

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
4 May 2006 (04.05.2006)

PCT

(10) International Publication Number
WO 2006/047183 A2

(51) International Patent Classification:
C07H 21/02 (2006.01) C12N 9/22 (2006.01)

(21) International Application Number:
PCT/US2005/037607

(22) International Filing Date: 21 October 2005 (21.10.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/620,939 21 October 2004 (21.10.2004) US

(71) Applicant (for all designated States except US): **NEW ENGLAND BIOLABS, INC.** [US/US]; 240 County Road, Ipswich, MA 01938-2723 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **XU, Shuang-Yong** [US/US]; 15 Carol Lane, Lexington, MA 02420 (US). **ZHU, Zhenyu** [CN/US]; 21 Goldsmith Avenue, Beverly, MA 01915 (US). **CHAN, Shi-hong** [—/US]; 7 Sagamore Road, Ipswich, MA 01938 (US). **XU, Yan** [CN/US]; 24 Tally Ho Drive, Hamilton, MA 01982 (US).

(74) Agent: **STRIMPEL, Harriet, M.**; New England Biolabs, Inc., 240 County Road, Ipswich, MA 01938-2723 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

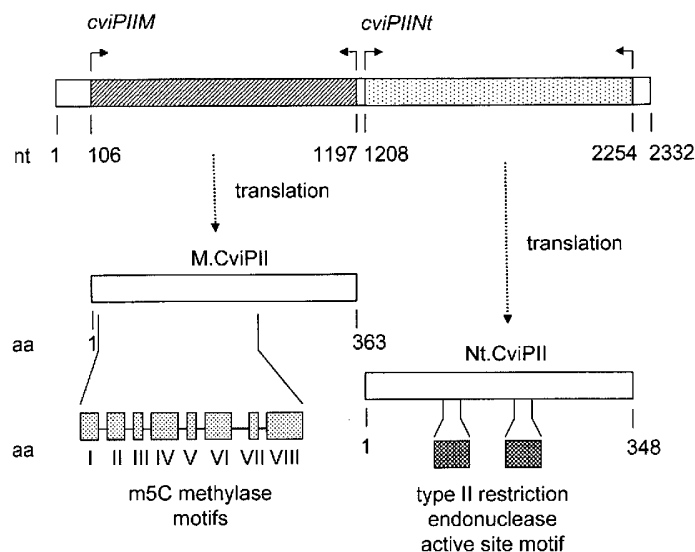
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: RECOMBINANT DNA NICKING ENDONUCLEASE AND USES THEREOF



(57) Abstract: Recombinant nicking endonucleases and associated methylases have been obtained and sequenced and their specificity has been defined. A mutant form of the nicking endonuclease has been cloned where the mutation includes deletion of amino acid sequences at the C-terminal end of the protein. The nicking enzymes have been used for a number of purposes including: amplifying DNA from as few cells as can be found in a single bacterial colony in the presence of a strand displacing polymerase; and for removing genomic DNA in a biological preparation where it is deemed to be a contaminant.

WO 2006/047183 A2

Recombinant DNA Nicking Endonuclease and Uses Thereof

BACKGROUND OF THE INVENTION

5 DNA nicking endonucleases cleave one strand of DNA in a sequence-specific and strand-specific manner. Although there are over 240 type II restriction endonucleases with unique specificities isolated from bacterial and viral sources, only a few site-specific nicking endonucleases are currently commercially available (Roberts et al. *Nucl. Acids Res.* 31:418-420 (2003); REBASE). Efforts to develop more nicking endonucleases consist of either genetic engineering from existing restriction endonucleases or screening from bacterial and viral sources. The nicking endonuclease 10 N.BstNBI and N.BstSEI (5' GAGTCN₄^ 3') were found in strains of *Bacillus stearothermophilus* (Morgan et al. *Biol. Chem.* 381:1123-1125 (2000); Abdurashitov et al. *Mol. Biol. (Mosk)* 30:1261-1267 (1996)) and the nicking endonuclease N.BsrDI (5' ^CATTGC 3'), the large subunit of BsrDI, was found in the strain *B. stearothermophilus* D70. Two natural 15 nicking endonucleases Nt.CviPII (^CCD) and Nt.CviQXI (R^AG) from chlorella viruses have been described (Xia et al. *Nucl. Acids Res.* 16:9477-87 (1988); Zhang et al. *Virology*, 240: 366-75 (1998)).

25 The low quantities of Nt.CviPII in NYs-1 infected lysate limited the potential application of this nicking endonuclease in DNA manipulation. To overcome this limitation, it would be desirable to clone and express this nicking-modification 30 system.

SUMMARY OF THE INVENTION

5 In an embodiment of the invention, an isolated DNA segment encodes a protein with DNA cleavage activity where the protein has an amino acid sequence that has at least about 25% amino acid sequence identity with SEQ ID NO:29. In a further embodiment, the protein is capable of cleaving, at a specific site, a single DNA strand in a duplex where the
10 specific cleavage site is for example, CCA, CCG or CCT.

The isolated DNA segment may be further characterized as having a DNA sequence with at least about 40% DNA sequence identity with SEQ ID NO:28.

15 In another embodiment, the isolated DNA segment encodes a protein with DNA cleavage activity where the DNA segment has at least about 10 contiguous bases identical to sequences contained in SEQ ID NO:28. Preferably, a protein of this type is capable of cleaving, at a specific site, a single
20 DNA strand in a duplex where the specific cleavage site is for example, CCA, CCG or CCT.

25 In another embodiment, the isolated DNA segment encodes a protein with DNA methylase activity, which shares at least about 47% amino acid sequence identity with SEQ ID NO:31. The sequence of the DNA segment shares at least about 53% DNA sequence identity with SEQ ID NO:30.

In additional embodiments, a recombinant nicking endonuclease is provided that has an amino acid sequence sharing at least about 25% sequence identity with SEQ ID NO:29 and a recombinant DNA methylase, is provided which shares at least about 47% amino acid sequence identity with SEQ ID NO:31.

In an additional embodiment, a recombinant nicking endonuclease is provided wherein the endonuclease is a mutant having a deletion at the C-terminal end. For example, mutants with deletions of about 51 and 19 amino acid residues at the C-terminus end retain their specificity for CCA, CCG and CCT.

In an embodiment of the invention, a vector is provided that includes a segment of DNA, which has a sequence that has at least about 10 contiguous bases identical to sequences contained in SEQ ID NO:28. A host cell containing this vector is also provided.

In an additional embodiment of the invention, a method is provided for amplification of DNA, that includes the steps of incubating the DNA with a DNA polymerase capable of strand displacement and a recombinant nicking endonuclease (as described above) and obtaining amplified DNA. This amplification method can be performed isothermally. An additional amplification step may optionally be added to the method in which random primers and a strand displacement polymerase are added to the amplified DNA to enhance the yield of the amplification by another round of amplification.

Examples of polymerases for use in the method include Bst polymerase, *Thermomicrobium roseum* pol I and *E. coli* DNA polymerase large (Klenow) fragment. An example of the recombinant nicking endonuclease is Nt.CviPII. The DNA may be obtained from a single bacterial colony.

In an embodiment of the invention, a method is provided for eliminating DNA from a sample of biological material, that includes (a) adding Nt.CviPII nicking endonuclease or mutant thereof to the sample of biological material; and (b) allowing the nicking endonuclease or mutant thereof to cleave the DNA so as to eliminate the DNA from the sample of biological material.

In an embodiment of the invention, a method for cloning a toxic nicking endonuclease is provided that depends on removing a C-terminal sequence from the DNA encoding the toxic nicking endonuclease; and cloning the truncated gene in a suitable host cell such as *E.coli*. This approach has worked effectively for toxic enzymes cloned from Chlorella viruses.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1A shows a cartoon of the CviPII nicking-modification system. This system contains 2332 nucleotides and two complete open reading frames (ORFs). The methyltransferase contains 8 motifs typically found in m5C DNA methyltransferases but lacks motif IX and X that are also typical of methyltransferases from organisms other than

chlorella. The nicking endonuclease has two active site motifs that are characteristic of some restriction endonucleases.

Figure 1B shows a small amount of Nt.CviPII purified from *E. coli* analyzed by SDS-PAGE using 4-10% gel. Lane 1 is Nt.CviPII while lane 2 is a molecular weight marker.

Figure 2A shows an alignment of the M.CviPII sequence (SEQ ID No 1) with another chlorella methyltransferase (M.CviPI) sequence (SEQ ID. No. 2) and a bacterial methyltransferase (M.HhaI) sequence (SEQ. ID. No 3).

Motifs I through X of m5C methyltransferase are marked. Conserved residues in the motifs are indicated by dots. Sequences that are identical are shown in black boxes. Conserved but not identical residues are shown in grey boxes. Asterisks indicate catalytic residues and hashes indicate S-adenosyl-L-methionine binding residues.

Figure 2B shows the amino acid sequences for Nt.CviPII (SEQ ID NO:4) and restriction endonuclease CviJI (SEQ ID NO:5). Sequences that are identical are shown in black boxes. Conserved but not identical residues are shown in grey boxes.

Two type II restriction endonuclease active site motifs (P-D/E-X_n-D/E/S-Z-K/E where Z = hydrophobic residue) are found in Nt.CviPII sequences namely (i) Ser126-Asp127-X₁₂-Glu139-Ile140-Lys141 and (ii) V189-Glu190-X₂₁-Glu211-Val212-Lys213. The latter motif can be partially aligned to the

proposed active site of CviJI. Asterisks indicate conserved residues of the active site motif.

Figure 3A shows the results of the experiment that probes the putative M-CviPII methylation site. The DNA is pUC-*cviPIIM* expressing M-CviPII. The DNA was isolated and digested with MspI or ScrFI. The first cytosine at the Nt-CviPII cleavage site of CCD was shown to be methylated because the plasmid was cleaved by ScrFI but was resistant to MspI digestion.

Figure 3B is a chromatogram showing target sequences for methylation by M-CviPII. PCR products derived from sodium bisulfide-treated pUC-*cviPIIM* are compared to those from un-treated pUC-*cviPIIM*. The change of cytosine to thymidine corresponds to unmodified cytosine, whereas the presence of cytosine in both sodium bisulfide-treated and un-treated DNA indicates C⁵-methylcytosine. Results for recognition sites CCT, CCG (SEQ ID NOS:8 and 9), CCA (SEQ ID NOS:10 and 11), CCGG (SEQ ID NOS:12 and 13) show that the first Cs are methylated in CCN triplets. The result for CCAA (SEQ ID. No. 14 and 15) show relaxed specificity towards the second C of CCAA. The modified nucleotides are indicated by down arrows.

Figure 4A compares the results from IPTG-induced Nt.CviPII expression and un-induced cell extract under a *lac* promoter on pUC19 DNA. Supernatant of the lysate of induced (upper panel) and un-induced culture (lower panel) from equivalent cultures were loaded on a SP FF column and eluted

with linear gradient of 0.1-1 M NaCl. High DNA nicking activity was obtained with IPTG induction. N= relaxed circle, L=linear and S= supercoiled.

5 Figure 4B shows that Nt.CviPII is surprisingly thermostable being active at least to temperatures of 60C. 0.5 µg of pUC19 was incubated with 1 unit of Nt.CviPII for 1 hour at various temperatures (lanes 2-10). The reactions were stopped and analyzed by electrophoresis on 1.5%
10 agarose gel. Lane 11 is linearized pUC19 DNA, lane 12 is supercoiled pUC19 DNA and lanes 1 and 13 are 100 base pair size markers.

15 Figure 4C shows the results of Nt.CviPII-cleaved pUC19 and single stranded M13 phage DNA by electrophoresis on 6% poly-acrylamide gel with 7 M urea.

 Lane 1: size marker.

 Lane 2: pUC19 (double strand DNA) with 1 unit of Nt.CviPII.

20 Lane 3: M13 (single strand DNA) with 1 unit of Nt.CviPII.

 Lane 4: M13 (single strand DNA) with 0.5 units of Nt.CviPII.

25 Lane 5: M13 (single strand DNA) with 0 units of Nt.CviPII.

 Lane 6: LMW: low molecular weight DNA size marker.

 Figure 5 shows that CCA and CCG are the preferred substrates (SEQ ID NOS:18 and 19) for Nt-CviPII cleavage. Nt.CviPII does not cleave at CCC but instead cleaves at the
30

overlapping CCA site (SEQ ID NO:20). CCT shows low-level cleavage by Nt-CviPII (SEQ ID NO:21).

The upper schematic shows the sequence that was
nicked by Nt.CviPII and read as the reverse-complement by
the sequencing primer. Therefore, TGG corresponds to CCA,
CGG to CCG, GGG to CCC and AGG to CCT. Triplet sequences
in boxes are CCN sites designed on the substrate DNA. Native
CCA sequences of pUC19 are underlined. The arrowheads
indicate the cleavage site. The arrows under the
chromatographs and the bracketed "a" in the schematic
indicate the adenine added by the template-independent
activity of *Taq* DNA polymerase used in sequencing reactions
at the cleavage site.

Figure 6 shows a schematic outline of isothermal
amplification using Nt.CviPII and Bst DNA polymerase I large
fragment.

Figure 7A shows that nicking-endonuclease mediated
isothermal amplification of DNA prefers a DNA polymerase
with strand displacement activity. Purified *E. coli* DNA was
incubated with Nt.CviPII in combination with Bst DNA
polymerase I large fragment (Bst), *T. roseum* (*Tro*), Vent DNA
polymerase, *T. aquaticus* (*Taq*) or *E. coli* DNA polymerase I
Klenow fragment (Klenow) in the presence of dNTPs. Only Bst
polymerase and *Tro* polymerase amplified the DNA
significantly. Klenow also amplified low levels of DNA. No
detectable amplification was found using Vent or *Taq* or no
polymerase.

Figure 7B shows the results of isothermal amplification of purified DNA from *T. thermophilus* and λ phage. *T. thermophilus* and λ DNAs were amplified with Nt.CviPII and Bst DNA polymerase I large fragment. The amplified products were analyzed by 1.5% agarose gel electrophoresis. Lanes 2 and 4 shows the results of amplification only in the presence of both enzymes. Lane 5 is a marker. Lanes 1 and 3 show results from an amplification reacton containing only Bst DNA polymerase I large fragment and no Nt.CviPII. No detectable DNA was amplified without Nt.CviPII.

Figure 7C shows that on 10-20% polyacrylamide gel with 7 M urea, the amplified DNA from Figure 7B are resolved into single-stranded DNA from < 25 to over 500 nt.

Figure 7D shows that DNA can be amplified from a single bacterial colony subjected to heat treatment to release the DNA. Only a fraction of DNA was amplified without heating the cells (lane 6).

Lane 1 is the heat-treated cells from a single colony incubated with Bst DNA polymerase I large fragment, CviPI and Nt.CviPII.

Lane 2 is the heat-treated cells from a single colony incubated with Bst DNA polymerase I large fragment, and Nt.CviPII.

Lane 3 is the heat-treated cells from a single colony incubated with Bst DNA polymerase I large fragment only.

Lane 4 is the heat-treated cells from a single colony incubated with Bst DNA polymerase I large fragment, MspI and Nt.CviPII.

Lane 5 is Bst DNA polymerase I large fragment, and Nt.CviPII with no DNA template.

Lane 6 is the non-heat treated cells from a single colony incubated with Bst DNA polymerase I large fragment, and Nt.CviPII.

Figure 8 shows DNA nicking activity of two Nt.CviPII truncation mutants. Four-fold dilutions of NPN297 or NPN329 (1 unit, 0.25 units, 0.06 units and 0.02 units) were incubated with 1 µg of pUC19 DNA at 37°C for 1 hour. The cleavage products were analyzed on a 1.5% agarose gel in TBE. Supercoiled (SC) and linearized (L) pUC19 DNA and a marker were included for comparison.

Figure 9 shows the results of Nt.CviPII truncation mutant-mediated DNA amplification. NPN297 (0.25 units) was used in combination with Bst DNA polymerase I large fragment (2 units) in the presence of 0.2 mM dNTPs and designated amount of λ DNA. The reactions were carried out at 55°C for 30 minutes.

Figure 10 shows the efficient removal of genomic DNA by Nt.CviPII in reverse transcription reactions.

Lane M: 2-log DNA ladder (New England Biolabs, Inc., Ipswich, MA).

Lane 1: reverse transcription (RT) without M-MuLV reverse transcriptase;

Lanes 2 and 4: RT with M-MuLV reverse transcriptase;

Lane 3: RT without M-MuLV reverse transcriptase, but in the presence of 0.5 unit of Nt.CviPII;

5 Lane 5: RT without M-MuLV reverse transcriptase, but in the presence of 2 units of Nt.CviPII.

Figure 11A shows the DNA sequence of the *CviPINt.* gene (SEQ ID NO:28).

10

Figure 11B shows the amino acid sequence of the *CviPIINt.* gene (SEQ ID NO:29).

Figure 12A shows the DNA sequence of the CviPIIM gene (SEQ ID NO:30).

15

Figure 12B shows the amino acid sequence of the CviPIIM gene (SEQ ID NO:31).

20

DETAILED DESCRIPTION OF THE INVENTION

The cloning and expression of CviPII nicking-modification system, purification of Nt.CviPII, and its utility in generating DNA oligonucleotides for random DNA
25 amplification are described here. A significant aspect of this work is overcoming a toxicity problem caused by the frequent DNA nicking activity of the enzyme so as to produce a sufficient amount of recombinant Nt.CviPII, for its use as a molecular biology reagent. Nt.CviPII is of interest for a
30 number of reasons including its ability to recognize a three-base sequence (CCD; D = A, G or T) of double strand DNA at

the 5' end of the first C of the top strand. It is also a naturally occurring frequently cutting nicking endonuclease. These properties have been exploited in a number of ways including a method of primer independent isothermal DNA amplification.

The nicking endonuclease Nt.CviPII described here has been cloned from a chlorella virus referred to as NYs-1. The CviPII nicking and modification system was cloned and expressed in *E. coli*. Initially, the *cviPIIM* gene was cloned in *E. coli* by the methyltransferase selection method. The adjacent ORFs were sequenced directly from the viral DNA. A downstream ORF showed some amino acid sequence identity to a restriction endonuclease CviJI (RG[^]CY) in a BlastP search of all known genes in GeneBank.

An ORF for Nt.CviPII was identified, amplified by PCR and transcribed and translated in an *in vitro* transcription and translation system. The cell free extract showed DNA nicking activity when it was compared to the native Nt.CviPII nicking endonuclease. However, it proved difficult to produce large amounts of protein for further characterization. Therefore, more efforts were made to express *cviPIIM* and *cviPIINT* genes in *E. coli*. *E. coli* expression host ER2683 was pre-modified by introduction of plasmid pUC-*cviPIIM*, and the *cviPIINT* gene was cloned in the expression vector pR976, a low copy number plasmid with P_{tac} promoter. Extra codons were added to the 5' end of *cviPIINT* gene to incorporate a tag of 6 histidine residues to facilitate purification of the protein.

5 To minimize the toxicity of the nicking endonuclease associated with its ability to nick DNA every 50-60 bp, the nicking endonuclease gene was positioned 18 nucleotides downstream of the ribosome-binding site to reduce the level of translation. Recombinant Nt.CviPII production was induced by the addition of IPTG to the culture media. Nt.CviPII was purified through affinity-tag purification and traditional chromatography such as metal-chelation chromatography, heparin FF and ion-exchange chromatographies. The purified Nt.CviPII revealed a nicking site preference of CCR>CCT (R = A or G) and was active through a wide temperature range with highest activity at a temperature range of 30-45°C while retaining activity at 55°C and 60°C.

15 M.CviPII was shown to modify the first Cs in CCA, CCG, CCT and CCC triplet sequence (Example 1). It also modified the first two Cs in the sequence of CCAA (Fig. 3B).

20 To further increase the amount of Nt.CviPII obtained by recombinant technology, two C-terminal truncation mutants of Nt.CviPII fused to an intein and chitin-binding domain were generated. The combination of truncation and fusion decreased the toxicity of the nicking endonuclease to the host cells so that the fusion protein was over-expressed in *E. coli* strain ER2566 (New England Biolabs, Inc., Ipswich, MA). The fusion proteins were purified by chitin column chromatography and the fusion part was removed by self-cleavage activity of the intein induced by reducing agent. The cleaved Nt.CviPII truncation mutants were further purified by standard chromatographic steps. The truncation mutants of

25

30

Nt.CviPII were found to possess the same sequence specificity but lower specific activity than the wild-type enzyme.

5 Uses of Nt.CviPII

(a) Isothermal amplification

10 Due to the high frequency of Nt.CviPII cleavage sites and its partial duplex cleavage product, Nt.CviPII was used in conjunction with several DNA polymerases in isothermal random DNA amplification. An assay system was developed to determine conditions for isothermal amplification. This assay system is described in Example 3. Using this approach, it is possible to show amplification of a DNA using Nt.CviPII and DNA polymerases possessing strand-displacement activity. Moreover, Figure 6 shows that DNA can be amplified from a single bacterial colony.

20 Nt.CviPII may also be used in prior art methods of isothermal amplification that utilize nicking endonucleases. These include: strand displacement amplification, exponential DNA amplification (EXPAR) or nick translation with DNA polymerases such as Klenow fragment, Bst DNA polymerase, 25 *Thermomicrobium roseum* DNA pol I large fragment, or phi 29 DNA polymerase to replicate DNA at the nicked sites.

(b) *Amplifying DNA from single colonies*

DNA can be amplified from a single *E. coli* colony using Nt.CviPII and a strand displacement DNA polymerase (Figure 7D). The cells in the single colony have been broken by heat to release DNA but any other method in the art can be used where preferably no additional purification steps are required before performing amplification.

(c) Removing genomic DNA from RNA or protein preparations

DNase I is the most commonly used enzyme in DNA contaminant removal from RNA samples. DNase I is a non-specific nicking endonuclease that works on single-stranded DNA, double-stranded DNA, and DNA-RNA hybrids. After DNase I treatment, the enzyme must be removed from the RNA sample before other applications such as RT-PCR. However, DNase I is heat-resistant and therefore phenol extraction is required to remove DNase I completely (Aguila et al. *BMC Molecular Biology* 6:9 (2005)). In contrast, Nt.CviPII is a sequence-specific nicking endonuclease that recognizes double-stranded DNA only. Therefore, the DNA contaminant removal can be done by Nt.CviPII simultaneously with the reverse transcription reaction so that no extra purification steps are required. By choosing a different frequent nicking endonuclease, a different digestion pattern can also be achieved.

(d) *Creation of gaps for assembling DNA molecules and for purifying DNA*

The nicking endonucleases described herein may be used for creating single-stranded regions in duplex nucleic acids. Such single-stranded regions can take the form of gaps interior to the duplex, or terminal single-stranded regions. Single-stranded termini can be crafted to allow linkage of various elements via base-pairing with elements containing a complementary single-stranded region. This joining is useful, for example, in an ordered, oriented assembly of DNA modules to create cloning or expression vectors. This joining is also useful in attaching detection probes and purifying DNA molecules containing the single-stranded region. Gaps are useful in similar applications, including attaching detection or purification probes (U.S. Patent No. 6,660,475 and U.S. Patent Publication No. 2003-0194736 A1 herein incorporated by reference).

(e) Labeling DNA

The nicking endonucleases described herein can be used to label DNA. At first, nicks are introduced into DNA by Nt.CviPII. Then DNA polymerases with strand displacement activity can be used to replicate DNA. Radioactive dNTP, biotinylated dNTP, or dNTP with fluorophore modification can be added in the DNA extension reaction. The newly synthesized DNA should be labeled according to the dNTP used.

(f) Detecting mutations

Similarly to restriction fragment polymorphism to detect genetic alterations, nicking fragment DNA polymorphism can be used to detect gene mutations if the point mutation takes place within the nicking site recognition sequence.

5

(g) Creating relaxed circles from supercoiled DNA

The nicking endonucleases described herein can be used to prepare relaxed circular DNA under limited nicking conditions, e.g., using diluted Nt.CviPII. Supercoiled plasmid DNA is first nicked by Nt.CviPII provided that the plasmid contains at least one CCD site. The supercoiled DNA should be converted to nicked-open circular DNA, which can be gel-purified. The purified nicked DNA is treated with DNA ligase to generate relaxed circular DNA.

10

15

The references cited above and below as well as U.S. provisional application serial number 60/620,939 are hereby incorporated by reference herein.

20

Example 1: Cloning and identification of *cviPIIM* and *cviPIINt*

Chlorella virus NYs-1 genomic DNA was digested partially with Sau3AI and ligated to a BamHI-digested and CIP-treated pUCAC (a derivative of pUC19 by inserting a PCR-amplified chloramphenicol resistant gene into AflIII site of pUC19) and the ligated DNA was used to transform ER1992 competent cells to construct a Sau3AI genomic DNA library.

25

30

Approximately 10^5 ampicillin resistant transformants were pooled and plasmid DNA was prepared. Clones that expressed M.CviPII methylase were selected by digesting pooled ampicillin and chloramphenicol resistant plasmids with MspI (cleaves CCGG and C^mCGG sequences but not ^mCCGG sequence). Eighteen plasmids from the Sau3AI genomic library were found to be partially resistant to MspI digestion. The inserts of six isolates were sequenced, which revealed an identical open reading frame (ORF, 1092 bp) that had 45.2% amino acid (aa) identity to the NYs-1 encoded M.CviPI (recognition sequence GC) and 41% amino acid identity to chlorella virus IL-3A encoded M.CviJI (recognition sequence RGCY) (Xu et al. *Nucl. Acids Res.* 26:3961-66 (1998); Shields et al. *Virology* 176:16-24 (1990)) (Fig. 2A). The presence of most of the conserved motifs of m5C methylases in the primary structure of M.CviPII indicates that M.CviPII is a m5C methylase (Chan et al. *Nucl. Acids Res.* 32:6187-6199 (2004)).

The putative *cviPIIM* gene was amplified by PCR and ligated into pUC19 at the SphI and SalI sites and transferred into *E. coli* ER2502. *In vivo* activity of M.CviPII was tested by challenging the plasmid isolated from ER2502 [pUC-*cviPIIM*]. The plasmid was incubated with MspI (C[^]CGG) or ScrFI (CC[^]NGG) at 37°C for 1 hour in NEBuffer 2 (10 mM Tris-HCl, 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, pH 7.9), New England Biolabs, Inc., Ipswich, MA. The digested DNA was analyzed by agarose gel electrophoresis. The plasmid pUC-*cviPIIM* was partially resistant to MspI but not to ScrFI (Fig. 3A). MspI cleaves C^mCGG, but not ^mCCGG, whereas ScrFI cleaves

^mCCNGG, but not C^mCNGG (<http://rebase.neb.com/rebase>). This indirect evidence (resistant to MspI digestion) indicated that M.CviPII modifies the first cytosine of the CCGG sequence, which is consistent with the result obtained from the native enzyme (Xia et al. *Nucl. Acids Res.* 16:9477-9487 (1988)).

To determine the site of modification, the plasmid pUC-*cviPIIM* used in methylase protection assay was treated with sodium bisulfide (EZ DNA Methylation Kit, Zymo Research Corporation, Orange, CA). The sodium bisulfide-treated DNA was purified by Qiaprep spin-columns (Qiagen, Valencia, CA) and used for PCR using primers that amplified the *cviPIIM* gene (MP-SphI-F and MP-SalI-F).

MP-SphI-F

CGGTAAGCSATGCATGCGTACAAAGTATCGTATTTTTATCGTATGT
ATG (SEQ ID NO:7)

MP-SalI-R

GAAGTCGACTTAATAATGCATAAGATCACCCAAATATTCGTGCG
(SEQ ID NO:22)

Untreated plasmid was also amplified by the same pair of primers as control. Methylated cytosines were protected from sodium bisulfide that converted un-modified cytosines to uracils, which are amplified as thymidines in PCR. Thus, by comparing with control PCR product, cytosine residues that became thymidines were un-modified while those that remained cytosines were methylated.

Sequencing of PCR-amplified DNA from the sodium bisulfide-treated pUC-*cviPIIM* indicated that most of the first but not the second cytosine of CCG, CCA and CCT are modified (Fig. 3B). Surprisingly, the first two cytosines in CCCG and CCAA were also modified (Fig. 3B, Table 2). In CCCG where CCC and CCG overlap, methylation of the first C in CCCG indicated that M.CviPII also methylates CCC site, a result of relaxed methyltransferase activity. Modification of the first and second Cs in CCAA suggested that M.CviPII modifies cytosine in relaxed recognition sequence under over-expressed condition (the *cviPIIM* gene was cloned in a high-copy-number plasmid and IPTG-induced).

NYS-1 DNA adjacent to *cviPIIM* gene was sequenced and an ORF of 349 codons was identified (in the same orientation) that began 12 nucleotides downstream of the *cviPIIM* stop codon (Fig. 1A). The protein expressed from the ORF had low but significant amino acid sequence identity to chlorella virus IL-3A encoded restriction endonuclease CviJI (RG/CY, with 18.3% identity (Fig. 2B)) which may reflect overlapping recognition sequence. The half site of CviJI (CY, Y = C or T) is similar to Nt.CviPII recognition sequence CCD. However, Nt.CviPII does not show significant similarity to other restriction endonucleases or nicking endonucleases with overlapping recognition sequence in GenBank. The pI value of Nt.CviPII was calculated to be 9.82 by ProtParam Tool of ExPASy Proteomics server of Swiss Institute of Bioinformatics.

Example 2: Expression and purification of Nt.CviPII

Due to the frequent Nt.CviPII nicking sites, difficulties in cloning the *cviPIINT* gene in *E. coli* were encountered.

Initially, Nt.CviPII was expressed using *in vitro* transcription and translation system. A low level of nicking activity was detected in the lysate in comparison with the native Nt.CviPII. However, it was difficult to achieve a clear digestion pattern. To achieve sufficient enzyme for purification, the Nt.CviPII system was modified for expression in *E. coli*. The expression host ER2683 was pre-modified by expression of M.CviPII via introduction of pUC-*cviPIIM*. Additional measures were taken to construct a stable expression clone: (i) A low copy number plasmid pR976 with P_{tac} (p15A replication origin) was used as the cloning vector for the *cviPIINT* gene; (ii) The *cviPIINT* gene was inserted 18 nucleotides downstream of the ribosome-binding site so as to reduce the expression level of the enzyme. The efforts to express M.CviPII in pACYC184 (pACYC-*cviPIIM* gave marginal modification of host genomic DNA] and Nt.CviPII in pET21a vector failed to generate a stable expression clone.

The expression strain ER2683 [pUC-*cviPIIM*, pR976-*cviPIINT*] was successfully constructed. M.CviPII that was constitutively expressed under the control of *lac* promoter on pUCAC protected the host DNA from basal expression of the Nt.CviPII. The Nt.CviPII expression plasmid alone could not transform *E. coli* cells, indicating that the residual expression of Nt.CviPII in the absence of the cognate methyltransferase is lethal to the host. To determine the basal level of

expression, induced or un-induced cultures of the expression strain ER2683 [pUC-*cviPIIM*, pR976-*cviPIINT*] were grown and partially purified by anion-exchange chromatography. The fractions were tested for DNA nicking activity.

5 A single colony of ER2683 [pUC-*cviPIIM*, pR976-*cviPIINT*] was grown to mid-log phase in 2 liters of rich medium (10 g/l Tryptone, 5 g/l yeast extract, 5 g/l NaCl, pH adjusted to 7.2 with NaOH) containing Amp (0.1 mg/ml),
10 kanamycin (Km, 0.05 mg/ml) and tetracycline (Tc, 0.01 mg/ml) at 30°C at 280 rpm. One liter of culture was induced with 0.25 mM IPTG and the other was not. Both cultures were incubated at 16°C for 18 hr and cells were harvested by centrifugation. The cell pellets (wet weight of 3.9 g for the
15 induced culture and 5.0 g for the un-induced culture) were sonicated in 100 ml of 20 mM sodium phosphate, 0.1 M NaCl, pH 7.4. After centrifugation, the soluble fractions were loaded on a SP FF (25 ml bed volume, Amersham Biosciences, now
20 GE Healthcare, Uppsala, Sweden) column and eluted with a linear gradient of 0.1-1 M NaCl. Two μ l of the fractions were incubated with 0.5 mg of pUC19 at 37°C for 1 hr. Reactions were stopped by adding 25 mM EDTA and analyzed by
25 agarose gel electrophoresis. DNA nicking activity was distinguished from non-specific host nucleases. Non-specific nucleases produce a smear while Nt.CviPII produces a characteristic banding pattern of the digested DNA.

30 In the presence of IPTG, fractions eluting from a SP FF column exhibited significantly higher DNA nicking activity (Fig. 4A, top panel) than fractions from the un-induced culture

(Fig. 4A, bottom panel). The un-induced culture, nevertheless, showed significant DNA nicking activity, indicating that there was basal expression of Nt.CviPII in the absence of IPTG. IPTG induction of cells grown at 16°C produced more Nt.CviPII activity than cells grown at higher temperatures.

To purify the recombinant His-tagged Nt.CviPII, the supernatant obtained after sonication was loaded into a nickel-charged HisTrap column (HisTrap HP, 5 ml bed volume, Amersham Biosciences, now GE Healthcare, Uppsala, Sweden) and protein was eluted with a step gradient of 50, 100, 200, 300, 400 and 500 mM imidazole in 20 mM NaHPO₄, 0.5 M NaCl, pH 7.4. DNA nicking activity was detected in fractions eluting in 50 mM and 300 mM imidazole. The 50 mM imidazole fraction was further purified on a heparin FF (25 ml bed volume, Amersham Biosciences, now GE Healthcare, Uppsala, Sweden) column and a SP FF (25 ml bed volume, Amersham Biosciences, now GE Healthcare, Uppsala, Sweden) column. The 300 mM imidazole fraction was concentrated and assayed. All purified protein preparations were concentrated by VIVASPIN 20 (10,000 MWCO, VIVASCIENCE) and stored in 10 mM NaHPO₄, 250 mM NaCl, pH 7.0 with 50% glycerol at -20°C.

N-terminal sequencing of the purified protein eluted in 50 mM imidazole revealed a MSTPQAKTKYY sequence (SEQ ID NO:6), which corresponds to amino acids 5 to 15 in Nt.CviPII. Thus this fraction contains a protein initiated at the fifth codon of *cviPIInt*, which is an ATG. Mass spectrometry showed that

the mass of this protein is 34,110 Da, compared to the predicted value of 40,069 Da. This preparation was designated His⁻-Nt.CviPII.

5 Protein that eluted at 300 mM imidazole was concentrated and two bands of ~34-36 kDa were observed by SDS-PAGE. N-terminal sequencing established that the upper band was Nt.CviPII with a 6-Histidine tag while the lower band was an *E. coli* 2-dehydro-3-deoxyphosphoheptonate aldolase contaminant. The two proteins were separated
10 cleanly on a SP FF column, resulting in pure Nt.CviPII as judged by SDS-PAGE (Fig. 1B). This protein was designated His⁺-Nt.CviPII. Because this latter preparation was cleaner than His⁻-Nt.CviPII, it was used in all subsequent
15 experiments.

Nt.CviPII nicking endonuclease activity

Nt.CviPII activity was measured at 16°C, 20°C, 25°C,
20 30°C, 37°C, 45°C, 55°C, 60°C and 65°C on 0.5 ug of pUC19 substrate in NEBuffer 4 (20 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM DTT, pH 7.9, New England Biolabs, Inc., Ipswich, MA) (Fig. 5A). DNA nicking activity was highest at 30-45°C, whereas the activity
25 at 20-25°C was higher than that at 55-60°C. Nt.CviPII showed lowest activity at 65°C, probably due to partial thermal denaturation. At 30-45°C, the cleavage product of pUC19 appears to be ~200 bp in 1.5% agarose gel electrophoresis (Fig. 4B). A time course experiment at 37°C
30 established that a stable cleavage pattern was reached by 1

hr incubation. Longer incubation times or increasing the enzyme concentration did not alter the appearance of this pattern. Also, the pattern is not the result of Nt.CviPII inactivation at 37°C because pre-incubation of Nt.CviPII at 37°C for 90 min did not alter subsequent enzyme activity. Extended incubation of pUC19 with both batches of Nt.CviPII (His⁺ and His⁻) for 16 hours indicated that there is no significant contamination of non-specific nuclease. One unit of Nt.CviPII activity is defined as the amount of enzyme needed to cleave pUC19 into ~200 bp products in 1 hr at 37°C, as judged by electrophoresis in 1.5% agarose gels. Protein concentrations were determined by the Bradford assay (Bio-Rad Protein Assay, Bio-Rad Laboratories, Hercules, CA) using BSA as a standard. The specific activity of His⁺-Nt.CviPII is 9410 units/mg of protein.

Analyzing the pUC19 cleavage products by electrophoresis on poly-acrylamide gel containing 7 M urea indicated that the single-stranded DNA fragments were smaller than 150 nucleotides in size (Fig. 4C). There are 252 CCD sites on both strands of pUC19. With 2686 base pair and assuming that nicking sites are equally distributed in both strands, complete cleavage of pUC19 gives fragments of ~20 nucleotides long. However, the end products of pUC19 are single-stranded DNA ranged from 25 to 150 nt as shown in Fig. 4B, suggesting that some of the sites are not cleaved. Therefore, some Nt.CviPII cleavage sites in pUC19 are apparently less susceptible to cleavage than others (see below).

The ability of Nt.CviPII to cleave single-stranded DNA was also tested. 250 ng of single-stranded M13 phage DNA was incubated with 0.5 or 1 unit of Nt.CviPII at 37°C for 1 hour. Electrophoresis on poly-acrylamide gel containing 7 M urea showed that the single-stranded DNA was partially cleaved by Nt.CviPII. The low cleavage activity of Nt.CviPII on single-stranded DNA is likely due to the nicking activity on the transient duplex form of the phage DNA instead of the cleavage of single-stranded DNA per se.

Nt.CviPII cleavage specificity

Double-stranded DNA substrates of 189 bp containing single CCA, CCT, CCC or CCG sites at nucleotides 160-162 (Table 1) were constructed by PCR. The substrates contain an internal XhoI site (C[^]TCGAG) as a control to monitor cleavage. The substrate DNAs also contain EcoRI and HindIII sites at the 5' and 3' ends, respectively, for ligation into pUC19. The substrate DNA was ligated to pUC19 and the inserted DNA sequence was confirmed by sequencing. Two ug of pUC19-substrate was incubated with Nt.CviPII at 37°C for 1 hour. Reactions were terminated by adding 25 mM EDTA and DNA samples were purified by QIAprep spin columns (Qiagen, Valencia, CA). One-eighth dilutions of the purified cleaved products were sequenced with custom primers that anneal at the 5' end of the substrate DNA.

When a nick occurs in a double-stranded DNA, sequencing reaction on the nicked strand stops after the nicked base and the peaks in the sequencing chromatogram

diminish sharply (Run-off sequencing). An extra adenine is added to the 3' end of the newly synthesized single-stranded DNA due to the template-independent polymerase activity of *Taq* DNA polymerase used in the sequencing reaction. The adenine peak also helps identify the cleavage site. Consistent with a previous report (Xia et al. Nucl. Acids Res. 16:9477-87 (1988)), run-off sequencing showed that the DNA was cleaved 5' of the first cytosine in CCA, CCG and CCT sequences, but not at the CCC site (Fig. 5). However, cleavage is less favorable at the CCT site than the CCA and CCG sites as only a small amount of the CCT substrate was cleaved (Fig. 5, bottom panel). A small peak of adenine following the AGG triplet indicates that some cleavage occurred at the CCT site. Cleavage occurred between the first and the second cytosine in the C[^]CCA sequence. No cleavage took place before the first C in CCCA sequence. In all experiments, a minor peak corresponding to adenine signified the end of the template.

Example 3. Truncation mutants of Nt.CviPII: Cloning, expression, and purification of C-terminal truncation mutants of Nt.CviPII

DNA oligos were designed such that they served as primers for PCR amplification of two truncation mutants of Nt.CviPII. The primers also added NdeI and SapI sites at the 5' and 3' end of the amplified DNA, respectively, for cloning purposes: NPN297 (NP NdeI-F and NPN297-SapIR) and NPN829 (Np NdeI-F and NPN329-SAPIR).

NP-NdeI-F

5'

ACCGTTGAGCATATGTATATATATGTCTACTCCGCAGGC

AAAG 3' (SEQ ID NO:23)

5

NPN297-SapI-R

5' GGTGGTTGCTCTTCCGCAACAGGAAGAAGAAATTAATTCTATTT

TATTTTCAAAACATCATCGAT 3' (SEQ ID NO:24)

10

NPN329-SapI-R

5' GGTGGTTGCTCTTCCGCAGCATTTTGGTGGACTTGTTATTTTCT

TTGATTTTGG 3' (SEQ ID NO:25)

15

Mutants NPN297 and NPN329 were generated such that they contains the first 297 aa (C-terminal deletion of 51 aa residues) and 329 aa (C-terminal deletion of 19 aa residues) of Nt.CviPII, respectively. The amplified DNA was ligated to pTXB1 (New England Biolabs, Inc., Ipswich, MA) at NdeI and SapI sites. The mutant proteins were expressed as C-terminal fusion to intein Mxe GyrA followed by a chitin-binding domain. The ligated DNA was sequenced to confirm that there was no secondary mutation. The constructs pTXB1-NPN297 and pTXB1-NPN329 were transferred to *E. coli* strain ER2566 (New England Biolabs, Inc., Ipswich, MA) and grown in LB media or agar plates containing 100 µg/ml of Amp.

20

25

30

For purification of the truncated mutant nicking endonucleases, the following protocol was used for NPN297 and may also be used for NPN329. Single colony was inoculated to a starter culture of 100 ml of LB media

containing 100 µg of Amp and grown at 37°C with 250 rpm for 12-16 hours. Ten ml of the starter culture was inoculated to 1 liter of fresh LB media containing 100 µg of Amp for 6 liters of media. The culture was incubated at 37°C with 250 rpm until
5 OD600 reached 0.6-0.9. IPTG was added to the culture at a final concentration of 0.25 mM/L and incubation was continued for 3 hours at 37°C with 250 rpm. The cultures were centrifuged at 3,550 g at 4°C for 15 minutes. The cell pellets were stored at -20°C. The frozen cell pellets from each
10 liter of culture were resuspended in 15 ml of 20 mM Tris-HCl, 0.5 M NaCl, 0.1% Triton X-100, 1 mM EDTA, pH 8.5 (chitin column buffer).

The resuspended cells were lysed by sonication on ice.
15 The lysate was centrifuged at 26,700g for 20 min at 4°C. The supernatant of the lysate was loaded to a chitin column (20 ml bed volume) at 4°C. After washing with 200 ml of chitin column buffer, the column was flushed with 40 ml of chitin column buffer with 40 mM of DTT within 10 minutes to induce
20 cleavage of the intein. The column was incubated at 25°C for 10-16 hours. The cleaved protein was collected by washing the column with 40 ml of chitin column buffer without DTT. The eluted protein solution was diluted one-fourth using 20 mM Tris-HCl, pH 7.7 such that the sodium chloride
25 concentration decreased to ~ 125 mM.

The diluted protein solution was loaded to a Heparin HyperD M column (BioSeptra, Inc., Fremont, CA) bed volume = 30 ml) at 4°C. The column was washed with 300 ml of a
30 buffer containing 20 mM Tris-HCl, pH 7.7 and then with a

linear gradient of 0 – 1 M NaCl in 20 mM Tris-HCl, pH 7.7. Fractions of 5 ml were collected and analyzed on SDS-PAGE. NPN297 was eluted in 0.6 M or higher concentration of NaCl. Fractions that contains the protein was pooled and dialyzed against 4 L of 10 mM potassium phosphate buffer, 50 mM NaCl, 1 mM EDTA, pH 7.0 (HTP column buffer) at 4°C for 12-16 hours. The dialyzed protein solution was loaded to a hydroxyapatite column (Bio Rad Bio-gel HTP, bed volume = 15 ml, Bio-rad Laboratories, Rockford, IL) at 2 ml/min. The column was washed with 150 ml of HTP column buffer and eluted by a linear gradient of 0 – 500 mM potassium phosphate buffer, 50 mM NaCl, 1 mM EDTA, pH 7.0. Fractions were collected for every 5 ml and analyzed using SDS-PAGE. NPN297 was found to elute at 0.25 M potassium phosphate. Fractions that contained NPN297 were pooled and dialyzed against 2 L of 40 mM Tris-HCl, 200 mM NaCl, pH 8.0 at 4°C for 12-16 hours. The dialyzed protein solution was concentrated using VivaSpin concentrator (molecular weight cut-off = 10 kDa). Equal volume of 100% glycerol was added to the concentrated protein solution. The protein preparation was stored at -20°C.

DNA nicking activity of the truncation mutants

DNA nicking activity and cleavage specificity of NPN297 and NPN329 were assayed essentially the same way as for the wild-type Nt.CviPII (Example 2). The truncation mutants were found to be active with the same cleavage specificity as the wild-type Nt.CviPII. The specific activity of NPN297 and NPN329 are estimated to be 2,600 units/mg and 1,400

units/mg, respectively, compared to 9,410 U/mg of the wild-type Nt.CviPII. Although the specific activity of the truncation mutants are lower than the wild-type Nt.CviPII, the yield of the truncation mutants are much higher such that the truncation mutant generates more units of activity from the same volume of culture than the wild type.

Example 4: Isothermal amplification of DNA using Nt.CviPII and DNA polymerases with strand displacement activity

Because of the high frequency of cleavage sites and its single-stranded cleavage product, Nt.CviPII and NPN297 were used in conjunction with several DNA polymerases in isothermal random DNA amplification.

(a) Isothermal amplification of purified DNA

The following experiments were conducted with various purified DNA showing that the amplification method is generally applicable.

Two hundred ng of λ , *E. coli* or *Thermus thermophilus* genomic DNAs were incubated with 1 unit of Nt.CviPII and 16 units of Bst DNA polymerase large fragment, 8 units of *Taq* DNA polymerase or 4 units of Vent DNA polymerase in ThermoPol reaction buffer at 55°C (20 mM Tris-acetate, 10 mM KCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 2 mM MgSO_4 , 0.1% Triton X-100, pH 8.5), or 10 units of Klenow fragment of *E. coli* DNA polymerase I in EcoPol buffer (10 mM Tris-HCl, 5 mM MgCl_2 ,

7.5 mM DTT, pH 7.5) at 37°C for 30 min supplemented with 0.1 mM dNTP. The amplified DNAs were analyzed by electrophoresis on either 1.5%-2% agarose gel or 6% polyacrylamide gel containing 7 M urea in 1X TBE buffer.

E. coli DNA was incubated with Nt.CviPII and various DNA polymerases and dNTPs at various temperatures for 30 min. With the Nt.CviPII, Bst DNA polymerase I large fragment generated the highest yield of DNA at 55°C. The large fragment of DNA polymerase I from *Thermomicrobium roseum* (U.S. patent number 5,962,296) also synthesized significant amounts of DNA while the addition of *Taq* DNA polymerases and Vent DNA polymerase did not result in any detectable DNA synthesis (Fig. 7A). Klenow fragment of *E. coli* DNA polymerase with Nt.CviPII generated a small amount of amplified DNA. Random DNA amplification can be achieved from *T. thermophilus* HB27 and λ DNA in the presence of Nt.CviPII/Bst DNA polymerase I large fragment but not with Bst DNA polymerase (PolI) large fragment alone (Fig. 7B). In Fig. 7C, the DNA having a PolI site in the range of < 50 to 200 nucleotides efficiently amplified. The DNA mass was amplified approximately 50-fold with 10 ng input DNA generating ~500 ng amplified small DNA fragments. The Nt.CviPII/Bst DNA pol amplified products can be further amplified by adding [N]₆, [N]₉, or [N]₁₂ random primers together with Bst DNA pol large fragment to the amplification product by incubation at 50°C for 60 min resulting in a second round of amplification which is expected to generate major DNA products in the range of 1 to 2 kb.

Although the amplification steps described above were done at 55°C, other temperatures can be used as long as denaturation of double-stranded DNA is favored. Nt.CviPII makes frequent cuts on the DNA and produces single-
5 stranded products or partial duplex DNA with 5' overhang (3' recessed ends) (Fig. 6). Bst DNA polymerase I large fragment then fills in at the 3' end until it reaches the end of the template. The extended DNA acted as a substrate for Nt.CviPII which, in turn, provided a new substrate for the
10 polymerase, allowing linear amplification. The size of the amplified DNA can be increased or decreased by altering the amount (units) of Nt.CviPII in the amplification reactions.

From the collection of DNA polymerases tested, only Bst
15 DNA polymerase I large fragment and *Thermomicrobium roseum* (*Tro*) DNA polymerase I large fragment produced significant amounts of amplified products. Klenow fragment also generated small amount of amplified DNA. These polymerases possess relatively high strand displacement
20 activity among the polymerases tested. While not wishing to be limited by theory, strand displacement activity may be involved in removing the nicked fragment and revealing a recessive 3' end for template-dependent amplification.

25 It was also demonstrated that incubation of random DNA oligonucleotides and fresh Bst DNA polymerase large fragment and dNTPs with the amplified product can result in DNA amplification. Alternatively, the amplified DNA can be purified and used as primers for direct amplification of

genomic DNA through isothermal or thermocycling procedures.

Randomly amplified DNA has been used as a highly sensitive probe for arrays of DNA oligonucleotides carrying "signature sequences" of pathogenic biological agents such as *E. coli* O157:H7 (Vora et al. *Appl. Environ. Microbiol.* 70:3047-54 (2004)). The DNA amplification method presented here does not require synthesis of primers and can generate large quantities of single-stranded DNA from a single bacterial colony within a short time frame (e.g. 10 to 30 minutes).

Unlike rolling circle amplification that can be used to generate high coverage of the genome, this DNA amplification method may not necessarily cover the entire genome. The amplified DNA can be used as a probe to detect target DNA by Southern blotting.

The procedure can be adapted for environmental or clinical samples and labels such as biotin or fluorescein can be incorporated into the amplified product by using modified deoxy-nucleotides. Development of timely, sensitive and specific detection methods to identify important pathogens is of great importance in bio-defense and public health.

(b) Isothermal amplification of DNA from a single bacterial colony

Experiments were also performed to amplify DNA from a single *E. coli* colony using Nt.CviPII and Bst DNA polymerase I large fragment (Fig. 7D, lane 2). Single colonies of *E. coli* (~ 0.2 mm in diameter) were suspended in 50 ml of water, heated at 94°C for 8 min, centrifuged, and 20 ml of the supernatant was used for amplification. Heating the colony at 94°C for 8 min was necessary to release the DNA for optimal amplification (Fig. 7D, lane 6). Including a degenerate 4-base cutting restriction endonuclease CviTI (RG[^]CY) in the amplification reaction generated slightly shorter amplified DNA fragments as judged by agarose gel electrophoresis (Fig. 7B, lane 1). However, including a 4-base cutting restriction endonuclease MspI (C[^]CGG) made no difference in the size of the amplified product.

Example 5: The use of frequent nicking endonucleases such as Nt.CviPII and Nt.CviQXI to eliminate contaminant DNA from RNA samples

Nt.CviPII recognizes [^]CCD (D = A, T, or G), which occurs at every ~21 bp. Nt.CviQXI recognizes R[^]AG (R = A or G), which occurs at every 32-64 bp. When used alone or together, the frequent nicking endonuclease(s) can degrade almost any larger DNA into very small pieces. An example of this application is to use frequent nicking endonucleases to remove genomic DNA contamination from RNA samples before reverse transcription and RT-PCR.

About 500 ng rat liver total RNA was mixed with 2 μ l dT₂₃VN (50 μ M, New England Biolabs, Inc., Ipswich, MA) and 7 μ l dH₂O. After denaturation at 70°C for five minutes, it was left on ice. A 10 μ l mix containing 100 mM Tris-HCl, pH 8.3, 150 mM KCl, 6 mM MgCl₂, 20 mM DTT, 8 units of M-MuLV (New England Biolabs, Inc., Ipswich, MA), 40 units of RNase inhibitors, 100 nmole dNTP, with or without 0.5 unit wild-type Nt.CviPII or 2 units of truncated Nt.CviPII was added. After one hour incubation at 42°C, 30 μ l dH₂O was added to dilute the cDNA product into a 50 μ l solution, from which 2 μ l was used in 35-cycle PCR amplification using GAPDH-specific primers:

5' TGCMTCTGCACCACCAACT 3' (forward primer)
(SEQ ID NO:26)
5' YGCCTGCTTCACCACCTT C 3' (reverse primer)
(SEQ ID NO:27)

Following RT PCR, eight μ l was analysed on a 1% agarose gel (Figure 10). In the presence of M-MuLV reverse transcriptase, the RT-PCR reaction gave a specific product (lanes 2 and 4). In the absence of M-MuLV reverse transcriptase, the RT-PCR reaction also gave a specific product due to the presence of contaminating genomic DNA in the RNA sample (lane 1). When 0.5 unit of Nt.CviPII was added to the RT reaction, a significant amount of genomic DNA contaminant was removed, leading to a much weaker band when M-MuLV reverse transcriptase was not added (lane 3). When two units of Nt.CviPII was added to the RT reaction, almost all genomic DNA contaminant was degraded,

leading to a non-detectable signal when M-MuLV reverse transcriptase was left out (lane 5). The results indicated that Nt.CviPII can be included in the reverse transcription step to efficiently remove genomic DNA (template) contamination.
5 (Similar results have been obtained using Nt.CviQXI.)

1. An isolated DNA segment encoding a protein with DNA cleavage activity, wherein the encoded protein has an amino acid sequence which has at least 20% amino acid sequence identity with
5 SEQ ID NO 29.

2. An isolated DNA segment according to claim 1, having at least 40% DNA sequence identity with SEQ ID
10 NO:28.

3. An isolated DNA segment according to claim 1, wherein the protein is capable of cleaving, at a specific site, a single DNA strand in a duplex.
15

4. An isolated DNA segment according to claim 3, wherein the specific cleavage site is selected from CCA, CCG and CCT.

20 5. An isolated DNA segment encoding a protein with DNA cleavage activity, the DNA segment having a sequence characterized by at least 10 contiguous bases identical to sequences contained in SEQ ID NO:28.

25 6. An isolated DNA segment according to claim 5, wherein the protein is capable of cleaving at a specific cleavage site on a single DNA strand in a duplex.

7. An isolated DNA segment according to claim 6, wherein the specific cleavage site is selected from CCA, CCG and CCT.

5 8. An isolated DNA segment encoding a protein with DNA methylase activity, wherein the protein has at least 47% sequence identity with SEQ ID NO:31.

10 9. An isolated DNA segment according to claim 8, wherein the DNA segment has a DNA sequence with at least 53% sequence identity with SEQ ID NO:30.

15 10. A recombinant nicking endonuclease, comprising an amino acid sequence with at least 25% sequence identity with SEQ ID NO:29.

20 11. A recombinant DNA methylase, comprising an amino acid sequence with at least 47% sequence identity with SEQ ID NO:31.

25 12. A recombinant nicking endonuclease according to claim 10, wherein the endonuclease is a mutant having a sequence truncation at the C-terminal end compared with a native nicking endonuclease.

 13. A recombinant nicking endonuclease according to claim 12, wherein 51 amino acid residues at the C-terminus have been removed.

14. A recombinant nicking endonuclease according to claim 12, wherein the 19 amino acid residues at the C-terminus have been removed.

5 15. A recombinant nicking endonuclease according to claim 12, having a cleavage specificity selected from CCA, CCG and CCT.

10 16. A recombinant nicking endonuclease according to claim 12, having a substantially similar cleavage activity to the native endonuclease.

15 17. A vector comprising a segment of DNA, the DNA further comprising at least 10 contiguous bases identical to sequences contained in SEQ ID NO:28.

18. A host cell comprising the vector of claim 17.

20 19. A method for amplification of DNA, comprising:
 (a) incubating the DNA with a DNA polymerase capable of strand displacement and a recombinant nicking endonuclease according to claims 10,12, 13 or 14; and
 (b) obtaining amplified DNA.

25 20. A method according to claim 19, wherein the amplification is isothermal.

30 21. A method according to claim 19, further comprising:
 (c) subjecting the amplified DNA from (b) to an additional amplification step in the presence of random primers and a

strand displacement polymerase to enhance the yield of the amplification.

22. A method according to claims 19 or 21, wherein the DNA polymerase is selected from Bst polymerase, *Thermomicrobium roseum* pol I and *E. coli* DNA polymerase large (Klenow) fragment.

23. A method according to claims 19, 21 or 22, wherein the recombinant nicking endonuclease is Nt.CviPII.

24. A method according to claims 19, 21 or 22, wherein the DNA is obtained from a single bacterial colony.

25. A method for eliminating genomic DNA from a sample of biological material, comprising:

(a) adding Nt.CviPII nicking endonuclease or mutant thereof to the sample of biological material; and

(b) allowing the nicking endonuclease or mutant thereof to cleave the genomic DNA so as to eliminate the genomic DNA from the sample of biological material.

26. A method for cloning a toxic nicking endonuclease; comprising:

removing a C-terminal sequence from the DNA encoding the toxic nicking endonuclease; and cloning the truncated gene in a suitable host cell.

27. A method according to claim 25, wherein the host cell is *E. coli*.

28. A method according to claim 25, wherein the toxic nicking endonuclease is derived from a Chlorella virus.

Fig. 1A

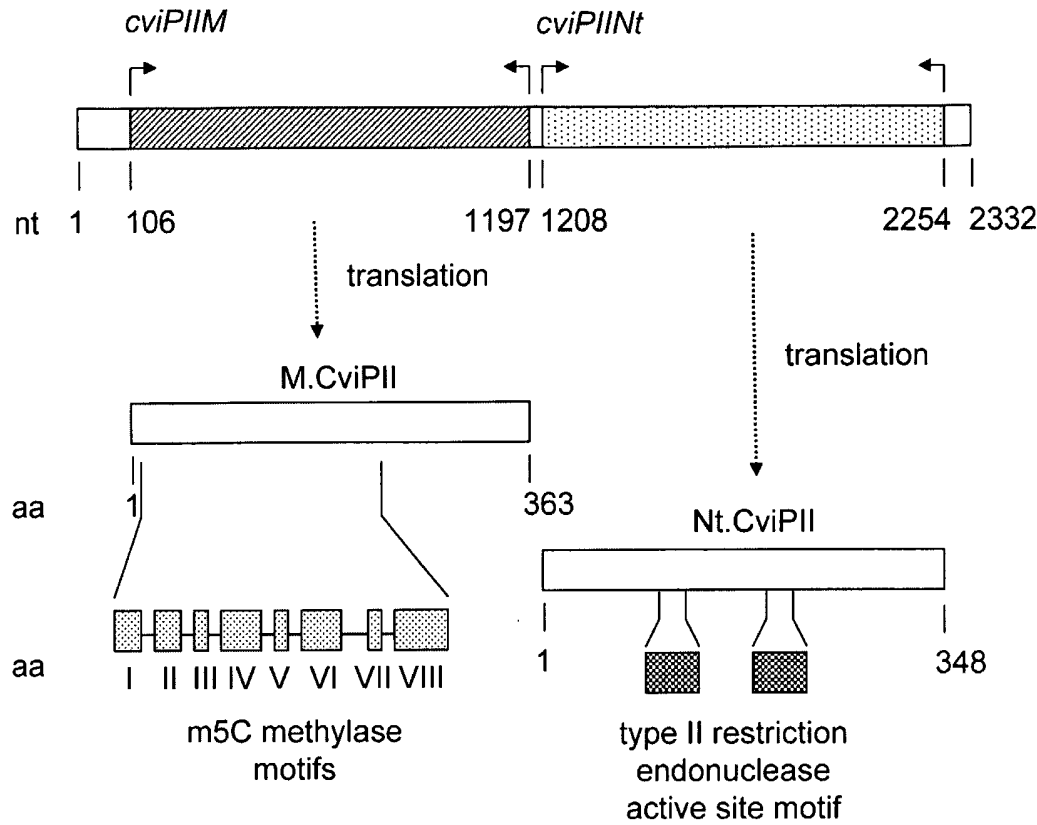
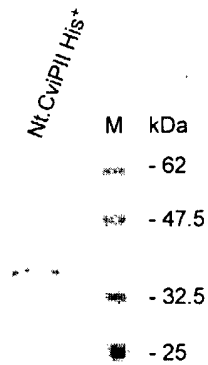
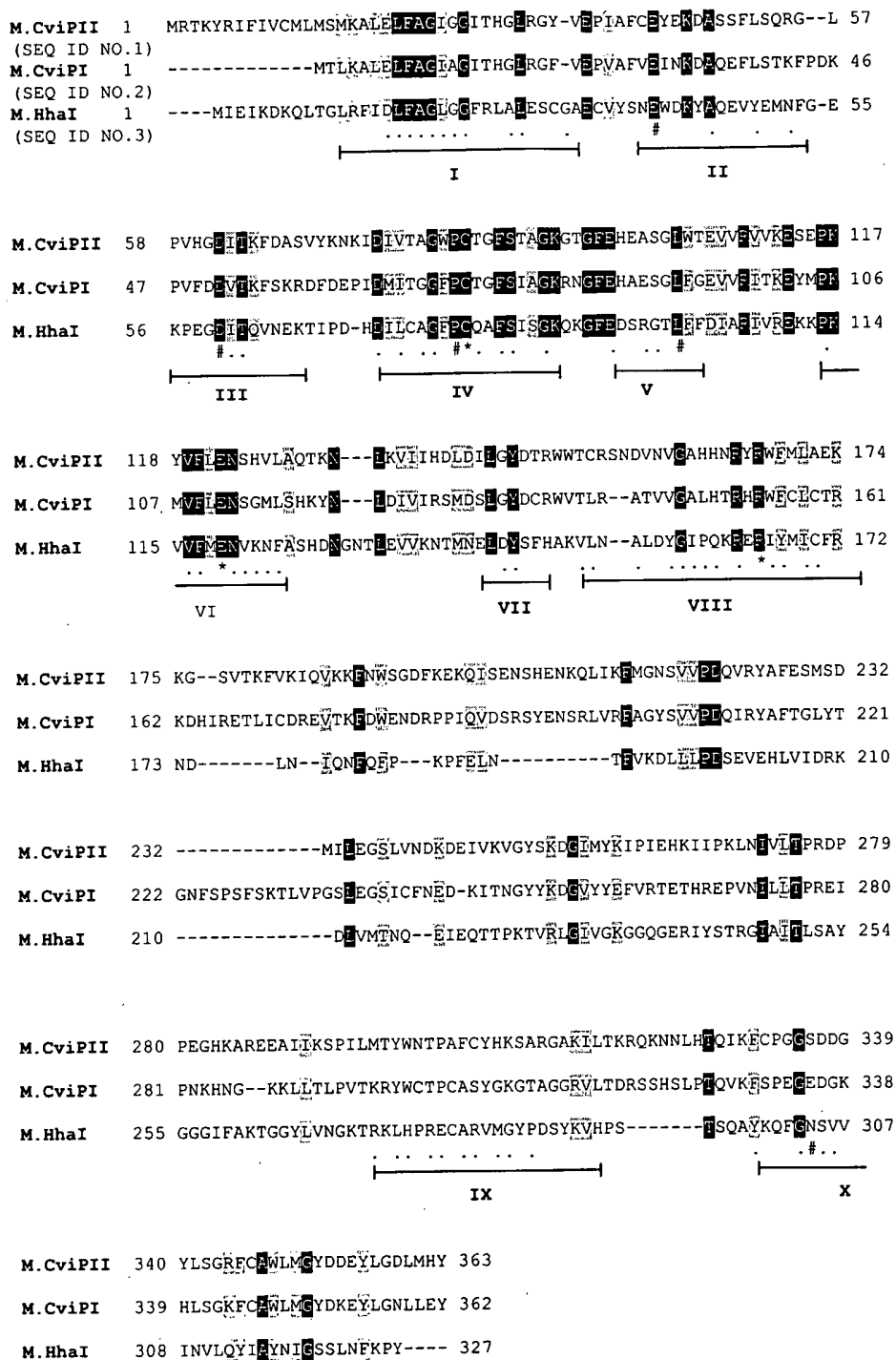


Fig. 1B



2/15

Fig. 2A



3/15

Fig. 2B

```

Nt.CviPII 1  -----MYT1YMSTPQAKT2YYE3QRFV4DFYK5EL6RN7KVS-----L8V9I10V11L12K13N 43
(SEQ ID NO:4)
CviJI      1  MEEKRLALIEKQ2IAEEKIASGR3IRKRIST4NATKH5RE6FV7KV8INSMFVGE9AT10FV11FV12I 60
(SEQ ID NO:5)

Nt.CviPII 44  L45IK46-Q47V48IC49NVSGV50E51IR52DKAN53AKSP54SKIKSE55LGRHVT56SKADIALFTEEKNG57TKV58DVA59N 102
CviJI      61  K62EN63KSRE64HN65VVR66FE67Q68IGSK69-AKSP70AYVDR71EYN---K72PKADIAVD---I73TK74DVA75N 112

Nt.CviPII 103  IS104HK105SNKDIHGKKITHA106OY107L108DAS109SDVMF110TKIG111Q112TK113EL114KD115FN116MI117SLSVPLTATKYC118NE 162
CviJI      113  IS114HK115AS-----EG116Y117O118Y119L120KIS121-GKNL122FTGKE123LE124V125LS126FR127V128VS129MAP---VSKI130NE 160

Nt.CviPII 163  KY164SGTSLR165IT166DD167V168KSTIL169MAIFG170VE171FG172AYSE173NNAN174IL175MC176DE177IL178EV179KDDKTI180IL181TT 222
CviJI      161  AN162IT163-----V164SP165IK166SN167L168IK169CAIFG170FDY171CKKPG172ED173VD174IL175GO176CR177IL178TKRGS---IL179YL 212

Nt.CviPII 223  KEN224G225FS226LANGFA227YIP228SK229DK230IL231FTK232PTSGKKT233VVDGKTIEGVSVWIIY234RS235V236AGSKN237EKI 282
CviJI      213  TFT214G215FS216ALNGHLE217NFTG218HE219V220FE221YVRTE-----RS222SS223SC224---E225 246

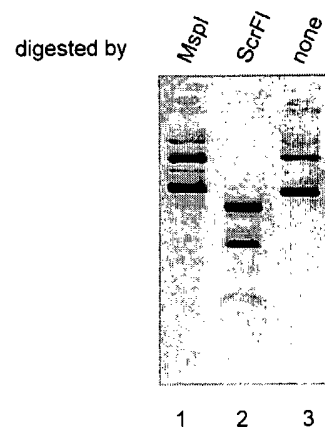
Nt.CviPII 283  DDVLKNKIELISSSC284SVKKKDN285FV286SI287MQSKKITS288EPKSKKITSPPKSKKITS289SPSKSKKIT 342
CviJI      246  -----S247ITTVV248NG249V250TYKNLRFFI251HE-----Y 267

Nt.CviPII 343  NF344FMK345IL346KYS347LSR 356
CviJI      268  NF269VSS270ST271ORIM--- 278

```

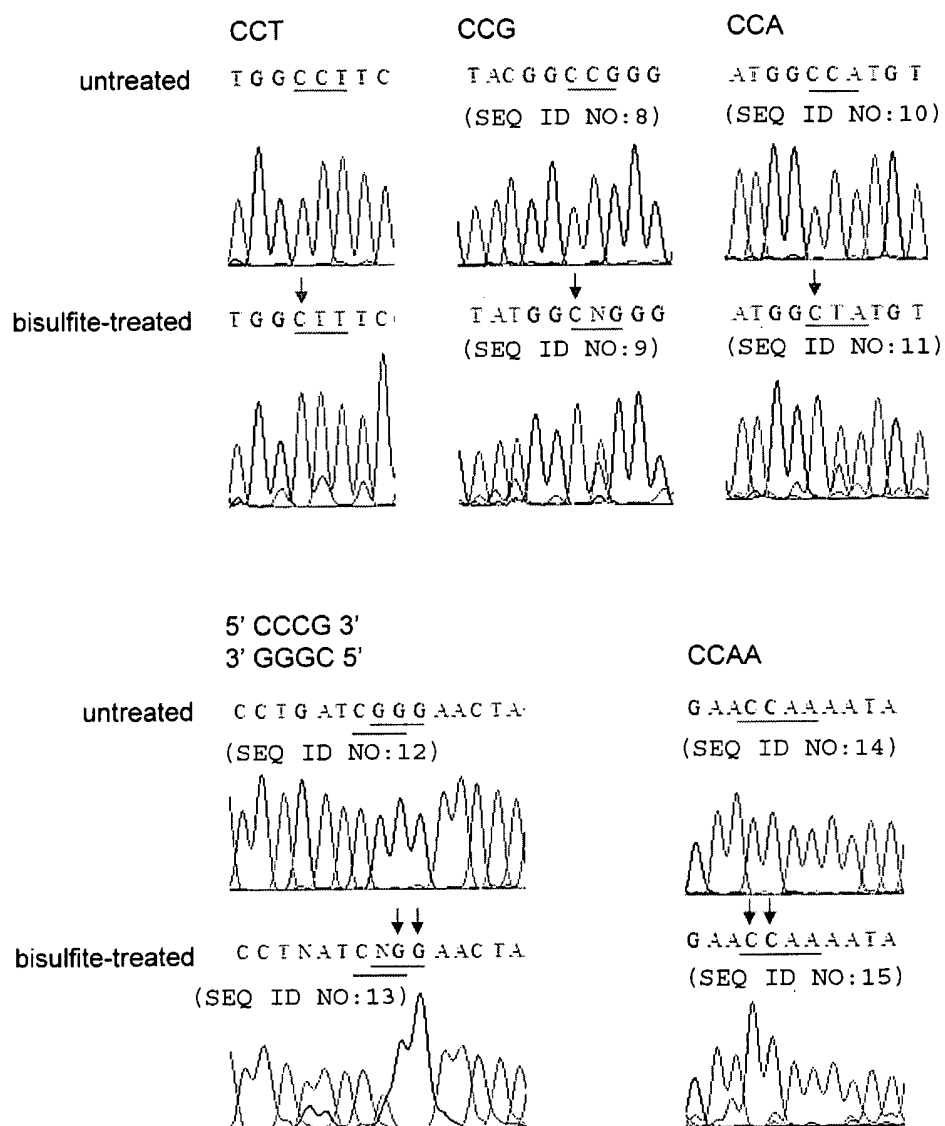
4/15

Fig. 3A



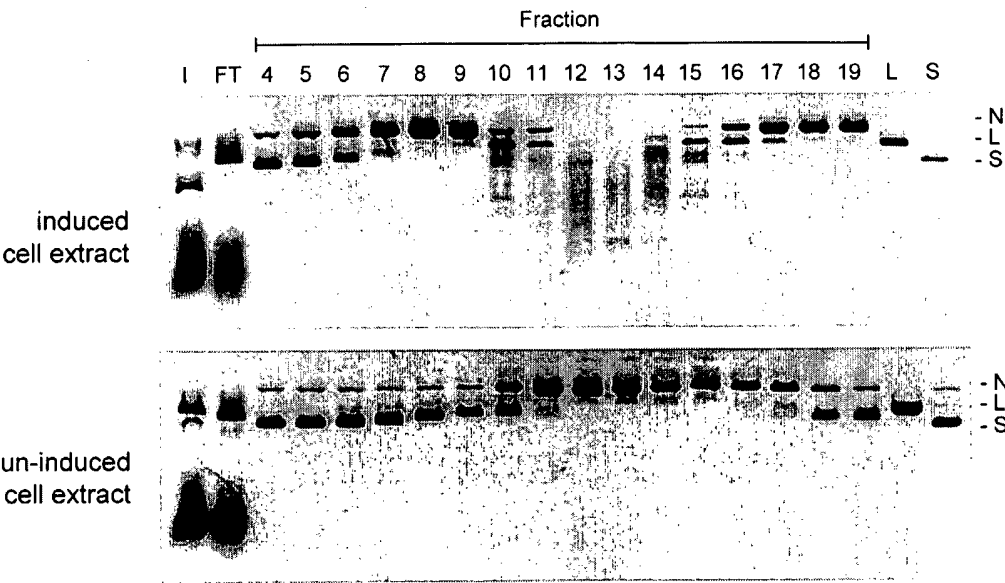
5/15

Fig. 3B



6/15

Fig. 4A



7/15

Fig. 4B

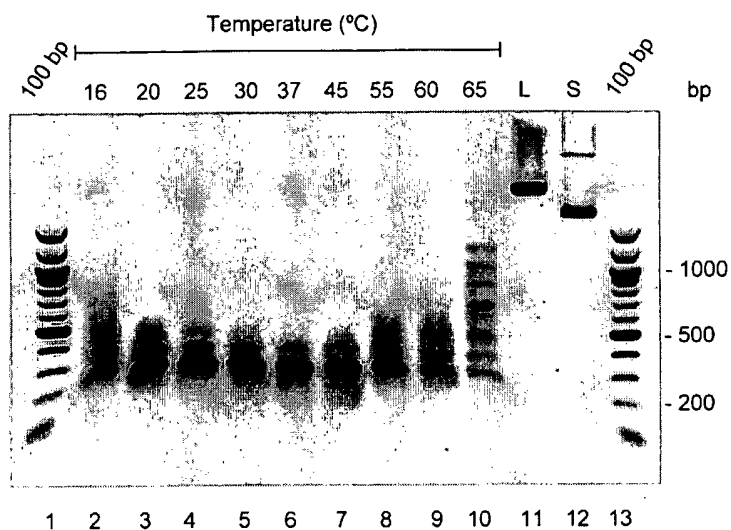
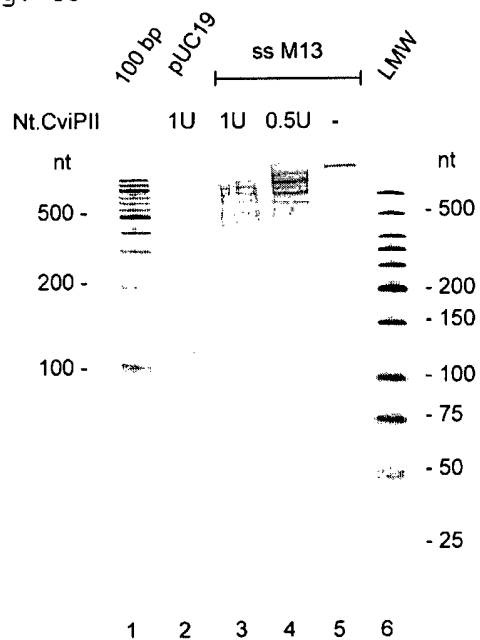
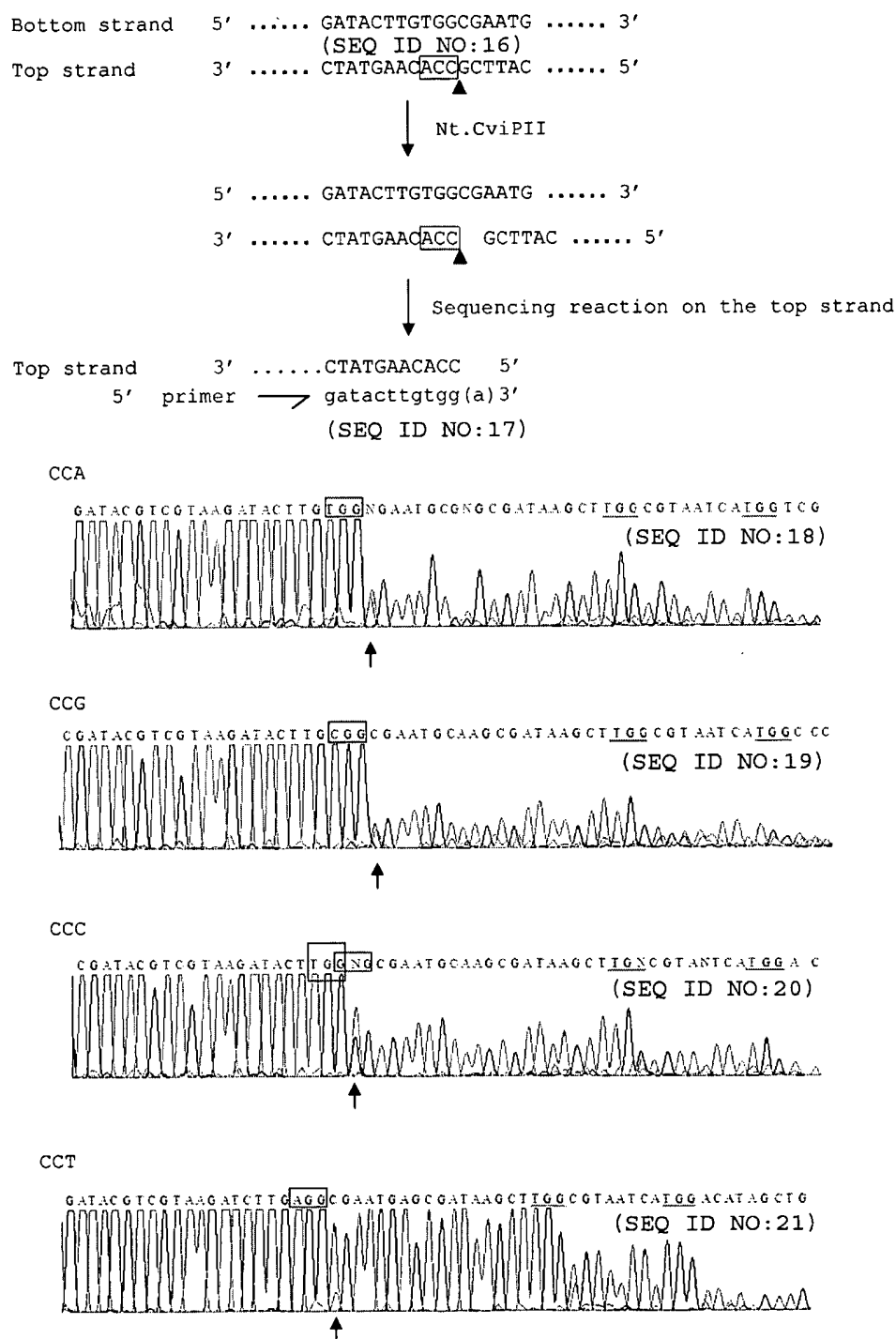


Fig. 4C



8/15

Fig. 5



9/15

Fig. 6

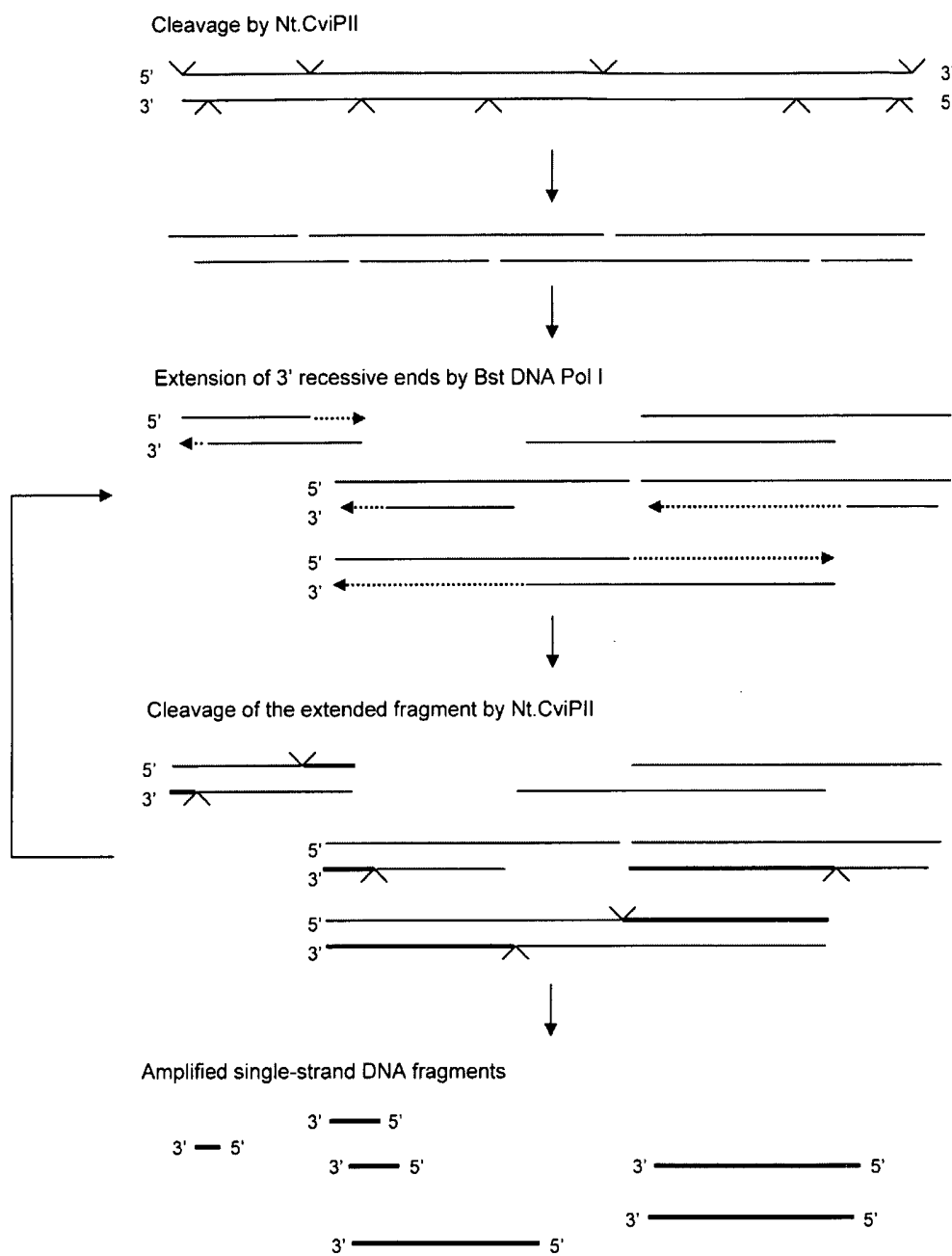


Fig. 7A

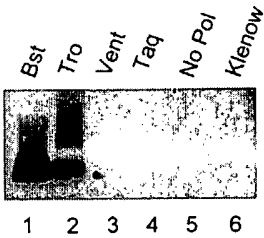


Fig. 7B

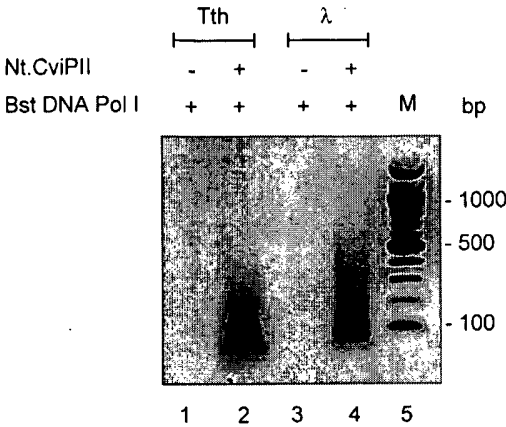


Fig. 7C

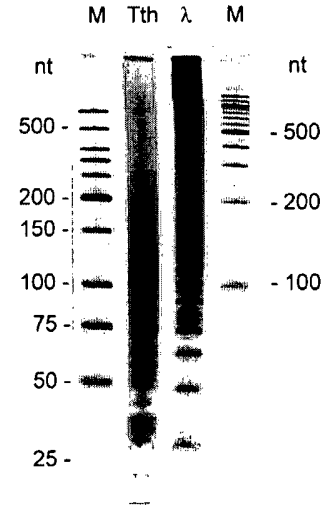
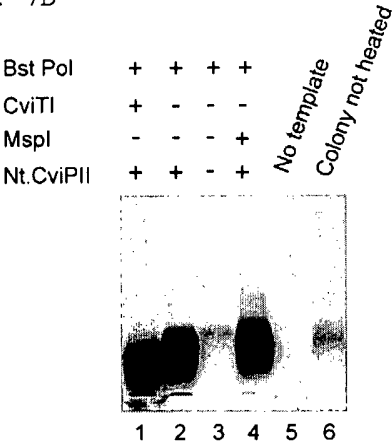


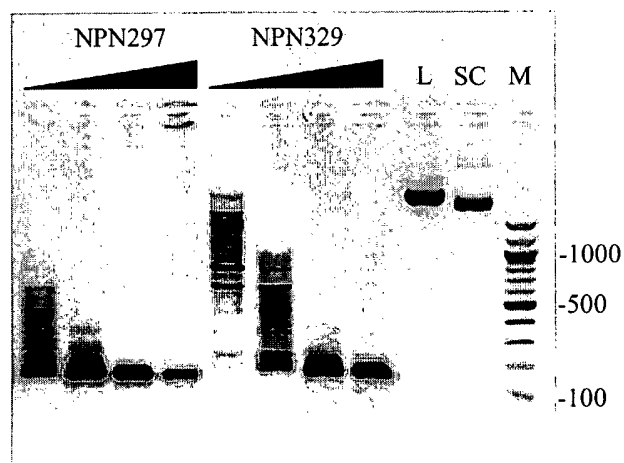
Fig. 7D



11/15

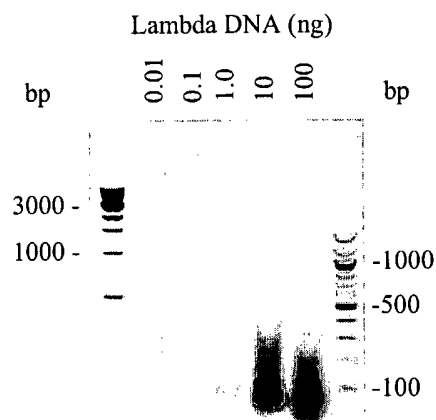
Fig. 8

bp



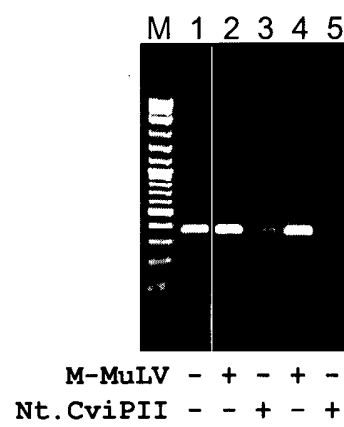
12/15

Fig. 9



13/15

Fig. 10



14/15

Figure 11A (SEQ ID NO:28)

ATGTATATATATATGTCTACTCCGCAGGCAAAGACCAAATATTATGAACAAAGATTTCGTAAACG
ATTTCTACAAAGAACTCGAACGTAATAAAGTTTCTCTGCCTGTCACAATAGTTTTGAAGGATAA
TCTGGGAATAAAACAAGTTATACAGAATGGGAGTGGTGTAGAGTTTTACGTGATAAAGCTAAT
GCGAAATCTCCGTCTAAGATAAAAAGTGAAGAATTAGGTAGACATGTGACATCGAAAGCGGATA
TAGCATTATTTACCGAGGAAAAGAATGGAACAAAAGTTGACGTTGCATGGATATCCCCCAATC
TCATAAAGATTTTCTTGAAAAAAGATAACTCCTGCTCAATATTTTGATGCCTCCTCAGATGTT
ATGTTTAAAACCAAATTTGGCCAACCAAAGGAAATAAAGGAGCTTAAGAATAAGATGATTTCTT
TGAGTGTTCTCTCACGGCAACGAAATATTGCTGGCCCAAGTACAAATCTGGAACATCTCTCAG
AATATGGGACGATGTACAAAGCACTATACTTATGAACATGGCAATATTTGGTGTGAGTTTGGA
AAGGCATATTGTCGAAATAATGCAAATATTTTGATGGTTGGAGATCCTCTTATAGAAGTAAAAG
ATGACAAAACGATAATTCTTACCACAAAAGAAAATGGATTTAGTCTGGCAAACGGATTTGCAGA
ATATATACCATCAAAGGATAAACCTATATTTTTCACTAAACCAACATCAGGAAAAAAAACAGTC
GTAGATGGAAAAACGATCGAGGGAGTGAGTGTATGGATAATTTATAGAAGTTATGCAGGATCAA
AGAATAGAAAAATCGATGATGTTTTGAAAAATAAAATAGAATTAATTTCTTCTTCTGTTCTGT
TAAAAAGAAAGATAACTTTGTTTCTATCATGCAATCAAAGAAAATAACAAGTCCACCAAATCA
AAGAAAATAACAAGTCCACCAAATCAAAGAAAATAACAAGTCCATCAAATCAAAGAAAATAA
CAAATTTTTTTATGAAAAATAA

Figure 11B (SEQ ID NO:29)

MYIYMSTPQAKTKYYEQRFVNDFYKELERNKVSLPVTIVLKDNLGKQVIQNGSGVRVLRDKAN
AKSPSKI KSEELGRHVTSKADIALFTEEKNGTKVDVAWIS PQSHKDFLGKKITPAQYFDASSDV
MFKTKIGQPKEIKELKNKMISLSVPLTATKYCWPKYKSGTSLRIWDDVQSTILMNMAIFGVEFG
KAYCRNNANILMVG DPLIEVKDDKTIILT KENGFS LANGFAEYIPSKDKPIFFT KPTSGKKTV
VDGKTIEGVSVWIIYRSYAGSKNRKIDDLKNKIELISSCSVKKKDNFVSIMQSKKITSPPKS
KKITSPPKSKKITSPSKSKKITNFFMKK

15/15

Figure 12A (SEQ ID NO:30)

ATGCGTACAAAGTATCGTATTTTTATCGTATGTATGTTAATGAGTATGAAAGCATT
AGAACTCTTCGCAGGTATAGGCGGGATCACTCATGGTCTTCGTGGTTACGTGGAAC
CAATTGCTTTTTGTGAATATGAAAAAGACGCGTCATCTTTTTTGTAGTCAACGTGGC
CTTCCGGTTCACGGAGATATCACGAAATTCGATGCTTCTGTGTACAAAAACAAGAT
CGATATCGTTACCGCTGGATGGCCATGTACCGGTTTCAGTACGGCCGGGAAAGGAA
CTGGTTTTCGAACATGAAGCGTCCGGTTTATGGACGGAAGTCGTCAGAGTGGTCAAA
GAAAGCGAACC AAAATATGTTTTTCTCGAGAATTCTCATGTGTTAGCTCAGACAAA
AAATCTCAAAGTAATCATTACGATCTCGACATTTTGGGATATGATACTCGTTGGT
GGACGTGTCGTTTGAACGATGTCAATGTAGGCGCACATCACAACAGATATAGATGG
TTCATGCTCGCCGAAAAGAAAGGTTCTGTGACAAAGTTCGTTAAAATTCAAGTGAA
AAAATTTAATTGGTCTGGTGATTTCAAAGAGAAACAAATTTAGAAAATTCTCATG
AAAACAAACAACTCATAAGTTTATGGGAAATAGTGTAGTTCCCGATCAGGTTTCA
TACGCTTTCGAATCGATGAGCGATATGATTCTCGAAGGATCTCTCGTGAACGATAA
GGACGAAATCGTGAAAGTAGGATATTCTAAAGATGGTATCATGTACAAGATACCAA
TAGAACATAAGATTATTCCAAAACCTTAATATCGTATTGACGCCTCGTGATCCTCCG
GAAGGTCACAAAGCTCGAGAAGAAGCAATCATCAATCTCCTATATTGATGACTTA
CTGGAATACTCCTGCATTCTGTTATCACAAATCCGCACGAGGAGCAAAAATTCTCA
CAAAACGACAGAAGAATAACTTACATACACAAATTAAATTCTGTCCTGGAGGTTCC
GATGATGGTTACTTGTCTGGAAGATTTTGCGCATGGCTCATGGGATATGACGACGA
ATATTTGGGTGATCTTATGCATTATTAA

Figure 12B (SEQ ID NO:31)

MRTKYRIFIVCMLMSMKALELFAGIGGITHGLRGYVEPIAFCEYEKDASSFLSQRG
LPVHGDITKFDASVYKNKIDIVTAGWPCTGFSTAGKGTGFEHEASGLWTEVVRVVK
ESEPKYVFLNSHVLAQTKNLKVIHDLIDILGYDTRWWTCRSNDVNVGAHNRYSRW
FMLAEKKGSVTKFVKIQVKKFNWSGDFKEKQISENSHENKQLIKFMGNSVVPDQVR
YAFESMSDMILEGSLVNDKDEIVKVGYSKDGIMYKIPIEHKIIPKLNIVLTPRDP
EGHKAREEAIKSPILMTYWNTPAFCYHKSARGAKILTKRQKNNLHTQIKFCPGGS
DDGYLSGRFCAWLMGYDDEYLGDLMHY

NEB-257-PCT.ST25.txt
SEQUENCE LISTING

<110> New England Biolabs, Inc.
Xu, Shuang-yong
Zhu, Zhenyu
Chan, Siu-hong
Xu, Yan

<120> Recombinant DNA Nicking Endonuclease and Uses Thereof

<130> NEB-257-PCT

<150> 60/620,939

<151> 2004-10-21

<160> 31

<170> PatentIn version 3.2

<210> 1

<211> 363

<212> PRT

<213> Chlorella virus NC64A

<400> 1

Met Arg Thr Lys Tyr Arg Ile Phe Ile Val Cys Met Leu Met Ser Met
1 5 10 15

Lys Ala Leu Glu Leu Phe Ala Gly Ile Gly Gly Ile Thr His Gly Leu
20 25 30

Arg Gly Tyr Val Glu Pro Ile Ala Phe Cys Glu Tyr Glu Lys Asp Ala
35 40 45

Ser Ser Phe Leu Ser Gln Arg Gly Leu Pro Val His Gly Asp Ile Thr
50 55 60

Lys Phe Asp Ala Ser Val Tyr Lys Asn Lys Ile Asp Ile Val Thr Ala
65 70 75 80

Gly Trp Pro Cys Thr Gly Phe Ser Thr Ala Gly Lys Gly Thr Gly Phe
85 90 95

Glu His Glu Ala Ser Gly Leu Trp Thr Glu Val Val Arg Val Val Lys
100 105 110

Glu Ser Glu Pro Lys Tyr Val Phe Leu Glu Asn Ser His Val Leu Ala
115 120 125

Gln Thr Lys Asn Leu Lys Val Ile Ile His Asp Leu Asp Ile Leu Gly
130 135 140

Tyr Asp Thr Arg Trp Trp Thr Cys Arg Ser Asn Asp Val Asn Val Gly
Page 1

NEB-257-PCT.ST25.txt

145 150 155 160
 Ala His His Asn Arg Tyr Arg Trp Phe Met Leu Ala Glu Lys Lys Gly
 165 170 175
 Ser Val Thr Lys Phe Val Lys Ile Gln Val Lys Lys Phe Asn Trp Ser
 180 185 190
 Gly Asp Phe Lys Glu Lys Gln Ile Ser Glu Asn Ser His Glu Asn Lys
 195 200 205
 Gln Leu Ile Lys Phe Met Gly Asn Ser Val Val Pro Asp Gln Val Arg
 210 215 220
 Tyr Ala Phe Glu Ser Met Ser Asp Met Ile Leu Glu Gly Ser Leu Val
 225 230 235 240
 Asn Asp Lys Asp Glu Ile Val Lys Val Gly Tyr Ser Lys Asp Gly Ile
 245 250 255
 Met Tyr Lys Ile Pro Ile Glu His Lys Ile Ile Pro Lys Leu Asn Ile
 260 265 270
 Val Leu Thr Pro Arg Asp Pro Pro Glu Gly His Lys Ala Arg Glu Glu
 275 280 285
 Ala Ile Ile Lys Ser Pro Ile Leu Met Thr Tyr Trp Asn Thr Pro Ala
 290 295 300
 Phe Cys Tyr His Lys Ser Ala Arg Gly Ala Lys Ile Leu Thr Lys Arg
 305 310 315 320
 Gln Lys Asn Asn Leu His Thr Gln Ile Lys Phe Cys Pro Gly Gly Ser
 325 330 335
 Asp Asp Gly Tyr Leu Ser Gly Arg Phe Cys Ala Trp Leu Met Gly Tyr
 340 345 350
 Asp Asp Glu Tyr Leu Gly Asp Leu Met His Tyr
 355 360

<210> 2
 <211> 362
 <212> PRT
 <213> Chlorella virus NC64A

<400> 2

Met Thr Leu Lys Ala Leu Glu Leu Phe Ala Gly Ile Ala Gly Ile Thr
 Page 2

NEB-257-PCT.ST25.txt

```

1              5              10              15
His Gly Leu Arg Gly Phe Val Glu Pro Val Ala Phe Val Glu Ile Asn
      20              25              30
Lys Asp Ala Gln Glu Phe Leu Ser Thr Lys Phe Pro Asp Lys Pro Val
      35              40              45
Phe Asp Asp Val Thr Lys Phe Ser Lys Arg Asp Phe Asp Glu Pro Ile
      50              55              60
Asp Met Ile Thr Gly Gly Phe Pro Cys Thr Gly Phe Ser Ile Ala Gly
      65              70              75              80
Lys Arg Asn Gly Phe Glu His Ala Glu Ser Gly Leu Phe Gly Glu Val
      85              90              95
Val Arg Ile Thr Lys Glu Tyr Met Pro Lys Met Val Phe Leu Glu Asn
      100              105              110
Ser Gly Met Leu Ser His Lys Tyr Asn Leu Asp Ile Val Ile Arg Ser
      115              120              125
Met Asp Ser Leu Gly Tyr Asp Cys Arg Trp Val Thr Leu Arg Ala Thr
      130              135              140
Val Val Gly Ala Leu His Thr Arg His Arg Trp Phe Cys Leu Cys Thr
      145              150              155              160
Arg Lys Asp His Ile Arg Glu Thr Leu Ile Cys Asp Arg Glu Val Thr
      165              170              175
Lys Phe Asp Trp Glu Asn Asp Arg Pro Pro Ile Gln Val Asp Ser Arg
      180              185              190
Ser Tyr Glu Asn Ser Arg Leu Val Arg Phe Ala Gly Tyr Ser Val Val
      195              200              205
Pro Asp Gln Ile Arg Tyr Ala Phe Thr Gly Leu Tyr Thr Gly Asn Phe
      210              215              220
Ser Pro Ser Phe Ser Lys Thr Leu Val Pro Gly Ser Leu Glu Gly Ser
      225              230              235              240
Ile Cys Phe Asn Glu Asp Lys Ile Thr Asn Gly Tyr Tyr Lys Asp Gly
      245              250              255

```

NEB-257-PCT.ST25.txt

Val Tyr Tyr Glu Phe Val Arg Thr Glu Thr His Arg Glu Pro Val Asn
 260 265 270

Ile Leu Leu Thr Pro Arg Glu Ile Pro Asn Lys His Asn Gly Lys Lys
 275 280 285

Leu Leu Thr Leu Pro Val Thr Lys Arg Tyr Trp Cys Thr Pro Cys Ala
 290 295 300

Ser Tyr Gly Lys Gly Thr Ala Gly Gly Arg Val Leu Thr Asp Arg Ser
 305 310 315 320

Ser His Ser Leu Pro Thr Gln Val Lys Phe Ser Pro Glu Gly Glu Asp
 325 330 335

Gly Lys His Leu Ser Gly Lys Phe Cys Ala Trp Leu Met Gly Tyr Asp
 340 345 350

Lys Glu Tyr Leu Gly Asn Leu Leu Glu Tyr
 355 360

<210> 3
 <211> 327
 <212> PRT
 <213> Haemophilus haemolyticus

<400> 3

Met Ile Glu Ile Lys Asp Lys Gln Leu Thr Gly Leu Arg Phe Ile Asp
 1 5 10 15

Leu Phe Ala Gly Leu Gly Gly Phe Arg Leu Ala Leu Glu Ser Cys Gly
 20 25 30

Ala Glu Cys Val Tyr Ser Asn Glu Trp Asp Lys Tyr Ala Gln Glu Val
 35 40 45

Tyr Glu Met Asn Phe Gly Glu Lys Pro Glu Gly Asp Ile Thr Gln Val
 50 55 60

Asn Glu Lys Thr Ile Pro Asp His Asp Ile Leu Cys Ala Gly Phe Pro
 65 70 75 80

Cys Gln Ala Phe Ser Ile Ser Gly Lys Gln Lys Gly Phe Glu Asp Ser
 85 90 95

Arg Gly Thr Leu Phe Phe Asp Ile Ala Arg Ile Val Arg Glu Lys Lys
 100 105 110

NEB-257-PCT.ST25.txt

Pro Lys Val Val Phe Met Glu Asn Val Lys Asn Phe Ala Ser His Asp
 115 120 125

Asn Gly Asn Thr Leu Glu Val Val Lys Asn Thr Met Asn Glu Leu Asp
 130 135 140

Tyr Ser Phe His Ala Lys Val Leu Asn Ala Leu Asp Tyr Gly Ile Pro
 145 150 155 160

Gln Lys Arg Glu Arg Ile Tyr Met Ile Cys Phe Arg Asn Asp Leu Asn
 165 170 175

Ile Gln Asn Phe Gln Phe Pro Lys Pro Phe Glu Leu Asn Thr Phe Val
 180 185 190

Lys Asp Leu Leu Leu Pro Asp Ser Glu Val Glu His Leu Val Ile Asp
 195 200 205

Arg Lys Asp Leu Val Met Thr Asn Gln Glu Ile Glu Gln Thr Thr Pro
 210 215 220

Lys Thr Val Arg Leu Gly Ile Val Gly Lys Gly Gly Gln Gly Glu Arg
 225 230 235 240

Ile Tyr Ser Thr Arg Gly Ile Ala Ile Thr Leu Ser Ala Tyr Gly Gly
 245 250 255

Gly Ile Phe Ala Lys Thr Gly Gly Tyr Leu Val Asn Gly Lys Thr Arg
 260 265 270

Lys Leu His Pro Arg Glu Cys Ala Arg Val Met Gly Tyr Pro Asp Ser
 275 280 285

Tyr Lys Val His Pro Ser Thr Ser Gln Ala Tyr Lys Gln Phe Gly Asn
 290 295 300

Ser Val Val Ile Asn Val Leu Gln Tyr Ile Ala Tyr Asn Ile Gly Ser
 305 310 315 320

Ser Leu Asn Phe Lys Pro Tyr
 325

<210> 4
 <211> 356
 <212> PRT
 <213> Chlorella virus NC64A
 <400> 4

NEB-257-PCT.ST25.txt

Met Tyr Ile Tyr Met Ser Thr Pro Gln Ala Lys Thr Lys Tyr Tyr Glu
 1 5 10 15
 Gln Arg Phe Val Asn Asp Phe Tyr Lys Glu Leu Glu Arg Asn Lys Val
 20 25 30
 Ser Leu Pro Val Thr Ile Val Leu Lys Asp Asn Leu Gly Ile Lys Gln
 35 40 45
 Val Ile Gln Asn Val Ser Gly Val Arg Val Leu Arg Asp Lys Ala Asn
 50 55 60
 Ala Lys Ser Pro Ser Lys Ile Lys Ser Glu Glu Leu Gly Arg His Val
 65 70 75 80
 Thr Ser Lys Ala Asp Ile Ala Leu Phe Thr Glu Glu Lys Asn Gly Thr
 85 90 95
 Lys Val Asp Val Ala Trp Ile Ser His Lys Ser Asn Lys Asp Ile His
 100 105 110
 Gly Lys Lys Ile Thr His Ala Gln Tyr Leu Asp Ala Ser Ser Asp Val
 115 120 125
 Met Phe Lys Thr Lys Ile Gly Gln Thr Lys Glu Ile Lys Asp Phe Lys
 130 135 140
 Asn Lys Met Ile Ser Leu Ser Val Pro Leu Thr Ala Thr Lys Tyr Cys
 145 150 155 160
 Trp Pro Lys Tyr Lys Ser Gly Thr Ser Leu Arg Ile Trp Asp Asp Val
 165 170 175
 Lys Ser Thr Ile Leu Met Asn Met Ala Ile Phe Gly Val Glu Phe Gly
 180 185 190
 Lys Ala Tyr Ser Arg Asn Asn Ala Asn Ile Leu Met Gly Gly Asp Pro
 195 200 205
 Leu Ile Glu Val Lys Asp Asp Lys Thr Ile Ile Leu Thr Thr Lys Glu
 210 215 220
 Asn Gly Phe Ser Leu Ala Asn Gly Phe Ala Glu Tyr Ile Pro Ser Lys
 225 230 235 240
 Asp Lys Pro Ile Phe Phe Thr Lys Pro Thr Ser Gly Lys Lys Thr Val
 245 250 255

NEB-257-PCT.ST25.txt

Val Asp Gly Lys Thr Ile Glu Gly Val Ser Val Trp Ile Ile Tyr Arg
 260 265 270

Ser Tyr Ala Gly Ser Lys Asn Arg Lys Ile Asp Asp Val Leu Lys Asn
 275 280 285

Lys Ile Glu Leu Ile Ser Ser Ser Cys Ser Val Lys Lys Lys Asp Asn
 290 295 300

Phe Val Ser Ile Met Gln Ser Lys Lys Ile Thr Ser Pro Pro Lys Ser
 305 310 315 320

Lys Lys Ile Thr Ser Pro Pro Lys Ser Lys Lys Ile Thr Ser Pro Ser
 325 330 335

Lys Ser Lys Lys Ile Thr Asn Phe Phe Met Lys Lys Leu Lys Tyr Ser
 340 345 350

Leu Leu Ser Arg
 355

<210> 5
 <211> 278
 <212> PRT
 <213> Chlorella virus NC64A

<400> 5

Met Glu Glu Lys Lys Arg Leu Ala Leu Ile Glu Lys Gln Arg Ile Ala
 1 5 10 15

Glu Glu Lys Ile Ala Ser Gly Arg Lys Ile Arg Lys Arg Ile Ser Thr
 20 25 30

Asn Ala Thr Lys His Glu Arg Glu Phe Val Lys Val Ile Asn Ser Met
 35 40 45

Phe Val Gly Pro Ala Thr Phe Val Phe Val Asp Ile Lys Gly Asn Lys
 50 55 60

Ser Arg Glu Ile His Asn Val Val Arg Phe Arg Gln Leu Gln Gly Ser
 65 70 75 80

Lys Ala Lys Ser Pro Thr Ala Tyr Val Asp Arg Glu Tyr Asn Lys Pro
 85 90 95

Lys Ala Asp Ile Ala Ala Val Asp Ile Thr Gly Lys Asp Val Ala Trp
 100 105 110

NEB-257-PCT.ST25.txt

Ile Ser His Lys Ala Ser Glu Gly Tyr Gln Gln Tyr Leu Lys Ile Ser
 115 120 125

Gly Lys Asn Leu Lys Phe Thr Gly Lys Glu Leu Glu Glu Val Leu Ser
 130 135 140

Phe Lys Arg Lys Val Val Ser Met Ala Pro Val Ser Lys Ile Trp Pro
 145 150 155 160

Ala Asn Lys Thr Val Trp Ser Pro Ile Lys Ser Asn Leu Ile Lys Asn
 165 170 175

Gln Ala Ile Phe Gly Phe Asp Tyr Gly Lys Lys Pro Gly Arg Asp Asn
 180 185 190

Val Asp Ile Ile Gly Gln Gly Arg Pro Ile Ile Thr Lys Arg Gly Ser
 195 200 205

Ile Leu Tyr Leu Thr Phe Thr Gly Phe Ser Ala Leu Asn Gly His Leu
 210 215 220

Glu Asn Phe Thr Gly Lys His Glu Pro Val Phe Tyr Val Arg Thr Glu
 225 230 235 240

Arg Ser Ser Ser Gly Arg Ser Ile Thr Thr Val Val Asn Gly Val Thr
 245 250 255

Tyr Lys Asn Leu Arg Phe Phe Ile His Pro Tyr Asn Phe Val Ser Ser
 260 265 270

Lys Thr Gln Arg Ile Met
 275

<210> 6
 <211> 11
 <212> PRT
 <213> Chlorella virus NC64A

<400> 6

Met Ser Thr Pro Gln Ala Lys Thr Lys Tyr Tyr
 1 5 10

<210> 7
 <211> 49
 <212> DNA
 <213> unknown

<220>
 <223> primer

NEB-257-PCT.ST25.txt

<220>
<221> misc_feature
<222> (0)..(9)
<223> s=g or c

<400> 7
cggttaagcsa tgcattgcgta caaagtatcg tttttttatc gtatgtatg 49

<210> 8
<211> 10
<212> DNA
<213> Chlorella virus NC64A

<400> 8
tacggccggg 10

<210> 9
<211> 10
<212> DNA
<213> Chlorella virus NC64A

<220>
<221> misc_feature
<222> (7)..(7)
<223> n is a, c, g, or t

<400> 9
tatggcnngg 10

<210> 10
<211> 10
<212> DNA
<213> Chlorella virus NC64A

<400> 10
atggccatgt 10

<210> 11
<211> 10
<212> DNA
<213> Chlorella virus NC64A

<400> 11
atggctatgt 10

<210> 12
<211> 15
<212> DNA
<213> Chlorella virus NC64A

<400> 12
cctgatcggg aacta 15

<210> 13

NEB-257-PCT.ST25.txt

<211> 15
 <212> DNA
 <213> Chlorella virus NC64A

<220>
 <221> misc_feature
 <222> (4)..(4)
 <223> n is a, c, g, or t

<220>
 <221> misc_feature
 <222> (8)..(8)
 <223> n is a, c, g, or t

<400> 13
 cctnatacngg aacta 15

<210> 14
 <211> 11
 <212> DNA
 <213> Chlorella virus NC64A

<400> 14
 gaaccaaata a 11

<210> 15
 <211> 11
 <212> DNA
 <213> Chlorella virus NC64A

<400> 15
 gaaccaaata a 11

<210> 16
 <211> 17
 <212> DNA
 <213> Chlorella virus NC64A

<400> 16
 gatacttgat gcgaatg 17

<210> 17
 <211> 12
 <212> DNA
 <213> unknown

<220>
 <223> primer

<400> 17
 gatacttgat ga 12

<210> 18
 <211> 59
 <212> DNA
 <213> Chlorella virus NC64A

NEB-257-PCT.ST25.txt

<220>
<221> misc_feature
<222> (24)..(24)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (32)..(32)
<223> n is a, c, g, or t

<400> 18
gatacgtcgt aagatacttg tggngaattgc gngcgataag cttggcgtaa tcatgggtcg 59

<210> 19
<211> 60
<212> DNA
<213> Chlorella virus NC64A

<400> 19
cgatacgtcg taagatactt gcggcgaatg caagcgataa gcttggcgta atcatggccc 60

<210> 20
<211> 59
<212> DNA
<213> Chlorella virus NC64A

<220>
<221> misc_feature
<222> (23)..(23)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (46)..(46)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (51)..(51)
<223> n is a, c, g, or t

<400> 20
cgatacgtcg taagatactt ggngcgaatg caagcgataa gcttgncgta ntcatggac 59

<210> 21
<211> 62
<212> DNA
<213> Chlorella virus NC64A

<400> 21
gatacgtcgt aagatcttga ggcgaatgag cgataagctt ggcgtaatca tggacatagc 60

tg 62

<210> 22
<211> 43
<212> DNA

NEB-257-PCT.ST25.txt

<213> unknown

<220>

<223> primer

<400> 22

gaagtcgact taataatgca taagatcacc caaatattcg tcg

43

<210> 23

<211> 45

<212> DNA

<213> unknown

<220>

<223> primer

<400> 23

accggtgagc atatgtatat atatatgtct actccgcagg caaag

45

<210> 24

<211> 66

<212> DNA

<213> unknown

<220>

<223> primer

<400> 24

ggtggttgct cttccgcaac aggaagaaga aattaattct attttatttt tcaaaacatc

60

atcgat

66

<210> 25

<211> 54

<212> DNA

<213> unknown

<220>

<223> primer

<400> 25

ggtggttgct cttccgcagc attttggtgg acttggtatt ttctttgatt ttgg

54

<210> 26

<211> 20

<212> DNA

<213> unknown

<220>

<223> primer

<220>

<221> misc_feature

<222> (4)..(4)

<223> m=a or c

<400> 26

tgcmtcctgc accaccaact

20

NEB-257-PCT.ST25.txt

<210> 27
 <211> 19
 <212> DNA
 <213> unknown

<220>
 <223> primer

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> y=c or t

<400> 27
 ygcctgcttc accaccttc

19

<210> 28
 <211> 1047
 <212> DNA
 <213> Chlorella virus NC64A

<400> 28
 atgtatatat atatgtctac tccgcaggca aagaccaaatt attatgaaca aagatttcgta 60
 aacgattttct acaaagaact cgaacgtaat aaagtttctc tgcctgtcac aatagttttg 120
 aaggataatc tgggaataaa acaagttata cagaatggga gtggtgttag agttttacgt 180
 gataaagcta atgcgaaatc tccgtctaag ataaaaagtg aagaattagg tagacatgtg 240
 acatcgaaag cggatatagc attatttacc gaggaaga atggaacaaa agttgacgtt 300
 gcatggatat ccccccaatc tcataaagat tttcttgga aaaagataac tcctgctcaa 360
 tattttgatg cctcctcaga tgttatgttt aaaacaaaa ttggccaacc aaaggaaata 420
 aaggagctta agaataagat gatttccttg agtgttcctc tcacggcaac gaaatattgc 480
 tggcccaagt acaaatctgg aacatctctc agaatatggg acgatgtaca aagcactata 540
 cttatgaaca tggcaatatt tgggtgttag tttggaaagg catattgtcg aaataatgca 600
 aatattttga tggttggaga tcctcttata gaagtaaaag atgacaaaac gataattctt 660
 accacaaaag aaaatggatt tagtctggca aacggatttg cagaatatat accatcaaag 720
 gataaaccta tatttttcac taaaccaaca tcaggaaaaa aaacagtcgt agatggaaaa 780
 acgatcgagg gagtgagtgt atggataatt tatagaagtt atgcaggatc aaagaataga 840
 aaaatcgatg atgttttgaa aaataaaata gaattaattt cttcttcctg ttctgttaaa 900
 aagaaagata actttgtttc tatcatgcaa tcaaagaaaa taacaagtcc accaaaatca 960
 aagaaaataa caagtccacc aaaatcaaag aaaataacaa gtccatcaaa atcaaagaaa 1020
 ataacaaatt tttttatgaa aaaataa 1047

NEB-257-PCT.ST25.txt

<210> 29

<211> 348

<212> PRT

<213> Chlorella virus NC64A

<400> 29

Met Tyr Ile Tyr Met Ser Thr Pro Gln Ala Lys Thr Lys Tyr Tyr Glu
 1 5 10 15

Gln Arg Phe Val Asn Asp Phe Tyr Lys Glu Leu Glu Arg Asn Lys Val
 20 25 30

Ser Leu Pro Val Thr Ile Val Leu Lys Asp Asn Leu Gly Ile Lys Gln
 35 40 45

Val Ile Gln Asn Gly Ser Gly Val Arg Val Leu Arg Asp Lys Ala Asn
 50 55 60

Ala Lys Ser Pro Ser Lys Ile Lys Ser Glu Glu Leu Gly Arg His Val
 65 70 75 80

Thr Ser Lys Ala Asp Ile Ala Leu Phe Thr Glu Glu Lys Asn Gly Thr
 85 90 95

Lys Val Asp Val Ala Trp Ile Ser Pro Gln Ser His Lys Asp Phe Leu
 100 105 110

Gly Lys Lys Ile Thr Pro Ala Gln Tyr Phe Asp Ala Ser Ser Asp Val
 115 120 125

Met Phe Lys Thr Lys Ile Gly Gln Pro Lys Glu Ile Lys Glu Leu Lys
 130 135 140

Asn Lys Met Ile Ser Leu Ser Val Pro Leu Thr Ala Thr Lys Tyr Cys
 145 150 155 160

Trp Pro Lys Tyr Lys Ser Gly Thr Ser Leu Arg Ile Trp Asp Asp Val
 165 170 175

Gln Ser Thr Ile Leu Met Asn Met Ala Ile Phe Gly Val Glu Phe Gly
 180 185 190

Lys Ala Tyr Cys Arg Asn Asn Ala Asn Ile Leu Met Val Gly Asp Pro
 195 200 205

Leu Ile Glu Val Lys Asp Asp Lys Thr Ile Ile Leu Thr Thr Lys Glu
 210 215 220

NEB-257-PCT.ST25.txt

Asn Gly Phe Ser Leu Ala Asn Gly Phe Ala Glu Tyr Ile Pro Ser Lys
 225 230 235 240

Asp Lys Pro Ile Phe Phe Thr Lys Pro Thr Ser Gly Lys Lys Thr Val
 245 250 255

Val Asp Gly Lys Thr Ile Glu Gly Val Ser Val Trp Ile Ile Tyr Arg
 260 265 270

Ser Tyr Ala Gly Ser Lys Asn Arg Lys Ile Asp Asp Val Leu Lys Asn
 275 280 285

Lys Ile Glu Leu Ile Ser Ser Ser Cys Ser Val Lys Lys Lys Asp Asn
 290 295 300

Phe Val Ser Ile Met Gln Ser Lys Lys Ile Thr Ser Pro Pro Lys Ser
 305 310 315 320

Lys Lys Ile Thr Ser Pro Pro Lys Ser Lys Lys Ile Thr Ser Pro Ser
 325 330 335

Lys Ser Lys Lys Ile Thr Asn Phe Phe Met Lys Lys
 340 345

<210> 30

<211> 1092

<212> DNA

<213> Chlorella virus NC64A

<400> 30

```

atgcgtacaa agtatcgtat ttttatcgta tgtatgttaa tgagtatgaa agcattagaa      60
ctcttcgcag gtataggcgg gatcactcat ggtcttcgtg gttacgtgga accaattgct      120
ttttgtgaat atgaaaaaga cgcgtcatct tttttgagtc aacgtggcct tccggttcac      180
ggagatatca cgaaattcga tgcttctgtg tacaaaaaca agatcgatat cgttaccgct      240
ggatggccat gtaccggttt cagtacggcc gggaaaggaa ctggtttcga acatgaagcg      300
tccggtttat ggacggaagt cgtcagagtg gtcaaagaaa gcgaaccaaa atatgttttt      360
ctcgagaatt ctcatgtgtt agctcagaca aaaaatctca aagtaatcat tcacgatctc      420
gacatttttg gatatgatac tcgttggtgg acgtgtcgtt cgaacgatgt caatgtaggc      480
gcacatcaca acagatatag atggttcattg ctgccgaaa agaaagggttc tgtgacaaag      540
ttcgttaaaa ttcaagtga aaaatttaatt tggctctggtg atttcaaaga gaaacaaatt      600
tcagaaaatt ctcatgaaaa caaacaactc ataaagttta tgggaaatag tgtagttccc      660
gatcagggtt gatacgcttt cgaatcgatg agcgatatga ttctcgaagg atctctcgtg      720
aacgataagg acgaaatcgt gaaagtagga tattctaaag atgggtatcat gtacaagata      780

```

NEB-257-PCT.ST25.txt

```

ccaatagaac ataagattat tccaaaactt aatatcgtat tgacgcctcg tgatcctccg      840
gaaggtcaca aagctcgaga agaagcaatc atcaaactc ctatattgat gacttactgg      900
aatactcctg cattctgtta tcacaaatcc gcacgaggag caaaaattct cacaaaacga      960
cagaagaata acttacatac acaaattaa ttctgtcctg gaggttccga tgatgggttac    1020
ttgtctggaa gattttgcg atggctcatg ggatatgacg acgaatattt gggatgatctt    1080
atgcattatt aa                                                           1092

```

<210> 31
 <211> 363
 <212> PRT
 <213> Chlorella virus NC64A

<400> 31

Met Arg Thr Lys Tyr Arg Ile Phe Ile Val Cys Met Leu Met Ser Met
 1 5 10 15

Lys Ala Leu Glu Leu Phe Ala Gly Ile Gly Gly Ile Thr His Gly Leu
 20 25 30

Arg Gly Tyr Val Glu Pro Ile Ala Phe Cys Glu Tyr Glu Lys Asp Ala
 35 40 45

Ser Ser Phe Leu Ser Gln Arg Gly Leu Pro Val His Gly Asp Ile Thr
 50 55 60

Lys Phe Asp Ala Ser Val Tyr Lys Asn Lys Ile Asp Ile Val Thr Ala
 65 70 75 80

Gly Trp Pro Cys Thr Gly Phe Ser Thr Ala Gly Lys Gly Thr Gly Phe
 85 90 95

Glu His Glu Ala Ser Gly Leu Trp Thr Glu Val Val Arg Val Val Lys
 100 105 110

Glu Ser Glu Pro Lys Tyr Val Phe Leu Glu Asn Ser His Val Leu Ala
 115 120 125

Gln Thr Lys Asn Leu Lys Val Ile Ile His Asp Leu Asp Ile Leu Gly
 130 135 140

Tyr Asp Thr Arg Trp Trp Thr Cys Arg Ser Asn Asp Val Asn Val Gly
 145 150 155 160

Ala His His Asn Arg Tyr Arg Trp Phe Met Leu Ala Glu Lys Lys Gly
 165 170 175

NEB-257-PCT.ST25.txt

Ser Val Thr Lys Phe Val Lys Ile Gln Val Lys Lys Phe Asn Trp Ser
 180 185 190

Gly Asp Phe Lys Glu Lys Gln Ile Ser Glu Asn Ser His Glu Asn Lys
 195 200 205

Gln Leu Ile Lys Phe Met Gly Asn Ser Val Val Pro Asp Gln Val Arg
 210 215 220

Tyr Ala Phe Glu Ser Met Ser Asp Met Ile Leu Glu Gly Ser Leu Val
 225 230 235 240

Asn Asp Lys Asp Glu Ile Val Lys Val Gly Tyr Ser Lys Asp Gly Ile
 245 250 255

Met Tyr Lys Ile Pro Ile Glu His Lys Ile Ile Pro Lys Leu Asn Ile
 260 265 270

Val Leu Thr Pro Arg Asp Pro Pro Glu Gly His Lys Ala Arg Glu Glu
 275 280 285

Ala Ile Ile Lys Ser Pro Ile Leu Met Thr Tyr Trp Asn Thr Pro Ala
 290 295 300

Phe Cys Tyr His Lys Ser Ala Arg Gly Ala Lys Ile Leu Thr Lys Arg
 305 310 315 320

Gln Lys Asn Asn Leu His Thr Gln Ile Lys Phe Cys Pro Gly Gly Ser
 325 330 335

Asp Asp Gly Tyr Leu Ser Gly Arg Phe Cys Ala Trp Leu Met Gly Tyr
 340 345 350

Asp Asp Glu Tyr Leu Gly Asp Leu Met His Tyr
 355 360