Self-splicing group I introns in eukaryotic viruses

Takashi Yamada*, Koichiro Tamura, Tadanori Aimi and Puttapon Songsri
Department of Fermentation Technology, Faculty of Engineering, Hiroshima University,
1-4-1 Kagamiyama, Higashi–Hiroshima 724, Japan

Received April 8, 1994; Revised and Accepted May 31, 1994
DDBJ accession nos*

ABSTRACT
We report the occurrence of self-splicing group I introns in viruses that infect the eukaryotic green alga Chlorella. The introns contained all the conserved features of primary sequence and secondary structure previously described for the group IB introns. The Chlorella viral introns (~400 nt) self-spliced in vitro, yielding the typical group I intron splicing intermediates and products. Contrasting to eukaryotic nuclear group I introns, all of which are located in the rRNA genes, these introns were inserted in genes encoding proteins. In one case, the exons encoded a protein showing significant homology to the eukaryotic transcription factor SII (TFIIS), which may be important for viral gene expression. In another case, the gene for the open reading frame (ORF) of a 14.2 kDa polypeptide with unknown functions contained the intron. Scattered distribution of these introns among the viral species and their structural similarity to the group I introns of algae and protists indicated horizontal intron transmission. These eukaryotic viral introns offer an opportunity to understand how group I introns reach organisms of different phylogenetic kingdoms.

INTRODUCTION
Group I introns have been shown to behave as mobile genetic elements, as demonstrated for the rRNA intron omega in yeast mitochondria (1), the alfa cyt-c oxidase subunit I intron of yeast mt DNA (2, 3), the nuclear rRNA intron from Physarum (4) and the td and sunY introns of bacteriophage T4 (5). In each case, mobility is mediated by a DNA endonuclease encoded by an ORF contained within the intron (for review, 6). Thus, intron mobility appears to occur by a DNA-based gene conversion event. A second possible mechanism for the group I intron transposition occurs at the RNA level and requires no such endonuclease to move the intron. As proposed by Woodson and Cech (7), reverse splicing could insert an intron into another RNA molecule, which could then be reverse-transcribed and reinserted into the genome. This mechanism requires the existence within the target RNA of a sequence which can pair with the internal guide sequence (IGS) of the intron.

Whichever mechanism actually operates, the intron sequence and a target site must come in contact with each other in a transposition event. In this context, it is puzzling that very similar introns have often been found in organisms of far distantly phylogenetic kingdoms; for example, the rRNA introns of green algae Chlorella (8, 9) and Ankistrodesmus (10) were shown to be extraordinarily similar to that of Tetrahymena (11). Here we propose a possible candidate, Chlorella viruses, for the mediator of intron spreading between algae and protists.

Chlorella viruses were first found in endosymbiotic algae in Paramecium bursaria and Hydra viridis (for review 12). The viruses are thought to maintain the endosymbiotic relationship between the algae and the protists; because immediately after isolating the algae from their host, the algae are lysed by the viruses. The Chlorella viruses have many unique features; a linear dsDNA genome with hairpin termini; DNA methylation and restriction enzymes; many structural proteins and a lipid component of the virion; and (iv) a bacteriophage-like infection process. The viruses are widespread world-wide (12–14).

In this paper, we describe the presence of group I self-splicing introns in Chlorella viruses isolated in Japan (13, 14). The structure and properties of these introns are very similar to those found in a variety of eukaryotic algae and protists.

MATERIALS AND METHODS
Viruses, vectors, and host strains
The growth of the host algae, Chlorella sp. strain NC64A on MBBM medium and the production and purification of Chlorella viruses, CVB11, CVK2, and CVU1 and the other species have been described (13–15). Escherichia coli JM109 and SURE served as hosts for plasmids pUC18, pUC19, and recombinant plasmids and for M13 mp18 and mp19.

DNA preparation and Southern blot analysis
DNA was isolated from the purified viruses by phenol extraction after treatment with proteinase K (0.1 mg/ml) and Sarkosyl (0.1%) at 50°C for 1h as described previously (13–15). Host algal DNAs (nuclear, chloroplast, and mitochondrial DNAs) were isolated and fractionated by ultracentrifugation in a CsCl–Hoechst 33258 gradient according to the previous method.

*To whom correspondence should be addressed

+D17367, D17368 and D29631
DNA was used for restriction enzyme digestion and Southern blot analyses. Following the transfer of DNA fragments separated by agarose gel electrophoresis onto nylon filters (Biodyne, Pall BioSupport), the filters were hybridized with probes that were labeled with nonradioactive digoxigenin-dUTP as previously described (13 – 15).

Cloning and sequencing

Restriction fragments of the viral DNAs were ligated to pUC19 and amplified in *E.coli* JM109 or SURE. Subcloning was performed using M13 mp18 and mp19 vectors to make overlapping clones. Single-stranded DNA was sequenced by the dideoxy method in an Automated Laser Fluorescent (ALF) DNA sequencer (Pharmacia). The EMBL/GenBank/DBJ data bases were searched for homologous sequences, using the FASTA computer program (17). Nucleotide sequence alignment and secondary structure analyses were performed using the GENETYX program (SDS Software Co. Ltd.).

In vitro splicing

A 726 nt *BamHl/DraI* fragment of the CVU1 TFIIIS that contained a 113 nt 5’ exon, 399 nt intron, and a 214 nt 3’ exon was ligated to *BamHl/SmaI* sites of pGEM-3Zf(+) (Promega Corp.) and propagated in *E.coli* JM109. From the recovered plasmid, uniformly labeled precursor RNA was transcribed. Template (1 µg) was in a 20 µl reaction mixture containing 40 mM Tris+HCl (pH 7.5), 6 mM MgCl2, 10 mM DTT, 2 mM spermidine, 0.5 mM each UTP, CTP and ATP, 0.2 mM GTP, [α-32P]GTP and 30 U SP6 RNA polymerase (Takara Shuzo Co.). The transcripts were separated by electrophoresis on a 4% polyacrylamide – 8 M urea gel. The full-length transcript was excised from the gel and eluted overnight in a solution containing 10 mM Tris+HCl (pH 7.5), 1 mM EDTA, and 0.25 M NaCl. Unlabeled precursor RNA was prepared in the same way, except for 0.5 mM GTP in the transcription mixture. The full length transcript was incubated under a standard *in vitro* splicing condition [50 mM Tris+HCl (pH 7.5), 10 mM MgCl2, 0.1 mM GTP, 2 mM spermidine at 37°C for 30 min] (8) for all but the parameter being examined. The reaction products and intermediates were examined by electrophoresis and autoradiography.

Synthetic oligonucleotides and cDNA synthesis

cDNA of the ligated exons was synthesized using a cDNA cloning kit (Pharmacia) according to the manufacturer’s instruction with a synthetic 26-mer oligonucleotide, 5’AATTCGAGGTCTGCTGACCAGATT- ACCAAAATGACC as a primer for the primary strand synthesis, that corresponded to the multi-cloning sites of pGEM-3Zf(+) and the 8 nt 3’ end of the insert. The secondary strand synthesis was with a synthetic 26-mer, 5’AAGCGTTCATGGCTGACGATGCGACTC as a primer that corresponded to the multi-cloning sites of pGEM-3Zf(+) . After digestion with HindIII and SacI, the fragment was ligated to M13 mp18 and cloned in *E.coli* JM109.

**RESULTS**

**Group I intron of Chlorella virus CVU1**

The viral intron sequences were first detected by Southern blot hybridization of restriction fragments of *Chlorella* virus DNAs probed with rDNA of *Chlorella ellipsoidea* that contained group I self-splicing introns in both small subunit (SSU) and large subunit (LSU) rRNA genes (8, 9). A hybridizing 3.0 kbp *EcoRV* fragment of CVU1, a viral species isolated in Hokkaido, Japan (14), was cloned and sequenced. The determined nucleotide sequence contained an ORF of 180 amino acid (aa) residues with a 399 nt typical group-I intron sequence between residues 164 and 165. Since the 399 nt insert of the CVU1 ORF contained

![Figure 1](image-url)

Figure 1. Secondary structure models for the CVU1 intron (A) and the CVB11 intron (B). Conventions for folding group I introns (11) were followed to generate these predicted secondary structures. Conserved stems P1 – P10 are indicated. Bold letters are core conserved nucleotides. Arrows indicate the 5’ and 3’ splice sites. The inlet shows the P1 and P10 interactions.
the typical catalytic core sequences P, Q, R, and S of the group I introns, a secondary structure model for this insert was constructed according to Cech (11). As shown in Fig. 1A, all P1-P10 motifs of the group I intron were identified and the overall secondary structure of the CVU1 intron fits the model for Group IB. The 48 nt viral sequence in the catalytic core (P, Q, R, and S) differed from those of the Tetrahymena intron (11), the Chlorella SSU rRNA intron (8) and the Chlorella LSU rRNA intron (9) in only 4, 4 and 5 nt, respectively.

The intron nature of this insert and the precise splicing sites were determined in three ways: (i) by in vitro splicing reactions described below, (ii) by cloning and sequencing of cDNA containing the mature mRNA, and (iii) by cloning and sequencing of a corresponding gene without the intron from another Chlorella viral species. We constructed cDNA from the ligated exons produced in vitro as described in Materials and Methods; the exons were digested with HindIII and SacI, ligated to M13 mp18, and cloned in E. coli JM109. The nucleotide sequence determined around the corresponding region of the cDNA clone (pCVU1) was 5'-GACGAACCTATGACGTT-3', and corresponded to the amino acid sequence DEP1MTV. This result indicated the splicing site and the internal guide sequence as shown in Fig. 1A. This splicing site was also confirmed by cloning and sequencing a corresponding gene of another virus, CVK2, that lacked the intron sequence. An ~3.0 kbp EcoRI/BamHI fragment of CVK2 DNA hybridized with the CVU1 ORF region (a 275 bp XhoI/EcoT22I fragment) as a probe. This fragment was cloned and sequenced. There was 91.7% nucleotide identity and 94.4% amino acid sequence identity between the CVU1 and CVK2 genes. The nucleotide sequence, 5'-GACGAACCTATGACGTT-3', of the CVK2 gene was exactly the same as that from the cDNA clone of the CVU1 ligated exons. These results again supported the model, which suggests that P1-P10 interactions in the CVU1 intron help to align the exons for splicing (Fig. 1A).

A comparison of the sequences in the databases with the protein sequence encoded by the ORF revealed a significant homology (FASTA scores of 133–223) to the eukaryotic transcription elongation factor II (TFIIS) (18). The amino acid sequence of the CVU1–TFIIS homolog is shown in Fig. 2A. An overall identity of 33.1% (This homology of 77.1%, when conserved replacements were considered.) was found between the sequences of the ORF and the yeast TFIIS (18). The similarity was greater in the carboxy-terminal half of the protein containing a putative zinc-finger domain. The intron was inserted within the zinc-finger region. In Fig. 2B, the region containing the zinc finger motif is compared among the Chlorella viral proteins and TFIISs of human (19), fruit fly (20), vaccinia virus (21), and African Swine Fever virus (22). As seen here, in addition to the four invariant cysteine residues which are probably involved in zinc binding, several other amino acids are well conserved in the sequences.

Southern blot analysis revealed that the TFIIS-homolog gene was highly conserved among Chlorella viral species (Fig. 3A). Under high stringency, the TFIIS gene probe strongly hybridized to a single EcoRI fragment for all the viral genomes tested in this study. Whereas with an intron sequence probe, a hybridizing signal at the same position of the TFIIS gene appeared for limited viral species even under low stringency; among five viral species tested, only CVU1 showed the presence of the intron in the TFIIS gene (Fig. 3B). These results indicated the optional nature of the

Figure 2. Comparison of the amino acid sequences of the Chlorella viral TFIIS-like polypeptides. (A) Alignment of the sequences deduced from the Chlorella viral genes with an intron (CVU1) and without any intron (CVK2). Arrow indicates the insertion site of the intron. Identical residues are indicated by asterisks. Cysteins proposed to coordinate zinc are boxed. (B) Multiple alignment of the putative zinc finger motifs of the Chlorella viral proteins and of various TFIIS proteins (18–22). Amino acids highly conserved in the sequences are boxed. The number of residues in the sequences are indicated. The consensus sequence is indicated below.

Figure 3. Southern blot analysis of Chlorella virus genomes. (A) EcoRI-digested fragments of viral DNA (10 μg DNA per lane) were separated by agarose gel electrophoresis and blotted to a nylon membrane. They were probed with a 284 bp XhoI/BamHI fragment of CVK2 DNA which contained a central part of the TFIIS-coding region. The probe was labeled with nonradioactive digoxigenin (DIG)–dUTP (Boehringer Mannheim). Hybridization was performed in a mixture containing 50% formamide, 5× SSC, 5% blocking reagent (Boehringer Mannheim), 0.1% Sarkosyl and 0.02% SDS for 20 h at 42°C. Lanes: 1–5, CVU1, CVCH1, CVB11, CVK2, and CVK1, respectively (13, 14). (B) The same blot of (A) was probed with a 385 bp Hael/BagII fragment of CVU1 DNA that contained only the intron sequence. The probe was labeled as above. Hybridization was performed in the same mixture (A) for 20 h at 30°C.
intron in TFII S gene expression and viral viability. Furthermore, they suggested that either precise intron loss from some viral species or lateral intron acquisition by the other viral species occurs.

**Group I intron of Chlorella virus CVB11**

It is also learned from Fig. 3 that a single band different from that of the TFII S gene hybridized with the intron probe in the cases of a few viruses such as CVB11 and CVCH1 (Fig. 3B, arrows). This suggested possible spreading of the intron to the genomes of different *Chlorella* viral strains.

To clarify the relationship between the CVU1 intron and other viral hybridizing sequences, a 2.3 kbp EcoRI fragment of CVB11 that hybridized with the CVU1 intron probe was cloned and sequenced. The identified 1000 bp sequence contained an ORF encoding 112 aa residues with a group I intron sequence of 400 nt between the first and second bases of the 15th codon (UUU for Phe). The intron sequence showed a 78% identity with that of the CVU1 intron; the nucleotide sequences in the catalytic core, P, Q, R, and S, were identical between the two introns. The predicted secondary structure for the CVB11 intron is shown in Fig. 1B. As seen here, the secondary structure motifs were also conserved between the two introns. These results strongly suggested horizontal transfer of this kind of intron from the gene of one *Chlorella* virus strain to another. Although the protein sequence encoded by the exon-ORF was compared with sequences in the data bases, no significantly homologous sequences were identified. Southern blot hybridization experiments, with this exon sequence as a probe, showed strong conservation of this protein among all tested *Chlorella* viruses from Japan (14) (data not shown). This indicated an important function for this protein. The splicing site predicted by aligning the CVB11 sequence with the CVU1 sequence was experimentally confirmed as follows: an ~2.3 kbp BamHI/EcoRI fragment of CVK2 DNA, which hybridized with the CVB11 exon probe but not with the CVB11 intron probe, was cloned and sequenced. The determined nucleotide sequence contained an ORF of 112 aa residues without any intron sequences. The deduced amino acid sequence was highly homologous with that of the CVB11 gene; in a primary alignment of overall sequences, there was 100% identity. The CVB11 sequence, 5'TCCT1TTACC, corresponding to the splicing site is precisely conserved in the CVK2 gene, supporting the P1–P10 interaction shown in Fig. 1B.

**Self-splicing of the CVU1 intron**

To verify that the intron found in the CVU1 TFII S gene is self-splicing, we cloned a 726 nt fragment containing a 113 nt 5'exon, the 399 nt intron and a 214 nt 3'exon using plasmid vector pGEM-3Zf(+) (Fig. 4A). The resulting plasmid was linearized with EcoRI and transcribed *in vitro* with SP6 RNA polymerase to yield a 792 nt primary transcript. When the radiolabeled transcript was incubated with GTP, spermidine and MgCl2 for 30 min at 37°C and fractionated by polyacrylamide/urea gel electrophoresis, products of approx. 792-, 636-, 399-, 393- and 156 nt were generated (Fig. 4B). However, only a precursor RNA band of 792 nt was observed in the absence of either GTP (lane 1) or Mg2+ (lane 2). The reaction also worked in the presence of 1.0 M NH4Cl (lane 4), 0.1 M (NH4)2SO4 (lane 5) or 1.0 M NaCl (lane 6) instead of spermidine. When the unlabeled primary transcript was incubated under splicing conditions in the presence of [α-32P]GTP, two bands of 636 nt and 399 nt appeared (data not shown). The products of the splicing reactions shown in Fig. 4 (lanes 3–6) corresponded (from the top to the bottom), to the full-length transcript (792 nt), to the intron–3'exon (636 nt), to the intron alone (399 nt), to the ligated 5'exon–3'exon (393 nt) and to free 5'exon (156 nt). Clearly the viral intron was capable of self-splicing *in vitro* via the group I self-splicing mechanism.
**Sequence comparison of group I introns**

The *Chlorella* viral introns are similar structurally to the eukaryotic nuclear group IB introns; they differ from bacteriophage introns in secondary structure (23, 24), which resemble mitochondrial and chloroplast introns belonging to the group IA (11). The nucleotide sequence of the viral introns were compared with those of nuclear rRNA introns of various eukaryotic algae. The highly conserved group of the viral and algal introns included the catalytic core (P, Q, R, and S elements), which is essential for group I-type splicing: of 48 nt positions of the CVU1 intron, 3, 5, 6, 4, 4, and 4 nt differed from those of the *D. salina* SSU rRNA [Ds(SSU)] (25), *D. parva* SSU rRNA [Dp(SSU2)] (25), *D. parva* SSURNA [Dp(SSU1)] (25), *Hildenbrandia rubra* [Hr(SSU)] (26), *Characium saccatum* SSU rRNA [Cs(SSU)] (9), and *C. ellipsosperma* SSU rRNA [Cs(SSU)] (8) introns, respectively (Fig. 5). In addition to these regions, several characteristic elements of group I introns, P3, P5, P5a, P5b, P5c, P6a, P6b, and P8 were relatively well conserved. But in contrast, regions of P1, P2, P2.1, P9, and P9.1 varied greatly. Because all nuclear introns thus far reported are located in the SSU or LSU rRNA genes, it would be interesting to compare the IGS of the viral introns with those of the SSU and LSU rRNA group I introns. Base pairing of the IGS of an intron with both the flanking 5' exon (forming helix P1) and with the flanking 3' exon (forming helix P10) presumably play a role in selection of cleavage sites for intron excision (27). The IGSs of the viral introns, 5'-UGGGA-3' [Cv(U1)] and 5'-GGGG-3'[Cv(B11)] show significant similarities to those of eukaryotic nuclear rRNA introns including 5'-UGUGG-3' [Ce(LSU)], 5'-AGGGA-3' [Pc(LSU)], 5'-AGGGGA-3' [Ca(LSU)], and 5'-UGGG-3' [Th(LSU)].

**DISCUSSION**

The optional nature and scattered distribution of the *Chlorella* viral group I introns suggest the possibility of their being transmitted horizontally. Contrasting to the bacteriophage introns, which contain an ORF encoding a site-specific endonuclease that paricipitates in intron transmission through DNA-based gene conversion (23, 24), the *Chlorella* viral introns are group IB introns without an ORF and are most similar to those found in eukaryotic algae (8-10, 25, 26) and protists (11). *Chlorella* viruses infect the algal cells residing in protists such as *Paramaecium* (12), this leads to the attractive speculation that this virus mediates horizontal transfer of group I introns between algae and protists, organisms of two far distant phylogenetic kingdoms. Our recent finding that two group I self-splicing introns with similar IGSs are in the same position of the same gene (SSU rDNA) in organisms of different phylogenetic kingdoms (algae and fungi) (8) would support the reverse splicing mechanism for intron-transposition proposed by Woodson and Cech (7). In this sense, it is noteworthy that the IGSs of the viral introns resembled those of rRNA introns of algae and other organisms. The viral introns could have come from rRNA introns in the host nuclear genes and might be going to move into a gene of another host organism by reverse splicing. To study the detailed mechanisms of horizontal transmission and molecular evolution of group I (especially of the IB type) self-splicing introns, the *Chlorella* viruses described in this paper should serve as an excellent experimental system.

The group I eukaryotic nuclear introns occur only in SSU and LSU rRNA genes, while chloroplast introns are found mainly in LSU rRNA and tRNA genes (11). Mitochondrial introns are in LSU rRNA genes and genes encoding components of the electron transport system (cytochrome b and subunits of cytochrome oxidase, ATPase, and NADH dehydrogenase). Bacteriophage introns are in genes encoding proteins involved in DNA metabolism (nucleotide diphosphate reductase, thymidine synthase, and DNA polymerase) (23, 24). The *Chlorella* viral introns were found in protein-coding genes: the TFIIS-homolog gene in the case of CVU1 and URF in CVB11. Thus, the CVU1 TFIIS-homolog gene is the first example of a transcription factor gene with a group I intron. The TFIIS-homolog and URF genes are highly conserved in *Chlorella* viruses isolated in Japan and are expressed during the viral infection cycle (Sonsri and Yamada, in preparation), suggesting important functions for these genes.

The presence of a group I intron in the viral genes suggests possible control of the viral infection with specific inhibitors of the group I self-splicing. We preliminarily examined the effects of streptomycin, kanamycin and pentamidine (28-31) on viral infection efficiency. Viral infection was significantly depressed by streptomycin (100 μg/ml) and kanamycin (100 μg/ml); the plaque forming efficiency declined to 1/3-1/2 of a control. Similar inhibitory effects were also observed with pentamidine: at a concentration of 50 μg/ml, the plaque forming efficiency was lower than 1/100 of the control. At these concentrations, none of the drugs tested significantly affected the growth of host cells. A consequence of inefficient intron splicing may be a reduction in the amount of TFIIS required for normal viral gene expression.

**Figure 5.** Comparison of conserved sequence elements of group I introns from *Chlorella* viruses and various organisms. (A) Catalytic core sequences of P, Q, R, and S (11). Stars indicate matching nucleotides. LSU and SSU indicate the genes for large subunit and small subunit rRNAs, respectively. Ca, *Candida albicans* (32); Ce, *Chlorella ellipsosperma* (8, 9); Cs, *Characium saccatum* (25); Cv, *Chlorella* virus; Dp, *Dunaliella parva* (25); Ds, *Dunaliella salina* (25); Hr, *Hildenbrandia rubra* (26); Pc, *Pneumocystis carinii* (33, 34); Tt, *Tetrahymena thermophila* (11).
ACKNOWLEDGEMENTS

We thank T. Miyakawa, E. Tsuchiya, and D. Hirata for helpful discussions. This work was supported in part by a grant from the Technical Research Center, Chugoku Electric Power Co. Inc.

REFERENCES