9-2006

Cloning of Nt.CviQII Nicking Endonuclease and Its Cognate Methyltransferase: M.CviQII Methylates AG Sequence

Siu-Hong Chan
New England Biolabs, Inc., Ipswich, MA

Zhenyu Zhu
New England Biolabs, Inc., 240 County Road, Ipswich, MA

David Dunigan
University of Nebraska-Lincoln, ddunigan2@unl.edu

James L. Van Etten
University of Nebraska-Lincoln, jvanetten1@unl.edu

Shuang-yong Xu
New England Biolabs, Inc., 240 County Road, Ipswich, MA

Published in *Protein Expression and Purification* **49**:1 (September 2006), pp. 138–150; doi 10.1016/j.pep.2006.04.002

Submitted February 3, 2006; revised March 28, 2006; published online April 25, 2006.

Cloning of Nt.CviQII nicking endonuclease and its cognate methyltransferase: M.CviQII methylates AG sequences

Siu-hong Chan,[1] Zhenyu Zhu,[1] David D. Dunigan,[2] James L. Van Etten,[2] and Shuang-yong Xu [1]

[1] New England Biolabs, Inc., 240 County Road, Ipswich, MA 01938, USA

[2] Department of Plant Pathology and Nebraska Center for Virology, University of Nebraska–Lincoln, Lincoln, NE 68583-0722, USA

Corresponding author — Shuang-yong Xu, email xus@neb.com

Abstract

Chlorella virus NY-2A has a large, highly methylated dsDNA genome (45% of the cytosines are 5-methylcytosine and 37% of the adenines are N^6-methyladenine). Here, we report the cloning, expression, and characterization of the NY-2A-encoded CviQII nicking-modification (N-M) system. The nicking endonuclease, Nt.CviQII, recognizes R ↓ AG (R = A or G, ↓ indicating cleavage site) sequences and cleaves the phosphodiester bond 5′ to the adenosine. Because of the difficulty in cloning and expressing the wild-type Nt.CviQII, C-terminal truncation mutants were generated and full-length Nt.CviQII was reconstructed by intein-mediated peptide ligation. The truncation mutants and the reconstructed full-length Nt.CviQII have the same recognition and cleavage specificity as the native enzyme. Full-length and truncated Nt.CviQII produced by a cell-free transcription/translation system have similar reaction rates. The methyltransferase, M.CviQII, was also cloned and expressed. It modifies the adenine in AG doublets of DNA in vitro and in vivo in *Escherichia coli*. To our knowledge, M.CviQII is the first adenine methyltransferase that recognizes a dinucleotide. Therefore, M.CviQII may be a useful reagent for blocking endonuclease cleavage when restriction sites overlap with AG sequences.

Keywords: Chlorella virus, intein-mediated peptide ligation, methyltransferases, nicking endonucleases, restriction-modification system

Abbreviations*:* AdoMet, *S*-adenosyl-methionine; BAC, bacteria artificial chromosome; CBD, chitin-binding domain; ds, double-strand; m5C, 5-methylcytosine; m6A, 6-methyladenine; MTase, methyltransferase; NEase, nicking endonuclease; N-M, nicking-modification system; Nt, top-strand nicking; REase, restriction endonuclease; R-M, restriction-modification; ss, single-strand; wt, wild-type.

The chlorella viruses are large, icosahedral, plaque-forming, genetically diverse but morphologically similar viruses. They have linear dsDNA genomes (315–380 kb) and belong to the family *Phycodnaviridae*, genus *Chlorovirus* [1]. The prototype chlorella virus, *Paramecium bursaria* chlorella virus (PBCV-1) has a 331-kb genome that contains ~366 protein-encoding genes of 65 or more codons and a polycistronic tRNA operon encoding 11 tRNAs. The chloroviruses infect certain fresh water, unicellular, eukaryotic chlorella-like green algae, which normally exist as endosymbionts in the protozoan *P. bursaria*.

One distinctive feature of the chlorella viruses is that they contain genes encoding Type II DNA site-specific restriction endonucleases (REases) and their companion DNA methyltransferases (MTases), collectively referred to as restriction-modification (R-M) systems. REases recognize and cleave specific nucleotide sequences, while their companion DNA MTases protect DNA from cleavage by methylating one of the nucleotides in the same nucleotide sequence. The number of MTase genes varies from virus to virus. For example PBCV-1 has five MTase-encoding genes, two of which have companion REases, whereas virus NY-2A has at least 10 functional MTase-encoding genes [2]. As a result, the NY-2A genome is heavily methylated and its genomic DNA is resistant to many REases [3].

Two virus-encoded site-specific DNA endonucleases have been isolated from NY-2A infected chlorella

cells. R.CviQI creates double-strand breaks 5′ to the T in the palindromic sequence G ↓ TAC [4] and NY-2A nickase makes single-strand breaks 5′ of the A in R ↓ AG sequences [2]. The NY-2A nickase, later named Nt.CviQII, is interesting because it recognizes three bases with 2-fold degeneracy in the first base, the shortest recognition sequence among all known dsDNA cleaving and nicking endonucleases (NEase). Its companion MTase should also recognize RAG or shorter sequences. Only one other three-base NEase has been described: Nt.CviPII encoded by chlorella virus NYs-1 [5, 6]. Nt.CviPII recognizes ↓CCD sequences (D = A, G, or T; three bases with 3-fold degeneracy in the third base). Therefore, cloning and characterizing Nt.CviQII and its companion MTase M.CviQII should provide additional information on NEases that recognize and cleave DNA frequently. Here, we report expression and purification of C-terminal deletion variants of Nt.CviQII and reconstruction of the full-length enzyme by intein-mediated peptide ligation. Recombinant M.CviQII was also purified and its recognition sequence was found to be more degenerate than that of the cognate NEase.

Materials and Methods

Identification of the CviQII N-M genes

Genomic DNA from virus NY-2A was isolated from purified virions as described [3]. The methyltransferase selection method [7] was used to clone the *cviQIIM* gene. NY-2A genomic DNA was partially digested with 2-fold serial dilutions of *Sau*3AI and ligated into *Bam*HI-digested and calf intestinal alkaline phosphatase (CIP)-treated pUCAC (a derivative of pUC19 containing a chloramphenicol-resistant gene [6]). The ligated products were transformed into *Escherichia coli* ER1992 [*Δ(argF-lac)U169 glnV44 mcr-67 rfbD1? relA1? endA1 dinD2::MudI1734* (*Km*^{R}, *lacZ*$^+$) *spoT1? thi-1 Δ(mcrC-mrr)114::IS10*]. Clones that expressed M.CviQII were selected by digesting plasmid DNA from the transformed cells with *Mnl*I [CCTC(N_7)↓; complementary sequence ↓(N_6)GAGG, ↓ indicating cleavage site] at 37 °C for 2 h. *Mnl*I does not cleave G^{m}AGG sequences and so *Mnl*I-resistant plasmids should be expressing M.CviQII. The *Mnl*I-treated plasmid library was used to transform *E. coli* ER2502 [*fhuA2 ara-14 leu Δ(gpt-proA)62 lacY1 glnV44 galK2 rpsL20 endA1 R(zgb210::Tn10)Tc*^{S}*xyl-5 mtl-1 Δ(mcrC-mrr)*$_{HB101}$]. Clones partially resistant to *Mnl*I were selected, which could contain the *cviQIIM* gene. Inserts from the plasmids were sequenced using pUC universal forward and reverse primers as well as custom-designed primers. The sequences were subjected to a BLAST search against REBASE and GenBank [8, 9].

Genes encoding R-M systems are located adjacent to one another in most cases [REBASE, [8]]. Therefore, DNA flanking the *cviQIIM* gene was sequenced from the NY-2A genomic DNA using custom primers to look for *cviQIINt*.

Expression and purification of M.CviQII

The *cviQIIM* gene was amplified from NY-2A genomic DNA by PCR such that *Nde*I and *Sap*I sites were added to the 5′ and 3′ ends of the gene, respectively. The amplified DNA was inserted into pTXB1 at *Nde*I and *Sap*I sites so that M.CviQII was expressed as a fusion protein with a Mxe GyrA intein and chitin-binding domain (CBD) at its C-terminus [10]. *E. coli* ER2566 [*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 [dcm] R(zgb-210::Tn10–Tc*^{S}*) endA1 Δ(mcrC-mrr)114::IS10 R(mcr-73::miniTn10–Tc*^{S}*)2]* was transformed by the plasmid and expression was induced with 0.25 mM IPTG for 3 h at 37 °C. Cells were harvested, re-suspended in 20 mM sodium phosphate buffer, pH 6.5, sonicated on ice and the disrupted cells were centrifuged at 12,000*g* for 20 min at 4 °C. The fusion protein appeared as an inclusion body. The pellet was washed with 0.2 M NaCl, 1% deoxycholate and then with 1% Triton X-100, 1 mM EDTA. The washed inclusion bodies were solubilized in 25 ml of 6.5 M guanidine hydrochloride (GdnHCl), 50 mM Tris, 0.5 M NaCl, pH 8.0, and incubated at room temperature for 2 h. The solubilized protein was clarified by centrifugation (12,000*g* for 20 min at 4 °C) and added dropwise to a refolding buffer [55 mM MES (2-Morpholinoethanesulfonic acid), pH 6.5, 0.56 mM NaCl, 0.44 mM KCl, 1.1 mM EDTA, 1 mM DTT, 440 mM sucrose, 550 mM l-arginine, final pH 10.6 at 25 °C] with constant stirring at room temperature (FoldIt Screen protein refolding kit, Hampton Research). The solution containing the refolded protein was incubated at room temperature for 2 h and then with 50 mM DTT for 5 h to induce cleavage of the intein. The solution turned turbid after 2 h; however, SDS–PAGE analysis indicated that most of the cleaved M.CviQII protein remained in the soluble fraction. The soluble protein was dialyzed against 20 mM sodium phosphate buffer, 0.1 M NaCl, pH 6.5 at 4 °C and cleared by centrifugation. Refolded protein was loaded on an SP column (15 ml bed volume; Amersham) and eluted using a 0.1–1.0 M NaCl gradient; M.CviQII eluted at ~0.3 M NaCl. Fractions containing M.CviQII were pooled, concentrated, and stored in 50% glycerol (v/v) at −20 °C. Protein concentrations were determined by the Bradford method.

Cell-free transcription/translation of Nt.CviQII and its truncation mutants

Cell-free transcription/translation of Nt.CviQII was accomplished with the Pure System™ classic II kit (Post Genome Institute Co. Ltd., Tokyo, Japan). Oligonucleotides were designed to generate both full-length and truncation mutants of Nt.CviQII by PCR. Two-step PCR was used to incorporate a T7 promoter and ribosome-binding sequence at the 5′ end of the genes following the manufacturer's instructions. The PCR products were gel-purified and used as templates for in vitro transcription/translation according to the manufacturer's instructions. Aliquots of the translation reactions were used directly in DNA nicking assays.

Nt.CviQII truncation mutants and intein-mediated peptide ligation

C-terminal truncation mutants of Nt.CviQII, NQN509, and NQN528, which have deletions of 41 and 22 amino acids, respectively, were amplified from NY-2A genomic DNA by PCR. Ser509 of NQN509 and Ser528 of NQN528 were mutated to Cys residues to facilitate intein-mediated peptide ligation. Primers were used to add *Nde*I and *Sap*I restriction sites to the 5′ and 3′ ends of the PCR products, respectively. The PCR products were inserted into pTXB1 by the same method used to clone *cviQIIM*. After verifying the DNA sequence, the constructs were transformed into *E. coli* ER2566 for expression. Expression procedures were identical to those used for M.CviQII. Cells were harvested, re-suspended in chitin column buffer (20 mM Tris–HCl, 0.5 M NaCl, 0.1% Triton X-100, 1 mM EDTA, pH 8.0), sonicated and centrifuged. Both truncated fusion proteins were soluble and they were applied to a chitin column (20 ml bed volume). After washing the column using chitin column buffer, the fusion proteins were released by incubating the bound protein with 50 mM sodium 2-mercaptoethanesulfonate (MESNA) overnight at 4 °C; the released proteins were eluted from the column with two-bed volumes of chitin column buffer.

The truncated proteins were ligated to C-terminal peptides as follows: 0.01 μmol of protein (0.7 mg) from the chitin column fractions was incubated with 1 μmol of peptide in 1 ml of chitin column buffer with 50 mM MESNA at 4 °C or 25 °C overnight. The ligated proteins were purified through a heparin column (Amersham; 10 ml bed volume) using a NaCl gradient in 20 mM Tris–HCl, pH 7.9. The peptides were synthesized in-house at New England Biolabs (NEB).

The NQN509 protein was subjected to additional purification. Fractions containing NQN509 that eluted from the chitin column were diluted 4-fold with 20 mM Tris–HCl, pH 7.8 such that the NaCl concentration was 125 mM. The diluted protein solution was applied to a Heparin HyperD M column (BioSepra, ~30 ml bed volume), which was then eluted with a linear NaCl gradient [50 mM—1 M]. Fractions containing the truncated protein were pooled and dialyzed against 10 mM potassium phosphate buffer, 50 mM NaCl, 1 mM EDTA, pH 7.0, and applied to a hydroxyapatite column (Bio-Rad Bio-gel HTP; 15 ml bed volume). Proteins were eluted using a linear gradient of 10–500 mM potassium phosphate buffer. Peak fractions were assayed using pBR322 DNA as a substrate, pooled, concentrated, and stored in 20 mM Tris–HCl, 100 mM NaCl, pH 8.0, and 50% glycerol at −20 °C.

In vivo and in vitro M.CviQII MTase activity

Total DNA was isolated from pUCAC-*cviQIIM*-transformed ER2502 to measure M.CviQII MTase activity in vivo. One microgram of total DNA was incubated with 10–20 U of selected REases at 37 °C for 1 h in appropriate buffer conditions as suggested by the manufacturer (NEB). The DNA fragments were analyzed by electrophoresis on 0.8% agarose gels.

To test in vitro activity, 1 μg of pUC19 was incubated with 2 μg of M.CviQII and 80 μM of *S*-adenosyl-methionine (AdoMet) in NEB buffer 2 (10 mM Tris–HCl, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT, pH 7.9) at 37 °C for 1 h in 20 μl reactions. After heat-inactivation at 65 °C for 15 min, the reaction volume was increased to 50 μl in the appropriate buffers (NEB) and incubated with 10–20 U of selected REases at 37 °C for 1 h.

DNA nicking activity assay

DNA nicking activity of the truncated and ligated Nt.CviQII proteins were assayed using pBR322 or pUC19 DNA as substrates. Various amounts of protein were incubated with 1 μg of DNA in NEB buffer 2 (10 mM Tris–HCl, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT, pH 7.9) at specific temperatures for various times. Cleaved DNA products were analyzed by electrophoresis on 1.5% agarose gels. To determine the cleavage specificity of the Nt.CviQII truncation mutants, DNA cleavage fragments were filled-in using Klenow fragment in the presence of 33 μmol dNTPs according to the instructions of the manufacturer (NEB). These fragments were ligated into *Hin*cII- and CIP-treated pUC19. The clones were sequenced to identify the cleavage sites.

DNA nicking activity from full-length and truncated Nt.CviQII produced by cell-free transcription/translation was assayed as follows: one microgram of pUC19 DNA was incubated with 2 μl of the transcription/translation product in NEB buffer 2 and 1 μg of bovine serum albumin (BSA) in a total volume of 20 μl at 30 °C for 4 h. Four microliters samples were collected hourly and the reactions were stopped by adding 5 mM EDTA and 0.1% SDS. Samples were analyzed on 0.8% agarose gels and the intensities of the nicked DNA bands were quantified by Quantity One software (Bio-Rad). The results from two independent experiments were averaged and plotted against time. Curve fitting was done using KaleidaGraph (Synergy Software).

Identification of Nt.CviQII in NY-2A virions

Highly purified and protease K-treated PBCV-1 virions were solubilized and the virus proteins were separated by SDS–PAGE as described previously [11]. [Note: Protease K at room temperature has no effect on PBCV-1 infectivity (Van Etten, unpublished results).] The gel was stained with Sypro Ruby (Molecular Probes) and the revealed bands were excised. The proteins in the gel slices were hydrolyzed by in-gel trypsin digestion [12]. The resulting peptide fragments were eluted from the gel pieces and then analyzed using a Q-TOF Ultima tandem mass spectrometer (Micromass/Waters) with electrospray ionization. The instrument was calibrated using the fragment

ion masses of doubly protonated Glu-fibrinopeptide. The MS/MS data were processed using Masslynx software (Micromass) to produce peak lists for database searching. MASCOT (Matrix Science) was used as the search engine. Data were searched against the NY-2A non-redundant database (Fitzgerald et al., manuscript in preparation).

Results

Cloning and identification of cviQIIM

Clones of *cviQIIM* were obtained by challenging a genomic library of NY-2A with *Mnl*I [CCTC(7/6); G^{m}AGG sequences are not cleaved by *Mnl*I]. M.CviQII was predicted to modify the adenine in RAG sequences. *E. coli* cells containing plasmid DNA that express MTase M.CviQII should be resistant to *Mnl*I cleavage. Sequencing the plasmids from two *Mnl*I partially resistant clones led to the identification of a 269 codon ORF (Figure 1, predicted molecular weight of 31 kDa) whose deduced amino acid sequence indicates it is a m6A MTase [13]. In addition to motifs PXXGXG (motif I) and NPPF (motif IV) that are characteristic of m6A MTases (Figure 2), M.CviQII also contains motifs II, III, V, VI, VIII, and X that are shared by cytosine and adenine MTases. Furthermore, the linear order of the motifs, i.e. motif X followed by I to VIII indicates that M.CviQII is a member of the γ-subtype of m6A MTases (m6A γ MTase) [13].

Members of m6A γ MTases typically have relatively low sequence identity to each other [13, 14]. For example, M.*Taq*I, M.*Hin*cII, M.*Eco*KI, M.*Sbf*I, M.*Vsp*I, M.*Xmn*I, and M.*Pst*I only share about 10% amino acid identity. Sequence identity is also low among some of the chlorella virus m6A γ MTases. M.CviBIII (from chlorella virus NC-1A) and M.CviRI (from chlorella virus XZ-6E) only have 14% and 11% amino acid sequence identity, respectively, with M.CviQII. However, there is 35% amino acid sequence identity between M.CviBIII and M.CviRI. Among all known chlorella virus m6A MTases, M.CviQI (recognition sequence GTAC), also from virus NY-2A [2], is most similar to M.CviQII with 24% amino acid sequence identity.

A BLAST search of GenBank (as of October 2005) identified two putative m6A γ MTases from evolutionarily distant bacteria that resemble M.CviQII (Figure 2); an unidentified ORF from *Sphingopyxis alaskensis* RB2256 (ZP_00579458) and a putative Type II MTase from *Helicobacter pylori* J99 (NP_223729) have 34% and 23% amino acid sequence identity with M.CviXII, respectively. Like M.CviQII, these two putative bacterial m6A γ MTases lack motifs VII and IX, but contain all the other DNA MTase motifs.

Alignment of M.CviQII and M.CviQI amino acid sequences with the two putative bacterial m6A γ MTases revealed a conserved region between motifs VI and VIII (Figure 2). This region, termed motif VII*, contains consensus AF residues that are flanked by charged and hydrophobic residues, respectively. Motif VII* could serve the same function as motif VII which is involved in forming the catalytic region [15].

Identification of cviQIINt

Like bacteria, chlorella virus REase-encoding genes are always located near their cognate DNA MTase gene, although the spacing and relative orientation of the two

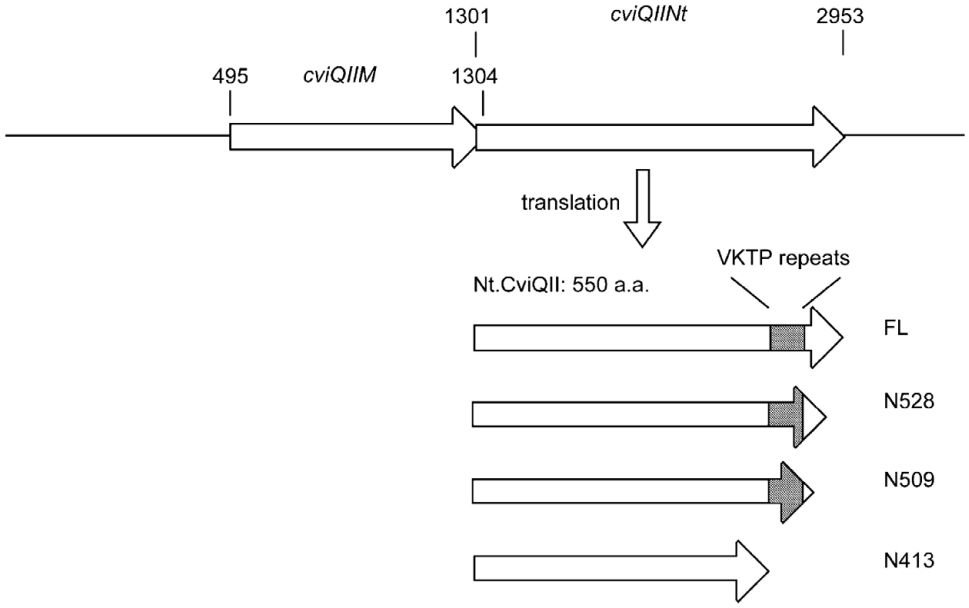

Figure 1. Gene organization of the CviQII N-M system. The MTase gene (*cviQIIM*) precedes the NEase gene (*cviQIINt*) in the same orientation. The M.CviQII ORF spans nucleotides 495–1304 and encodes a protein of 269 amino acids, whereas the Nt.CviQII ORF spans nucleotides 1301–2953 and encodes a protein of 550 amino acids. The first four nucleotides of *cviQIINt* (ATGA) overlap the last four nucleotides of *cviQIIM* such that the *cviQIINt* translational start codon overlaps the *cviQIIM* stop codon. The C-terminal deletion variants N528, N509, and N413 are diagrammed below the wt Nt.CviQII.

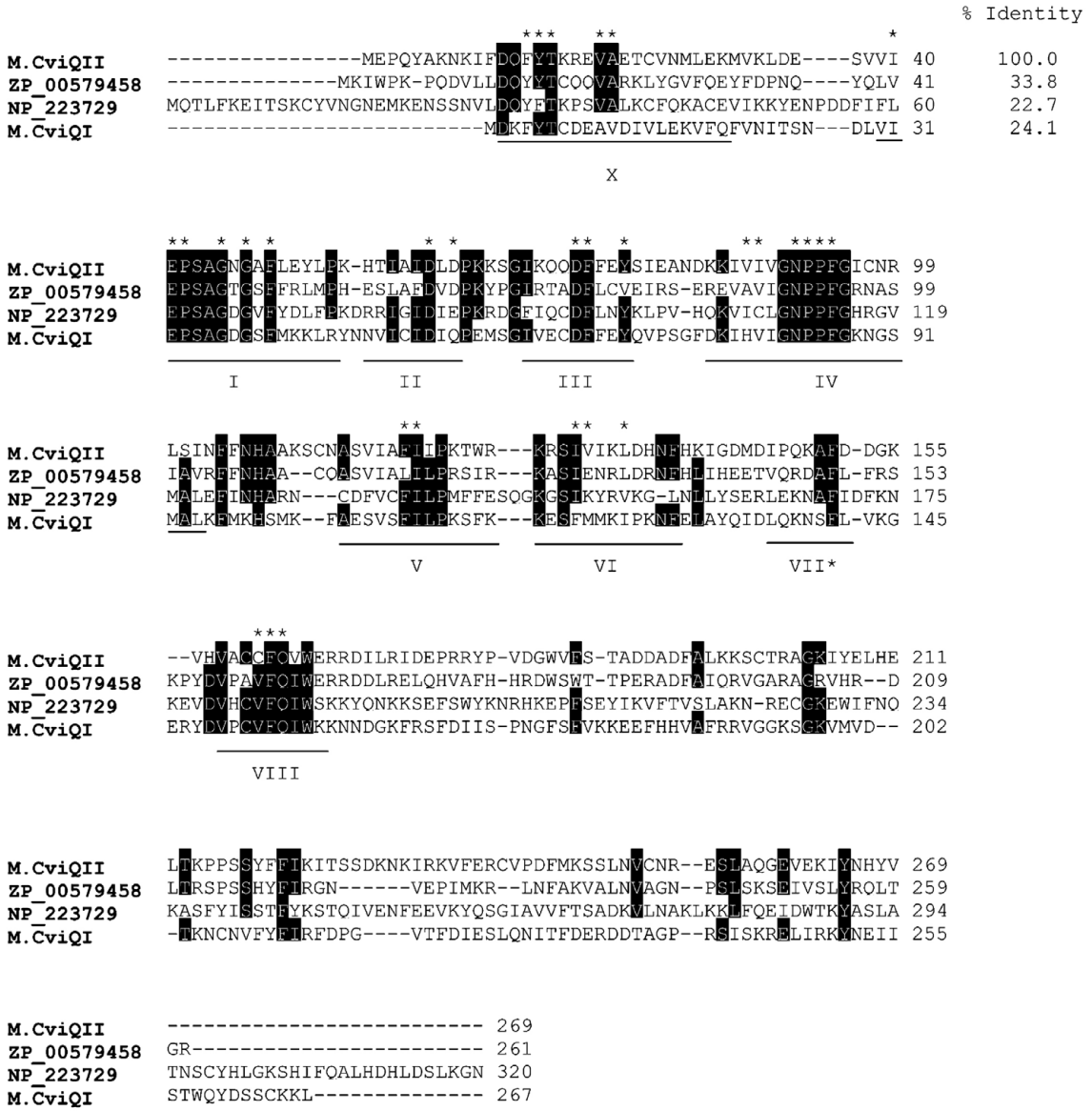

Figure 2. Alignment of M.CviQII amino acid sequence to its closest matches from GenBank. ZP_00579458 and NP_223729 are putative ORFs from *Sphingopyxis alaskensis* RB2256 and *Helicobacter pylori* J99, respectively. M.CviQI is a γ-subtype of m6A MTase isolated from chlorella virus NY-2A with a recognition sequence of GT^{m}AC. DNA MTase motifs are underlined. Motif VII is absent in all four proteins; however, all four proteins have a similar amino acid sequence, labeled motif VII*. Identical residues are shown in black boxes and conserved residues in MTase motifs are indicated by an asterisk.

genes can vary [14]. Therefore, we sequenced the NY-2A genomic DNA adjacent to *cviQIIM* and identified an ORF of 550 codons (Figure 1, in the same orientation as the *cviQIIM* gene but in a different reading frame) that overlaps *cviQIIM* by four nucleotides. As a result, the stop codon of M.CviQII (ATGA) overlaps the Nt.CviQII start codon (ATGA). The Nt.CviQII ORF (predicted molecular weight 63 kDa) has low sequence identity to all known site-specific endonucleases except two from chlorella viruses. Nt.CviQII has 34% and 23% amino acid sequence identity with Nt.CviPII and CviJI [CviJI short form, without the EEKKR repeats at the N-terminus, [16]], respectively (Figure 3). Nt.CviPII is a three-base NEase (↓CCD) from chlorella virus NYs-1 [6], whereas, CviJI, from chlorella virus IL-3A, produces dsDNA breaks (RG ↓ CY) with blunt ends [16] and [17]. There is one Type II endonuclease catalytic motif in Nt.CviQII, $P_{237}D_{238}$-$X_{[25]}$-$D_{264}GK_{266}$ (Figure 3) [18]. This catalytic motif can be aligned to one of the putative catalytic motifs of Nt.CviPII ($A_{234}E_{235}$-$X_{[22]}$-$D_{258}GK_{260}$) [6].

Nt.CviQII is at least 150 amino acids larger than all other known chlorella virus site-specific endonucleases. This increased length is due primarily to many tandem repeats located between amino acids 414 and 485: 14 repeats of VKTP, 3 repeats of VKSP, 1 VLTP, and 1 VKIL (Figure 3). Cell-free transcription/translation of a truncated mutant of Nt.CviQII indicates that the tandem repeat sequences are not required for DNA nicking activity (see below). The biological role of the short amino acid repeats, if any, remained to be studied.

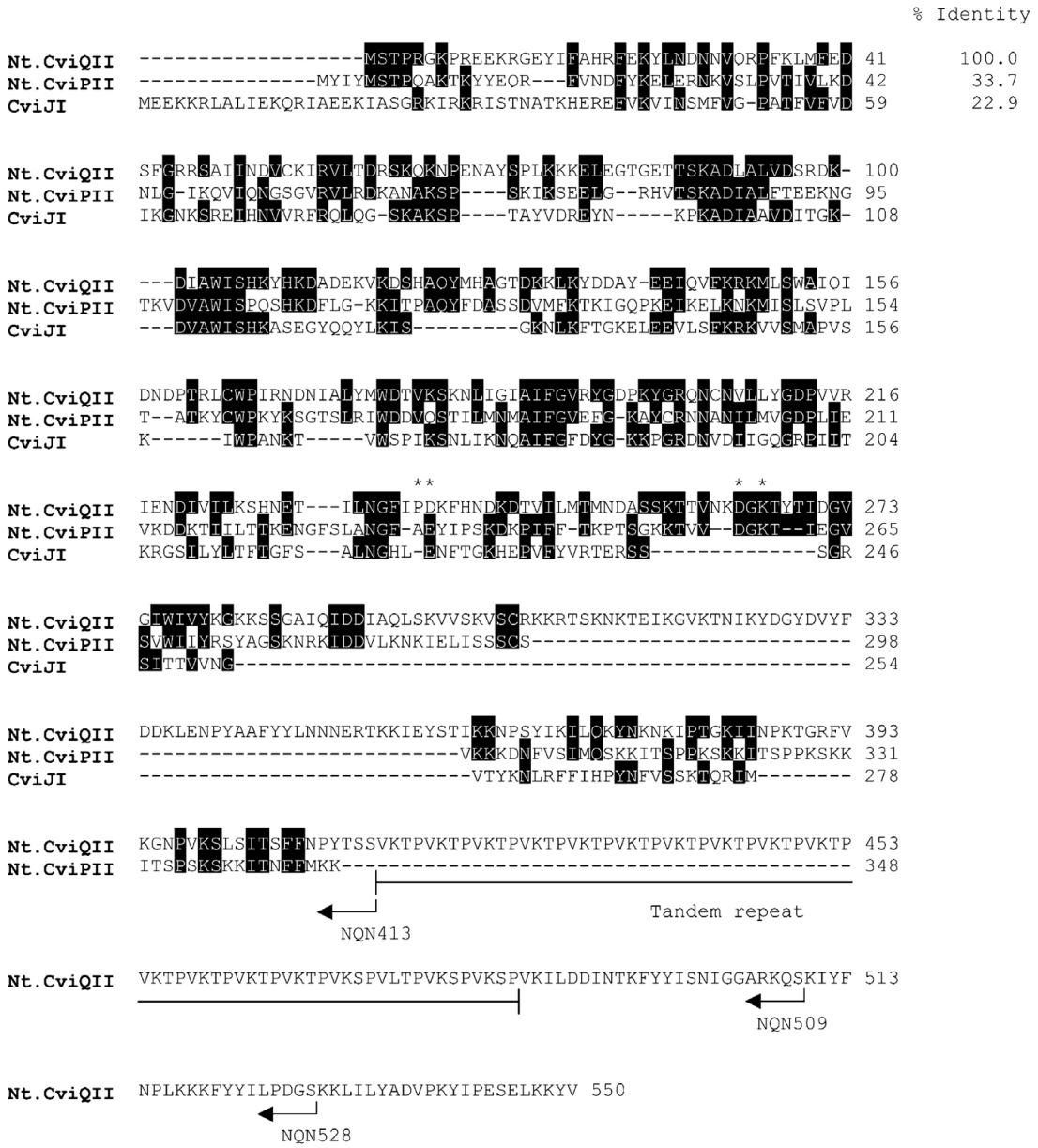

Figure 3. Alignment of Nt.CviQII amino acid sequences with its closest matches in GenBank. Nt.CviPII (↓CCD) and *Cvi*JI (RG ↓ CY) are from chlorella viruses NYs-1 and IL-3A, respectively. The putative active site residues in Nt.CviQII indicated by an asterisk. The asterisks also indicates the putative active site residues of Nt.CviPII, which can be aligned to that of Nt.CviQII. Tandem repeats of VKTP are indicated. The terminal residues of truncation mutants NQN413, NQN509, and NQN528 are indicated by arrows.

M.CviQII MTase activity

The genomic library of NY-2A was generated by ligating *Sau*3AI (↓ GATC) fragments of NY-2A genomic DNA to *Bam*HI (G↓ GATCC) cleaved pUCAC. Because there is a *Sau*3AI site within the ORF of *cviQIIM* near the 5′ end, the resulting pUCAC-*cviQIIM* clone generated a modified form of M.CviQII such that the N-terminal 12 amino acid residues of M.CviQII were replaced by the N-terminal 16 amino acid residues of LacZ α fragment. To determine if bacterial DNA was methylated by this modified M.CviQII in the transformed cells, total DNA (genomic and plasmid pUCAC-*cviQXIM* DNAs) was isolated from the transformed cells and incubated with REases containing recognition sequences overlapping NAG: *Hin*dIII (A ↓ AGCTT) and *Afl*II (C ↓ TTAAG) for AAG; *Mnl*I (↓N_6GAGG) and *Ple*I (GAGTCN_4↓) for GAG; *Pvu*II (CAG ↓ CTG) for CAG and *Xba*I (T ↓ CTAGA) for TAG; *Sca*I (AGT ↓ ACT) and *Eco*NI (CCTNN ↓ NNNAGG) for NAG. Two REases, *Dra*I (CACNN ↓ GTG) and *Xmn*I (GAANN ↓ NNTTC), with unrelated recognition sequences were also tested on the *E. coli* DNA as controls (Figure 4A). The DNA was largely resistant to *Hin*dIII, *Afl*II, *Mnl*I, *Ple*I, *Pvu*II, *Xba*I, *Sca*I, and *Eco*NI cleavage but not to *Dra*I and *Xmn*I. There-

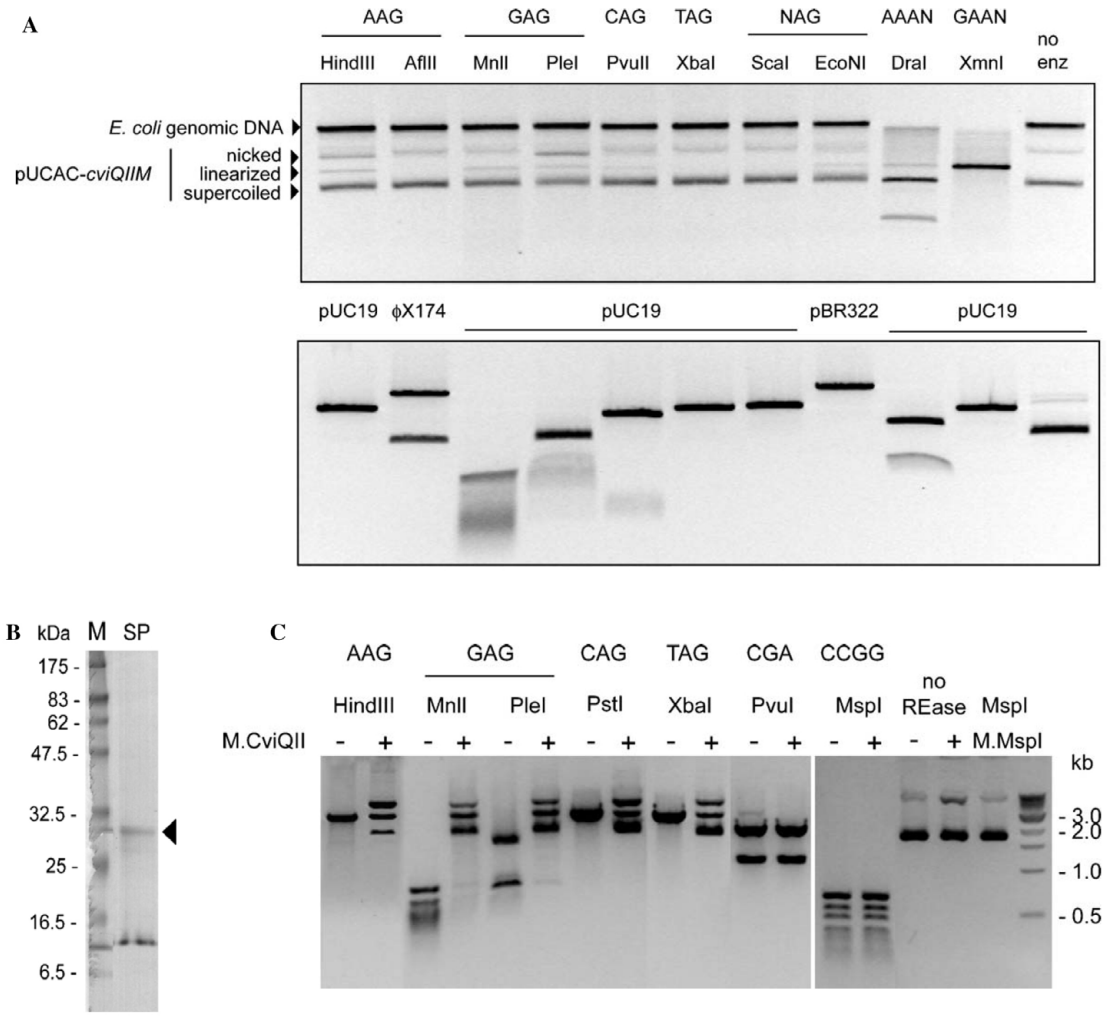

Figure 4. (A) In vivo protection of *E. coli* DNA by M.CviQII. Total DNA isolated from pUCAC-*cviQII*-transformed ER1992 was challenged with REases with NAG recognition sequences (upper panel). The *E. coli* DNA, nicked, linearized, and supercoiled forms of pUCAC-*cviQIIM* are indicated. The lower panel shows that the REases cleaved unmodified pUC19, pBR322, or φX174 ds-DNA. (B) Purification of M.CviQII. The insoluble M.CviQII-intein-CBD fusion protein was refolded and intein-mediated cleavage was induced by DTT. The protein solution was then loaded onto a SP column and eluted using a linear gradient of NaCl. The arrowhead indicates the band corresponds to M.CviQII on a 10–20% SDS–PAGE. (C) In vitro activity of M.CviQII. Substrate pUC19 DNA was incubated with or without M.CviQII in the presence of AdoMet followed by incubation with the indicated REases. Partial recognition sequences are shown on top of the corresponding REases. Substrate pUC19 DNA incubated with M.MspI, followed by *Msp*I digestion, served as a control.

fore, M.CviQII modifies the adenine in AAG, GAG, CAG, and TAG sequences, which means it methylates AG sequences. The presence of a small amount of linearized plasmid DNA in the *Hin*dIII, *Mnl*I, *Ple*I, and *Pvu*II reactions indicates that M.CviQII does not modify every AG site in *E. coli*.

Full-length M.CviQII was also cloned and expressed as a recombinant fusion protein. The *cviQIIM* gene was ligated to pTXB1 to produce a C-terminal fusion protein containing the Mxe GyrA intein and chitin-binding domain (CBD). However, the expressed fusion protein was insoluble. After denaturing and refolding, the fusion protein was cleaved by the intein-mediated reaction and the protein was purified by ion exchange chromatography to ~60% purity (Figure 4B). This partially purified protein was assayed for MTase activity using pUC19 DNA as a substrate. After incubating pUC19 DNA with the partially purified M.CviQII and AdoMet, the DNA product was incubated with appropriate REases to determine if specific sites were protected (Figure 4C). Similar to the results of the in vivo experiments, M.CviQII partially protected AAG (against *Hin*dIII), GAG (against *Mnl*I and *Ple*I), CAG (against *Pst*I) and TAG (against *Xba*I) sites from cleavage. These same REases cleaved control pUC19 into the expected fragments. *Msp*I (C ↓ CGG) and *Pvu*I (CGAT ↓ CG), which are not inhibited by m6A, cleaved the M.CviQII-modified pUC19 to completion. However, like the in vivo methylation experiments, M.CviQII protection was not complete and both nicked and linearized plasmid DNAs were observed in all of the reactions.

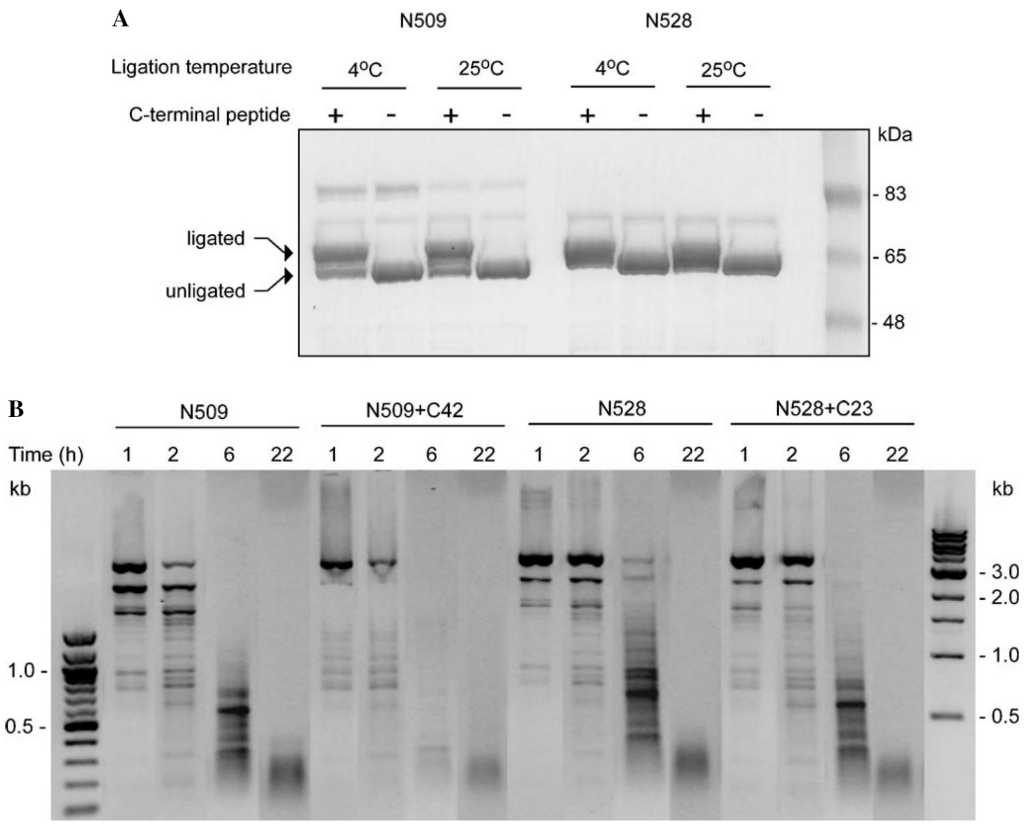

Figure 5. (A) Intein-mediated peptide ligation. Truncated Nt.CviQII mutants NQN509 and NQN528 were ligated to peptides C42 and C23, respectively, by intein-mediated peptide ligation at 4 °C or 25 °C for 22 h. Ligated and unligated proteins are indicated. (B) DNA nicking activity of the truncation mutants and ligated proteins incubated for various times at 30 °C. One microgram of pUC19 DNA was incubated with 0.3 µg of the proteins and analyzed on a 1.5% agarose gel.

Cloning Nt.CviQII and construction of deletion mutants

The standard procedure for cloning R-M systems is to protect the host DNA by expressing the MTase gene before cloning the REase [19]. However, due to the high number of Nt.CviQII nicking sites in the host genome, cloning *cviQIINt* was expected to be difficult. We attempted to protect the host DNA by over-expressing M.CviQII before introducing *cviQIINt* into *E. coli*. The *cviQIIM* gene containing an optimal ribosome binding site and spacing was ligated to pACYC184 or pLT7K [20] because these two plasmids are compatible with the plasmids used to express *cviQIINt*. However, plasmid DNA isolated from *E. coli* ER2566 transformed with these two *cviQIIM* containing plasmids was only partially protected against *Mnl*I cleavage (data not shown).

Full-length *cviQIINt* was amplified from NY-2A genomic DNA by PCR and ligated into several plasmids including: (i) pR976 (vector constructed by P. Riggs, NEB), a low copy plasmid with a P*tac* promoter used to successfully clone Nt.CviPII [6] in the presence of pUCAC-*cviQIIM*, (ii) pLT7K in the presence of pACYC184-*cviQIIM*, and (iii) a single-copy plasmid with inducible copy number that was derived from pBAC with a P*tac* promoter (pBACtac, unpublished results, J.C. Samuelson, NEB). All attempts to clone *cviQIINt* into *cviQIIM* containing cells were unsuccessful, either no inserts were obtained or the inserts had altered *cviQIINt* sequences.

We then tried to express C-terminal truncation mutants of Nt.CviQII in *E. coli* and reconstruct the full-length protein by ligating C-terminal peptide sequences to the recombinant proteins using an intein-mediated peptide ligation system (NEB). Two Nt.CviQII truncation mutants were constructed. The C-terminal 41 and 22 amino acid residues were deleted in NQN509 (509 amino acids) and NQN528 (528 amino acids), respectively (Figure 3). Ser509 of NQN509 and Ser528 of NQN528 were mutated to Cys to facilitate intein-mediated peptide ligation. The two proteins were expressed as C-terminal fusions to intein Mxe GyrA followed by a CBD, which increased the size of the recombinant truncated proteins by ~26 kDa. The resulting constructs pTXB1-NQN509 and pTXB1-NQN528 were expressed in *E. coli* ER2566. The expression levels were relatively high and independent of the presence of the *cviQIIM* gene in the host cells. Proteins that eluted from the chitin column were ligated to peptides with an N-terminal Cys in the presence of reducing agent (50 mM MESNA) at 4 °C and 25 °C. About 50% of the truncated proteins were ligated to the peptides (Figure 5A). The ligated proteins were further purified by heparin column chromatography and assayed for DNA nicking activity.

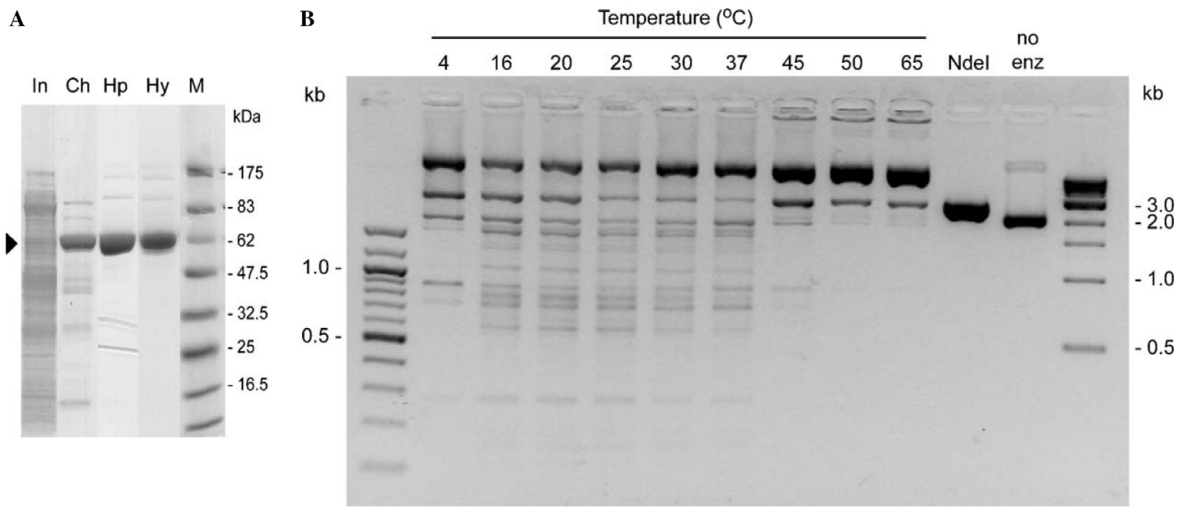

Figure 6. Purification and DNA nicking activity of truncation mutant NQN509. (A) The lysate supernatant from 2 L of expression culture (In) was loaded onto a chitin column. After intein-mediated cleavage, the mutant protein was eluted (Ch). The eluted protein was diluted, and loaded onto a heparin column. The eluted protein was pooled (Hp) and dialyzed against potassium phosphate buffer. The dialyzed protein solution was loaded onto a hydroxyapatite column and the eluted protein was concentrated and stored (Hy). The arrowhead indicates the band that corresponds to NQN509 on a 4–20% SDS–PAGE. (B) One hundred units of the purified NQN509 were incubated with 1 μg of pUC19 at the indicated temperatures for 1 h. pUC19 linearized by *Nde*I or without enzyme treatment served as controls for linearized and supercoiled DNA, respectively.

DNA nicking activity of Nt.CviQII and its truncation mutants

DNA nicking activity of the Nt.CviQII truncated mutants and reconstructed wild-type proteins were assayed using the 2686 bp closed circular pUC19 DNA as a substrate; the two DNA strands of the plasmid contain 180 RAG sites. Complete cleavage of the plasmid should result in ssDNA fragments of ~50 nucleotides. One microgram (0.56 pmol) of pUC19 DNA was incubated with 4 μg (60 pmol) of protein at 30 °C. All of the proteins exhibited DNA nicking activity; however, none of them cleaved pUC19 to ~50 nucleotides in 1 h (Figure 5B). Therefore, longer incubations were used. The ligated NQN509 + C42 protein had the highest DNA nicking activity in the first 2 h and its cleavage products became smaller with time (Figure 5B). Both ligated and unligated NQN528 proteins have similar activity. After incubating for 22 h, the cleavage products were ~100–200 bp, indicating that the reaction did not go to completion. Incubating the cleavage products with fresh NQN509 cleaved the DNA fragments into smaller pieces, indicating that Nt.CviQII was inactivated during extended incubation before the cleavage reactions were completed (data not shown).

Since a stable cleavage pattern was not obtained in 2 h, one enzyme unit of Nt.CviQII and its deletion mutants was defined as the amount of protein needed to give 50% cleavage of 1 μg of closed-circular pUC19 DNAs to a nicked circular form in 1 h. Using this definition, the specific activities of NQN509 and NQN509 + C42 were 107,000 U/mg and 213,000 U/mg, respectively. The specific activities of NQN528 and NQN528 + C23 were both 80,000 U/mg (results not shown). The lower specific activities of NQN528 and its ligated protein may be due to either altering protein conformation or the Ser528 to Cys substitution.

Protein NQN509 was chosen for further study because of its relatively high specific activity and it was easier to produce. The chitin column fractions of NQN509 were pooled and further purified through heparin and hydroxyapatite chromatography. A total of 15.6 mg of purified protein with 3×10^6 U were obtained (Table 1). The purification is only 1.4-fold compared to the chitin column chromatograph because relatively high purity can be obtained by the combination of chitin column and intein-mediated cleavage (Figure 6A). One hundred units of purified NQN509 were incubated with 1 μg of pUC19 at

Table 1. Purification of NQN509

Steps	Total protein[a] (mg)	Total activity[a] (U)[b]	Specific activity (U/mg)	Purification Yield (%)	(fold)
Lysate supernatant	350	—	—	—	—
Chitin column	29	4.0×10^6	138×10^3	100	1.0
Heparin	20	3.2×10^6	160×10^3	80	1.2
Hydroxyapatite	15.6	3.0×10^6	192×10^3	75	1.4

[a] The numbers are per Liter culture.
[b] One enzyme unit is defined as the amount of enzyme needed to convert half of 1 μg of closed-circular form of pUC19 to nicked circular form.

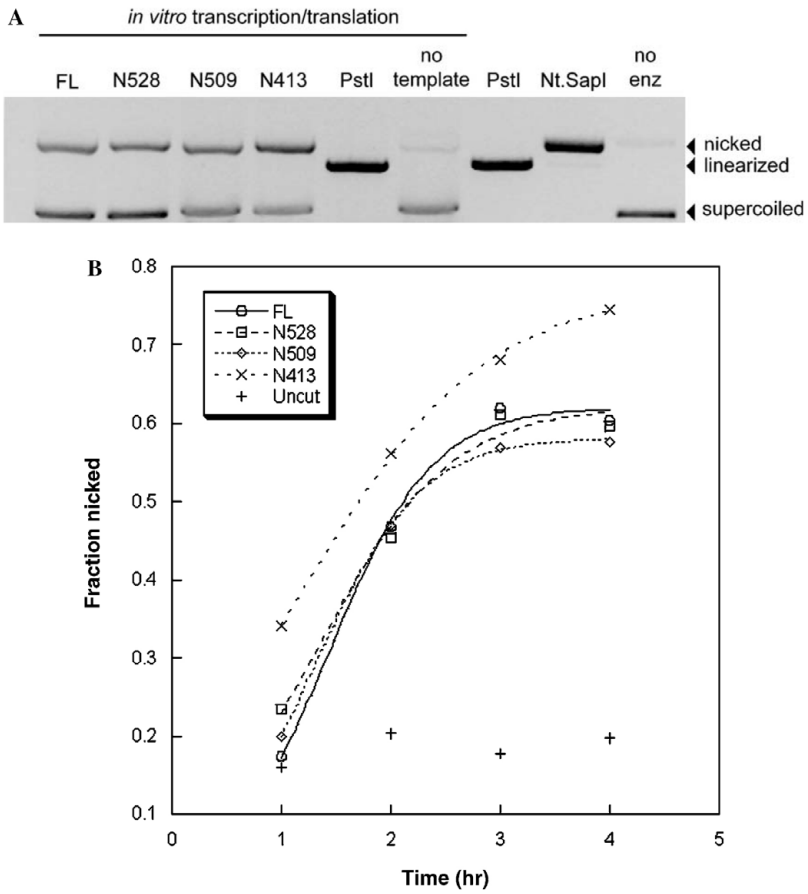

Figure 7. (A) DNA nicking activity of wt and truncation mutants of Nt.CviQII produced by cell-free transcription/translation. ORFs of full-length wt, truncation mutants NQN528 (N528), NQN509 (N509), and NQN413 (N413) were amplified from NY-2A genomic DNA in PCR. T7 promoter and a ribosome-binding sequence were added 5′ to the ORFs by appropriate primers. A *Pst*I ORF served as a positive control for the in vitro transcription/translation system. Cell-free transcription/translation products were incubated with pUC19 DNA at 30 °C for 2 h. pUC19 DNA was cleaved by *Pst*I and Nt.SapI to produce linearized and nicked DNA, respectively. (B) Comparison of reaction rates of wt Nt.CviQII, NQN528, NQN509, and NQN413. The DNA nicking activity assay was continued for 4 h and the intensities of nicked and supercoiled DNA from two independent experiments were averaged. Fractions of nicked DNA were plotted against time and curve fitting was done using KaleidaGraph.

temperatures from 4 to 65 °C for 1 h (Figure 6B). The enzyme exhibited optimal activity at 16–25 °C with reduced activity above 45 °C.

To compare the activity of the truncation mutants to wt Nt.CviQII, a small quantity of wt enzyme was produced in a cell-free transcription/translation system, because we were unable to clone and express wt Nt.CviQII in *E. coli*. Three Nt.CviQII truncation mutants were produced by the same system in parallel: NQN528, NQN509, and NQN413 (with the last 137 codons deleted). The synthesized proteins contained a His-tag and their presence in the resulting transcription/translation reaction was confirmed by immuno-blotting using an anti-His-tag antibody (data not shown). NQN528 and NQN509 have similar reaction rates as wt Nt.CviQII over 4 h (Figure 7). Unexpectedly, NQN413 has a slightly higher initial rate than the wt enzyme. Because NQN413 lacks the entire C-terminal tandem repeat sequences, we conclude that the Nt.CviQII C-terminal 137 amino acids are not required for DNA nicking activity.

Nt.CviQII nicking sites

To determine the Nt.CviQII cleavage sites, plasmid pBR322 was cleaved with dilutions of NQN509 and NQN509 + C42 at 30 °C for 1 h. The cleaved fragments were filled-in using the Klenow fragment and cloned into pUC19. Sequencing the junctions of 10 cloned pBR322 fragments indicated that the two enzymes cleaved the plasmid 5′ to the middle adenine at AAG or GAG sites (data not shown). The length of the cleavage products ranged from 60 to 300 bp with an average of 165 and 172 bp for NQN509 and NQN509 + C42, respectively (data not shown). The 5′ end always started with AG and the 3′ end always ended with CT (the reverse complement of AG) in all the cloned fragments. These results indicate that both NQN509 and NQN509 + C42 only cleaved plasmid DNA in one of the strands rather than both because there would be a junction of TY … RA for staggered end cleavage or Y … R for blunt end cleavage. Moreover, only the RAG containing strand (not YTC) was cleaved. Oth-

erwise, some fragments would end as TY … RA. The results agree with experiments reported previously where 2 RAG sites were sequenced [2].

Nt.CviQII is packaged in NY-2A virions

Attempts to identify all of the virus-encoded proteins that are packaged in the NY-2A virion (Dunigan et al., manuscript in preparation) indicated that Nt.CviQII is packaged in the virion, whereas its corresponding MTase M.CviQII is not. The identified peptide fragments cover 32% of the Nt.CviQII protein with Mowse scores of high confidence. One concern was that the enzyme might be a contaminate protein on the surface of the virions. However, virions-treated with protease K during their purification did not remove Nt.CviQII. [Note: Most proteases have little or no effect on PBCV-1 infectivity (Van Etten, unpublished results)]. These results suggest that Nt.CviQII has a physiological function during an early phase of virus infection. Separate studies indicate that NY-2A infection causes rapid degradation of host chlorella chromosomal DNA (Agarkova et al., submitted for publication). Nt.CviQII could be involved in this degradation process which may aid in either shutting off host transcription during infection, allow for a recycling of nucleotides into virus progeny DNA, and/or prevent subsequent secondary virus infections by destroying the competing viral DNA.

Discussion

One unusual property of chlorella viruses is that they encode multiple Type II DNA R-M systems [1, 21]. Virus NY-2A has the most heavily methylated DNA among chlorella viruses (45% of the cytosines are m5C and 37% of the adenines are m6A) and consequently its DNA is resistant to most REases [3]. In the current manuscript we demonstrated that M.CviQII methylates AG sequences in vitro and in vivo in a heterologous host (*E. coli*). Because M.CviQII modifies a dinucleotide sequence, it can account for ~25% of the adenine methylation in NY-2A. To our knowledge M.CviQII is the first m6A MTase to recognize a dinucleotide sequence. This property of M.CviQII can be used to study the sensitivity of restriction endonucleases with overlapping AG sites.

After most of the experiments described in this manuscript were completed, the 370 kb NY-2A genome sequence became available (http://greengene.uml.edu). Amazingly, NY-2A contains 18 putative MTase-encoding genes (Table 1) (Fitzgerald et al., manuscript in preparation). In addition to the CviQII N-M system described in this report, NY-2A encodes another R-M system, named CviQI. CviQI is a typical Type II REase that cleaves 5′ of the T in G ↓ TAC sequences [2, 4]. Its cognate MTase M.CviQI is also a m6A γ MTase that modifies adenine in its recognition sequence. The *cviQIR* gene was cloned and expressed in *E. coli* before the viral genome was sequenced (Zhu and Xu, unpublished results).

Previously the methylation sites of eight other NY-2A MTases were identified by either cloning the gene (M.CviQIII, M.CviQV, M.CviQVI, and M.CviQVII) or deducing the site from the resistance of NY-2A DNA to selected REases (M.CviQIV, M.CviQVIII, M.CviQIX, and M.CviQX) [2]. In fact, the M.CviQII methylation site was inferred originally from the sensitivity/resistance of NY-2A DNA to selected REases; the predicted site (RAR) was nearly correct. Now knowing that NY-2A has 18 MTases-encoding genes, it is likely that the methylation site attributed to M.CviQII previously results from two or more MTases with overlapping sequence specificities [2].

In addition to the CviQI R-M and CviQII N-M systems, the 16 remaining MTases, which are located throughout the NY-2A genome, may be orphan MTases because no ORFs near these MTase-encoding genes resemble known REases or NEases (Table 2) . The NY-2A gene number assigned to the 18 MTases, as well as their closest homologs from other chlorella viruses, are listed in Table 1. Eleven of the 18 MTases are predicted to be m6A MTases and seven are predicted to be m5C MTases. Of the 11 m6A MTases, seven are m6A γ MTases, three are m6A α MTases, and one is a m6A β MTase. M.CviQXV resembles M.Hpy188I more than chlorella MTases.

Twelve of the 18 NY-2A MTases have >45% amino acid identity with MTases from other chlorella viruses (Table 1). In particular, M.CviQVIII and M.CviQIV have 99% and 89% amino acid identity to M.CviPI and M.CviPII, respectively, from chlorella virus NYs-1 [2, 22]. The high sequence identity among MTases from different viruses suggests that the MTase genes have been shuffled among the viruses during the course of evolution. Gene duplication within a virus, however, is not very common because none of the NY-2A MTases have more than 50% amino acid identity with each other.

Including the two site-specific endonucleases, at least 20 of NY-2A's putative 400 protein-encoding genes (5% of the genes) produce proteins involved in R/N-M functions; presumably they must serve some physiological function(s). The Nt.CviQII enzyme is packaged in the virion and may be involved in degrading host chromosomal DNA, which occurs almost immediately after NY-2A infection (Agarkova et al., submitted for publication). However, other related viruses that infect the same chlorella host survive with fewer R-M encoding genes. For example, the prototype chlorella virus, PBCV-1 has five MTase-encoding genes, two of which have companion REases (CviAI and CviAII) [1].

NY-2A not only encodes many putative genes involved in R-M functions, it also has 28 putative homing endonuclease-encoding genes and six putative transposase-encoding genes (Fitzgerald et al., manuscript in preparation). Currently, it is not known if any of these genes encode functional enzymes. Assuming that at least some of these genes encode functional proteins, one might expect the NY-2A genome to be unstable. However, we have grown NY-2A for about 20 years in the laboratory and have not observed any evidence of genome instability.

Table 2. DNA methyltransferases from virus NY-2A

Name	NY-2A ORF[d]	Recognition sequence	Chlorella virus homologs	aa identity (%)	Type
M.CviQI[a]	B543L	GT^{m}AC[a]	–	–	m6Aγ
M.CviQII[a]	B359R	^{m}AG[a]	–	–	m6Aγ
M.CviQIII[a]	B16L	TCG^{m}A[a]	M.CviBIII	75	m6Aγ
M.CviQIV[b]	B774R	G^{m}ATC[b]	M.CviAI	81	m6Aα
M.CviQV[a]	B230L	TGC^{m}A[a]	M.CviRI	79	m6Aγ
M.CviQVI[a]	B399R	G^{m}ANTC[a]	M.CviBI	82	m6Aα
M.CviQVII[a]	B236L	C^{m}ATG[a]	M.CviAII	80	m6Aα
M.CviQVIII[b]	B769R	G^{m}C[b]	M.CviPI	99	m5C
M.CviQIX[b]	B681L	^{m}CCD[b]	M.CviPII	89	m5C
M.CviQX[c]	B8R	^{m}CGR[b]	–	–	m5C
M.CviQXI[c]	B10R	–	–	–	m6Aβ
M.CviQXII[c]	B88L	–	M.CviAIII	49	m5C
M.CviQXIII[c]	B411L	G^{m}C[c]	M.CviPI	45	m5C
M.CviQXIV[c]	B416R	TGC^{m}A[c]	M.CviSI	65	m6Aγ
M.CviQXV[c]	B418R	TCNG^{m}A[e]	–	–	m6Aγ[e]
M.CviQXVI[c]	B566R	–	–	–	m5C
M.CviQXVII[c]	B567L	TGC^{m}A[c]	M.CviSI	85	m6Aγ
M.CviQXVIII[c]	B697R	–	M.CviAIV	85	m5C

[a] Proteins cloned and shown to have the indicated activity.
[b] Recognition sequences deduced from the sensitivity/resistance of NY-2A DNA to selected REases.
[c] MTases are putative.
[d] The DNA sequences of the ORFs are available at http://greengene.uml.edu/.
[e] The recognition sequence and subtype of M.CviQXV are deduced from its homologous MTase M.Hpy188I, which shows 41% sequence identity with M.CviQXV.

Recombinant M.CviQII does not methylate every AG sequence in substrate DNA in vitro because all eight REases that are inhibited by m6A in AG sequences can partially cleave M.CviQII-treated DNA. This lack of complete protection is unlikely to be due to sub-optimal reaction conditions (such as buffer composition or concentration of AdoMet) because permutations of various reaction conditions did not improve the protection (results not shown). Likewise, both plasmid and chromosomal DNAs isolated from *E. coli* transformed with the *cviQIIM* gene on high copy plasmids and controlled by a constitutive *lac* promoter (pUCAC-*cviQIIM*) were only partially resistant to these same eight REases. However, NY-2A genomic DNA is completely resistant to Nt.CviQII. Therefore, either M.CviQII methylates all of the AG sites of the virus DNA during virus replication in its natural hosts (chlorella cells), or other NY-2A MTases methylate overlapping AG sites. For example, the recognition sequence of M.CviQVI has been deduced to be G<u>A</u>NTC [2].

Nt.CviQII is 150–200 amino acids larger than all the other chlorella virus site-specific nucleases described to date. This increased size is due primarily to a tetrapeptide tandem repeat VKTP sequence located near the C-terminus (amino acids 414–485). A BLAST search of the repeat sequence against proteins in GenBank revealed some similar, but not identical, tandem repeats in uncharacterized ORFs from the fungus *Gibberella zeae* pH 1, the cellular slime mold *Dictyostelium discoideum*, a putative ATP-dependent *Acyl*-CoA synthetase from protozoan *Plasmodium falciparum* 3D7 and a hypothetical protein from chimpanzee *Pan troglodytes*.

The function of the Nt.CviQII tandem repeat is unknown. In vitro transcription/translation experiments established that it is not required for NEase activity because Nt.CviQII truncation mutants have similar NEase activity as the full-length protein. However, after writing this manuscript, a clone of full-length Nt.CviQII fused to the intein and CBD was obtained after a few more attempts of cloning. To our surprise, DNA nicking activity was detected in the membrane fraction of the induced cells. DNA nicking activity also occurred in the membrane fraction of the truncated mutants NQN509 and NQN528, albeit to a lesser extent. Finding the recombinant full-length Nt.CviQII in the membrane fraction suggests that the C-terminal sequence (the tetrapeptide repeats and the proceeding sequence) may contribute to the membrane localization of the protein. Since mutants NQN509 and NQN528 also have all of the tetrapeptide repeats, the C-terminal sequence may target the protein to the membrane. The tetrapeptide repeats could function to insert the protein into the membrane. Additional mutagenesis experiments will be conducted to explore this possibility.

Nt.CviQII is part of a short list of natural DNA nicking enzymes. It joins Nt.CviPII from chlorella virus NYs-1 in recognizing and nicking at very short recognition sites. It is interesting that Nt.CviQII has 34% amino acid identity with Nt.CviPII even though they are sensitive to different methylated nucleotides, i.e. Nt.CviQII is blocked by m6A and Nt.CviPII by m5C. Most REases and NEases from chlorella viruses recognize and cleave short (3–5 bp) recognition sequences (REBASE). Naturally occurring nicking endonucleases from other organisms, such as

Nt.BstNBI (GAGTCN$_4$↓) [23], Nb.BsrDI(↓CATTGC) and Nb.BtsI (↓CACTGC) from *Bascillus* sp. recognize and nick 5–6 bp recognition sequences (Xu and Zhu, unpublished results). Other nicking endonucleases such as Nt.SapI (GCTCTTCN↓) and Nt.AlwI (GGATCN$_4$↓) have been created by genetically engineering known REases [24, 25].

The G + C content of the *cviQIIM* gene is 40%, that is the same as the NY-2A genome. However, the G + C content of the *cviQIINt* gene is 36% and if the tandem repeats are removed the G + C content is 33%. A similar phenomenon was observed with the CviPII N-M system in which *cviPIINt* is 32% G + C and *cviPIIM* is 43% G + C. In general, the G + C content of most chlorella virus genes is about 40%. The relatively low GC contents of the two NEase genes suggest that they might have been acquired separately from their companion MTase genes.

Acknowledgments

We thank Richard Roberts, Iain Murray, and James Samuelson for helpful discussions and critical reading of the manuscript, Tom Evans for advice on the construction of C-terminal deletion variants, Jim Gurnon for isolating NY-2A DNA, Don Comb for support and encouragement. This work was partly supported by NIH National Human Genome Research Institute phase II STTR Grant 9R42HG003976-02. The CviQII N-M DNA sequences have been deposited in GenBank with Accession No. DQ355510.

References

[1] J. L. Van Etten, Unusual life style of giant chlorella viruses, *Annu. Rev. Genet.* **37** (2003), pp. 153–195.

[2] Y. N. Zhang, M. Nelson, J. Nietfeldt, Y. Xia, D. Burbank, S. Ropp and J. L. Van Etten, Chlorella virus NY-2A encodes at least 12 DNA endonuclease/methyltransferase genes, *Virology* **240** (1998), pp. 366–375.

[3] A. M. Schuster, D. E. Burbank, B. Meister, M. P. Skrdla, R. H. Meints, S. Hattman, D. Swinton and J. L. Van Etten, Characterization of viruses infecting a eukaryotic chlorella-like green alga, *Virology* **150** (1986), pp. 170–177.

[4] Y. Xia, K. E. Narva and J. L. Van Etten, The cleavage site of the *Rsa*I isoschizomer, *Cvi*II, is G ↓ TAC, *Nucleic Acids Res.* **15** (1987), p. 10063.

[5] Y. N. Xia, R. Morgan, I. Schildkraut and J. L. Van Etten, A site-specific single strand endonuclease activity induced by NYs-1 virus infection of a chlorella-like green alga, *Nucleic Acids Res.* **16** (1988), pp. 9477–9487.

[6] S. H. Chan, Z. Zhu, J. L. Van Etten and S. Y. Xu, Cloning of CviPII nicking and modification system from chlorella virus NYs-1 and application of Nt. CviPII in random DNA amplification, *Nucleic Acids Res.* **32** (2004), pp. 6187–6199.

[7] I. Szomolanyi, A. Kiss and P. Venetianer, Cloning the modification methylase gene of *Bacillus sphaericus* R in *Escherichia coli*, *Gene* **10** (1980), pp. 219–225.

[8] R. J. Roberts, T. Vincze, J. Posfai and D. Macelis, REBASE–restriction enzymes and DNA methyltransferases, *Nucleic Acids Res.* **33** (2005), pp. D230–D232.

[9] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* **25** (1997), pp. 3389–3402.

[10] T. C. J. Evans, J. Benner and M. Q. Xu, Semisynthesis of cytotoxic proteins using a modified protein splicing element, *Protein Sci.* **7** (1998), pp. 2256–2264.

[11] M. P. Skrdla, D. E. Burbank, Y. Xia, R. H. Meints and J. L. Van Etten, Structural proteins and lipids in a virus, PBCV-1, which replicates in a chlorella-like alga, *Virology* **135** (1984), pp. 308–315.

[12] A. Shevchenko, M. Wilm, O. Vorm and M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, *Anal. Chem.* **68** (1996), pp. 850–858.

[13] T. Malone, R. M. Blumenthal and X. Cheng, Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes, *J. Mol. Biol.* **253** (1995), pp. 618–632.

[14] G. G. Wilson and N. E. Murray, Restriction and modification systems, *Ann. Rev. Genet.* **25** (1991), pp. 585–627.

[15] X. Cheng, DNA modification by methyltransferases, *Curr. Opin. Struct. Biol.* **5** (1995), pp. 4–10.

[16] N. Swaminathan, D. A. Mead, K. McMaster, D. George, J. L. Van Etten and P. M. Skowron, Molecular cloning of the three base restriction endonuclease R. CviJI from eukaryotic chlorella virus IL-3A, *Nucleic Acids Res.* **24** (1996), pp. 2463–2469.

[17] P. M. Skowron, N. Swaminathan, K. McMaster, D. George, J. L. Van Etten and D. A. Mead, Cloning and applications of the two/three-base restriction endonuclease R. CviJI from IL-3A virus-infected chlorella, *Gene* **157** (1995), pp. 37–41.

[18] A. Pingoud, M. Fuxreiter, V. Pingoud and W. Wende, Type II restriction endonucleases: structure and mechanism, *Cell Mol. Life Sci.* **62** (2005), pp. 685–707.

[19] S. Y. Xu, J. P. Xiao, L. Ettwiller, M. Holden, J. Aliotta, C. L. Poh, M. Dalton, D. P. Robinson, T. R. Petronzio and L. Moran, Cloning and expression of the *Apa*LI, *Nsp*I, *Nsp*HI, *Sac*I, *Sca*I, and *Sap*I restriction-modification systems in *Escherichia coli*, *Mol. Gen. Genet.* **260** (1998), pp. 226–231.

[20] H. Kong, L. F. Lin, N. Porter, S. Stickel, D. Byrd, J. Posfai and R. J. Roberts, Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome, *Nucleic Acids Res.* **28** (2000), pp. 3216–3223.

[21] M. Nelson, D. E. Burbank and J. L. Van Etten, Chlorella viruses encode multiple DNA methyltransferases, *Biol. Chem.* **379** (1998), pp. 423–428.

[22] M. Xu, M. P. Kladde, J. L. Van Etten and R. T. Simpson, Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC, *Nucleic Acids Res.* **26** (1998), pp. 3961–3966.

[23] L. S. Higgins, C. Besnier and H. Kong, The nicking endonuclease N. BstNBI is closely related to type IIs restriction endonucleases *Mly*I and *Ple*I, *Nucleic Acids Res.* **29** (2001), pp. 2492–2501.

[24] J. C. Samuelson, Z. Zhu and S. Y. Xu, The isolation of strand-specific nicking endonucleases from a randomized *Sap*I expression library, *Nucleic Acids Res.* **32** (2004), pp. 3661–3671.

[25] Y. Xu, K. D. Lunnen and H. Kong, Engineering a nicking endonuclease N. AlwI by domain swapping, *Proc. Natl. Acad. Sci. USA* **98** (2001), pp. 12990–12995.