# Nucleotide Sequence Analysis of *p10* Gene of *Antheraea pernyi* Nucleopolyhedrovirus and Construction of Two Transfer Vector Plasmids

Ken-ichi Maegawa, Xue Ying Wang, Jun Kobayashi and Tetsuro Yoshimura

*Reprinted from*
*Int. J. Wild Silkmoth & Silk*
Vol. 9, 2004

*Int. J. Wild Silkmoth & Silk* **9**, 53–60 (2004)


# Nucleotide Sequence Analysis of *p10* Gene of *Antheraea pernyi* Nucleopolyhedrovirus and Construction of Two Transfer Vector Plasmids

Ken-ichi Maegawa[1], Xue Ying Wang[2], Jun Kobayashi[1,*] and Tetsuro Yoshimura[1]

[1] *Faculty of Engineering, Mie University, Tsu, Mie 514-8507, Japan.*

[2] *Bioscience and Technology College of Shenyang Agricultural University, Shenyang 110161, China.*

* *Present Address; Faculty of Agriculture, Yamaguchi University, Yamaguchi 753-8515, Japan.*

**Abstract** **The *p10* gene of *Antheraea pernyi* nucleopolyhedrovirus (AnpeNPV) was identified in 4.8-kbp *Pst*I I fragment of viral DNA by comparing genome maps between AnpeNPV and *Orgyia pseudotsugata* multinucleocapsid NPV. The nucleotide sequence analysis revealed that the AnpeNPV *p10* gene consists of the promoter sequence (55 bp) and the coding sequence (264 bp) encoding 87 amino acids. The predicted amino acid sequence of the AnpeNPV *p10* gene product showed the highest identity (93%) to that of *Choristoneura fumiferana* NPV among other NPVs by homology search using the BLAST program. Two transfer vector plasmids, *p10* locus-based pApp10 and *polyhedrin* locus-based pApCH3, were constructed for the foreign gene expression under the *p10* gene promoter. Among them, pApCH3 is a dual expression vector, which contains a copy of the *p10* promoter and polyadenylation signal inserted in tandem, downstream of the *polyhedrin* promoter, but opposite orientation.**

**Key words:** ***Antheraea pernyi*, nucleopolyhedrovirus, *p10*, transfer vector**

## Introduction

Previously, we have constructed a transfer vector plasmid pApCH1 using the strong *polyhedrin* promoter of *Antheraea pernyi* NPV (AnpeNPV) and established a baculovirus expression vector system (Kobayashi *et al.*, 2001). Using the system, we have demonstrated the high-level expression of foreign genes in cultured A. *pernyi* cells (AnPe) and diapausing pupae (Huang *et al.*, 2001), and the mammalian-like biantennary, complex N-glycan addition to a recombinant glycoprotein in AnPe cells (Nagaya *et al.*, 2002, 2003). Because of these advantageous characteristics for the protein production, we consider the AnpeNPV vector system is an ideal natural bioreactor (Kobayashi, 2002). We have also succeeded to suppress proteolytic degradation and liquefaction of virus-infected AnPe cells and pupae by deleting the virus-encoded *cathepsin* gene (Huang and Kobayashi, 2003).

In addition to the *polyhedrin* promoter, the promoter of another baculovirus very late gene, *p10*, have been frequently used for the high level expression of foreign genes not only for single gene expression but also for dual gene expression in other systems using *Autographa californica* NPV (AcNPV) and *Bombyx mori* NPV (BmNPV) (Vlak *et al.*, 1990; Weyer, U *et al.*, 1990; Tomita *et al.*, 1995). Therefore, it is

expected that the utilization of the *p10* promoter could enhance potential capability and usability to produce recombinant proteins by the AnpeNPV vector system.

In this paper, we describe the identification and nucleotide sequence of the *p10* gene of AnpeNPV, and construction of two transfer vector plasmids containing the *p10* promoter, which enable us to insert foreign genes in the AnpeNPV genome at the *p10* locus for single expression and at the *polyhedrin* locus for dual expression, respectively.

## Materials and Methods

*Virus, Bacterium and Plasmid*

The 4.8-kbp *Pst*I I fragment of DNA genome of AnpeNPV A strain (Huang *et al.*, 2002) was used for the *p10* gene analysis and transfer vector construction. For the vector construction, the 1.8-kbp *Pst*I b fragment (Huang *et al.*, 2002) was also used. Competent *E. coli* strain XL1-Blue cells (Stratagene) were used for plasmid DNA transformations. Plasmids, pBluescriptII (Stratagene), pCR 2.1 (Invitrogen) and pApCH1 (Kobayashi *et al.*, 2001) were used for construction of the transfer vector plasmids pApp10 and pApCH3 as described later.

*DNA manipulations*

All plasmid DNA recombination techniques were essentially as described by Sambrook *et al.* (1989). Restriction enzymes and other DNA modifying enzymes were purchased from Takara-Shuzo Co. Ltd.

*DNA sequencing and sequence analysis*

The nucleotide sequence of the AnpeNPV *p10* gene region was determined by using DNA sequencer (ABI PRISM 310 Genetic Analyzer, Applied Biosystems) and BigDye terminator cycle sequencing kit (Applied Biosystems) with the following oligonucleotide primers.

M13 forward, 5'-CGTTGTAAAACGACGGCCAG-3'
M13 reverse, 5'-CAGGAAACAGCTATGACCAT-3'
P26DN, 5'-TTACCACGTGTGCAACAAGC-3'
P10UP, 5'-AGATCTGGTAAAATATTATAACTATT-3'
P10DN, 5'-AGATCTCAAGTCAATGGTTTTAAAAT-3'
P74DN, 5'-CAGATTTTGACGAGGCGACG-3'

At each 5' end of P10UP and P10DN, a *Bgl*II restriction site (AGATCT) was added to create a foreign gene-cloning site just downstream of the *p10* promoter by PCR.

Homology search of predicted amino acid sequence of the AnpeNPV *p10* gene product was carried out using the basic local alignment search tool (BLAST) program provided by the GenomeNet WWW Server (Bioinformatics Center, Kyoto University). The GENETYX program (Genetyx) was used for both sequence data analysis and phylogenetic tree construction by the unweighted pair-group method with arithmetic mean (UPGMA).

*Nucleotide sequence accession number*

The 1873-bp nucleotide sequence within *Pst*I I fragment containing AnpeNPV *p26*, *p10* and *p74* (partial) has been submitted to the DDBJ/EMBL/Genbank databases under accession number AB106130.

## Results and Discussion

*Identification and sequence analysis of the AnpeNPV* p10 *gene*

By comparing genome maps between AnpeNPV, and OpMNPV, the *p10* gene of AnpeNPV was predicted to exist within *Pst*I I fragment (Huang *et al.*, 2002). By constructing several deletion mutants and using the various primers described in Materials and Methods, 1873-bp nucleotide sequence within 4.8-kbp *Pst*I I fragment was determined. As shown in Fig. 1, three ORFs (2 complete and 1 partial ORFs) were found in the sequenced region. Homology search revealed that these 3 ORFs were homologs of OpMNPV ORFs 132 (*p26*), 133 (*p10*) and 134 (*p74*) (Ahrens *et al.*, 1997), respectively. Relative positions and directions of these OpMNPV ORF homologs were conserved in the AnpeNPV genome, except *p74* whose relative direction was reverse as found in several group I NPVs such as AcNPV (Kuzio *et al.*, 1989), BmNPV (Gomi *et al.*, 1999) and *Choristoneura fumiferana* NPV (CfNPV) (Hill

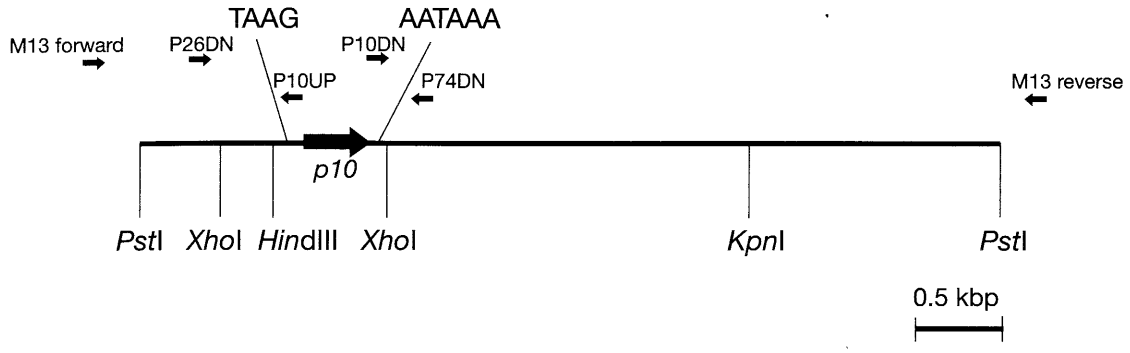


Fig. 1. Physical map of 4.8-kbp *Pst*I I fragment of AnpeNPV genome. Positions of consensus late promoter transcription start signal (TAAG) and polyadenylation signal (AATAAA) are indicated. Several restriction sites are also indicated. Small arrows and a large arrow indicate positions and directions of oligonucleotide primers and *p10* coding region, respectively. A scale bar represents 500 bp.

*et al.*, 1993).

Sequence analysis of the *p10* gene region revealed the consensus late promoter transcription start signal (TAAG), translation start codon (ATG), translation stop codon (TAG) and putative polyadenylation signal (AATAAA), respectively (Fig. 2), indicating that the AnpeNPV *p10* gene consisted of a promoter sequence of 55 bp and a coding sequence of 264 bp encoding 87 amino acids.

Homology search for predicted amino acid sequence of the AnpeNPV *p10* gene product using BLAST program revealed very high identity (93%) to that of CfNPV and relatively high identities (45 ~ 52%) to those of other group I NPVs, while those of group II NPVs showed lower identities (31 ~ 36%), except 47% of *Buzura suppressaria* NPV (BusuNPV). A phylogenetic tree constructed on the basis of multiple alignments of p10 sequences, however, supported the separation of the group I and II NPVs (Fig. 3). Therefore, the relatively high identity of p10 sequence between AnpeNPV and BusuNPV is likely to be an overestimation derived by the BLAST program algorithm for sequence comparison.

All of the results indicated that AnpeNPV possesses a *p10* gene homolog, which is highly identical to the CfNPV *p10* and has no obvious sequence deficiency disturbing its expression at the very late phase of virus infection.

*Construction of transfer vector pApp10 and pApCH3*

Two transfer vector plasmids, pApp10 and pApCH3, were constructed for generating recombinant AnpeNPVs expressing foreign genes under the control of the *p10* promoter as follows (Fig. 4).

First, 0.5 kbp of 5' upstream fragment containing the *p10* promoter and 0.3 kbp of 3' downstream fragment containing the polyadenylation signal were amplified by PCR using the *Pst*I I fragment as a template and two different primer pairs, P26DN/P10UP and P10DN/P74DN, respectively. Then, after trimming the amplified upstream fragment with restriction enzymes, *Hin*dIII and *Bgl*II, and the downstream fragment with *Bgl*II and *Xho*I, a 0.4-kbp *Hin*dIII-*Xho*I fragment was obtained by ligation of the trimmed upstream and downstream fragments (0.2 kbp each) at their *Bgl*II end.

For the construction of *p10* locus-based transfer vector plasmid, the 0.4-kbp *Hin*dIII-*Xho*I fragment was cloned into pBluescriptII together with the 1.8-kbp *Pst*I b fragment and 2 subfragments derived from the *Pst*I I fragment, 0.8-kbp *Pst*I-*Hin*dIII and 2.0-kbp *Xho*I-*Kpn*I fragments (Fig. 4a). The resulting pApp10 (7.9 kbp) contains both 5' and 3' flanking regions of the *p10* gene for homologous recombination with AnpeNPV DNA at the *p10* locus and between them 1 unique restriction site (*Bgl*II) for insertion of a foreign gene just downstream of the *p10* promoter.

```
   1 CTGCAGCGTCAGCAGCGCGGCGCGATTAAACTCACGGTCGTTGTACGTCGACGCCGAGTCGTCCGTAGCATACACGTAATACTTGTACAT  90
     PstI
  91 GTTGGCGAATGATGCGGCTGCACTTTATATACTGCTGTGCTCCGCAAACCTCGCATAAACACACGTGATGGATAACAAAATGTACGTGCG 180
                                                                          M  D  N  K  M  Y  V  R
                                                                      p26 --->
 181 CATGGAGATTAAGTTTGACGAGGACACGGGCCGTTTGCAAATTGGTGACCGGGACGTGTTTGTGCGCGTGTTTGAGCCCGGCCAAGAGGT 270
       M  E  I  K  F  D  E  D  T  G  R  L  Q  I  G  D  R  D  V  F  V  R  V  F  E  P  G  Q  E  V

 271 GTTTGACGAAACGTTGGACCAATACCACCAGTTTCCCGGCGTGGCCACAGACGTTGTGTTTCCGCAAGTGGACACGGGCGCGACGGTGAG 360
       F  D  E  T  L  D  Q  Y  H  Q  F  P  G  V  A  T  D  V  V  F  P  Q  V  D  T  G  A  T  V  S
                                                                  P26DN --->
 361 CGTGCACACCGCGGCCGGCGCGTACTCAGGGCCGCTAACCGCCAGCTGCTTCAATTACCACGTGTGCAACAAGCGCTTCGTGTTTGGTGT 450
       V  H  T  A  A  G  A  Y  S  G  P  L  T  A  S  C  F  N  Y  H  V  C  N  K  R  F  V  F  G  V

 451 TTTGCCGGCGCTCGAGGTGCCCGCCGACGTAAGCCCGCACCTGCGCGTCGGCGCGCCCATCCTCTGCAACGAGCAATTGGTGTCCGTGGT 540
       L  P  A  L  E  V  P  A  D  V  S  P  H  L  R  V  G  A  P  I  L  C  N  E  Q  L  V  S  V  V
                XhoI
 541 GACGGCCGTGCACGAGCGCGCCGACGGCGTCTGGCTGGTGCCCGTGACGGGCGTGCGCGGGCCGCACCAGGTGTCGGGACACGCGCGCGT 630
       T  A  V  H  E  R  A  D  G  V  W  L  V  P  V  T  G  V  R  G  P  H  Q  V  S  G  H  A  R  V

 631 GTGCAATGGCGTGCGCGCAGAGCGCCTGCGCGCCGGGCGGTCGGTGTACGGCGCGGTGCAACTACCATATGACAAACTGAAGACGCACGC 720
       C  N  G  V  R  A  E  R  L  R  A  G  R  S  V  Y  G  A  V  Q  L  P  Y  D  K  L  K  T  H  A

 721 GCTATCGCAAACGGCGCCACACGCTGAAGCTTCCGAATCGTGCGCGCTGTTTTACAATGACTCTGAAGTGCGCATTACTTTTAACAAAGG 810
       L  S  Q  T  A  P  H  A  E  A  S  E  S  C  A  L  F  Y  N  D  S  E  V  R  I  T  F  N  K  G
                                HindIII
 811 CAGTTTTGAATTGATGCACTGGCGATTGCCGGGGCCTTTTGTTGCCACAGCGTTAAATAAGCATTATTGACAATTTATTATTATATCATT 900
       S  F  E  L  M  H  W  R  L  P  G  P  F  V  A  T  A  L  N  K  H  Y  *
              <--- P10UP
 901 GAAATAGTTATAATATTTTACCATGTCCAAACCAAGCATTTTACAACAAATTTTGACAGCTGTGCAAGATGTCGACACAAAAGTTGACGC 990
                           M  S  K  P  S  I  L  Q  Q  I  L  T  A  V  Q  D  V  D  T  K  V  D  A
                        p10 --->
 991 GTTGCAAGCGCAACTGACTGAATTGGACGGCAAAGTTCAGCCGTTAGATGGTTTGTCCGAGCAACTGACCGCTTTGGATACTAAAGTGAC1080
       L  Q  A  Q  L  T  E  L  D  G  K  V  Q  P  L  D  G  L  S  E  Q  L  T  A  L  D  T  K  V  T

1081 CACAATTCAAGACATACTAGGTGGAGCGGAAATTCCCGATATTCCCGATGTGCCTCTACCCGACAATCCTTTGAACAAAACCCGCAGTCA1170
       T  I  Q  D  I  L  G  G  A  E  I  P  D  I  P  D  V  P  L  P  D  N  P  L  N  K  T  R  S  Q

              *  F  L  L  D  I  T  K  F  Y  E  Y  P  E  K  I  L  Y  Y  L  A  I  M  L  F  I  A  F
1171 AGCAAAATTAAAATAGCAAGTCAATGGTTTTAAAATATTCGTACGGTTCCTTGATTAAATAGTACAATGCAATCATTAAAAATATCGCAA1260
       A  K  L  K  *        P10DN --->

       Y  V  F  F  T  L  E  K  Y  L  I  F  A  A  V  A  A  G  T  N  L  L  F  L  G  A  V  V  R  N
1261 AGTACACAAAAAATGTAAGCTCTTTGTACAATATAAAGGCCGCAACGGCCGCCCCCGTGTTCAGCAAAAATAAACCCGCTACCACCCTAT1350

       L  K  N  N  N  E  N  M  D  L  L  T  N  Q  R  F  T  Y  Q  M  F  E  L  R  T  Y  L  S  S  S
1351 TTAATTTGTTATTATTTTCGTTCATGTCCAATAGCGTGTTTTGCCTAAACGTGTACTGCATAAACTCGAGACGCGTGTACAGCGAGCTGC1440
                                                                          XhoI
       A  L  A  Q  G  V  L  T  A  E  D  F  D  T  I  Q  D  S  E  D  L  Q  L  M  Q  G  D  S  N  V
1441 TGGCCAGCGCTTGGCCCACCAGCGTCGCCTCGTCAAAATCTGTAATCTGGTCACTTTCATCCAATTGTAGCATTTGACCGTCCGAATTGA1530
                              <--- P74DN
       E  L  A  A  V  Y  D  L  L  H  Y  L  S  Q  F  M  A  D  D  D  T  E  V  F  D  A  F  F  E  P
1531 CCTCAAGCGCCGCCACGTAGTCGAGCAGGTGGTACAGCGATTGGAACATGGCGTCGTCGTCTGTCTCTACAAAATCGGCAAAAAATTCGG1620

       L  F  E  I  I  E  R  S  S  N  A  D  L  T  E  F  Y  A  T  L  F  T  R  S  L  D  D  P  F  E
1621 GCAAAAACTCTATGATTTCGCGCGAGCTGTTGGCGTCCAGCGTTTCAAAGTAGGCGGTAAGAAACGTGCGCGACAAGTCGTCGGGAAACT1710

       R  P  F  M  N  S  Y  G  F  P  D  W  L  A  L  V  L  D  A  L  T  F  L  I  L  V  I  G  V  V
1711 CGCGCGGAAACATGTTGCTGTACCCAAACGGGTCCCACAGCGCCAGCACCAAATCGGCCAGCGTGAACAAAATCAAGACTATGCCCACCA1800
                                                                      <--- p74
       S  S  A  K  I  A  I  R  T  L  A  K  A  A  T  T  L  T  K  I  A  I  R  N
1801 CGGAGCTGGCCTTGATGGCGATGCGCGTTAGGGCTTTGGCTGCCGTGGTGAGCGTCTTGATGGCAATTCGGTT   1873
```

Fig. 2. Nucleotide sequence of 1873 bp within *Pst*I I fragment containing the AnpeNPV *p26, p10* and *p74* (partial). The consensus late promoter transcription start signal (TAAG), translation start codon (ATG), translation stop codon (TAG) and putative polyadenylation signal (AATAAA) of the *p10* gene are underlined. Restriction sites of *Pst*I, *Hin*dIII and *Xho*I are double underlined. Positions of oligonucleotide primers P26DN, P10UP, P10DN and P74DN are boxed, and arrows indicate their directions. Deduced amino acid sequences are shown below (*p26* and *p10*) or above (*p74*) the corresponding nucleotide sequences, and arrows indicate their directions.

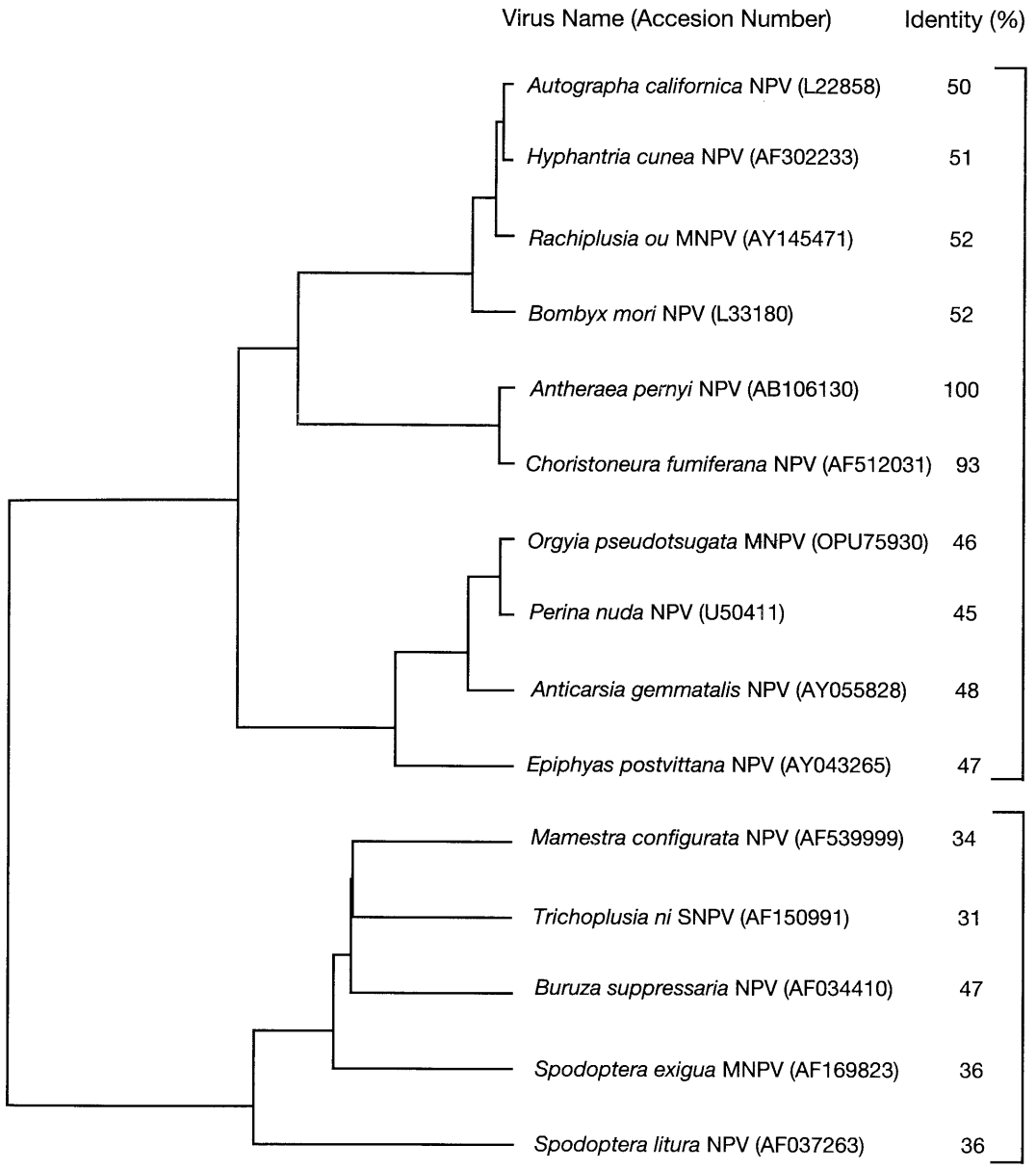


Fig. 3. A phylogenetic tree of NPVs based on the deduced amino acid sequences of *p10* gene product available in the databases. Virus name, accession number in the Genbank database and amino acid sequence identity (%) to the *p10* gene product of the AnpeNPV revealed by BLAST search are indicated. The division of NPV group I and II, is also indicated.

For the construction of *polyhedrin* locus-based transfer vector plasmid, the 0.4-kbp *Hin*dIII-*Xho*I fragment was blunt-ended and cloned into *Bam*HI site, which was also blunt-ended, of pApCH2 (6.4 kbp) constructed by inserting 3.7-kbp *Hin*dIII-*Pst*I fragment of pApCH1 (6.4 kbp) into truncated pCR2.1 with *Hin*dIII and *Pst*I digestion (Fig. 4b). The resulting pApCH3 (6.8 kbp) contains both 5' and 3' flanking regions of the *polyhedrin* gene for homologous

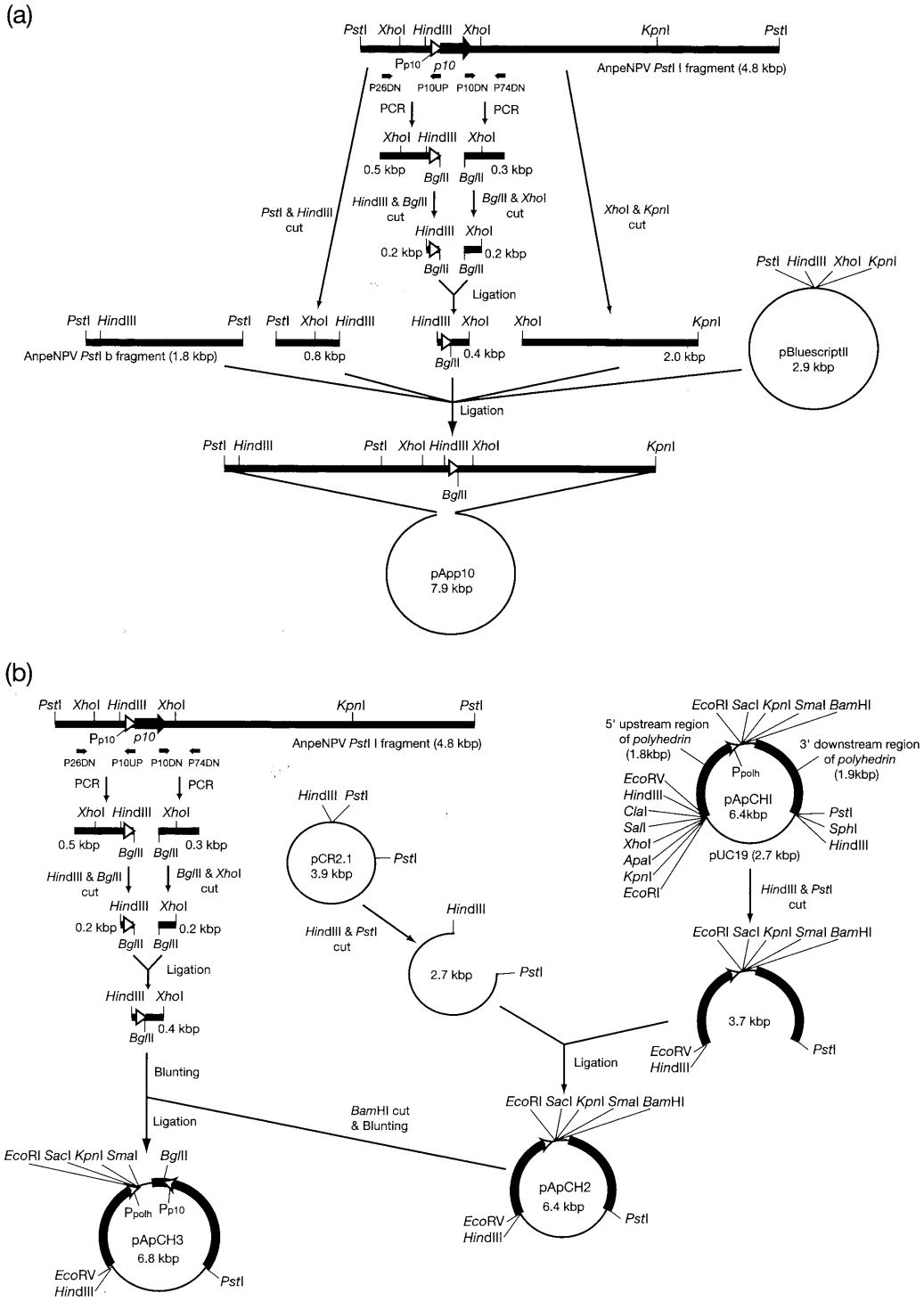


Fig. 4. Schematic diagram of procedure used for the construction of transfer vector plasmids, pApp10 (a) and pApCH3 (b). $P_{p10}$ and $P_{polh}$ indicate the AnpeNPV *p10* and *polyhedrin* promoters. The details are described in Results and Discussion.

recombination with AnpeNPV DNA at the *polyhedrin* locus and between them, in addition to 4 unique restriction sites (*EcoR*I, *Sac*I, *Kpn*I and *Sma*I) for insertion of a foreign gene just downstream of the *polyhedrin* promoter, 1 unique restriction site (*Bgl*II) for insertion of one more foreign gene just downstream of the *p10* promoter. Therefore, pApCH3 enable to construct recombinant AnpeNPVs expressing two foreign genes (dual expression) under control of the *polyhedrin* and *p10* promoters, which were aligned in tandem but opposite orientation.

We are now constructing recombinant AnpeNPVs using these transfer vectors to evaluate the protein production performance of the *p10* promoter as well as effectiveness of dual expression in both cell culture and diapausing pupae.

## Acknowledgments

This work was partly supported by a grant for Insect Factory Research Project from National Institute of Agrobiological Sciences, Japan.

## References

Ahrens, C. H., R. L. Russell, C. J. Funk, J. T. Evans, S. H. Harwood and G. F. Rohrmann, 1997. The sequence of the *Orgyia pseudotsugata multinucleocapsid* nuclear polyhedrosis virus genome. Virology, **229**: 381-399.

Gomi, S., K. Majima and S. Maeda, 1999. Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. J. Gen. Virol., **80**: 1323-1337.

Hill, J. E., J. Kuzio, J. A. Wilson, E. A. MacKinnon and P. Faulkner, 1993. Nucleotide sequence of the p74 gene of a baculovirus pathogenic to the spruce budworm, *Choristoneura fumiferana* multicapsid nuclear polyhedrosis virus. Biochim. Biophys. Acta, **1172**: 187-189.

Huang, Y. J. and J. Kobayashi, 2003. *Antheraea pernyi* nucleopolyhedrovirus mutants lacking functional cathepsin and/or chitinase genes suppress proteolysis and liquefaction of virus-infected diapausing *A. pernyi* pupae. Int. J. Wild Silkmoth & Silk, **8**: 43-50.

Huang, Y. J., J. Kobayashi and T. Yoshimura, 2002. Genome mapping and gene analysis of *Antheraea pernyi* nucleopolyhedrovirus for improvement of baculovirus expression vector system. J. Biosci. Bioeng., **93**: 183-191.

Huang, Y. J., X. Y. Wang, S. Miyajima, T. Yoshimura and J. Kobayashi, 2001. Efficiency of *Antheraea pernyi* nucleopolyhedrovirus-mediated protein production in both an established cell line and diapausing pupae of *A. pernyi*. Int. J. Wild Silkmoth & Silk, **6**: 59-71.

Kobayashi, J., R. Ando, X. Y. Wang, Y. J. Huang and S. Miyajima, 2001. Nucleotide sequence analysis of the polyhedrin gene of *Antheraea pernyi* nucleopolyhedrovirus and construction of a transfer vector plasmid. Int. J. Wild Silkmoth & Silk, **6**: 53-58.

Kuzio, J., R. Jaques and P. Faulkner, 1989. Identification of p74, a gene essential for virulence of baculovirus occlusion bodies. Virology, **173**: 759-763.

Nagaya, M., J. Kobayashi, N. Takahashi, K. Kato and T. Yoshimura, 2003. Two-dimensional high performance liquid chromatography mapping of sugar chains demonstrated the biantennary, complex *N*-glycan addition to the recombinant glycoprotein produced by baculovirus-infected *Antheraea pernyi* insect cells. J. Insect Biotechnol. Sericol., **72**: 79-86.

Nagaya, M., J. Kobayashi and T. Yoshimura, 2002. Evaluation of N-glycosylation property of cultured *Antheraea pernyi* cells in comparison with four other lepidopteran cell lines used in baculovirus expression vector systems. Int. J. Wild Silkmoth & Silk, **7**: 59-68.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, New York, NY.

Tomita, S., T. Kanaya, J. Kobayashi and S. Imanishi, 1995. Isolation of p10 gene from *Bombyx mori* nuclear polyhedrosis virus and study of its promoter activity in recombinant baculovirus vector system. Cytotechnology, **17**: 65-70.

Vlak, J. M., A. Schouten, M. Usmany, G. J. Belsham, E. C. Klinge-Roode, A. J. Maule,

J. B. F. van Lent and D. Zuidema, 1990. Expression of cauliflower mosaic virus gene I using a baculovirus vector based upon the p10 gene and a novel selection method. Virology, **179**: 312-320.

Weyer, U., S. Knight and R. D. Possee, 1990. Analysis of very late gene expression by *Autographa californica* nuclear polyhedrosis virus and the further development of multiple expression vectors. J. Gen Virol., **71**: 1525-1534.