Thesis for the Degree of Master of Science

Functional implication of tRNA genes encoded in the PBCV-1 genome

by

Da young Lee

Department of Microbiology

The Graduate School

Pukyong National University

February 2006

Functional implication of tRNA genes encoded in the PBCV-1 genome

(PBCV-1 genome에서의 tRNA gene의 기능적 의미)

A thesis submitted in partial fulfillment of the requirements

for the degree of

Master of Science

in Department of Microbiology, The Graduate School,

Pukyong National University

February 2006

이다영의 이학석사 학위논문을 인준함.

2006년 2월 24일

주 심 이학박사 김 영 태 인

위 원 이학박사 송 영 환 인

위 원 이학박사 최 태 진 인

Functional implication of tRNA genes encoded in the PBCV-1 genome

A dissertation

by

Da young Lee

Approved by :

Chairman : Young Tae Kim, Ph. D.

Member : Young Hwan Song, Ph. D.

Member : Tae Jin Choi, Ph. D.

Member :

Member :

February 24, 2006

CONTENTS

Abstract ····· 1

Introduction ····· 2

Material and Method ····· 4

Virus culture and purification ····· 4

Extraction of viral genomic DNA ····· 4

PCR amplification and cloning of the a189/192r gene ····· 5

Expression of recombinant protein ····· 8

Western blot ····· 8

Result and Discussion ····· 10

Expression of recombinant protein in *E. coli* BL21(DE3) SI strain ····· 10

Expression using *E. coli* BL21 CodonPlus (DE3) RIL strain ····· 13

Comparison of codon usage between PBCV-1 and chlorella NC64A ····· 17

Codon usage in the major capsid protein and DNA polymerase genes ····· 20

Implications for adaptation in different hosts ····· 23

PBCV-1 에서의 tRNA gene의 기능적 의미(국문 초록) ····· 24

감사합니다! ····· 25

Reference ····· 27

Functional implication of tRNA genes encoded in the PBCV-1 genome

Da young Lee

Department of Microbiology, The Graduate School,
Pukyong National University

Abstract

The prototype chlorella virus PBCV-1 encodes 11 tRNA genes and over 350 protein-encoding genes in its 330kbp genome. Initial attempts to overexpress the recombinant A189/192R protein, a putative virus attachment protein, in *E. coli* strain BL21(DE3) SI were unsuccessful, and multiple protein bands were detected on Western blots. However, the full-length A189/192R recombinant protein or fragments derived from it were detected when they were expressed in *E. coli* BL21 CodonPlus (DE3) RIL, which contains extra tRNAs. Codon usage analysis of the a189/192r gene showed highly biased usage of the AGA and AUA codons compared to genes encoded by *E. coli* and chlorella. In addition, there were biases of XXA/U (56%) and XXG/C (44%) in the codons recognized by the viral tRNAs, which correspond to the codon usage bias in the PBCV-1 genome of XXA/U (63%) over those ending in XXC/G (37%). Analysis of the codon usage in the major capsid protein and DNA polymerase showed preferential usage of codons that can be recognized by the viral tRNAs. The Asn (AAC) and Lys (AAG) codons whose corresponding tRNA genes are duplicated in the tRNA gene cluster were the most abundant (i.e., preferred) codons in these two proteins. The tRNA genes encoded in the PBCV-1 genome seem to play a very important role during the synthesis of viral proteins through supplementing the tRNAs that are frequently used in viral proteins, but are rare in the host cells. In addition, these tRNAs would help the virus to adapt to a wide range of hosts by providing tRNAs that are rare in the host cells.

Introduction

Paramecium bursaria chlorella virus (PBCV-1) is the prototype of large, icosahedral, plaque-forming dsDNA viruses belonging to the family *Phycodnaviridae*. This virus infects an exsymbiotic, unicellular, eukaryotic chlorella-like green algae, NC64A (Van Etten, 2003). Its genome size is 330 kbp, which is about 60% that of the smallest known microorganism genome, *Mycoplasma genitalium* (580 kb) (Fraser *et al.*, 1995). The viral genome of chlorella virus PBCV-1 contains about 370 protein-encoding genes and 11 tRNA genes (Li *et al.*, 1995, 1997; Lu *et al.*, 1995, 1996; Kutish *et al.*, 1996). The viral genome encodes various proteins, including restriction/modification enzymes, transcription factors, translation factors, topoisomerase, chitinase, hyaluronan synthase, and enzymes that act on sugars and lipids (Xia *et al.*, 1986; Zhang *et al.*, 1998; Sun *et al.*, 1999; Graves *et al.*, 1999; Lavrukhin *et al.*, 2000). Van Etten (2003) showed that about 50% of over 370 PBCV-1 gene products could be identified and some seem irrelevant to virus replication.

A remarkable feature of the PBCV-1 genome is that it contains a cluster of 11 tRNA genes. One of the characteristics that distinguish viruses from other organisms is that viruses depend on the host machinery for protein synthesis, including ribosomes and tRNAs. Genes encoding tRNA have also been found in murine gamma herpesvirus 68 (Bowden *et al.*, 1997), bacteriophages T4s and T5 (Desai *et al.*, 1986; Calender, 1988), mycobacteriophage D29 (Ford *et al.*, 1998), phage 933W (Plunkett *et al.*, 1999; Kanjo and Inokuchi, 1999), and Streptomyces phage phiC31 (Smith *et al.*, 1999). Kunisawa suggested (2002) that these phage tRNA genes exist for their supplementary role in the efficient synthesis of phage protein. A tRNA gene cluster has also been found in the genome of Japanese chlorella virus strain CVK2 (Yamada *et al.*, 1993; Nishida *et al.*, 1998). Nishida *et al.* (1999) found that tRNA genes are generally found in chlorella viruses, and they suggested that the tRNA genes of chlorella viruses are involved in viral protein synthesis and overcome the codon usage barriers between host and virus. Chlorella virus strains KH-1, KH-2, SS-1, and SS-2 isolated in Korea also encode 14 to 16 tRNA genes in their genomes (Cho *et al.*, 2002).

Two of the PBCV-1 proteins, A140/145R and A189/192R, are thought to be the attachment proteins of the mature virions. Micrographs of PBCV-1 attaching to the chlorella cell wall by hair-like fibers suggest that the tips of these hair-like fibers are responsible for the initial recognition attachment of the virus to the host receptor (Van Etten *et al.*, 1991). A plaque inhibition assay with monoclonal antibodies that do not react to the major capsid protein of PBCV-1, followed by Western blot analysis and mass spectrometric analysis of the positive protein bands identified the proteins encoded by the *a140/145r* and *a189/192r* genes of the PBCV-1 genome as possible attachment proteins (unpublished data).

As the first step in characterizing A189/192R protein, the production of recombinant A189/192R protein was attempted in *Escherichia coli*. Although there were no nucleotide sequence changes, multiple bands were identified in Western blot analysis with antiserum against the histidine tag attached to the N-terminus of the recombinant protein (unpublished data).

In order to verify whether the fragmentation was due to degradation of expressed protein or incomplete translation, full length or overlapping fragments were expressed in *E. coli* strain that contains extra tRNA genes. The codon usages in *E. coli* genome, PBCV-1 genome, host chlorella NC64A, and that of A189/192R protein were compared to find the roles of tRNA genes encoded in the PBCV-1 genome.

Material and Method

Virus culture and purification

Chlorella strain NC64A was cultured in modified Bold's basal medium (MBBM), as described (Van Etten *et al.*, 1983). One hundred milliliters of actively growing chlorella strain NC64A were inoculated with virus at a multiplicity of infection (MOI) of 0.01 and incubated until completely lysed. The lysate was centrifuged in a Sorvall GS-3 rotor at 5,000 rpm for 5 min at 4°C. Triton-X100 was added to the supernatant at a 0.1% final concentration and stirred for 20 min at 4°C. The virus particles were pelleted by centrifugation in a Sorvall T-880 rotor at 20,000 rpm, for 60 min. The pellet was suspended in 50 mM Tris-HCl (pH 7.8) and centrifuged through a 10-40% discontinuous sucrose gradient (20,000 rpm, 20 min, 4°C). The virus band was collected from the 30-40% interface, and then pelleted for 3 hrs at 27,000 rpm with a T-880 rotor. The pellet was resuspended in 50 mM Tris-HCl (pH 7.8) (Van Etten *et al.*, 1983).

Extraction of viral genomic DNA

Purified virus was mixed with 10 TEN (100 mM Tris-HCl (pH 7.4), 10 mM EDTA, 1 M NaCl) buffer (60 μl), 1% Na-sarcosyl (60 μl), 60% (w/w) CsCl (0.6 ml), and a trace amount of EtBr. After heating at 75°C for 15 min, the mixture was loaded on a preformed 40~60% (w/w) CsCl gradient and centrifuged in a Sorvall TH-641 rotor at 3,500 rpm, for 18 hrs, at 25°C (Van Etten *et al.*, 1981).

PCR amplification and cloning of the a189/192r gene

The A189/192R protein was initially thought to be two proteins made by two adjacent genes, but later reconfirmation of its sequence clarified that the a189/192r ORF encodes one protein of molecular weight 130 kDa. The full length or part of the a189/192r ORF was cloned into pMAL-2Cx (New England Biolabs, USA) and pET23a vector (Novagen, Germany) for recombinant protein expression in *Escherichia coli*. The primers used for PCR amplification are listed in Table 1. The full-length a189/192r ORF was amplified using primer set 189FECO/189RXHO with purified genomic DNA (0.5 μg) as template. The PCR reaction was performed with a KOD Hot Start kit (Novagen) and the reaction conditions were as follows: an initial 5 min denaturation at 95°C, 40 cycles of 1 min at 95°C, 1 min at 50°C, and 1.5 min at 72°C, followed by a 7 min final extension at 72°C. The PCR products were confirmed on an agarose gel, digested with BamHI and HindIII, and cloned into the pMAL-2Cx and pET23a vectors.

The N- and C-terminal halves of the a189/192r ORF were amplified using the 189FECO/189NTR and 192FECO/189RXHO primer sets, respectively, and cloned into pMAL-2Cx. In addition, the a189/192r gene was cloned into four different overlapping clones (Fig. 1). The fragment 1 clone was obtained from a clone containing the N-terminal half of the a189/192r gene by digestion with BglII and XhoI, and subsequent cloning into pGEX-5T-1 vector (Amersham Bioscience, USA). Fragments 2, 3, and 4 were amplified from the genomic DNA with primer sets 189N2ECO/189NTR, 189CTECO/192CXHO, and 192CTECO/189RXHO, respectively. The PCR reaction conditions were as follows: an initial 2 min denaturation at 94°C, 40 cycles of 45 sec at 94°C, 30 sec at 50°C, and 4 min at 68°C, followed by a final 10 min extension at 68°C. The PCR products were analyzed by electrophoresis in 1% agarose gels and cloned into expression vectors. The PCR products for the N- and C-terminal halves were digested with BamHI/PstI and BamHI/HindIII, respectively, and cloned into appropriately digested pMAL-2Cx vector. The PCR products for fragments 2, 3, and 4 were cloned into pGEX-5T-1 vector digested with EcoRI and XhoI.

Table. 1. Oligonucleotide primers used for the PCR of a189/192r gene

primer	sequence
189FECO	5'- CTATATACTAAGAATTCATGTCAAGCTC -3'
189RXHO	5'- GGGTGCACTCGAGTTCGTAATTGC -3'
189NTR	5'- CTTCACTCGAGCGTCTCTACC -3'
192FECO	5'- CAATCTTCAAGAATTCGAGGAAGTGTAC -3'
189N2ECO	5'- CACCAATAGAATTCTATATGCCATACG -3'
189CTECO	5'- ACACAGGAATTCAACGTTCTAAAG -3'
192CTECO	5'- CGGCTGACGAATTCGTCGGAAAGCTC -3'
192CXHO	5'- GATATCAACTCGAGCTTTCTTGAACAG -3'

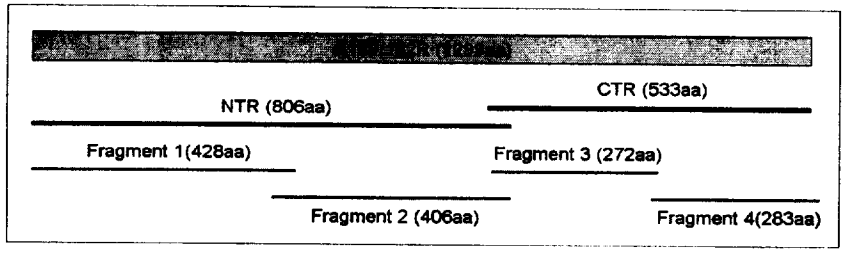


Fig. 1. Schematic description of the a189/192r genes of PBCV-1 and the relative locations of the fragments expressed in *E. coli*. The relative locations and sizes of the fragments are shown with the total number of amino acid residues in parenthese.

Expression of recombinant protein

The ligate was transformed into *E. coli* DH5α strain. The nucleotide sequence of the inserted DNA were confirmed by sequencing. Plasmid DNA was then transformed into *E. coli* strains BL21(DE3) SI (Invitrogen, USA) and BL21 codonPlus (DE3) RIL (Stratagene, USA).

To express the A189/192R protein from pMAL-2Cx vector, a single colony was inoculated into 3 ml of LBON (Luria Bertani medium without sodium) broth containing ampicillin (50㎍/ml) and cultured for 4 hrs at 25℃. When the cells reached OD600=0.5, 300㎕ of 3M NaCl were added and cultured 4 hours.

To express the A189/192R protein from pET23a vector *E. coli* BL21(DE3), a single colony was inoculated into 10 ml of double yeast tryptone (DYT) broth containing ampicillin (50㎍/ml) and chloramphenicol (50㎍/ml) and cultured at 37℃ until reaching OD600=1.0. The cells were induced with IPTG at a final concentration of 1mM and cultured for 4 hours. *E. coli* cells containing the four fragments from the a189/192r gene were cultured in DYT broth containing ampicillin (50㎍/ml), and the recombinant proteins were induced as described above.

cultured cell were collected by centrifugation, and the pellets were resuspended in sample loading buffer (1 mM EDTA, 250 mM Tris-HCl (pH 6.8), 4% SDS, 2% β-mercaptoethanol, 0.2% bromophenol blue, 50% glycerol), boiled at 100℃ for 5 min, chilled in ice, and subjected to 12% SDS-PAGE.

Western blot

The recombinant proteins expressed from pMAL-2Cx vector were detected by Western blot analysis. After electrophoresis, the separated protein bands were transferred onto nitrocellulose membrane electrophoretically. The membrane was inoculated in TTBS (0.8% NaCl, 0.2% KCl, 20 mM Tris-HCl (pH 7.4), 0.05% Tween-20) containing 5% skim milk

for 30 min, and then reacted with primary antibody for 1.5 hrs. Monoclonal antibodies against maltose binding protein (MBP), His-taq, and glutathione-S-transferase (GST) were used as the primary antibodies to detect the recombinant protein expressed from the pMAL-2Cx, pET23a, and pGEX-5T-1 vectors, respectively. The membrane was washed three times for 5 min in TTBS, and incubated with peroxidase-conjugated anti-mouse IgG (1:30,000) for 1 hr. After three 5 min washes, the protein bands were detected using a chemiluminescent detection kit (Pierce, USA).

Result and Discussion

Expression of recombinant protein in *E. coli* BL21(DE3) SI strain

The full-length a189/192r gene was cloned into pMAL-2Cx and pET23a vectors, resulting in clones pMAL189/192 and pET189/192, respectively. The N- and C-terminal halves were cloned into *E. coli* expression vector pMAL-2Cx, resulting in clones pMAL189NTR and pMAL189CTR, respectively. The sequence of the insert was confirmed, and the purified plasmids were introduced into *E. coli* BL21(DE3) SI strain for recombinant protein expression. The expressed protein was detected using Western blot analysis and the result is shown in Fig. 2.

The molecular weight of MBP in pMAL-2Cx vector was 42 kDa, while the full-length A189/192R, N-terminal half, and C-terminal half had molecular weights of 144, 88, and 60 kDa, respectively. Therefore, recombinant proteins with molecular weights of 130, and 102 kDa were expected from these clones. Although proteins of the expected molecular weights were detected, multiple smaller protein bands were also detected from all of the constructs (Fig. 2).

There are several reasons for the appearance of multiple bands. One possibility is the degradation of expressed recombinant proteins within *E. coli*. Misfolded proteins resulting from nonsense or missense mutations, mistakes in translation, or gene fusion, or that fail to associate with other proteins, are degraded via the energy-dependent proteolytic pathway (Goldberg, 1992; Hershko and Ciechanover, 1998). It is also possible that incomplete translation of the recombinant proteins occurred. During the elongation step of translation, the ribosome may pause because of rare codons, a limited supply of certain aminoacyl-tRNA species, or the formation of stable structures in certain regions of the mRNA, which can result in incomplete protein synthesis. One problem with the expression of recombinant proteins in *E. coli* occurs when the codon usage in the recombinant gene differs from the codon usage in the host cells. High-level expression of a gene with codons that are rarely used by *E. coli* depletes the internal tRNA pools and can result in incomplete translation. This problem has been thoroughly documented for the arginine codons AGA and AGG, which are the rarest codons in *E. coli* (Chen and Inouye, 1994). In addition, the

codons for arginine (CGA), isoleucine (AUA), leucine (CUA), and proline (CCC) can affect the amount and quality of protein produced in *E. coli* hosts (Deana *et al.*, 1998; Jiang *et al.*, 2001). *E. coli* strains with extra tRNAs have been developed to solve this problem (Carstens *et al.*, 2002).

In order to determine whether the appearance of multiple bands is due to a codon usage difference, the constructs of the a189/192r gene were transformed into *E. coli* BL21 CodonPlus (DE3) RIL strain (Stratagene, USA).

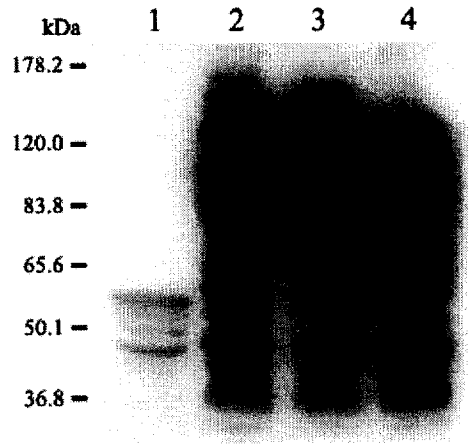


Fig. 2. . The expression of recombinant A189/192R protein from *E. coli* BL21(DE3) strain. Lane 1, pMAL-2Cx vector; lane 2, the full-length A189/192R protein from pMAL189/192; lane 3, N-terminal region of A189/192R protein from pMAL189NTR; lane 4, C-terminal region of A189/192R protein from pMAL189CTR.

Expression using *E. coli* BL21 CodonPlus (DE3) RIL strain

Escherichia coli BL21 CodonPlus (DE3) RIL strain is a modified form of *E. coli* BL21(DE3) that possesses extra copies of the tRNA genes argU (AGA, AGG), ileY (AUA), and leuW (CUA), which encode the tRNAs that most frequently limit the translation of heterologous proteins in *E. coli*. Therefore, they are suitable for the high-level expression of proteins affected by rare codon usage. *E. coli* BL21 CodonPlus (DE3) RIL strain has been used for the efficient expression of various proteins that are difficult to express in *E. coli*, such as recombinant peanut allergens (Kleber-Janke and Becker, 2000), delta-endotoxins from *Bacillus thuringiensis* in plants (Kumar *et al.*, 2005), and tobacco anionic peroxidase (Hushpulian *et al.*, 2003).

In contrast to the multiple recombinant protein bands detected from the *E. coli* BL21(DE3) SI strain, a single protein band of about 150 kDa was detected on Western blot analysis of the *E. coli* BL21 CodonPlus (DE3) RIL strain, transformed with the pET189/192 clone (data not shown). This suggested that the multiple bands seen with the *E. coli* BL21(DE3) SI strain result from incomplete translation. This was confirmed with the clones containing fragments of the a189/192r ORF, and the results for fragments 1 and 3 are shown in Fig. 3. Although several protein bands were detected from both *E. coli* strains, proteins of the expected molecular weight were detected from the BL21 CodonPlus (DE3) RIL strain. In particular, more of the full-length molecular weight protein and fewer smaller protein bands were detected from fragment 3 (lanes 4 and 5).

The difference between *E. coli* strains BL21 CodonPlus (DE3) RIL and BL21(DE3) SI is that the former contains extra tRNA genes, while the latter does not. Therefore, the results shown in Fig. 3 indicate that the multiple bands detected from *E. coli* BL21(DE3) SI strain result from incomplete translation because of a codon usage difference rather than the degradation of the expressed protein. In order to clarify this, the codon usage difference between the a189/192r gene and *E. coli* was compared. As shown in Fig. 4, the codons for Arg (AGA, AGG), Ile (AUA), and Leu (CUA) that are rarely used in *E. coli* and are present in BL21 CodonPlus (DE3) RIL strain as extra copies are preferentially used for the a189/192r gene. The codon usage patterns of *E. coli* genes are closely related to tRNA abundance (Kleber-Janke and Becker, 2000). Therefore, we

concluded that the production of multiple protein bands from the a189/192r gene in strain BL21(DE3) SI resulted from an incomplete translation codon usage difference between the viral protein and host bacteria.

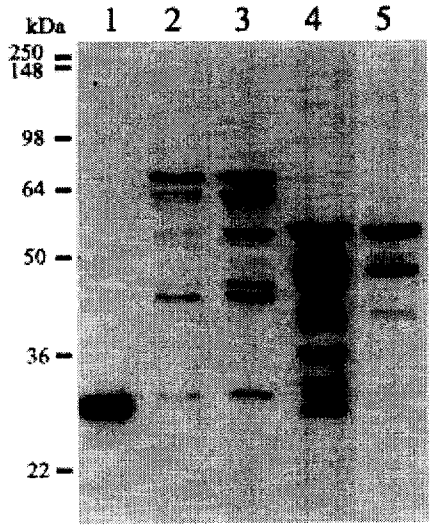


Fig. 3. The expression of fragments 1 and 3 of A189/192R protein cloned in pGEX-5X-1 vector. Lane 1, pGEX-5X-1 vector; lane 2, fragment 1 in BL21(DE3) SI strain; lane 3, fragment 1 in BL21 CodonPlus (DE3) RIL strain lane 4, fragment 3 in BL21(DE3) SI strain; lane 5, fragment 3 in BL21 CodonPlus (DE3) RIL strain.

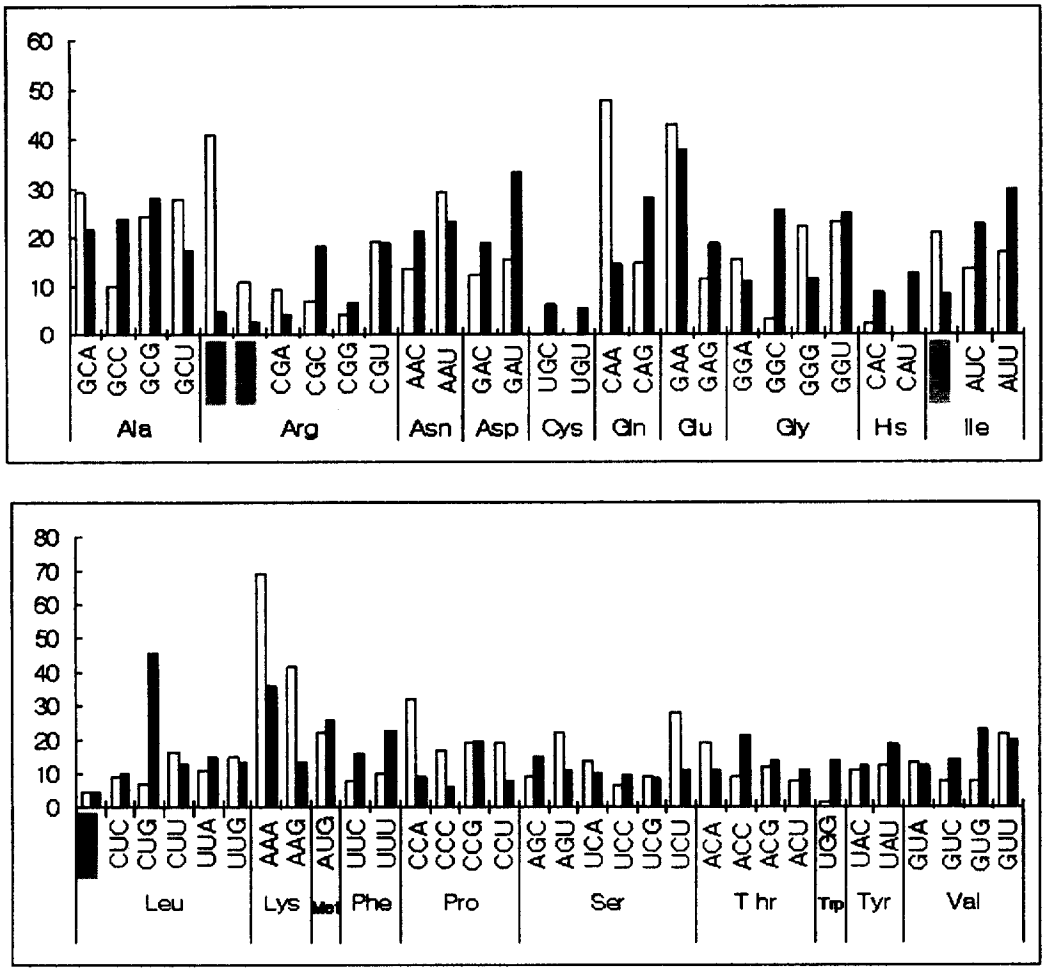


Fig. 4. Comparison of codon usage between *E. coli* and the a189/192r gene of the PBCV-1 genome. The analysis was performed using the program Codon Usage Tabulated from GenBank (CUTG). ▭, codon frequency in the a189/192r gene; ■, codon frequency in *E. coli* ▬, tRNA codons supplemented in the BL21 CodonPlus (DE3) RIL strain.

Comparison of codon usage between PBCV-1 and chlorella NC64A

The PBCV-1 genome contains 11 tRNA genes (Van Etten, 1991), which are thought to be a way to compensate for insufficient tRNAs in the host. Generally, viruses depend on their hosts for protein production, and viruses encoding tRNA genes are rare. However, a few viruses encoding tRNA genes have been reported. The three tRNAs identified in the phage 933W genome help with the efficient synthesis of viral Shiga toxin encoded by rare codons in the host (Plunkett *et al.*, 1999; Kanjo and Inokuchi, 1999). Virulent mycobacteriophage D29 encodes five tRNA genes, which are thought to replace host isoacceptor tRNA species that are inappropriate for the translation of viral protein (Kunisawa, 2000). Bacteriophage T4 also encodes eight tRNAs that are rare in *E. coli* (Cowe and Sharp, 1991). Kunisawa reported (2002) that all eight tRNAs in T4 phage can be found in the host *E. coli*, and the phage tRNAs do not carry any novel anticodon species. The frequency of synonymous codons read by phage tRNAs is always higher in phage genes than in host genes (Kunisawa, 2002). Therefore, it was hypothesized that phage tRNAs could serve to supplement host tRNAs present in minor amounts, thereby enhancing the efficiency of translation of phage genes.

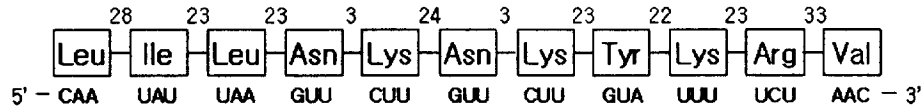
The codon usage between the host, chlorella-like green algae NC64A, and PBCV-1 was compared (Fig. 5). Since only limited information is available for genes encoded by the host, the ribulose 1,5-bisphosphate carboxylase/oxygenase gene that encodes the ribulose 1,5-bisphosphate carboxylase large subunit and the S14 ribosomal protein gene (Amberg and Meints, 1991) were compared as representative NC64A genes. As the arrows in Fig. 5B indicate, codon preference in the viral genes was observed for five out of the nine tRNA genes encoded in the viral genome. Considering the limited information available on codon usage by hosts, this is very intriguing and more genetic information might reveal preferential codon usage in the viral genome, as observed in bacteriophage T4 (Kunisawa, 2002).

Another consideration is the G+C contents of the host and viral genomes. Kunisawa (2002) showed that T4 genes tend to use codons ending in U or A because of the low G+C content, while *E. coli* genes use codons ending in G or C rather than codons ending in U or A, because they tend

to use codons recognized by tRNA species that exist in great quantities. Therefore, Kunisawa (2002) concluded that T4 supplies its own tRNAs to supplement the host tRNA populations that are present in minor amounts for more efficient production of phage proteins.

The G+C content of chlorella-like alga NC64A nuclear DNA is 67%, while that of the PBCV-1 genome is 41.68% (Van Etten *et al.*, 1991). In addition, codon usage by PBCV-1 is strongly biased toward codons ending in A or U (63%) over those ending in C or G (37%) (Schuster *et al.*, 1990). By contrast, the codon usage of the host algal tubulin gene exhibits a bias toward codons ending in C or G (67%) (Van Etten *et al.*, 1991; Yamada *et al.*, 1993). Although a bias in the codons recognized by PBCV-1 tRNAs was not obvious, as in the host versus PBCV-1, there was a bias of 56% XXA/U and 44% XXG/C in the codons recognized by the viral tRNAs. Therefore, the tRNA genes encoded in the PBCV-1 genome might help the virus to overcome the codon usage barrier between the virus and host by supplementing codons for replication.

(A)



(B)

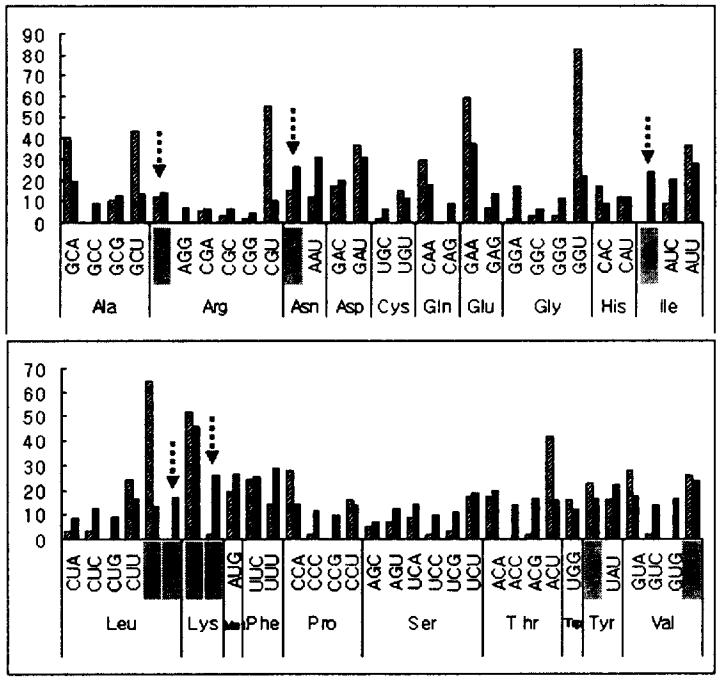


Fig. 5. . The tRNA gene cluster encoded in the PBCV-1 genome (A) and a comparison of codon usage between PBCV-1 and its host NC64A (B). The codon usage analysis was performed using the program Codon Usage Tabulated from GenBank (CUTG). (A). The numbers between the tRNAs indicate the number of nucleotides. The sequence below each tRNA represents the anticodon of the corresponding tRNA. (B). ▬, codon frequency in the ribulose 1,5-bisphosphate carboxylase/oxygenase gene and the S14 ribosomal protein of NC64A; ▬, codon frequency in PBCV-1; ▬, tRNA codon in PBCV-1. Arrows indicate the codons encoded by the viral genome and used more frequently in PBCV-1 than in the host.

Codon usage in the major capsid protein and DNA polymerase genes

The viral tRNAs would be conducive to the predominant translation of viral proteins during viral replication. Therefore, the quantities of tRNAs should be related to the frequencies of those codons in the genes. This was analyzed for two proteins encoded by PBCV-1. Vp54 protein is the major capsid protein (MCP, GenBank accession no. M85052) of PBCV-1 and comprises about 40% of the total PBCV-1 structural proteins (Songsri *et al.*, 1997). Unlike many dsDNA viruses that use the host DNA polymerase, PBCV-1 encodes its own DNA polymerase (GenBank accession no. M86836), which would be specific to viral DNA replication.

Of the 438 codons present in the MCP gene of PBCV-1, the frequencies of the codons recognized by the virus-encoded tRNA genes were as follows: 13 of 13 Lys codons, 32 of 34 Asn codons, 25 of 29 Tyr codons, 6 of 25 Val codons, 1 of 22 Ile codons, 2 of 37 Leu codons and 0 of 18 Arg codons (Table 2). Similarly, the frequencies of the tRNA codons in the DNA polymerase gene (914 codons) were 90 of 90 Lys, 25 of 39 Tyr, 22 of 38 Asn, 23 of 61 Leu, 25 of 63 Val, 3 of 14 Arg, and 2 of 54 Ile (data not shown). One interesting feature is the frequent usage of codons whose corresponding tRNAs have two copies in the tRNA gene cluster. As shown in Fig. 5A, there are two copies of the Asn (AAC) and Lys (AAG) tRNA genes. The AAC Asn codon is used more frequently in the PBCV-1 genome than in the host (Fig. 5B), and it is preferred in both the MCP (32 AAC vs. 2 AAU codons) and polymerase (22 AAC vs. 16 AAU codons) proteins. The AAG Lys codon is rarely used in the host as compared to the PBCV-1 genome (Fig. 5B). The AAG Lys codon is the most frequent codon (51 of 913 codons) and is the preferred Lys codon (51 AAG vs. 39 AAA codons) in the polymerase protein. In addition, the AAG Lys codon is the preferred codon (11 AAG vs. 2 AAA codons) in the major capsid protein. Accordingly, there is a relationship between the duplication of a tRNA gene in the tRNA gene cluster and the frequency or preference of the codon in the viral proteins.

Although the involvement of tRNAs encoded in the viral genome in the synthesis of viral protein has not been proved directly, several factors indicate that this is the case. Nishida *et al.* (1999) showed that the tRNA gene cluster of chlorella virus CVK2 was transcribed as one unit and

processed into small tRNAs. In addition, actual aminoacylation of the tRNA has been observed with the tRNAs recovered from CVK2-infected host cells (Nishida, 1999). Therefore, the tRNA genes encoded in the PBCV-1 genome seem to play a very important role during the synthesis of viral proteins by supplementing the tRNAs that are rare in the host cells.

Table 2. The frequencies of codon usage in the major capsid protein of PBCV-1

codon	count	per 1000	codon	count	per 1000	codon	count	per 1000	codon	count	per 1000
TTT-Phe	1	2.3	TCT-Ser	3	6.8	TAT-Tyr	4	9.1	TGT-Cys	0	0.0
TTC-Phe	20	45.7	TCC-Ser	9	20.5	[redacted]	25	57.1	TGC-Cys	3	6.8
[redacted]	1	2.3	TCA-Ser	2	4.6	TAA-***	0	0.0	TGA-***	1	2.3
[redacted]	1	2.3	TCG-Ser	1	2.3	TAG-***	0	0.0	TGG-Trp	4	9.1
CTT-Leu	6	13.7	CCT-Pro	3	6.8	CAT-His	0	0.0	CGT-Arg	8	18.3
CTC-Leu	8	18.3	CCC-Pro	14	32.0	CAC-His	4	9.1	CGC-Arg	10	22.8
CTA-Leu	4	9.1	CCA-Pro	0	0.0	CAA-Gln	8	18.3	CGA-Arg	0	0.0
CTG-Leu	17	38.8	CCG-Pro	0	0.0	CAG-Gln	19	43.4	CGG-Arg	0	0.0
ATT-Ile	6	13.7	ACT-Thr	12	27.4	AAT-Asn	2	4.6	AGT-Ser	2	4.6
ATC-Ile	15	34.2	ACC-Thr	30	68.5	[redacted]	32	73.1	AGC-Ser	6	13.7
[redacted]	1	2.3	ACA-Thr	0	0.0	[redacted]	2	4.6	[redacted]	0	0.0
ATG-Met	6	13.7	ACG-Thr	2	4.6	[redacted]	11	25.1	AGG-Arg	0	0.0
[redacted]	6	13.7	GCT-Ala	8	18.3	GAT-Asp	5	11.4	GGT-Gly	21	47.9
GTC-Val	6	13.7	GCC-Ala	30	68.5	GAC-Asp	12	27.4	GGC-Gly	12	27.4
GTA-Val	1	2.3	GCA-Ala	3	6.8	GAA-Glu	8	18.3	GGA-Gly	4	9.1
GTG-Val	12	27.4	GCG-Ala	1	2.3	GAG-Glu	6	13.7	GGG-Gly	0	0.0

* The codons whose corresponding tRNAs are present in the viral genome are enclosed in gray boxes. The actual counts in the major capsid protein (403 amino acids) were converted into the frequency per 1,000 amino acid

Implications for adaptation in different hosts

During replication in the host, chlorella viruses use some components of the host synthesis machinery and a special set of tRNAs introduced into the host by the virus. In this way, chlorella viruses can alter the existing host system. This strategy seems to be required when viruses adapt to a wide range of host organisms with various codon usages (Nishida *et al.*, 1999). Chlorella viruses are found throughout the natural environment, but little is known about their natural hosts or origin (Van Etten *et al.*, 1991; Yamada *et al.*, 1991). Some virus-encoded genes are closely related to those of bacteria, fungi, and yeasts (Li *et al.*, 1995, 1997; Lu *et al.*, 1995, 1996; Kutish *et al.*, 1996). PBCV-1 encodes enzymes that are required for the decomposition of chitin and chitosan, polymers of N-acetylglucosamine that are normal components of fungal cell walls and the exoskeletons of crustaceans and insects, and expresses them during infection (Lu *et al.*, 1996; Yamada *et al.*, 1997). Moreover, the genes for the viral major capsid protein, Vp54, of PBCV-1 constitute a gene family (Lu *et al.*, 1995). This implies that the surface of the virus might vary(Nishida *et al.*, 1999). Consequently, Nishida *et al.* (1999) suggested that chlorella viruses have a wide potential range of hosts. From experiments with mutant T4 phages defective in the tRNA gene, Wilson (1972) indicated that phage tRNAs were not essential for viral replication. Mutant T4 phages that could not synthesize tRNA replicated well in *E. coli* B strain. However, when other strains of *E. coli* were used as hosts, these same phage mutants did not replicate. These findings suggest that during evolution, the bacteriophages obtained components, including tRNA genes, that are necessary for their replication in host cells whose biochemical machinery might not include a complete set of the tRNAs necessary for the synthesis of viral protein. This suggests that viruses have evolved to possess tRNA genes for their own replication. Although only a few algae strains are known to host chlorella viruses, they could have had many more host species in the past, or many may remain to be discovered. Consequently, the presence of tRNA genes in many chlorella viruses could be the evolutionary result of adaptation to a wide range of host organisms.

PBCV-I 에서의 tRNA gene의 기능적 의미

이 다 영

부경대학교 대학원 미생물학과

요 약

클로렐라 바이러스, PBCV-1은 330kbp의 genome size를 가지고 350개 이상의 단백질을 코드화 하고 있으며, 11개의 tRNA gene을 가지고 있다. 바이러스의 attachment protein이라고 추정되어지는 A189/192R protein을 *E. coli* strain BL21(DE3) SI에서 발현시켜 Western blot으로 확인해 본 결과 여러 개의 밴드가 나타났다. 그러나 A189/192R protein의 full length와 네 개의 조각으로 쪼갠 각각의 gene을 추가적으로 tRNA가 포함된 *E. coli* BL21 CodonPlus (DE3) RIL strain에서 발현시킨 결과, 발현이 제대로 이루어지는 것을 확인하였다. a189/192r gene의 codon usage 분석에서 *E. coli*와 chlorella에 코드화 되어진 유전자와 비교해보았을 때, AGA와 AUA codon의 사용에 치중하는 경향을 보였다. 이에 더해서, 바이러스의 tRNA codon은 XXA/U (56%) 와 XXG/C (44%) 의 codon usage 경향을 보이는데, 이것은 PBCV-1에서의 codon usage에서 XXC/G (37%) codon 보다 XXA/U (63%) codon을 많이 사용하는 것과 유사하다. 바이러스의 major capsid protein과 DNA polymerase에서의 codon usage 분석에서도 바이러스의 tRNA에 의해 인식되어 질 수 있는 codon을 사용하는 경향이 높다는 것을 보여준다. 바이러스의 tRNA gene cluster에서 두 개씩 존재하는 tRNA gene인 Asn(AAC)와 Lys(AAG)는 viral protein인 MCP와 DNA polymerase에서 다수로 사용되고 있다. PBCV-1의 genome에 coding되어진 tRNA gene은 바이러스가 host에서 replication하는 동안, host cell에는 다주 존재하지 않으면서 바이러스의 protein 합성에 종종 사용되어 지는 tRNA들을 보충함으써 매우 중요한 역할을 하는 것 같다. 그리고 이러한 tRNA들은 호스트에 의존할 수 없는 tRNA를 가짐으로써 다양한 범위의 host에 적응하는데 도움을 줄 것이라 추측 되어진다.

감사합니다!

지금의 내가 여기까지 올 수 있도록 추진력을 팍팍 불어넣어 주시며 지도해 주신 최태진 교수님께 제일 먼저 감사의 말씀드립니다. 모자란 절 받아주시고, 다채로운(?) 실수도 감싸주시며, 석사 생활 동안 연구 외에도 학회 참석이나 MT 등 잊지 못할 추억 거리를 주셨습니다. 많은 것을 얻고 배웠습니다. 감사드립니다. 교수님!

바쁜 일정 속에서도 논문을 검토해 주신 김영태 교수님과 송영환 교수님께도 깊은 감사의 말씀드립니다. 학부 그리고 대학원 생활 동안 많은 관심과 가르침을 주신 이원재 교수님, 김진상 교수님, 이훈구 교수님, 이명숙 교수님, 김군도 교수님께도 감사드립니다.

아무것도 모르는 암흑상태의 저를 '실험이란 이런 것이야!'하며 불꽃을 튀겨주시고, 앞을 밝혀주신데 큰 공헌을 해 주신 윤홍묵 선배님께도 감사의 말씀드립니다. 지금와서 생각해보니, 선배님 주름살에 한 몫 단단히 한 것 같아서 죄송스럽게 생각하고 있습니다. 귀중한 닉네임도 선사하시고, 다정한 옆집 아저씨 같은, 완벽 실존주의 영욱선배 (열심히 하겠습니다^^V), 푸근한 종오선배 (제가 sonicator 때문에 많이 괴롭혔죠?), 독특한 말투와 재치 넘치는 귀여운 현향언니 (논문 봐주셔서, 이 논문을 완성하는데 큰 도움이 되었습니다), 장난기 가득하면서도 진지한 경규선배, 힘들었던 석사생활 같이 헤엄쳐(?) 나갔던 재경선배, 그리고 경우선배 (함께여서 든든하였습니다), 선배이자 친구로써 함께 생활했던 혜경이, 혜진이, 사고 뭉치라도 나랑 비슷한 점이 많은 후배인 베트남형 다영이(?) tham, 실험실 생활하랴, 학과 수업 들으랴 무지 바쁜 학부생 후배님들, 경용이, 상은이, 민정이 모두 감사드립니다.

10년 넘게(....!) 7호관을 지키고(^^;) 계신 두 맏형, 대성선배 그리고 용배선배 (힘든 고비, 고비 마다 많은 힘이 되어 주셨습니다), 이쁘니 윤숙이 언니, 그리고 석사 과정이면서 자칭 지석사라 우기는 근호... 말로만이 아니라, 진정한 4층 family로 여겨졌습니다. 감사합니다.

해맑은 웃음으로 격려해주신 영주언니, 과 업무로 바쁘다가도 나의 실없는 농담에 웃어준 미영언니, 그리고 인영이오빠 (힘든 박사 과정, 잘 해낼꺼야. 힘내~) 정민 선배님, 재형 선배님. 또 오며가며 '안녕하세요~'하고 인사해준 귀엽고 깜찍한 미생물학과

후배님들..모두 모두 감사드립니다.

같이 공부하고 불확실한 미래를 함께 고민하고, 내 생활의 자극제가 되어 준 친구 지영이, 희진이..고마워~ 그리고 전화할 때마다 바쁘다면서 안 만나준다고 왕따시킬꺼라고 협박한 내 오랜 친구들. 다영이 왕따 안시켜줘서 정말 고맙다~^^; 명선아, 미영아, 진화야, 혜정아 무진장 사랑한다~ ^_____^

술 먹고 다닌다고, 늦게 다닌다고, 엄마 신경 쓰이게 하지 말라고... 잔소리의 대가로 거듭난 나의 분신 사랑하는 언니. 언제나 언니의 그 화려한 명언(?)들은 많은 위로가 되었고, 왔다갔다 갈피를 못 잡는 나에게 방향을 제시해 주려고 애써준 덕에 정신적 안정에 큰 힘이 되었어. 언니~고마워~! 마지막으로 내 뒤에서 항상 든든한 버팀목으로 자리 잡고 있는 엄마. 걱정스럽기만한 딸 투, 엄마의 말없는 믿음과 변함없는 성원에 항상 감사드려요. 사랑합니다~

Reference

Amberg, S. M. and Meints, R. H. 1991. Nucleotide sequence of the genes for ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit and ribosomal protein S14 from a chlorella-like green alga. *J. Phycol.* 27:753-758.

Bowden, R. J., Simas, J. P., Davis, A. J. and Efstathiou, S. 1997. Murine gammaherpesvirus 68 encodes tRNA-like sequences which are expressed during latency. *J. Gen. Virol.* 78:1675-1687.

Carstens, C.-P., Bonnardel, J., Allen, R. and Anna Waesche, A. 2002. BL21 CodonPlus cells correct expression problems caused by codon bias. *Strategies* 14:50-51.

Chen, G. T. and Inouye, M. 1994. Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli. Genes Dev.* 8:2641-52.

Cowe, E. and Sharp, P. M. 1991. Molecular evolution of bacteriophages: discrete patterns of codon usage in T4 genes are related to the time of gene expression. *J. Mol. Evol.* 33:13-22.

Calendar, R. 1988. The bacteriophages. In "the viruses" (H. Fraenkel-Conrat, and R. R. Wagner, Eds.). P.562. Plenum Press, New York.

Cho, H. H., Park, H. H., Kim, J. O. and Choi, T. J. 2002. Isolation and characterization of chlorella viruses from freshwater sources in Korea. *Mol. Cells.* 14:168-176.

Deana, A., Ehrlich, R. and Reiss, C. 1998. Silent mutations in the *Escherichia coli* ompA leader peptide region strongly affect transcription and translation in vivo. *Nucleic Acids Res.* 26:4778-82.

Desai, S. M., Vaughan, J. and Weiss, S. B. 1986. Identification and location of nine T5 bacteriophage tRNA genes by DNA sequence analysis. *Nucleic Acids Res.* 14:4197-4205.

Ford, M. E., Sarkis, G. J., Belanger, A. E., Hendrix, R. W. and Hatfull, G. F. 1998. Genome structure of mycobacteriophage D29: implication for phage evolution. *J. Mol. Biol.* 279:143-164.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A.R., Sutton, G., Kelly, J. M., Fritchman, J. L., Weidman, J. F., Small, K.

V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, J. F., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J. F., Dougherty, B. A., Bott, K. F., Hu, P. C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison, C. A., Ⅲ. and Venter, J. C. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science*. 270: 397-403.

Hershko, A. and Ciechanover, A. 1998. The ubiquitin system. *Annu. Rev. Biochem*. 67:425479.

Hushpulian, D. M., Savitski, P. A., Rojkova, A. M., Chubar, T. A., Fechina, V. A., Sakharov, I. Y., Lagrimini, L. M., Tishkov, V. I. and Gazaryan, I. G. 2003. Expression and refolding of tobacco anionic peroxidase from *E. coli* inclusion bodies. *Biochemistry(Mosc)* 68:1189-1194.

Goldberg, A. L. 1992. The mechanism and functions of ATP-dependent proteases in bacterial and animal cells. *Eur. J. Biochem*. 203:9-23.

Graves, M. V., Burbank, D. E., Roth, R., Heuser, J., DeAngelis, P. L. and Van Etten, J. L. 1999. Hyaluronansynthesis in virus PBCV-1 infected chlorella-like green algea. *Virology* 257:15-23.

Jiang, X., Nakano, H., Kigawa, T., Yabuki, T., Yokoyama, S., Clark, D. S. and Yamane, T. 2001. Dosage effect of minor arginyl- and isoleucyl-tRNAs on protein synthesis in an *Escherichia coli* in vitro coupled transcription / translation system. *J. Biosci. Bioeng.* 91:53-57.

Kanjo, N. and Inokushi, H. 1999. Genes for tRNA-Arg located in the upstream region of the Shiga toxin Ⅱ operon in enteromorrhagic *Escherichia coli* O157:H7. *DNA Res*. 6:71-73.

Kleber-Janke, T. and Becker, W. M. 2000. Use of modified BL21(DE3) *Escherichia coli* cells for high-level expression of recombinant peanut allergens affected by poor codon usage. *Protein Expression and Purification* 19:419-424.

Kumar, S., Birah, A., Chaudhary, B., Burma, P. K., Gupta, G. P. and Pental, D. 2005. Plant codon optimized cry genes of *Bacillus thuringiensis* can be expressed as soluble proteins in *Escherichia coli* BL21 Codon Plus strain as NusA-Cry protein fusions. *J. Invertebr. Pathol.* 88:83-86.

Kunisawa, T. 2000. Functional role of mycobacteriophage transfer RNAs. *J. Theor. Biol.*

205:167-170.

Kunisawa, T. 2002. Functional role of bacteriophage transfer RNAs: codon usage analysis of genomic sequences stored in the GENBANK/EMBL/DDBJ tadabases. *Data Science Journal.* 1:216-228.

Kutish, G., Li, Y., Lu, Z., Furuta, M., Rock, D. L. and Van Etten, J. L. 1996. Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: Map position 182 to 258. *Virology* 223:303-317.

Lavrukhin, O. V., Fortune, F. M., Wood, T. G., Burbank, D. E., Van Etten, J. L., Osheroff, N. and Lloyd, R. S. 2000. Topoisomerase II from chlorella virus PBCV-1. Characterization of the smallest known type II topoisomerase. *J. Biol. Chem.* 275:6915-6921.

Li, Y., Lu, Z., Burbank, D. E., Kutish, G. F. and Van Etten, J. L. 1995. Analysis of 43 kb of the chlorella virus PBCV-1 330 kb genome: Map position 45 to 88. *Virology* 212:134-150.

Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G. F., Rock, D. L. and Van Etten, J. L. 1997. Analysis of 74 kb of DNA located at the right end of the 330-kb virus PBCV-1 genome. *Virology* 237:360-377.

Lu, Z., Li, Y., Zhang, Y., Kutish, G. F., Rock, D. L. and Van Etten, J. L. 1995. Analysis of 45kb of DNA located at the left end of the chlorella virus PBCV-1 genome. *Virology* 206:339-352.

Lu, Z., Li, Y., Que, Q., Kutish, G. F., Rock, D. L. and Van Etten, J. L. 1996. Analysis of 94 kb of the chlorella virus PBCV-1 330 kb genome: Map position 88 to 182. *Virology* 216:102-123.

Nishida, K., Suzuki, S., Kimura, Y., Nomura, N., Fujie, M. and Yamada, T. 1998. Group I introns found in chlorella viruses : Biological implications. *Virology* 242:319-326.

Nishida, K., Kawasaki, T., Fujie, M., Usami, S. and Yamada, T. 1999. Aminoacylation of tRNAs encoded by chlorella virus CVK2. *Virology* 263:220-229.

Plunkett, III, G., Rose, D. J., Durfee, T. J. and Blattner, F. R. 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7:Shiga toxin as a phage late-gene product. *J. Bacteriol.* 181:1767-1778.

Schuster, A. M., Graves, M., Korth, K., Ziegelbein, M., Brumbaugh, J., Grone, D. and Meints, R. H. 1990. Transcription and sequence studies of a 4.3kbp fragment from a dsDNA eukaryotic algal virus. *Virology* 176:515-523.

Smith, M. C. M., Burns, R. N., Wilson, S. E. and Gregory, M. A. 1999. The complete genome sequence of the Streptomyces temperate phage ΦC31: evolutionary relationships to other viruses. *Nucl. Acids Res.* 27:2145-2155.

Songsri, P., Hiramatsu, S., Fujie, M. and Yamada, T. 1997. Proteolytic processing of chlorella virus CVK2 capsid proteins. *Virology* 227:252-254.

Sun, L., Adams, B., Gurnon, J. R., Ye, Y. and Van Etten, J. L. 1999. Characterization of two chitinase genes and one chitosanase gene encoded by chlorella virus PBCV-1. *Virology* 263:376-387.

Van Etten, J. L., Meints, R. H., Burbank, D. E., Kuczmarski, D., Cuppels, D. A. and Lane, L. C. 1981. Isolation and characterization of a virus from the intracellular green alga symbiotic with Hydra viridis. *Virology* 113:704-711.

Van Etten, J. L., Burbank, D. E., Kuczmarski, D. and Meints, R. H. 1983. Virus infection of culturable chlorella-like algea and development of a plaque assay. *Science* 219:994-996.

Van Etten, J. L., Lane, L. C., and Meints, R. H. 1991. Viruses and viruslike particles of eukaryotic algae. *Microbiological reviews* 55:586-620.

Van Etten, J. L. 2003. Unusual life style of giant chlorella viruses. *Annual review of Genetics* 37:153-195.

Wilson, J. 1972. Ph.D. thesis, California Institute of Technology, Division of Biology, Pasadena, Calif.

Xia, Y. and Van Etten, J. L. 1986. DNA methyltransferase induced by PBCV-1 virus infection of a chlorella-like green algea. *Mol. Cell. Biol.* 6:1440-1445.

Yamada, T., Higashiyama, T. and Fukuda, T. 1991. Screening of natural waters for viruses infect chlorella cells. *Appl. Environ. Microbiol.* 57:3433-3437.

Yamada, T., Fukuda, T., Tamura,K., Furukawa, S. and Songsri, P. 1993. Expression of the gene encoding a translational elongation factor 3 homolog of chlorella virus CVK2. *Virology*

197:742-750.

Yamada, T., Hiramatsu, S., Songsri, P. and Fujie, M. 1997. Alternative expression of a chitosanase gene produces two different proteins in cells infected with chlorella virus CVK2. *Virology* 230:361-368.

Zhang, Y., Burbank, D. E. and Van Etten, J. L. 1998. Chlorella viruses isolated in China. *Appl. Environ. Microbiol.* 54:2170-2173.