

Functional Implication of the tRNA Genes Encoded in the *Chlorella* Virus PBCV-1 Genome

Da Young Lee¹, Michael V. Graves², James L. Van Etten³ and Tae-Jin Choi^{1*}

¹Department of Microbiology, Pukyong National University, Busan 608-737, Korea

²Department of Biological Sciences, University of Massachusetts-Lowell, Lowell, MA 01854, USA

³Department of Plant Pathology, University of Nebraska, Lincoln, NE 68583-0722, USA

(Received on October 8, 2005; Accepted on November 11, 2005)

The prototype *Chlorella* virus PBCV-1 encodes 11 tRNA genes and over 350 protein-encoding genes in its 330 kbp 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.

Keywords : adaptation, *chlorella* virus, codon bias, tRNA

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 (Kutish et al., 1996; Li et al., 1995, 1997; Lu et al., 1995, 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 (Graves et al., 1999; Lavrukhin et al., 2000; Sun et al., 1999; Xia et al., 1986; Zhang et al., 1998). 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 (Calender, 1988; Desai et al., 1986), mycobacteriophage D29 (Ford et al., 1998), phage 933W (Plunkett et al., 1999; Kanjo and Inokuchi, 1999), and *Streptomyces* phage ϕ C31 (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 (Nishida et al., 1998; Yamada et al., 1993). 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

*Corresponding author.

Phone) ++82-51-620-6367, FAX) +82-51-611-6358

E-mail) choitj@pknu.ac.kr

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 Methods

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

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

Primer	Sequences
189FECO	5'- CTATATACTAAGAATTCATGTCAAGCTC -3'
189RXHO	5'- GGGTGCACCTCGAGTTCGTAATTGC -3'
189NTR	5'- CTTCACTCGAGCGTCTCTACC -3'
192FECO	5'- CAATCTTCAAGAATTCGAGGAAGTGAC -3'
189N2ECO	5'- CACCAATAGAATTCTATATGCCATACG -3'
189CTECO	5'- ACACAGGAATTCAACGTTCTAAAG -3'
192CTECO	5'- CGGCTGACGAATTCGTCGAAAGCTC -3'
192CXHO	5'- GATATCAACTCGAGCTTTCTTGAACAG -3'

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

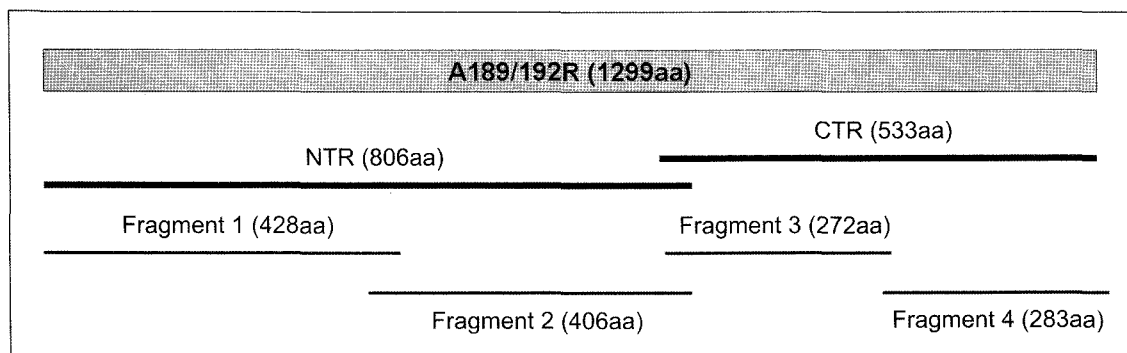


Fig. 1. Schematic description of the *a189/192r* genes of PBCV-1 and the relative locations of the fragments expressed in *E. coli*. The relative locations and sizes of the fragments are shown with the total number of amino acid residues in parentheses.

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 *Bam*HI and *Hind*III, 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 *Bgl*III and *Xho*I, 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 *Bam*HI/*Pst*I and *Bam*HI/*Hind*III, 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 *Eco*RI and *Xho*I.

Expression of recombinant protein. The ligate was transformed into *E. coli* DH5α strain. The nucleotide sequences 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 µg/ml) and cultured for 4 hrs at 25°C. When the cells reached OD₆₀₀=0.5, 300 µl of 3 M 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 µg/ml) and chloramphenicol (50 µg/ml) and cultured at 37°C until reaching OD₆₀₀=1.0. The cells were induced with IPTG at a final concentration of 1 mM 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 µg/ml), and the recombinant proteins were induced as described above.

Cultured cells 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°C for 5 min, chilled in ice, and subjected to 12% SDS-PAGE.

Western blot analysis. 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 incubated 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-tag, 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 186, 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

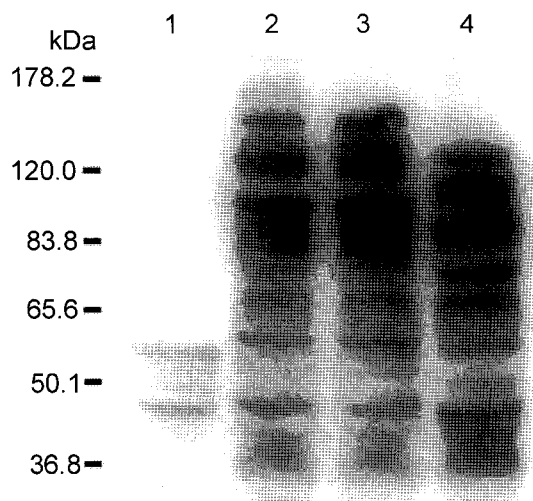


Fig. 2. The expression of recombinant A189/192R protein from *E. coli* BL21(DE3) strain. Lane 1, pMAL-2Cx vector; lane 2, the full-length A189/192R protein from pMAL189/192; lane 3, N-terminal region of A189/192R protein from pMAL189NTR; lane 4, C-terminal region of A189/192R protein from pMAL189CTR.

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).

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 (Hushpalian 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

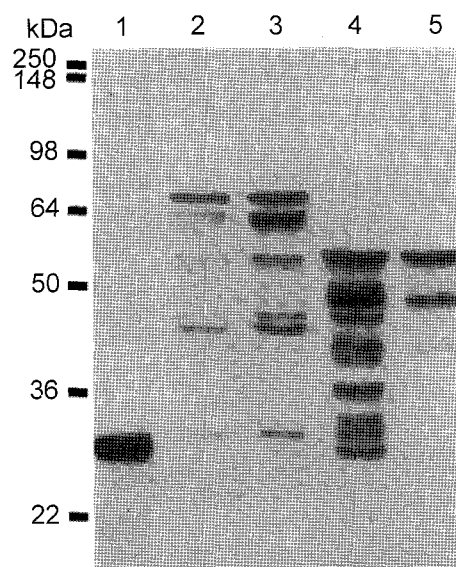


Fig. 3. The expression of fragments 1 and 3 of A189/192R protein cloned in pGEX-5X-1 vector. Lane 1, pGEX-5X-1 vector; lane 2, fragment 1 in BL21(DE3) SI strain; lane 3, fragment 1 in BL21 CodonPlus (DE3) RIL strain; lane 4, fragment 3 in BL21(DE3) SI strain; lane 5, fragment 3 in BL21 CodonPlus (DE3)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

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-Janek 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.

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 (Kanjo and Inokuchi, 1999; Plunkett et al., 1999). Virulent mycobacteriophage D29 encodes five tRNA genes, which are thought to replace host iso-acceptor 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

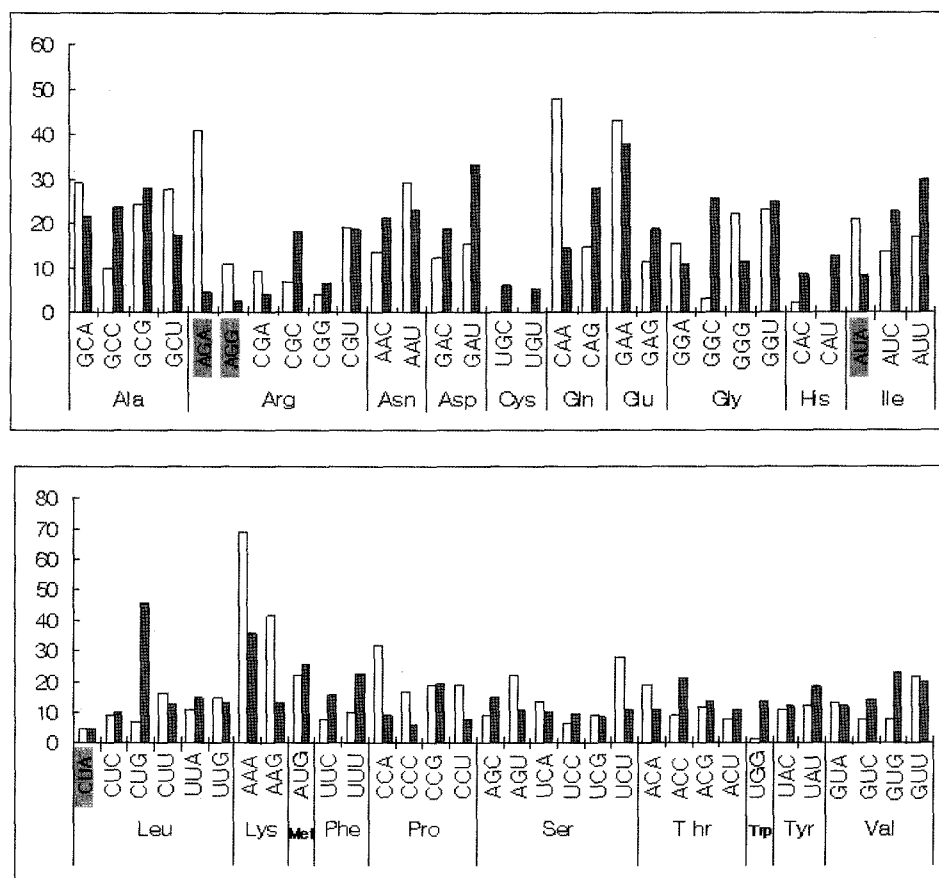


Fig. 4. Comparison of codon usage between *E. coli* and the *a189/192r* gene of the PBCV-1 genome. The analysis was performed using the program Codon Usage Tabulated from GenBank (CUTG). □, codon frequency in the *a189/192r* gene; ■, codon frequency in *E. coli*; ▨, tRNA codons supplemented in the BL21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

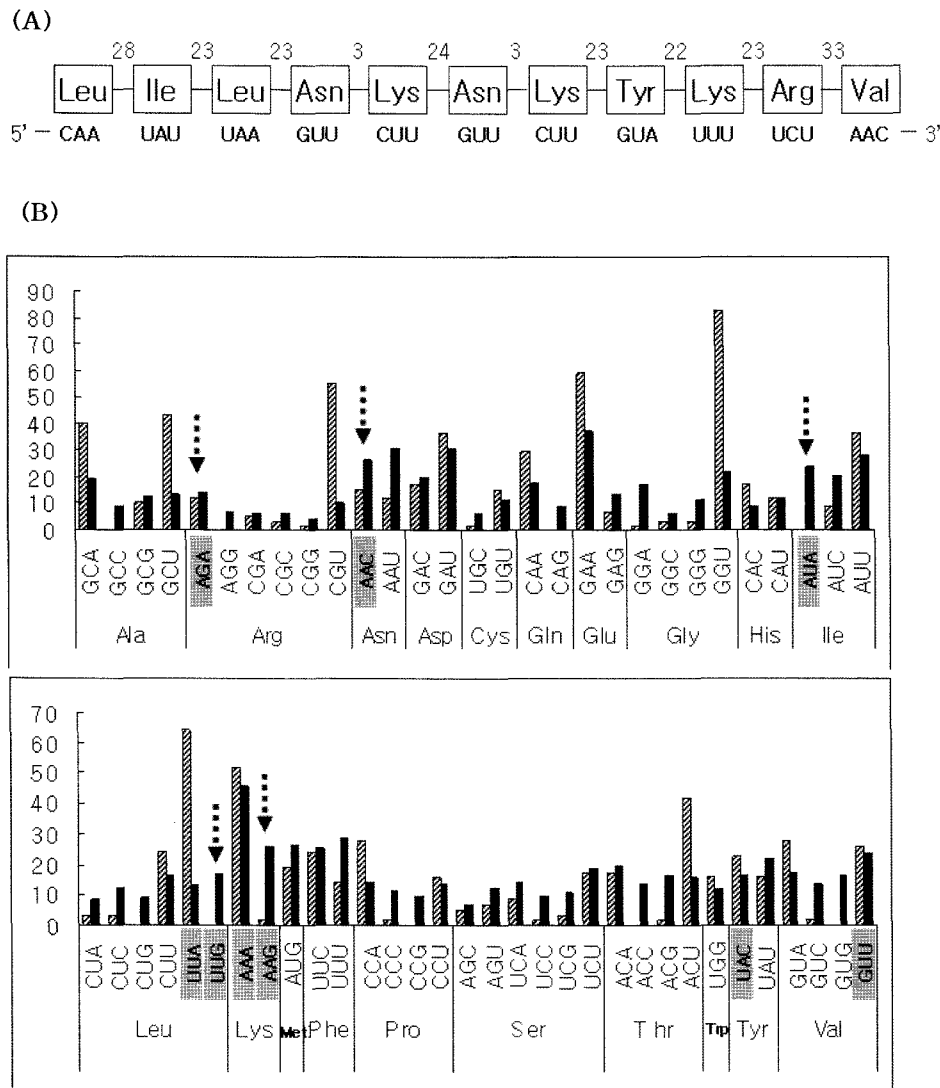


Fig. 5. The tRNA gene cluster encoded in the PBCV-1 genome (A) and a comparison of codon usage between PBCV-1 and its host NC64A (B). The codon usage analysis was performed using the program Codon Usage Tabulated from GenBank (CUTG). (A). The numbers between the tRNAs indicate the number of nucleotides. 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.

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.

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,

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

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	TAC-Tyr	25	57.1	TGC-Cys	3	6.8
TTA-Leu	1	2.3	TCA-Ser	2	4.6	TAA-***	0	0.0	TGA-***	1	2.3
TTG-Leu	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	AAC-Asn	32	73.1	AGC-Ser	6	13.7
ATA-Ile	1	2.3	ACA-Thr	0	0.0	AAA-Lys	2	4.6	AGA-Arg	0	0.0
ATG-Met	6	13.7	ACG-Thr	2	4.6	AAG-Lys	11	25.1	AGG-Arg	0	0.0
GTT-Val	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

^aThe 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 acids.

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.

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 (Kutish et al., 1996; Li et al., 1995, 1997; Lu et al., 1995, 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.

Acknowledgment

This research was supported by a grant (P-2004-06) from Marine Bioprocess Research Center of the Marine Bio 21 Center funded by the Ministry of Maritime Affairs & Fisheries, Republic of Korea.

References

- 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*, ed. by H. Fraenkel-Conrat and R. R. Wagner, Plenum Press, New York. 56 pp.
- 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*. *Nucl. 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. *Nucl. 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.* 67:425-479.
- Hushpuliian, 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. Hyaluronan synthesis in virus PBCV-1 infected chlorella-like green algae. *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 enteromorphagic *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. *Prot. Exp. 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 databases. *Data Science J.* 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 *Chlorella* 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 ϕ C31: 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 algae 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. *Microbiol. Rev.* 55:586-620.
- Van Etten, J. L. 2003. Unusual life style of giant *chlorella* viruses. *Annu. Rev. Gen.* 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 alga. *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.