Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 15 (2005) 5402–5406

Selection of 5′-untranslated sequences that enhance initiation of translation in a cell-free protein synthesis system from wheat embryos

Nami Kamura,[a,b,c] Tatsuya Sawasaki,[a,b,c] Yuko Kasahara,[a] Kazuyuki Takai[a,b,c] and Yaeta Endo[a,b,c,*]

[a]*Department of Applied Chemistry, Faculty of Engineering, Ehime University, 3, Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan*
[b]*Cell-Free Science and Technology Research Center, Ehime University, 3, Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan*
[c]*Venture Business Laboratory, Ehime University, 3, Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan*

Received 26 July 2005; revised 31 August 2005; accepted 1 September 2005
Available online 5 October 2005

Abstract—Random libraries of mRNA 5′-leader sequences were screened to obtain some sequences that can stimulate the translation initiation in a cell-free translation system from wheat embryos as efficiently as the Ω sequence from tobacco mosaic virus. Several sequences that are as useful as the Ω sequence and are homologous to no known sequences survived the screening. We expect that these sequences add useful options to the cell-free protein synthesis system that is becoming a powerful tool in the post-genomic researches.

The Ω sequence from tobacco mosaic virus is one of the most frequently used mRNA 5′-leader sequences that can enhance downstream translation in in vivo and in vitro experiments.[1–7] It is a 68-nt sequence that includes three direct repeats and a C/A-rich region usually denoted by $(CAA)_n$ or poly(CAA). We have also used the sequence for the development of a high-performance cell-free protein synthesis system from washed wheat embryos,[8,9] in which mRNAs with a 5′-untranslated region (5′-UTR) with a GAA trinucleotide at the 5′ terminus followed by the Ω sequence (GAA-Ω) can be translated efficiently, as well as the capped mRNAs. The use of GAA-Ω has facilitated the production of many different proteins encoded by PCR-generated DNA fragments in parallel and large-scale synthesis of proteins encoded by a gene on a pEU plasmid.[9,10] Thus, the method is becoming a powerful tool for the post-genomic protein functional and structural analyses.[11–15]

The 5′-UTR sequence may be easier to handle if it is shorter, and an alternative to the Ω sequence for the 5′-UTR for the use in the cell-free protein synthesis system may add some flexibility to the application of the method. Thus, some sequences that can be used as substitutes for GAA-Ω would be beneficial.

It has been suggested that the function of the $(CAA)_n$ region overlaps with the cap and the poly(A) tail,[7] which is consistent with our observation that the polysomes are formed with the GAA-Ω-containing mRNAs that lacks the cap and poly(A) in our cell-free translation system.[8] The function of the tandem repeats overlaps and is weaker than the function of $(CAA)_n$.[4] These functional redundancies might mean that the sequence has survived natural situations under which a much more efficient translation is required than under normal experimental conditions. If so, many different and/or shorter sequences are expected to enhance translation initiation in the cell-free protein synthesis system. Thus, in the present study, we selected random libraries of 5′-UTR sequences with the use of the cell-free protein synthesis system.

To begin with, we prepared three DNA pools having a 22-nt stretch of B (an equimolar random mixture of T, C, and G), V (A, G, and C), or H (A, T, and C) bases upstream of a luciferase coding sequence and a 1.6 kb 3′-UTR, which could be transcribed under the control

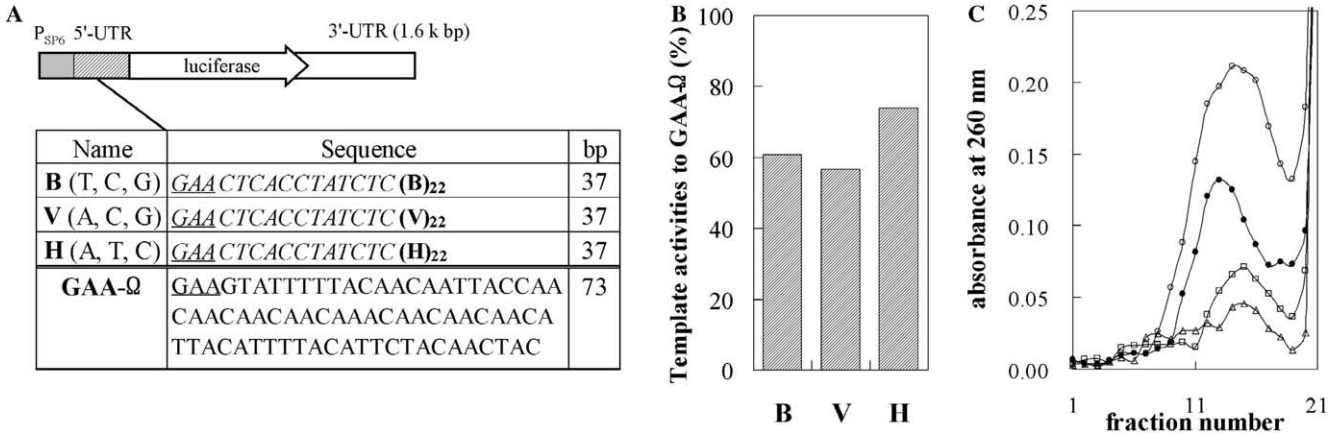
Keywords: Cell-free protein synthesis; 5′-Untranslated region; Random library; Polysome; Translational enhancer.
* Corresponding author. Tel.: +81 89 927 9936; fax: +81 89 927 9941; e-mail: yendo@eng.ehime-u.ac.jp

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.09.013

of an SP6 RNA polymerase promoter (Fig. 1A). No initiation codon should appear in the random region because at least one of the three nucleotides constituting the AUG codon is lacking. The library size is 3×10^{10}, which is much smaller than the number of molecules in the initial DNA pool. The mRNA pools produced from these DNAs have a 5′-UTR of 37 nucleotides. These were translated in the cell-free system, and the incorporation of [^{14}C]Leu into luciferase was measured (Fig. 1B). The result showed that all of the three libraries have significant template activities, and that the H library is the most efficient. The polysome profiles of these translation mixtures (Fig. 1C) also showed that the H library resulted in the formation of large polysomes with the highest efficiency.

We have not identified the reason why the H library as a whole had a much higher polysome-forming activity than the other libraries, while it is possible that the RNA molecules from the B and V libraries had a much higher probability to form secondary structures that may inhibit translation initiation.[16] We have shown that the 5′-terminal trinucleotide affects the translation efficiency with the Ω sequence, and the GAA sequence is the best.[9] We adopted the trinucleotide also in the present study for convenience, and it worked.

Then, the polysomes from the translation mixture with the H library were isolated, and the mRNA was recovered. The selected sequences were reverse-transcribed and amplified by PCR. This completed a cycle of selection (Fig. 2A). In the same way, three more selection cycles were performed. During these selection cycles, the template activity of the mRNA pools increased (Fig. 2B), and the polysome fraction during the cell-free translation grew larger (Fig. 2C). We stopped the



Name	Sequence	bp
B (T, C, G)	*GAA CTCACCTATCTC* **(B)$_{22}$**	37
V (A, C, G)	*GAA CTCACCTATCTC* **(V)$_{22}$**	37
H (A, T, C)	*GAA CTCACCTATCTC* **(H)$_{22}$**	37
GAA-Ω	GAAGTATTTTTACAACAATTACCAA CAACAACAACAAACAACAACAACA TTACATTTTACATTCTACAACTAC	73

Figure 1. Random libraries and their template activities in the wheat germ cell-free protein synthesis system. (A) The DNA pools used in this study. The B, V, and H libraries have the 5′-UTR sequence shown in the table, where '(B)$_{22}$,' '(V)$_{22}$,' and '(H)$_{22}$' represent random 22-nt stretches of the three bases other than A, T, and G, respectively. GAA-Ω was the same as in a previous study.[9] (B) Template activities of the transcripts from the B, V, and H libraries as compared with that of the transcript from GAA-Ω, determined from the incorporation of [^{14}C]Leu into luciferase during the cell-free translation essentially as described.[9] (C) Polysome profiles during the cell-free translation of the library mRNAs. The peaks appeared before fraction 19 correspond to the polysomes. Open circles show the data for GAA-Ω-containing sequence, and open squares, open triangles, and filled circles show the data for the B, V, and H libraries, respectively.

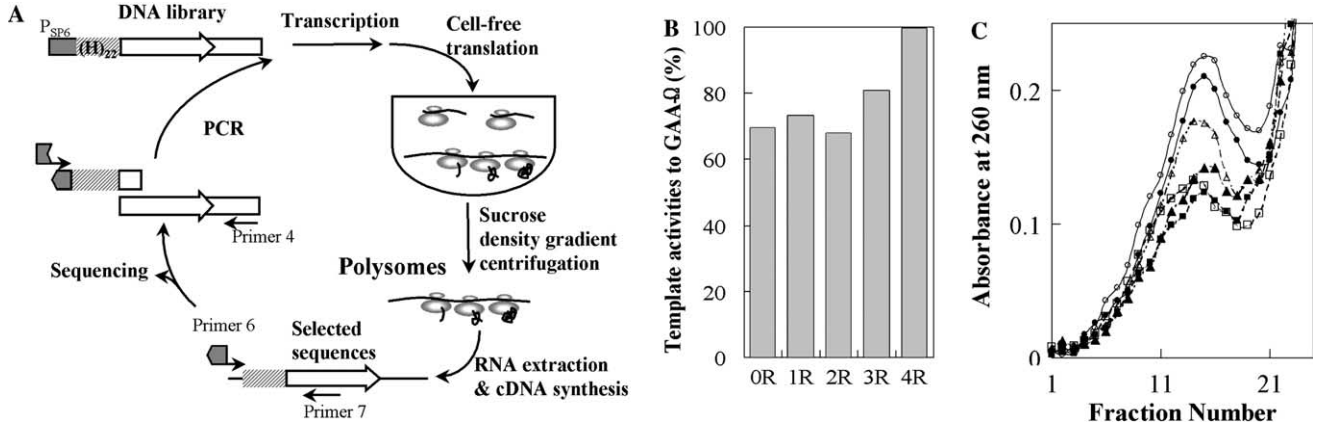


Figure 2. Polysome-mediated selection of 5′-UTRs that enhance the initiation of translation efficiently. (A) Schematic representation of the selection cycle. (B) The template activities of the mRNA pools after each selection round. (C) Polysome profiles of the 30-min translation mixture from each round. Open circles and squares represent the data for the GAA-Ω-containing sequence and the H library, respectively. Filled squares, filled triangles, open triangles, and filled circles show the profiles for the first, second, third, and fourth round pools, respectively.

selection after the 4th round because the template activity reached almost the same level as that of the GAA-Ω mRNA.

This is the first demonstration of the selection of functional sequences from a random library by use of the protein synthesis system prepared from the washed wheat embryos. This was possible probably because the extract contains only trace amounts of translation inhibitors originating from endosperm, such as tritin, thionin, and degrading enzymes. It was observed that the efficiency of protein synthesis in this system depends primarily on the efficiency of the initiation reaction, which is dependent on the sequence in the UTR of the added mRNA molecule.[9]

We sequenced 95 clones from the 4th-round pool and found that all of them were different from each other (see Supplementary material 2). Thus, a large number of different sequences survived after the selection as expected. It is obvious that a large fraction in the initial H library had high activities. The template activities of the individual clones were measured. We found here that the best clone (G11) had an activity lower than that of the GAA-Ω mRNA (Table 1, A), even though the activity of the mixture in the selected pool was almost the same as that of the GAA-Ω-containing mRNA. Therefore, we tested if the template activity could be improved by some combinations of the selected sequences, which should lengthen 5′-UTR (Table 1B–F). We found that a combination of G01 and G11 had almost the same activity as GAA-Ω (Table 1D). Therefore, some minimal length or functional redundancy may be required for the full activity. This is analogous to the case of the Ω sequence in which the two copies of the 25-nt $(CAA)_n$ elements had the full activity, while one copy does not, in an in vivo experiment.[4] The G01–G11 sequence will be denoted hereafter by E02.

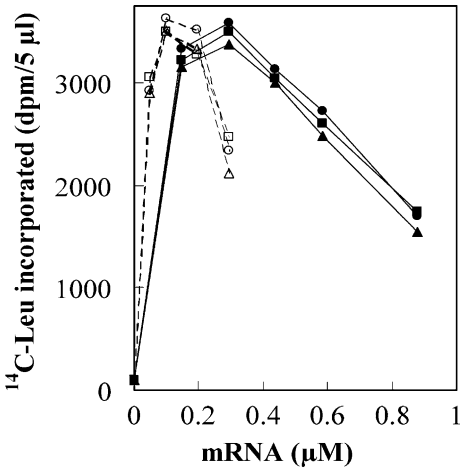


Figure 3. Incorporation of [^{14}C]Leu into GFP during 3-h cell-free syntheses as related to the added concentration of mRNA. Open symbols show the data for the capped mRNAs that were synthesized with a cap analog as described.[8] Filled symbols show the data for the mRNAs with no cap structure. Squares, GAA-Ω; triangles, E01; circles, E02.

Table 1. Relative translation enhancer activities of the selected sequences as compared to that of GAA-Ω

	Name	5′-UTR sequence	Length (nt)	Translational activities to GAA-Ω (%)
A	G11	GAACUCACCUAUCUC AUACAACUUUCAACUUCCUAUU	37	69
B	G11-E08	GAACUCACCUAUCUC AUACAACUUUCAACUUCCUAUU AUAUUAUUAACCCUUUUCAAAU	59	87
C	E08-G11	GAACUCACCUAUCUC AUAUUAUUAACCCUUUUCAAAU AUACAACUUUCAACUUCCUAUU	59	91
D	G11-G01	GAACUCACCUAUCUC AUACAACUUUCAACUUCCUAUU UCUACACAAAACAUUUCCCUAC	59	101
E	G01-G11 (E02)	GAACUCACCUAUCUC UCUACACAAAACAUUUCCCUAC AUACAACUUUCAACUUCCUAUU	59	105
F	C05-E08-G11	GAACUCACCUAUCUC UACACAUACAAUCUAAUUCCCU AUAUUAUUAACCCUUUUCAAAU AUACAACUUUCAACUUCCUAUU	81	98
G	6 (E01)	GAACUCACCUAUCUCCCCAACACC UAAUAACAUUCAAUCACUCUUUCCAC UAACCACCUAUCUACAUCACCAA	73	100
H	40	GAACUCACCUAUCUCCUAUAAACC CACCUUACCAAUCUCCACAUUCAAUA UCUCUCCCCUUACCCUCAUCACA	73	97
I	91	GAACUCACCUAUCUCCCAACACCA AUACCAACUCCACUCACCUAUCUCCA CCUCACACACACUUUUCCAUCCA	73	90

We also selected a longer library with a 57-nt random region containing A, C, and T (the H′ library). This resulted, after a 3-round selection of polysome-generating mRNA molecules, in a sequence that enhances the translation as efficiently as the GAA-Ω and E02 sequences (Table 1G–I). This time, no manual combination of the most efficient sequences was required. This also suggested that some minimal length in 5′-UTR may be required. The most effective sequence (Table 1G) will be denoted hereafter by E01.

We could not find any similarity to Ω or other known 5′-UTR sequences in the selected sequences. Thus, E02 and E01 may be the first non-natural sequences that stimulate downstream translation in a plant system.

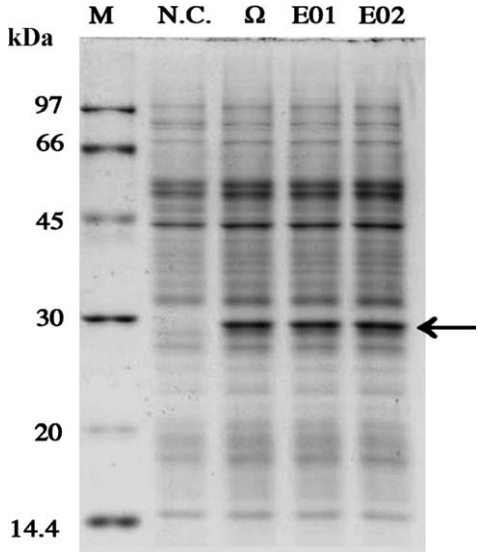


Figure 4. An SDS gel showing the products of a large-scale 48-h cell-free protein synthesis. The mRNA was prepared by transcription of the plasmid harboring each 5′-UTR sequence upstream the GFP gene. Transcription and translation in the dialysis mode was performed as described previously.[9] The arrow shows the bands of GFP.

The optimal range of the concentration of the GAA-Ω-containing mRNAs in the translation mixture is generally much broader than in the case of the capped mRNAs.[9] This property is essential for the parallel production of many different proteins from cDNA libraries because it minimizes the effort to measure the concentrations of the transcription products before adding them to the translation reactions. Thus, we tested the concentration-dependence of [^{14}C]Leu incorporation into a protein with the E01 and E02 mRNAs (Fig. 3). The result showed that the dependences are almost the same as that for the GAA-Ω mRNA. Therefore, these sequences have the prerequisite for the high-throughput protein production.

Then, we constructed plasmid vectors harboring the E01 and E02 sequences (pEUE01 and pEUE02, respectively) similar to pEU with the GAA-Ω sequence.[9] These were used to synthesize green fluorescent protein by a dialysis method, and the products were analyzed on an SDS gel (Fig. 4). The result showed that both of these plasmids are as useful as the original pEU in the large-scale protein production.

It was also revealed that the sequence of the 3′-UTR was not important in the cell-free translation with the GAA-Ω-containing mRNAs for the high efficiency of translation, while the length should be larger than 1.6 kb.[9] It has been considered that this feature of the GAA-Ω sequence is advantageous because a DNA fragment that can be used to obtain a sufficiently active mRNA preparation through in vitro transcription could be synthesized by a PCR with a cDNA plasmid used for DNA sequencing. Thus, we tested if the E01 and E02 sequences are also compatible with different sequences with the length of about 1.6 kb as the 3′-UTR (Table 2A). The results showed that the sequences other than that from pEU could also be used.

Table 2. Relative translational activities of different mRNAs

3′-UTR	Length (nt)	Relative activities to GAA-Ω (%)	
		E01	E02
A. Luciferase mRNAs with different 3′-UTRs			
pT7-Blue 96-1696	1600	104	109
pT7-Blue 1383-96	1600	105	115
pEU 814-2467	1653	96	107
Protein	Molecular weight	Relative activities to GAA-Ω (%)	
		E01	E02
B. mRNAs encoding different proteins			
Homo sapiens			
Neuron-specific γ-2 enolase	47,266	99	100
ζ-Crystallin/quinone reductase	35,205	98	101
Enolase 3	46,956	104	111
Arabidopsis thaliana			
Flowering locus T	19,808	99	97
Flowering locus F	21,864	98	100
Chlorophyllase 2	34,902	99	110

With the GAA-Ω sequence, the mRNA templates could be produced from cDNA libraries by a split-primer PCR method, which facilitates parallel production of many different proteins. We now performed the split-primer procedure with the use of the selected sequences (Table 2B). The results showed that E01 and E02 had the same performance as GAA-Ω, no matter what is the ORF sequence.

In conclusion, E01 and E02 can be used as substitutes for GAA-Ω: they can be used both for the PCR-based high-throughput protein synthesis and for the large-scale massive protein production. It was also demonstrated that our cell-free protein synthesis system could be applied to a screening of functional molecules from random libraries. We expect that the selected sequences will be used to explore the world of proteins and that the screening method will be utilized for seeking out new functions of biological macromolecules.

Acknowledgments

This work was in part supported by the Special Coordination Funds for Promoting Science and Technology by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Supplementary data

The Materials and methods in the present study are provided online in Supplementary material 1. A list of the 95 sequences selected after the 4th-round selection is also provided in Supplementary material 2. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2005.09.013.

References and notes

1. Jobling, S. A.; Gehrke, L. *Nature* **1987**, *325*, 622.
2. Gallie, D. R.; Sleat, D. E.; Watts, J. W.; Turner, P. C.; Wilson, T. M. A. *Nucleic Acids Res.* **1987**, *15*, 8693.
3. Gallie, D. R.; Sleat, D. E.; Watts, J. W.; Turner, P. C.; Wilson, T. M. A. *Nucleic Acids Res.* **1987**, *15*, 3257.
4. Gallie, D. R.; Walbot, V. *Nucleic Acids Res.* **1992**, *20*, 4631.
5. Tanguay, R. L.; Gallie, D. R. *J. Biol. Chem.* **1996**, *271*, 14316.
6. Wells, D. R.; Tanguay, R. L.; Le, H.; Gallie, D. R. *Genes Dev.* **1998**, *12*, 3236.
7. Gallie, D. R. *Nucleic Acids Res.* **2002**, *30*, 3401.
8. Madin, K.; Sawasaki, T.; Ogasawara, T.; Endo, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 559.
9. Sawasaki, T.; Ogasawara, T.; Morishita, R.; Endo, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14652.
10. Sawasaki, T.; Hasegawa, Y.; Tsuchimochi, M.; Kamura, N.; Ogasawara, T.; Kuroita, T.; Endo, Y. *FEBS Lett.* **2002**, *514*, 102.
11. Morita, E. H.; Sawasaki, T.; Tanaka, R.; Endo, Y.; Kohno, T. *Protein Sci.* **2003**, *12*, 1216.
12. Endo, Y.; Sawasaki, T. *Biotechnol. Adv.* **2003**, *21*, 695.
13. Endo, Y.; Sawasaki, T. *J. Struct. Funct. Genom.* **2004**, *5*, 45.
14. Sawasaki, T.; Hasegawa, Y.; Morishita, R.; Seki, M.; Shinozaki, K.; Endo, Y. *Phytochemistry* **2004**, *65*, 1549.
15. Vinarov, D. A.; Lytle, B. L.; Peterson, F. C.; Tyler, E. M.; Volkman, B. F.; Markley, J. L. *Nat. Methods* **2004**, *1*, 149.
16. Lawson, T. G.; Ray, B. K.; Dodds, J. T.; Grifos, J. A.; Abramsoni, R. D.; Merricki, W. C.; Betsch, D. F.; Weithll, H. L.; Thach, R. E. *J. Biol. Chem.* **1986**, *261*, 13979.