Nucleic Acids Research, 1996, Vol. 24, No. 13 **2463–2469**

Molecular cloning of the three base restriction endonuclease R.*C*viJI from eukaryotic *Chlorella* virus IL-3A

Neela Swaminathan, David A. Mead+, Karolyn McMaster, David George, James L. Van Etten[1] and Piotr M. Skowron*

CHIMERx, 3802 Packers Avenue, Madison, WI 53704, USA and [1]Department of Plant Pathology, University of Nebraska, Lincoln, NE 68583-0722, USA

Received May 13, 1996; Accepted May 14, 1996

EMBL accession no. U09001

ABSTRACT

R.*C*viJI is unique among site-specific restriction endonucleases in that its activity can be modulated to recognize either a two or three base sequence. Normally R.*C*viJI cleaves RGCY sites between the G and C to leave blunt ends. In the presence of ATP R.*C*viJI* cleaves RGCN and YGCY sites, but not YGCR sites. The gene encoding R.*C*viJI was cloned from the eukaryotic *Chlorella* virus IL-3A and expressed in *Escherichia coli*. The primary *E.coli cviJIR* gene product is a 278 amino acid protein initiated from a GTG codon, rather than the expected 358 amino acid protein initiated from an in-frame upstream ATG codon. Interestingly, the 278 amino acid protein displays the normal restriction activity but not the R.*C*viJI* activity of the native enzyme. Nine restriction and modification proteins which recognize a central GC or CG sequence share short regions of identity with R.*C*viJI amino acids 144–235, suggesting that this region is the recognition and/or catalytic domain.

INTRODUCTION

Most Type II restriction endonucleases function as homodimers and recognize a symmetric sequence of 4–8 bases (1,2). They precisely cleave both DNA strands within or adjacent to their recognition sites in the presence of magnesium. Restriction enzymes are extremely specific, catalyzing incorrect cleavage fewer than once in 10^7 binding events in the case of *Eco*RI (3) and *Taq*I (4). The specificity of some endonucleases can be 'relaxed' [also known as 'star' (*) activity] by modified assay conditions such as low ionic strength, high pH, high glycerol concentration, the presence of organic solvents, high ratio of enzyme to DNA or substituting other divalent cations for magnesium (5). These star recognition sequences usually differ from the canonical sequence by a single base pair, although strand preferences can be detected with the appropriate assay (4). A unifying element of star-inducing reaction conditions is a net decrease in bound water activity, which alters protein–DNA contacts and relaxes recognition and cleavage (6).

The phycodnavirus IL-3A (7–9), which infects a unicellular eukaryotic *Chlorella*-like green alga, encodes at least one site-specific restriction endonuclease, R.*C*viJI (10), as well as a cognate cytosine 5-methyltransferase, M.*C*viJI (11). The normal specificity of R.*C*viJI, RGCY (R, purine; Y, pyrimidine) resembles a bacterial Type II restriction endonuclease in requiring only magnesium for activity. However, in the presence of magnesium and ATP, cleavage specificity relaxes to include RGCN (N, any nucleotide) and YGCY sites (10,12). This relaxed specificity is similar to the star activity reported for bacterial Type II restriction endonucleases. R.*C*viJI* activity is unique, though, in requiring an adenine nucleotide cofactor, which suggests a functional domain is involved in this ATP-dependent modulation of the restriction specificity. This feature of R.*C*viJI, as well as its short recognition sequence, distinguishes it from other Type II endonucleases.

Because R.*C*viJI and R.*C*viJI* cleave DNA frequently they can be used for a variety of molecular biology applications. For example, R.*C*viJI* generates numerous sequence-specific oligonucleotides from anonymous DNA samples (12) or produces quasi-random distributions of DNA fragments (13). Although R.*C*viJI can be isolated from virus IL-3A-infected *Chlorella*, this process is inefficient because the host *Chlorella* grows slowly and the virus-infected alga contains many nucleases. This report describes the cloning, sequencing and expression of the *cviJIR* gene in *Escherichia coli.*

MATERIALS AND METHODS

Viruses, vectors and host strains

The growth of the host alga, *Chlorella* strain NC64A, on MBBM medium, the production and purification of virus IL-3A (7) and the isolation of IL-3A DNA (14) have been described. The molecular cloning of *cviJIM* in plasmid pIL-3A.22.8 (11) has been described. *Escherichia coli* strain DH5αMCR (Life Technologies, Gaithersburg, MD) served as a host for recombinant plasmids and bacteriophage M13mp18 (15).

Reagents

Restriction and modification enzymes were obtained from Molecular Biology Resources (Milwaukee, WI). The Minute

* To whom correspondence should be addressed

+Present address: BioRad Laboratories, Hercules, CA 94547, USA

Miniprep DNA purification kit and Sequal DNA sequencing kit were products of CHIMERx (Madison, WI). Oligonucleotides were obtained from Synthetic Genetics (San Diego, CA), M13mp18 and the 100 bp DNA ladder were from Life Technologies and the Muta-Gene *in vitro* mutagenesis kit and protein molecular weight markers were purchased from BioRad (Hercules, CA). *Thermus aquaticus* DNA polymerase for the PCR was obtained from Applied Biosystems (Foster City, CA). Ampicillin, X-Gal, IPTG and XOMAT AR X-ray film were from Sigma (St Louis, MO). [γ-^{32}P]ATP was purchased from Du Pont (Boston, MA).

R.*Cvi*JI restriction of DNA

One microgram of DNA was digested under R.*Cvi*JI conditions (10 mM $MgCl_2$, 50 mM NaCl, 50 mM Tris–HCl, pH 8.0) or R.*Cvi*JI* conditions (1 mM ATP, 20 mM DTT, 10 mM $MgCl_2$, 50 mM NaCl, 50 mM Tris–HCl, pH 8.0) at 37°C for 1 h. Restriction fragments were electrophoresed on 1.5% agarose gels in Tris–acetate/EDTA buffer (16).

DNA and amino acid sequencing

The *cviJIR* gene was located on plasmid pCJH1.4 and both DNA strands were sequenced by the dideoxy chain termination method (17) using three approaches: (i) primer walking; (ii) cloning restriction endonuclease digests into M13mp18; (iii) sequencing restriction endonuclease deletion derivatives of pCJH1.4. To ensure the accuracy of the *cviJIR* gene sequence, PCR primers were synthesized to flank both sides of the gene and the amplified region from IL-3A virus was cloned into a λ promoter-containing vector and maintained in the presence of a compatible plasmid expressing the cloned *cviJIM* gene. The gene was sequenced again from two such cloned samples. The N-terminal 15 amino acids of R.*Cvi*JI purified from *E.coli* was sequenced at the Protein and Nucleic Acid Shared Facility of the Medical College of Wisconsin (Milwaukee, WI) by the Edman degradation method using an Applied Biosystems 477A Liquid Phase Protein Sequencer with an on-line 120A PTH Analyzer (18). The computer-translated sequence of R.*Cvi*JI was compared with the public sequence databases using MacDNASIS Pro software (Hitachi Software, San Bruno, CA).

Site-specific oligonucleotide mutagenesis

The *lacZ′* Shine–Dalgarno sequence and the *cviJIM* ATG start codon were separated by 215 nt on plasmid p710. These 215 nt were deleted by site-specific oligonucleotide mutagenesis using the oligomer 5′-CAATTTCACACAGGAAACAGCTATGTCT-TTTCGCACGTTAGAAC-3′ as indicated in Figure 1A. The mutagenesis was facilitated by converting the double-stranded plasmid DNA to single-stranded DNA by co-infecting p710-containing *E.coli* with the helper phage R408 (19). The oligonucleotide was annealed to the single-stranded plasmid, extended in the presence of T4 DNA polymerase, ligated using T4 DNA ligase and transformed into *E.coli* DH5αMCR. The transformed cells were grown overnight in liquid culture, the DNA was purified, *Xho*I digested, dephosphorylated with calf intestinal alkaline phosphatase (CIAP) and transformed into *E.coli* DH5αMCR cells. *Xho*I digestion served to enrich for the desired mutation because a single *Xho*I site was located in the sequence being deleted (Fig. 1A).

Purification of 32.5 kDa recombinant R.*Cvi*JI

Lysate preparation. Escherichia coli DH5αMCR[pCJH1.4] was grown at 30°C with vigorous aeration in TB broth medium (20) supplemented with 0.5 mM $MgCl_2$ and 100 μg/ml ampicillin. When the OD_{595} reached 0.5, IPTG to 0.5 mM and maltose to 0.4% were added and the culture was grown for an additional 2 h. Cells (60 g) were concentrated by centrifugation and resuspended in three volumes of cold lysis buffer A [30 mM Tris–HCl, pH 7.9 (4°C), 200 mM NaCl, 2 mM EDTA, 10 mM β-mercaptoethanol, 50 μg/ml PMSF, 20 μg/ml benzamidine, 2 μg/ml *o*-phenantroline, 0.7 μg/ml pepstatin, 0.7 μg/ml leupeptin A]. The suspension was passed through a Manton–Gaulin cell disrupter and the cells were removed by centrifugation. Polyethyleneimine solution (10%, pH 7.5) was added to the supernatant to a final concentration of 1% and stirred for 30 min. The resulting precipitate was removed by centrifugation and the nucleic acid-free solution was precipitated overnight with ammonium sulfate (0.5 g/ml). Precipitated proteins were pelleted by centrifugation, dissolved in 100 ml of buffer A, desalted on a Sephadex G-25 column (5 × 30 cm) and equilibrated in buffer B (10 mM K/PO_4, pH 7.2, 30 mM NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol, 10% glycerol, 0.05% Triton X-100, 50 μg/ml PMSF, 20 μg/ml benzamidine, 2 μg/ml *o*-phenantroline, 0.7 μg/ml pepstatin, 0.7 μg/ml leupeptin A).

Phosphocellulose chromatography. The desalted R.*Cvi*JI extract was applied to a 5 × 12 cm phosphocellulose P11 (Whatman) column equilibrated in buffer B. After washing with 4 l of buffer B, the protein was eluted with 4 l of a 0–1.2 M NaCl linear gradient in buffer B. Fractions were assayed by incubating 1 μg pBR322 DNA with 2 μl aliquots of each fraction. The endonuclease eluted at ~0.7 M NaCl.

Heparin–Sepharose chromatography. Active fractions from the phosphocellulose column were diluted 7-fold in buffer C [20 mM Tris–acetate, pH 7.2 (22°C), 60 mM potassium acetate, 2 mM EDTA, 10 mM β-mercaptoethanol, 10% glycerol, 0.01% Triton X-100, 50 μg/ml PMSF, 20 μg/ml benzamidine, 2 μg/ml *o*-phenantroline, 0.7 μg/ml pepstatin, 0.7 μg/ml leupeptin A] and applied to a heparin–Sepharose (Pharmacia) column (5 × 10 cm) that had been equilibrated with buffer C. The column was washed with 1.5 l of buffer C and R.*Cvi*JI was eluted with 3 l of a 0–2.5 M potassium acetate gradient in buffer C. The endonuclease activity eluted at 0.9 M potassium acetate.

Hydroxyapatite HTP chromatography. The pooled R.*Cvi*JI-containing fractions were diluted 5-fold in buffer D (10 mM K/PO_4, pH 7.0, 50 mM NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol, 10% glycerol, 0.05% Triton X-100, 50 μg/ml PMSF, 20 μg/ml benzamidine, 2 μg/ml *o*-phenantroline, 0.7 μg/ml pepstatin, 0.7 μg/ml leupeptin A) and applied to a hydroxyapatite HTP (BioRad) column (1.5 × 3 cm) equilibrated in buffer D. The column was washed with 200 ml of buffer D and the enzyme was eluted with a 0–0.7 M K/PO_4, pH 7.0, linear gradient in buffer D. R.*Cvi*JI activity was recovered at 0.2 M K/PO_4.

Molecular sieving. Active fractions from the hydroxyapatite HTP column were concentrated to 3 ml using ultrafiltration membrane YM10 (Amicon) and subjected to molecular sieving on a Sephadex G-100 (Pharmacia) column (2.5 × 100 cm) that had been equilibrated in buffer E [20 mM Tris–acetate, pH 8.0 (4°C), 60 mM potassium acetate, 2 mM EDTA, 10 mM β-mercaptoethanol, 10%

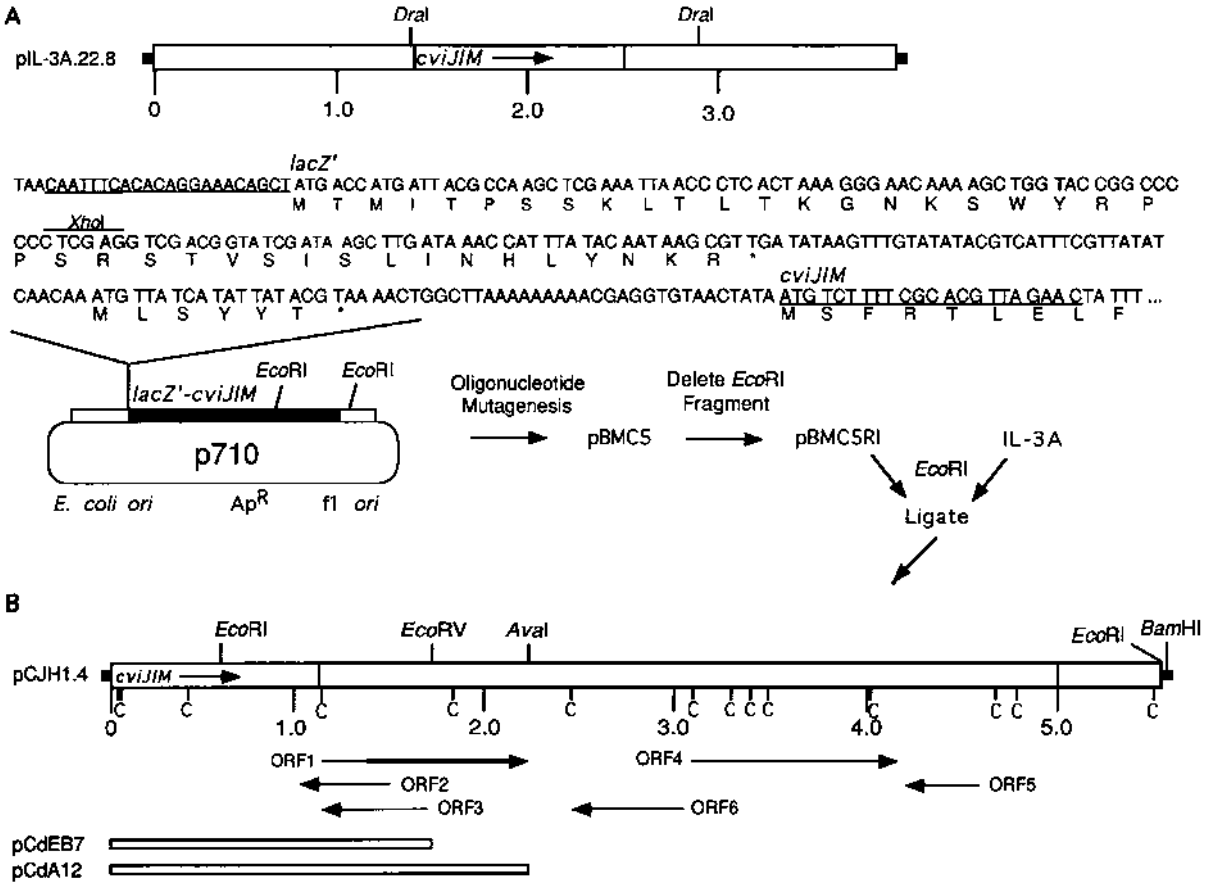

Figure 1. Molecular cloning and gene organization of *Chlorella* virus IL-3A DNA adjacent to the *cviJIM* gene. (**A**) Schematic diagram for cloning R.*C*viJI. A 1.6 kb *Dra*I fragment encoding the *cviJIM* gene was cloned into a *lacZ′* vector, resulting in p710. The translational start codons for the *lacZ′* and *cviJIM* genes are indicated. The relative position of the *cviJIM* gene is indicated by the shaded region of the circular map. Improved expression of M.*C*viJI was achieved by site-specific oligonucleotide mutagenesis, using an oligomer equivalent to the underlined nucleotides to remove the intervening sequence, resulting in pBMC5. The restriction fragment between the *Eco*RI sites was subsequently deleted, resulting in pBMC5RI, to force the cloning of the *cviJIR* gene (see text for details). Abbreviations: Ap^R, ampicillin resistance gene; f1 *ori*, phage f1 intergenic region; *lacZ′*, truncated β-galactosidase gene; *cviJIM*, gene coding for M.*C*viJI methyltransferase; *E.coli ori*, origin of plasmid replication. (**B**) Restriction map and ORFs of 5497 bp of cloned and sequenced IL-3A viral DNA. The 1101 bp (367 amino acids)*cviJIM* gene is on the far left. The directions of the transcripts are indicated by the arrows. The shaded region is vector DNA and the open region is IL-3A DNA. Six ORFs are indicated. The portion of ORF1 expressed in *E.coli* is marked in bold. pCdA12 and pCdEB7 are restriction endonuclease deletion derivatives of pCJH1.4 containing the IL-3A sequence indicated. C represents R.*C*viJI restriction sites.

glycerol, 0.01% Triton X-100, 50 µg/ml PMSF, 20 µg/ml benzamidine, 2 µg/ml *o*-phenantroline, 0.7 µg/ml pepstatin, 0.7 µg/ml leupeptin A]. Peak fractions containing homogeneous R.*C*viJI were pooled, dialyzed against storage buffer S [20 mM Tris–acetate, pH 8.0 (4°C), 200 mM potassium acetate, 0.5 mM EDTA, 0.1 mM DTT, 50% glycerol, 0.02% Triton X-100, 50 µg/ml PMSF, 0.7 µg/ml pepstatin, 0.7 µg/ml leupeptin A] and stored at –70°C.

RESULTS AND DISCUSSION

Improved expression of the *cviJIM* gene

The successful cloning of the endonuclease portion of a restriction–modification (R-M) gene system usually requires methyltransferase expression (1). The *cviJIM* gene from the virus IL-3A genome was previously cloned, expressed and sequenced from plasmid pIL-3A.22.8 (11). Unpublished results indicated that the DNA following the 3′-end of the *cviJIM* gene was not co-linear with the viral genomic DNA. Therefore, to remove extraneous DNA and to improve M.*C*viJI expression, a 1.6 kb *Dra*I fragment from plasmid pIL-3A.22.8, which includes the *cviJIM* gene (367 codons), was subcloned into the *Eco*RV site of a *lacZ′* vector. The resultant plasmid, p710 (Fig. 1A), had the *cviJIM* gene in the same translational orientation as the *lacZ′* gene and expressed M.*C*viJI activity (RG^mCY). The ability of this construct to methylate plasmid DNA was tested by incubating p710 with either *Hae*III (GG\CC), *Sac*I (GAGCT\C) or *Hin*dIII (A\AGCTT). Efficient M.*C*viJI expression would make these sites sufficiently resistant to these enzymes. However, the plasmid was only 90% resistant to *Hae*III and *Sac*I and 25% resistant to *Hin*dIII digestion. Therefore, it was necessary to increase M.*C*viJI expression before attempting to clone the *cviJIR* gene.

Improved translation of the *cviJIM* gene was obtained by deleting 215 nt between the *lacZ′* and *cviJIM* ATG start codons, as detailed in Materials and Methods. This deletion created a perfect fusion of the *lacZ′* Shine–Dalgarno sequence with the *cviJIM* initiation codon (Fig. 1A). Plasmid DNA from five clones was >95% resistant to *Hae*III, *Hin*dIII, *Pvu*II (CAG\CTG) or *C*viJI restriction. Four of the plasmids were sequenced and the results confirmed the expected deletion. One plasmid, pBMC5, was chosen for further modification.

Molecular cloning of the *cviJIR* gene

Two observations suggested that the *cviJIR* gene on the IL-3A virus genome was 3′ of the *cviJIM* gene: (i) *E.coli* containing cloned DNA 5′ of the *cviJIM* gene lacked R.*C*viJI activity; (ii) attempts to clone the DNA 3′ of the *cviJIM* gene resulted in deletions/rearrangements of this region. The *cviJIR* gene was cloned using the following strategy. First, an *Eco*RI deletion derivative of pBMC5, which lacks the 3′-half of the *cviJIM* gene, was created (pBMC5RI; Fig. 1A). Second, both pBMC5RI and IL-3A genomic DNA were cleaved with *Eco*RI, the plasmid was dephosphorylated and the DNAs were mixed with T4 DNA ligase and used to transform DH5αMCR. Third, the pooled pBMC5RI/IL-3A transformants were grown overnight in liquid culture. Plasmid DNA was isolated from the cell mixture and digested with *Hae*III, dephosphorylated with CIAP and transformed into DH5αMCR cells. Six colonies grew; the DNA from one, pCJH1.4, was resistant to *Hae*III. *Escherichia coli* containing pCJH1.4 also expressed R.*C*viJI activity.

Gene organization and sequence analysis

The *Eco*RI fragment cloned into pCJH1.4 is 4901 bp in length and the entire fragment was sequenced. Seven discrepancies were noted in the 273 bases following the previously reported *cviJIM* gene (11) and that reported here. The sequence data shown here was obtained from three independently derived clones from the viral genome (one plasmid genomic library clone, pCJH1.4, and two PCR-derived clones) and carefully checked for inaccuracies. The insert in pCJH1.4 contains six open reading frames (ORFs) of 1074 (ORF1), 468 (ORF2), 555 (ORF3), 1086 (ORF4), 396 (ORF5) and 579 bp (ORF6), which could code for polypeptides containing 358 (41.4 kDa), 156 (19.4 kDa), 185 (20.3 kDa), 362 (38.9 kDa), 132 (14.5 kDa) and 193 (21.9 kDa) amino acids respectively, assuming the most commonly used start codon ATG is the translational initiation site (Fig. 1B). ORF4–ORF6 do not encode R.*C*viJI, because *E.coli* containing pCdA12, which lacks DNA between the *Ava*I and *Bam*HI sites (Fig. 1B), produces R.*C*viJI activity. *Escherichia coli* containing pCdEB7, which lacks DNA between the *Eco*RV and *Bam*HI sites, does not produce R.*C*viJI activity. Thus, ORF1 or ORF3 are the most likely candidates to encode R.*C*viJI.

R.*C*viJI purified from *E.coli* containing pCJH1.4 produced a single band on 10% SDS–PAGE with an apparent size of 32.5 kDa (Fig. 2). The purified recombinant R.*C*viJI protein was sequenced and the first 15 N-terminal amino acids were (T, M, R or K)EEKKRLALIEKQRI. The ambiguous N-terminal amino acid sequence (T, M, R or K) reported here could be the result of partial post-translational processing by methionine aminopeptidase (21). However, the penultimate residue, glutamic acid, does not favor removal of methionine by this peptidase. Comparing the N-terminal sequence with that predicted from ORF1 revealed 14 out of 15 amino acids match if translation of the recombinant protein began at the GTG codon located at nt 299–301 (Fig. 3). Furthermore, the observed 32.5 kDa molecular weight of the recombinant enzyme agrees with this predicted molecular weight (31.6 kDa). These results indicate that the primary product of *cviJIR* expressed in *E.coli* is a GTG codon-initiated protein of 278 amino acids. Because of this unexpected result, the *cviJIR* gene and flanking sequence from *Chlorella* virus IL-3A DNA were PCR amplified using appropriate primers (22). The PCR product was cloned for

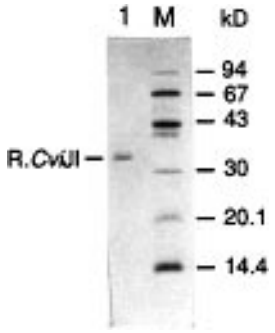

Figure 2. Electrophoresis of R.CviJI purified from *E.coli* on a 10% SDS–polyacrlyamide gel. The single band migrated with an apparent molecular weight of 32.5 kDa (lane 1). Lane M, protein standards.

redundant sequencing purposes. The same nucleotide sequence as that presented in Figure 3 was observed, eliminating the possibility of cloning artifacts (23).

Like many bacterial R-M systems (1) and another *Chlorella* virus R-M system, *C*viAII (24), the *cviJIM* gene has the same transcriptional orientation as the *cviJIR* gene (Fig. 3). Either 21 or 261 nt separate the *cviJIRM* genes, depending on whether the start codon at position 55–57 (ATG) or 299–301 (GTG) is utilized (Fig. 3).

It is not known if translation of the native R.*C*viJI mRNA from IL-3A-infected *Chlorella* also begins at the GTG codon or if it begins at the upstream ATG codon. Translation initiation at this in-frame ATG codon would produce a protein of 358 amino acids (Fig. 3). Approximately 8% of prokaryotic and some eukaryotic gene products are initiated with a GTG start codon rather than the usual ATG codon (25–27). It is interesting to speculate that the choice of a non-optimal start codon is one method to achieve low expression of a lethal protein such as R.*C*viJI. As indicated below, four observations suggest that translation of the native enzyme probably begins at the ATG codon.

(i) DNA cleavage activity of the recombinant and native enzymes were compared under normal and star reaction conditions. The native enzyme produced the expected fragments under both conditions (Fig. 4, lanes 1 and 2). Surprisingly, recombinant R.*C*viJI displays the normal restriction activity (Fig. 4, lane 3), but not the star activity (Fig. 4, lane 4). Thus, the specificity of the recombinant enzyme was not altered by ATP, which indicates a fundamental difference between the native and cloned enzymes.

(ii) During the initial stages of purifying the recombinant enzyme, a second peak of R.*C*viJI activity which elutes separately from the 32.5 kDa activity peak is consistently observed (data not shown). However, only trace amounts of this separate activity peak are available, making it difficult to determine its molecular weight. A comparison of the elution profile of native and recombinant R.*C*viJI on a molecular sieving gel also indicates that the native enzyme is slightly larger (data not shown).

(iii) Previous studies have established that the 50 bases preceding the translation start codon of functional *Chlorella* virus genes typically consist of at least 70% A+T (28). The 50 bases preceding the ATG translation start site of M.*C*viJI (11) and M.*C*viAII (23) contain 76 and 82% A+T respectively. The 50 bases preceding the GTG codon in the R.*C*viJI gene are 62% A+T. In contrast, the 50 bases preceding the upstream ATG codon are 80% A+T. The GTG codon may be favored in *E.coli* because

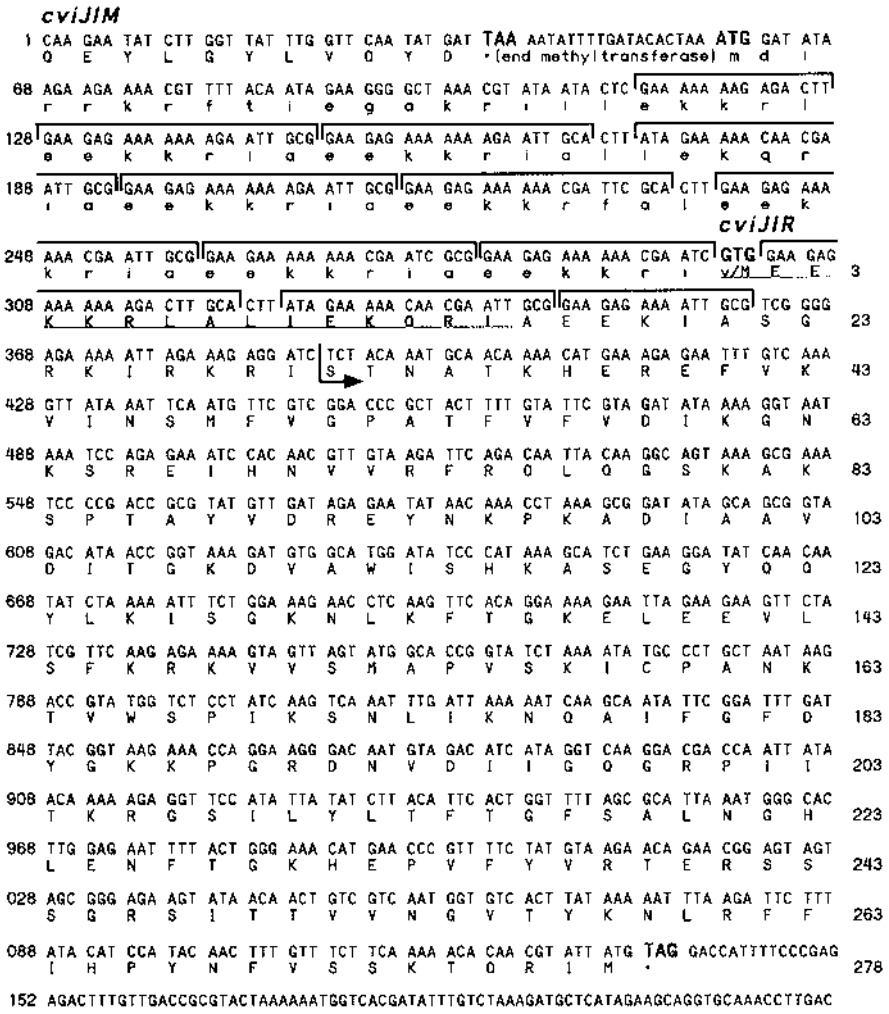

Figure 3. DNA sequence of the *cviJIR* gene and its flanking regions. The predicted amino acid sequence of the 31.6 kDa R.*C*viJI is indicated in capital letters. The N-terminal amino acid sequence of the *E.coli* purified 32.5 kDa enzyme, determined by chemical analysis, is underlined. The repeats of the amino acid sequence (E/I/–)EK(K/Q)R(I/L/F)(A/–) are overlined. The vertical line and associated arrow indicate where the DNA sequence from pCJH1.4 diverges from that of plL-3A.22.8 (11). The potential R.*C*viJI 5′ ATG start codon and predicted in-frame amino acid sequence are indicated in lower case letters.

an *E.coli*-like ribosome binding site (RBS; GAAAAAA) (29) with an optimal seven base spacing is located at –13 to –8 prior to this initiation site. The region preceding the ATG codon lacks an *E.coli*-like RBS.

(iv) The putative N-terminal 80 amino acids preceding the *E.coli* GTG start codon has a pI of 10.81. This domain contains nine repeats of the basic amino acid sequence (E/I/–)EK(K/Q)-R(I/L/F)(A/–) (Fig. 3). The GTG-initiated recombinant protein with a pI of 10.61 has three such repeats at its N-terminus. Such striking similarities of pI and basic amino acid repeats suggests that the N-terminal 80 amino acid domain is an integral part of the native R.*C*viJI protein.

Sucrose gradient-purified *Chlorella* virus IL-3A particles contain a significant amount of virally packaged R.*C*viJI (J.L. Van Etten, unpublished results). These basic amino acid repeats may function as a non-specific DNA binding domain and thus facilitate viral packaging. Regions containing basic amino acids are common in DNA binding proteins, including those containing helix–loop–helix and leucine zipper motifs (see for example 30,31). Protein nuclear targeting sequences also share a similar basic amino acid pattern (32). While there are alternative explanations, these observations suggest that the native ezyme is expressed from the upstream, in-frame ATG codon. This putative N-terminal 80 amino acid domain may function as a nucleic acid binding domain for viral packaging or it could be responsible for the ATP-dependent modulation of restriction specificity, or both. However, no recognizable ATP binding motifs (33) are present in the *cviJIR* gene.

Because of the observed functional differences between the native and recombinant enzyme, experiments to express the putative N-terminal region and better characterize the native enzyme were devised. The sequence preceding the putative ATG methionine codon was modified to include an optimal *E.coli* RBS (34). The resulting clone produced higher yields of the previously characterized 32.5 kDa protein, with no apparent ‘star’ activity or larger protein products. The codons for the ‘longer’ protein were PCR amplified and the product was ligated to three different regulated promoters and transformed into several *E.coli* strains. All constructs were sequenced to confirm the expected changes and in all three cases, the 32.5 kDa form of the enzyme was the only activity found. Additionally, the valine start codon was mutagenized to a non-start valine codon GTC in order to force the

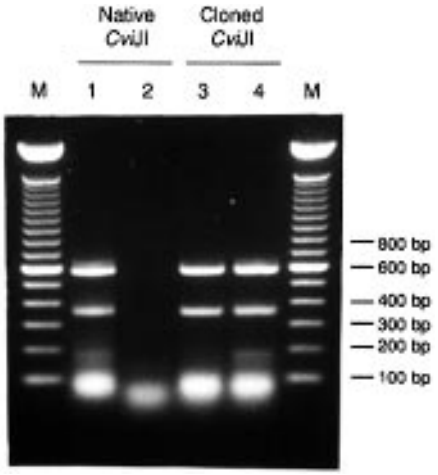

Figure 4. *C*viJI and *C*viJI* restriction activities of the native and cloned enzymes. One microgram of pUC19 DNA was restricted with 1 U R.*C*viJI either purified from virus IL-3A-infected *Chlorella* (lanes 1 and 2) or from the cloned gene in *E.coli* (lanes 3 and 4) for 1 h. The normal (lanes 1 and 3) and star (lanes 2 and 4) reaction conditions are given in Materials and Methods. Lane M, 100 bp DNA ladder.

use of the 5′ upstream ATG codon. Analysis of several hundred clones revealed that those with the desired GTC codon consistently had frameshift mutations at other regions in ORF1.

Additional efforts to purify the native enzyme were complicated by the rapid loss of enzyme activity with increasing purity. Thus, we cannot eliminate the possibility that the observed 'star' activity is associated with another subunit containing ATP-modulated endonuclease activity. However, that the native R.*C*viJI and R.*C*viJI* activities result from the same protein was suggested by their co-elution from heparin–Sepharose, Blue-Sepharose, Q-Sepharose, hydroxyapatite and phosphocellulose columns (10; our unpublished data), their overlapping recognition sequences (10,12,13) and by comparison of the time cource of their cleavage (10). Analysis of the putative N-terminal region by conventional techniques such as expression in *E.coli* and additional purification of the native enzyme have been unsuccesful. Tracking highly purified R.*C*viJI will require antibodies to the enzyme in order to further characterize the molecular weight of the native and recombinant enzymes. Expression of the eukaryotic *cviJIR* gene may also require a host other than *E.coli* to translate the putative longer protein.

Comparison of amino acid sequences

In spite of the common recognition sequence shared by cognate R-M enzymes, only two, *Eco*RII (35) and *Eco*57I (36), share significant amino acid similarities between the endonuclease and methyltransferase. The predicted amino acid sequences of R.*C*viJI and M.*C*viJI were compared for intermolecular homology and two small regions of identity were found (Fig. 5). One region (TGFS-_G_F_K_F) partially overlaps with a motif (motif IV) present in all cytosine 5-methyltransferases (37). Interestingly, the TGFS sequence is immediately adjacent to the invariant dipeptide PC in M.*C*viJI (11), which is the catalytic site of M.*Hha*I and other bacterial cytosine 5-methyltransferases (37). The other small region of identity is SF_RK_V_AP. Although these homology regions are limited (SF_RK_V_AP shares 54% identity over 13 residues and TGFS-_G_F_K_F shares 38% identity over 21 residues) and not colinear, the similarity between them may be significant and suggests a role in DNA recognition or cleavage.

Amino acids 144–235 of R.*C*viJI show limited homologies to several restriction and modification enzymes (Fig. 5), all of which contain a central GC or CG as part of their recognition sequence. R.*C*viJI has the highest homology with R.*Bsu*RI (GG\CC) from *Bacillus subtilis* (38), sharing 35% amino acid identity over 132 residues, a portion of which is shown in Figure 5. This homology is not colinear, but is segmented or epitopic. Seven other proteins which recognize a central GC or CG sequence also share limited identities with R.*C*viJI, including M.*Hgi*CII (GG[A/T]C^{m}C; 39) (63% identity/16 amino acids), R.*Hae*III (GG\CC; G.G.Wilson, unpublished results) (39%/61 amino acids), R.*Ngo*PII (GG\CC; 40) (28%/40 amino acids), R.*Pvu*II (CAG\CTG; 41) (29%/31 amino acids), M.*Pvu*II (CAG^{m}CTG; 42) (35%/78 amino acids), R.*Pst*I (CTGCA\G; 43) (36%/50 amino acids) and R.*Sma*I (CCC\GGG; 44) (35%/48 amino acids). A particularly interesting similarity is found between R.*Pvu*II and R.*C*viJI (Fig. 5). The crystal structure of R.*Pvu*II complexed with its cognate DNA sequence (45) indicates that E55 and D58 may represent the Mg^{2+} binding ligands and that E68 and K70 are the candidate active site residues (underlined residues in Fig. 5). R.*C*viJI appears to have the appropriate charged group equivalents in D191, D194, K205 and R206. The sequence adjacent to the active site of R.*Pvu*II, GFSTHHH, contains three residues involved in DNA sequence recognition, S81, H83 and H84. A partially similar sequence in R.*C*viJI is also present in the same location (212LTf-GFS). These similarities, as well as the coincidence of the nine central GC recognition R-M proteins having limited regions of identity to R.*C*viJI, suggest that amino acids 144–235 encompass the target recognition and/or cleavage domain.

The rationale for utilizing short epitopic and in some cases non-colinear, homology comparisons is based on several considerations. Endonucleases that share similar recognition sequences are statistically closely related, despite their short amino acid similarities (46). For example, five endonucleases that

```
R. PvuII                                                    PGRegnDavdnaqqeyelksinidLTk-GFSthhh
R. CviJI   144 SFKRKVVSMAPVSKICPANKTVWSPIKSNLIKNQAIFGFDYGKKPGRDNVDIIGQGRPIITKRGSILYLTFTGFSALNGHLENFTGKHEPVF 235
M. CviJI       SF RK V  AP                                                          TGFS-  G    F  K    F
R. BsuRI         KR V S  P /K     K  WS IK       /I  - YGK      V       P    /GSIL  T    /AL       T K
M. HgiCII                                                FDYG    RD V IIG
R. HaeIII      S KRK     PV/KI   NK     I   LI    I /FD  K    DN      /R    K   I Y  F
R. NgoPII                         K NL       G DY          I  Q    I   G I
M. PvuII         KRK       V   W   KT W-- KS   K  A /G D GK    DN    I  -  I        YL          H   F  K- P F
R. PstI          KRK-      V    P    /W   K  L      I  F-YGK          I         /G ILY
R. SmaI             V S     S   P        IK  L     A  /FD  K       DI                          /HLEN
```

Figure 5. Amino acid identity between R.*C*viJI and related proteins. Alignment of R.*C*viJI amino acids 144–235 with the indicated restriction endonucleases and methyltransferases. A / indicates a non-contiguous position in the alignment of the protein. A – indicates a gap or deletion introduced to maximize homology.

recognize hexanucleotide palindromic sequences and leave tetranucleotide 5′-overhangs were found to share two regions of similarity with the catalytic and recognition domain of *Eco*RI (47). The crystal structure of four restriction enzymes reveal common structural and functional elements between *Bam*HI (48), *Eco*RI (49), *Eco*RV (50) and *Pvu*II (45), despite the lack of significant sequence similarity. The catalytic and DNA recognition regions of these molecules are not contiguous, but interspersed. For those restriction enzymes related by similar recognition or cleavage sites, this type of analysis may reveal additional correlations.

CONCLUSION

The Phycodnaviridae family of double-stranded DNA viruses encode numerous restriction and modification enzymes. Based on the sensitivity/resistance of viral DNA to some 50 methylation-sensitive restriction endonucleases, a minimum of 12 modification phenotypes has been deduced (51). Some of these R-M enzymes have the same modification and cleavage specificity as bacterial Type II enzymes, whereas others have unique specificities. The R.*C*viJI restriction endonuclease from *Chlorella* virus IL-3A is unusual with respect to its short recognition sequence (RGCY) and the ATP-dependent relaxation of sequence specificity (RGCN, YGCY). The three base restriction activity of R.*C*viJI has been cloned in *E.coli* apparently separate from its two base cleavage specificity. As the recombinant R.*C*viJI displays the normal restriction activity, but not the ATP-dependent star activity, the putative 80 amino acid N-terminal domain may be responsible for the ATP-dependent modulation of R.*C*viJI restriction specificity. Alternatively, the basic amino acid region could function as a nucleic acid binding domain for viral packaging. It is also reasonable that a domain separate from the catalytic portion of the enzyme exists to carry out these functions. Uncovering the role of this putative region, if any, will contribute to a detailed understanding of the structure–function relationship of this unique enzyme.

ACKNOWLEDGEMENTS

We thank Dwight Burbank for growing and processing virus-infected cells, Geoffrey Wilson for generously providing us with amino acid sequences for R.*Hae*III and R.*Mth*TI prior to publication, Yanping Zhang for the p710 plasmid and Lily Marr and Jim Szablewski for technical assistance. This investigation was supported, in part, by Public Health Service grant GM-32441 to JVE from the National Institute of General Medical Sciences.

REFERENCES

1 Wilson,G.G. and Murray,N.E. (1991) *Annu. Rev. Genet.*, **25**, 585–627.
2 Roberts,R.J. and Macelis,D. (1992) *Nucleic Acids Res.*, **20**, 2167–2180.
3 Halford,S.E. and Johnson,N.P. (1980) *Biochem. J.*, **191**, 593–604.
4 Barany,F. (1988) *Gene*, **65**, 149–165.
5 Brooks,J.E. (1987) *Methods Enzymol.*, **152**, 113–129.
6 Robinson,C.R. and Sligar,S.G. (1993) *J. Mol. Biol.*, **234**, 302–306.
7 Van Etten,J.L., Burbank,D.E., Schuster,A.M. and Meints,R.H. (1985) *Virology*, **140**, 135–143.
8 Van Etten,J.L., Schuster,A.M., Girton,L., Burbank,D.E., Swinton,D. and Hattman,S. (1985) *Nucleic Acids Res.*, **13**, 3471–3478.
9 Schuster,A.M., Burbank,D.E., Meister,B., Skrdla,M.P., Meints,R.H., Hattman,S., Swinton,D. and Van Etten,J.L. (1986) *Virology*, **150**, 170–177.
10 Xia,Y., Burbank,D.E., Uher,L., Rabussay,D. and Van Etten,J.L. (1987) *Nucleic Acids Res.*, **15**, 6075–6090.
11 Shields,S.L., Burbank,D.E., Grabherr,R. and Van Etten,J.L. (1990) *Virology*, **176**, 16–24.
12 Swaminathan,N., George,D., McMaster,K., Szablewski,J., Van Etten,J.L. and Mead,D.A. (1994) *Nucleic Acids Res.*, **22**, 1470–1475.
13 Fitzgerald,M.C., Skowron,P., Van Etten,J.L., Smith,L.M. and Mead,D.A. (1992) *Nucleic Acids Res.*, **20**, 3753–3762.
14 Van Etten,J.L., Meints,R.H., Burbank,D.E., Kuczmarski,D., Cuppels,D.A. and Lane,L.C. (1981) *Virology*, **113**, 704–711.
15 Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
16 Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1987) *Current Protocols in Molecular Biology*. Wiley-Interscience, New York, NY.
17 Sanger,F., Nichlen,S. and Coulson,R.A. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
18 Strickler,J.E., Hunkapiller,M.W. and Wilson,K.J. (1984) *Anal. Biochem.*, **140**, 553–566.
19 Russel,M., Kidd,S. and Kelly,M.R. (1986) *Gene*, **45**, 333–338.
20 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
21 Ben-Bassat,A., Bauer,K., Chang,S.-Y., Myambo,K., Boosman,A. and Chang,S. (1987) *J. Bacteriol.*, **169**, 751–757.
22 Mullis,K.B. and Faloona,F.A. (1987) *Methods Enzymol.*, **155**, 335–350.
23 Barany,F., Slatko,B., Danzitz,M. Cowburn,D. Schildkraut,I. and Wilson,G. (1992) *Gene*, **112**, 91–95.
24 Zhang,Y., Nelson,N., Nietfeldt,J.W., Burbank,D.E. and Van Etten,J.L. (1992) *Nucleic Acids Res.*, **20**, 5351–5356.
25 Gualerzi,C.O. and Pon,C.L. (1990) *Biochemistry*, **29**, 5881–5889.
26 Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1–45.
27 Kozak,M. (1989) *J. Cell Biol.*, **108**, 229–241.
28 Schuster,A.M., Graves,M., Korth,K., Ziegelbein,M., Brumbaugh,J., Grone,D. and Meints, R.H. (1990) *Virology*, **176**, 515–523.
29 Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342–1346.
30 Ellenberger,T.E., Brandl,C.J., Struhl,K. and Harrison,S.C. (1992) *Cell*, **71**, 1223–1237.
31 Fisher,D.E., Parent,L.A. and Sharp,P.A. (1993) *Cell*, **72**, 467–476.
32 Bairoch,A. (1992) *Nucleic Acids Res.*, **20**, 2013–2018.
33 Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) *EMBO J.*, **1**, 945–951.
34 Gold,L. and Stormo,G.D. (1991) *Methods Enzymol.*, **185**, 89–93.
35 Kossykh,V.G., Repyk,A.V. and Hattman,S. (1993) *Gene*, **124**, 65–68.
36 Janulaitis,A., Vaisvila,R., Timinskas,A., Klimasauskas,S. and Butkus,V. (1992) *Nucleic Acids Res.*, **20**, 6051–6056.
37 Kumar,S., Cheng,X., Klimasauskas,S., Mi,S., Posfai,J., Roberts,R.J. and Wilson,G.G. (1994) *Nucleic Acids Res.*, **22**, 1–10.
38 Kiss,A., Posfai,G., Keller,C.C., Venetianer,P. and Roberts,R.J. (1985) *Nucleic Acids Res.*, **13**, 6403–6421.
39 Erdmann,D., Horst,G., Dusterhoft,A. and Kroger,M. (1992) *Gene*, **117**, 15–22.
40 Sullivan,K.M. and Saunders,J.R. (1989) *Mol. Gen. Genet.*, **216**, 380–387.
41 Athanasidadis,A., Gregoriu,M., Thanos,D., Kokkinidis,M. and Papamatheakis,J. (1991) *Nucleic Acids Res.*, **18**, 6434.
42 Tao,T., Walter,J., Brennan,K.J., Cotterman,M.M. and Blumenthal,R.M. (1989) *Nucleic Acids Res.*, **17**, 4161–4175.
43 Walder,R.Y., Walder,J.A. and Donelson,J.E. (1984) *J. Biol. Chem.*, **259**, 8015–8026.
44 Heidmann,S., Seifert,W., Kessler,C. and Domdey,H. (1989) *Nucleic Acids Res.*, **17**, 9783–9796.
45 Cheng,X., Balendiran,K., Schildkraut,I. and Anderson,J.E. (1994) *EMBO J.*, **13**, 3927–3935.
46 Jeltsch,A., Kroger,M. and Pingoud,A. (1995) *Gene*, **160**, 7–16.
47 Sisknys,V., Timinskas,A., Klimasauskas,S., Butkus,V. and Janulaitis,A. (1995) *Gene*, **157**, 311–314.
48 Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1994) *Nature*, **368**, 660–664.
49 Kim,Y., Grable,J.C., Love,R., Greene,P.J. and Rosenberg,J.M. (1990) *Science*, **249**, 1307–1309.
50 Winkler,F.K., Baner,D.W., Oefner,C., Tsernoglou,D., Brown,R.S., Heatman,S.P., Bryan,R.K., Martin,P.D. Petratos,K. and Wilson,K.S. (1993) *EMBO J.*, **12**, 1781–1795.
51 Nelson,M., Zhang,Y. and Van Etten, J.L. (1993) In Jost,J.P. and Saluz,H.P. (eds), *DNA Methylation: Molecular Biology and Biological Significance*. Birkhauser Verlag, Basel, Switzerland, pp. 186–211.