variants which could greatly improved the protein quantities, finally achieve that the protein could be expressed by artificial control.

Acknowledge

This work was financially supported by Liaoning Natural Science Foundation Project (project number: 20170520093, 20170540466), Liaoning Provincial Education Office Service Local Project (project number: 2017FWDF05).

References

1. O. Shimomura, F. H. Johnson, and Y. Saiga, "Extraction, purification, and properties of aequorin, a bioluminescent protein from the luminous hydromedusa, *Aequorea*," *J. Cell. Comp. Physiol.*, vol. 59, pp. 223-239 (1962)
2. D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier, "Primary structure of the *Aequorea victoria* green-fluorescent protein," *Gene*, vol. 111, pp. 229-233 (1992)
3. J. C. March, G. Rao, and W. E. Bentley, "Biotechnological applications of green fluorescent protein," *Appl Microbiol Biotechnol.*, vol. 62, pp. 303-315(2003)
4. A. Miyawaki and R. Y. Tsien, "Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein," *Methods Enzymol.*, vol. 327, pp. 472-500 (2000)
5. Y. Wu, Y. Zhou, J. Song, X. Hu, Y. Ding, and Z. Zhang, "Using Green and Red Fluorescent Proteins to Teach Protein Expression, Purification, and Crystallization," *Biochemistry and molecular biology education*, vol. 36, pp. 43-54 (2008)
6. J. Morschhauser, S. Michel, J. Hacker, "Expression of a chromosomally integrated, single copy GFP gene in *Candida albicans*, and its use as a reporter of gene regulation," *Mol Gen Genet.*, vol. 257, pp. 412-420, February (1998)
7. M. Chalfie, M. Y. Tu, G. Euskirchen, W. W. Ward, D.C. Prasher, "Green fluorescent protein as a marker for gene expression," *Science*, vol. 263, pp. 802-805 (1994)
8. N. H. Tolia and L. Joshua-Tor, "Strategies for protein coexpression in *Escherichia coli*," *Nature methods*, vol. 3, pp. 55-64 (2006)
9. G. L. Rosano and E. A. Ceccarelli, "Recombinant protein expression in *Escherichia coli*, advances and challenges," *Frontiers in microbiology*, vol. 5, pp. 1-17 (2014)
10. G. Hannig and S. C. Makrides, "Strategies for optimizing heterologous protein expression in *Escherichia coli*," *Trends in Biotech.*, vol. 16, pp. 54-60 (1998)
11. F. Baneyx, "Recombinant protein expression in *Escherichia coli*," *Curr Opin Biotechnol.*, vol. 10, pp. 411-421 (1999)
12. A. K. Panda, "Bioprocessing of therapeutic proteins from the inclusion bodies of *Escherichia coli*," *Adv Biochem Eng Biotechnol.*, vol. 85, pp. 43-93(2003)
13. L. A. Palomares, S. Estrada-Mondaca, and O. T. Ramirez, "Production of recombinant proteins, challenges and solutions," *Methods Mol Biol ch.2.*, pp. 15-52 (2004)
14. F. J. Mergulhao, D. K. Summers, and G. A. Monteiro, "Recombinant protein secretion in *Escherichia coli*," *Biotechnol Adv.*, vol. 23, no. 3, pp. 177-202 (2005)
15. J. Yin, G. Li, X. Ren, and G. Herrler, "Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes," *J Bacteriol.*, vol. 127, pp. 335-347 (2007)
16. R.A. Jefferson, T.A. Kavanagh, M.W. Bevan, "GUS fusions: glucuronidase as a sensitive and versatile gene fusion marker in higher plants," *EMBO J.*, vol. 6, pp. 3901-3907 (1987)
17. R.A. Jefferson, "Assaying chimeric genes in plants: the GUS gene fusion system," *Plant Mol Biol Rep.*, vol. 5, pp. 387-405 (1987)
18. G. Kudla, A. W. Murray, D. Tollervey, and J. B. Plotkin, "Coding-Sequence Determinants of Gene Expression in *Escherichia coli*," *Science*, vol. 324, pp. 255-257 (2009)

19. B. P. Cormack, R. Valdivia, and S. Falkow, "FACS-optimized mutants of the green fluorescent protein (GFP)," *Gene*, vol. 173, pp. 371-374 (1996)
20. C. Nolte, M. Matyash, T. Pivneva, C. G. Schipke, C. Ohlemeyer, U. K. Hanisch, F. Kirchhoff, and H. Kettenmann, "GFAP promoter controlled EGFP-expressing transgenic mice: a tool to visualize astrocytes and astrogliosis in living brain tissue," *Glia*, vol. 33, pp. 72-86 (2001)
21. H. J. Cha, M. Q. Pham, G. Rao, and W. E. Bentley, "Expression of Green Fluorescent Protein in Insect Larvae and Its Application for Heterologous Protein Production," *Biotechnology and bioengineering*, vol. 56, pp. 239-247 (1997)
22. R. Heim, A. B. Cubitt, and R. Y. Tsien, "Improved green fluorescence," *Nature*, vol. 373, pp. 663-664 (1995)
23. A. Fluitt, E. Pienaar, H. Viljoen, "Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis," *Comput Biol Chem*, vol. 31, pp. 335-346 (2007)
24. JF. Curran , M. Yarus, "Rates of aminoacyl-tRNA selection at 29 sense codons *in vivo*," *J Mol Biol*, vol. 209, pp. 65-77 (1989)
25. FJM. Mergulhão, DK. Summers, GA. Monteiro, "Recombinant protein secretion in *Escherichia coli*," *Biotechnol Adv*, vol. 23, pp. 177-202 (2005)
26. M. Kozak, "Point mutations close to the aug initiator codon affect the efficiency of translation of rat preproinsulin *in vivo*," *Nature*, vol. 308(5956), pp. 241-246 (1984)
27. DR. Cavener, "Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates," *Nucleic Acids Res*, vol. 15(4), pp. 1353-1361 (1987)
28. DR. Cavener , SC. Ray, "Eukaryotic start and stop translation sites," *Nucleic Acids Res*, vol. 19(12), pp. 3185-3192 (1991)
29. A. Ranjan, SE. Hasnain, "Influence of codon usage and translation initiation codon context in theAcNPV-based expression system: computer analysis using homologous and heterologous genes," *Virus Genes*, vol.9(2), pp. 149-153 (1991)
30. T. Sugio, H. Matsuura, T. Matsui, M. Matsunaga M, "Effect of the sequence context of the AUG initiation codon on the rate of translation in dicotyledonous and monocotyledonous plant cells," *J Biosci Bioeng*, vol.109(2), pp. 70-173 (2010)
31. S. Agarwal, S. Jha, I. Sanyal, DV. Amla, "Effect of point mutations in translation initiation context on the expression of recombinant human a1-proteinase inhibitor in transgenic tomato plants," *Plant Cell Rep*, vol.28(12), pp. 1791-1798 (2009)
32. L. Cherbas, P. Cherbas, "The arthropod initiator: the capsite consensus plays an important role in transcription," *Insect Biochem Mol Biol*, vol.23, pp. 81-90 (1993)
33. C. Finger, M. Gamer, S. Klunkelfuß, B. Bunk, R. Biedendieck, "Impact of rare codons and the functional coproduction of rate-limiting tRNAs on recombinant protein production in *Bacillus megaterium*," *Appl Microbiol Biotechnol*, vol. 99, pp. 8999-9010 (2015)
34. X. L. Hao, S. Inoue, and M. 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, pp.13-18 (2015)
35. J. Hammon, D. V. Palanivelu, J. Chen, C. Patel, and D. L. Minor, "A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms," *Protein Sci*, vol. 18, no. 1, pp. 121-133 (2009)
36. G. S. Waldo, B. M. Standish, J. Berendzen, and T. C. Terwilliger, "Rapid protein-folding assay using green fluorescent protein," *Nat Biotechnol*, vol. 17, pp. 691-695 (1999)
37. K. R. Siemerling, R. Golbik, R. Sever, and J. Haseloff, "Mutations that suppress the thermosensitivity of green fluorescent protein," *Current Biology*, vol. 6, pp. 1653-1663 (1996)