

678982

AUSTRALIA
Patents Act 1990

P/00/008
Section 20(1)
Regulation 3.1(2)

NOTICE OF ENTITLEMENT
(To be filed before acceptance)

I, Dr. HERBERT WALKER, being authorised by COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION (CSIRO), of 407 Royal Parade, Parkville, Victoria 3052 Australia, an applicant in respect of an application for a patent for an invention entitled "Insect viruses and their uses in protecting plants" filed under Australian Application No 46912/93, state the following:-

The person(s) nominated for the grant of the patent has, for the following reasons, gained entitlement from the actual inventor(s):-

At the time the invention was made the actual inventors were officers of CSIRO and the invention was made in the course of their official duties with CSIRO; the applicant is therefore entitled to apply by virtue of Section 54(1) of the Science and Industry Research Act 1949.

The person(s) nominated for the grant of the patent is an applicant(s) of the provisional application(s) listed in the Declaration under Article 8 of the PCT.

The person(s) nominated for the grant of the patent is the depositor(s) of the deposit(s) as listed hereafter:-

Deposit list (by number, deposit institution, date)

- N92/35575, Australian Government Analytical Laboratories, 5 August 1992.

Signed: Herbert Walker

Date: 16 August 1996

Status: Authorised Officer

F.B. RICE & CO. PATENT ATTORNEYS

NOTICE OF ENTITLEMENT
(To be filed before acceptance)

I, NORMAN HENRY DAVEY, being authorised by PACIFIC SEEDS
PTY LTD of 288 Anzac Avenue, Toowoomba, Queensland, 4350, Australia, an applicant
in respect of an application for a patent for an invention entitled "Insect viruses and
their uses in protecting plants" filed under Australian Application No 48912/83, state
the following:-

The person(s) nominated for the grant of the patent has, for the following reasons,
gained entitlement from the actual inventor(s):-

By assignment.

The person(s) nominated for the grant of the patent is an applicant(s) of the provisional
application(s) listed in the Declaration under Article 8 of the PCT.

The person(s) nominated for the grant of the patent has entitlement from the
depositor(s) of the deposit(s) as listed hereafter:-

By assignment.

Deposit list (by number, deposit institution, date)

- N92/35575, Australian Government Analytical Laboratories, 5 August 1992.

Signed: Norman Davey

Date: 21/10/96

Status: DIRECTOR

F.B. RICE & CO. PATENT ATTORNEYS



AU9346912

(12) PATENT ABRIDGMENT (11) Document No. AU-B-46912/93
(19) AUSTRALIAN PATENT OFFICE (10) Acceptance No. 678982

(54) Title
INSECT VIRUSES AND THEIR USES IN PROTECTING PLANTS

International Patent Classification(s)
(51)^s C12N 007/00 A01H 005/00 A01N 063/00 C12N 015/40
C12Q 001/68
(51)^g C12Q 001/70

(21) Application No. : 46912/93 (22) Application Date : 13.08.93

(87) PCT Publication Number : WO94/04660

(30) Priority Data

(31) Number	(32) Date	(33) Country
PL4081	14.08.92	AU AUSTRALIA
089372	08.07.93	US UNITED STATES OF AMERICA

(43) Publication Date : 15.03.94

(44) Publication Date of Accepted Application : 19.06.97

(71) Applicant(s)
COMMONWEALTH SCIENTIFIC & INDUSTRIAL RESEARCH ORGANISATION; PACIFIC SEEDS PTY.
LTD.

(72) Inventor(s)
PETER DANIEL CHRISTIAN; KARL HEINRICH JULIUS GORDON; TERRY NELSON HANZLIK

(74) Attorney or Agent
F B RICE & CO , 28A Montague Street, BALMAIN NSW 2041

(57) Claim

1. An isolated small RNA virus wherein the virus is up to 40nm in size.
is not occluded, infects insect species including Heliothis species, and
comprises a genome hybridizable with the nucleotide sequence of RNA 1
(SEQ ID No: 39) or RNA 2 (SEQ ID No: 47).

INSECT VIRUSES AND THEIR USES IN PROTECTING PLANTS

FIELD OF THE INVENTION

The present invention relates to insect viruses useful in control of insect attack on plants.

- 5 It particularly relates to biological insecticides, especially those comprised of insect viruses. In particular applications, the invention also provides recombinant viruses and transgenic plants.

BACKGROUND OF THE INVENTION

- 10 There is increasing awareness of the desirability of insect pest control by biological agents. Considerable effort in recent years has been devoted to the identification and exploitation of DNA viruses with large genomes, especially the baculoviruses. These viruses generally require extensive genetic manipulation to become effective insecticides, and the first such modified viruses are only now being evaluated.

- 15 In contrast, very little effort has been devoted to the study and use of small viruses with RNA genomes.

- 20 Four main groups of small RNA viruses have been isolated from insects. These include members of the Picornaviridae, the Nodaviridae, the Tetraviridae and the unclassified viruses. Descriptions of these groups can be found in the Atlas of Invertebrate Viruses (eds J.R. Adams and J. R. Bonami) (CRC Press, Boca Raton, 1991) and Viruses of Invertebrates (ed. E. Kurstak) (Marcel Dekker, New York, 1991). These disclosures relating to these viruses concern their pathology and biology, not their use in biological control.
- 25

SUMMARY OF THE INVENTION

- In a first aspect of the present invention there is provided an isolated small RNA virus capable of infecting insect species including *Heliothis* species, wherein the virus is up to
- 30 40nm in size and is not occluded.



- 2 -

In one particular embodiment, the present invention provides an isolated preparation of *Heliothis armigera* stunt virus referred to as "HaSV" herein.

In a further aspect of the present invention there is provided an isolated
5 nucleic acid molecule comprising a nucleic acid sequence hybridizable with RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47) described herein under low stringency conditions.

In still a further aspect the invention provides a vector comprising a nucleic
10 acid molecule, the sequence of which is hybridizable with RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47) as described herein. These vectors include expression and transfer vectors for use in animals including insect, plant and bacterial cells.

~~15 In a further aspect the invention provides an isolated protein or polypeptide preparation of the proteins or polypeptides derivable from the isolated virus of the present invention. The invention also extends to antibodies specific for the protein and polypeptide preparations.~~

~~20 In a yet further aspect the invention provides a recombinant insect virus vector incorporating all or a part of the isolated virus of the present invention.~~

~~In a still further aspect of the present invention there is provided a method of controlling insect attack in a plant comprising genetically manipulating said
25 plant so that it is capable of expressing HaSV or mutants, derivatives or variants thereof or an insecticidally effective portion of HaSV, mutants, variants or derivatives thereof and optionally other insecticidally effective agents such that insects feeding on the plants are deleteriously effected.~~

~~30 In another aspect of the present invention there is provided a preparation of HaSV or a mutant variant or derivative thereof, or an insecticidally effective portion of HaSV, mutant, variant or derivative thereof, suitable for application~~



In a further aspect the invention provides an isolated nucleic acid molecule of 15 or more nucleotides in length, said molecule comprising a nucleic acid sequence hybridisable with RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47) under low stringency conditions, preferably medium stringency conditions, most preferably high stringency conditions.

In a still further aspect the invention provides an isolated protein or polypeptide preparation of the proteins or polypeptides derivable from the isolated virus of the present invention. The invention also extends to antibodies specific for the protein and polypeptide preparations.

In a yet further aspect the invention provides a recombinant insect virus vector incorporating all or a part of the isolated virus of the present invention.

In a still further aspect of the present invention there is provided a method of controlling insect attack in a plant comprising genetically manipulating said plant so that it is capable of expressing HaSV or mutants, derivatives or variants thereof or an insecticidally effective portion of HaSV, mutants, variants thereof or derivatives thereof and optionally other insecticidally effective agents such that insects feeding on the plants are deleteriously effected.

In another aspect of the present invention there is provided a preparation of HaSV or a mutant variant or derivative thereof, or an insecticidally effective portion of HaSV, mutant, variant or derivative thereof, suitable for application



to plants, wherein the preparation is capable of imparting an insect protective effect.

In yet another aspect, the present invention provides a capsovector for use in controlling insect pests, comprising a capsid protein from a insect small RNA virus of up to 40nm

5 in size which is not occluded and infects insect species including *Heliothis* species, said capsid protein encapsidating;

(i) an insecticidal protein toxin, or

(ii) a nucleic acid molecule which is insecticidal or which encodes an insecticidal protein toxin,

10 such that the protein toxin/nucleic acid molecule is protected from inactivation in the gut of an insect following ingestion of the capsovector by the insect.



BRIEF DESCRIPTION OF FIGURES

Figure 1 is the complete sequence of RNA 1 (SEQ ID No: 39) and of major encoded polypeptide.

- 5 Figure 2 is the complete sequence of RNA 2 (SEQ ID No: 47) in the authentic version, and its encoded polypeptides (the RNA 2 variant (SEQ ID No: 51) called the "5C version" is also shown around nucleotide position 570 [the amino acid sequence encoded by the 5C version is not included in this Figure but this may be deduced from the nucleotide sequence given]).

- 10 Figure 3 is bioassay data showing HaSV-induced stunting of larvae.

Figure 4 is a schematic representation of the proteins encoded by RNA 1 (SEQ ID No: 39) and RNA 2 (SEQ ID No: 47).

Figure 5 is a schematic representation of the proteins expressed by RNA 2 (SEQ ID No: 47) in bacteria DNA fragments encoding P17 (SEQ ID No: 48), P71 (SEQ ID No: 50),

- 15 P64, P7 and the fusion protein P70 (SEQ ID No: 52) were synthesised by PCR. The flanking NdeI and BamHI sites used in cloning are indicated. (Note that P17 is followed by a BglII site, whose cohesive ends are compatible with those of BamHI).

Figure 6 illustrates the 3'-terminal secondary structure of HaSV RNAs. The tRNA-like structures at the 3' ends of RNAs 1 and 2 (SEQ ID Nos: 39 and 47) are shown.

- 20 Residues in bold are common to both sequences.

Figure 7 Expression strategies for HaSV cDNAs in insect cells. The upper part of the figure shows the genome organization of RNAs 1 and 2 (SEQ ID Nos: 39 and 47). The lower part shows insertion of cDNAs corresponding to these RNAs into a plasmid vector, between the heat shock protein (HSP70) promoter of *Drosophila* and a suitable polyadenylation (pA) signal. The HSP promoter was obtained by PCR using suitable primers, with a BamHI site inserted by PCR immediately upstream of the start of transcription, giving the following sequence:



- 4 -

GGATCCACAGnnn (SEQ ID No: 1), where the underlined residue is the transcription start site for either RNA. The cDNAs are terminated by *Cl*I sites, allowing direct linkage to ribozyme sequences as described in the text.

Figure 8 Ribozymes to yield correct 3' ends. The sequences of the ribozymes

5 inserted as short cDNA fragments into HaSV cDNA clones are shown. The ribozyme fragments were assembled and cloned as described in the text.

Designed self-cleavage points are indicated by bold arrows.

Figure 9 Immunoblots to map epitopes on HaSV. A. Detected with HaSV antiserum. Lane 1: pTP70delSP; lane 2: pTP70; lane 3: pTP17; lane 4: control;

10 lane 5: pTP70delN; lane 6: pTP70; lane 7: pTP71; lane 8: HaSV virions;

lane 9: molecular weight markers. B. Detected with HaSV antiserum. Lane 1:

pTP70delN; lane 2: pTP70delSPN; lane 3: pTP70. C. Detected with an

antiserum to the Bt toxin (CryIA(c)). lane 1: pTP70; lane 2: HaSV virions;

lane 3: control extract.

15 Figure 10 New field isolates of HaSV. The genomic organization of RNA 2 (SEQ ID No: 47) is shown at the top of the Figure. PCR using appropriate

primers with *Bam*HI restriction sites and in some cases altered context

sequences of the AUG initiating translation of the P17 (SEQ ID No: 48) or

P71 (SEQ ID No: 50) genes were used to make fragments for cloning into the

20 *Bam*HI sites of the expression vectors. Constructs 17E71 (SEQ ID No: 35)

and P71 (SEQ ID No: 50) have altered context sequences of the AUG

initiating translation of the P17 (SEQ ID No: 48) and P71 (SEQ ID No: 50)

genes respectively; these alterations correspond to the context derived from the

JHE gene (see text). All context sequences are given on the right of the

25 figure. R2 is a clone of the complete RNA sequence as a *Bam*HI fragment in the vector.

Figure 11 Maps of the expression constructs in baculovirus vectors.

Figure 12 a to e Various strategies utilising the present invention.

Figure 13 Expression of RNAs 1 and 2 (SEQ ID Nos: 39 and 47) from

30 baculovirus vectors. The full-length cDNA clone of HaSV RNA 1 or 2 (SEQ ID Nos: 39 and 47) was inserted as a *Bam*HI fragment into the

baculoexpression vectors. PCR was used to add *Bam*HI sites immediately



- 5 -

adjacent to the 5' and 3' termini of the RNA 1 (SEQ ID No: 39) sequence; sequences of the primers are given in the text. Constructs R1RZ and R2RZ carry cis-acting ribozymes immediately adjacent to the 3' end of the sequence of RNA 1 and 2 (SEQ ID Nos: 39 and 47) respectively.

- 5 Figure 14 Expression strategies for HaSV cDNAs in plant cells. The upper part of the Figure shows the genome organization of RNAs 1 and 2 (SEQ ID Nos: 39 and 47). The lower part shows insertion of cDNAs corresponding to these RNAs into a plasmid vector, between 35S promoter of cauliflower mosaic virus and the polyadenylation (pA) signal on plasmid pDH51 (Pietrzak et al, 1986). The cDNAs were obtained by PCR using suitable primers, with a BamHI site inserted by PCR immediately upstream of the start of each cDNA. The cDNAs are terminated by ClaI sites, allowing direct linkage to ribozyme sequences as described in the text.

15 DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

- A first aspect of the invention contemplates use of small RNA viruses for biological control of insects. In particular, in accordance with the first aspect of this invention there is provided an isolated small RNA virus, particularly *H. armigera* stunt virus or mutants, variants or derivatives thereof capable of
- 20 infecting insects, in particular the insect species such as *Helicoverpa armigera*. The small RNA virus isolate of the instant invention is insecticidal and in particular stunts the growth of insect larvae, for example *Helicoverpa armigera* larvae and inhibits or prevents development into the adult stage.
- 25 The small RNA viruses of the instant invention have insecticidal, anti-feeding, gut-binding or any synergistic property or other activity useful for insect control.

- In particular, *Helicoverpa armigera* stunt virus (HaSV) particles are isometric
- 30 and approximately 36 nm in diameter with a buoyant density on CsCl gradients of 1.36g/ml. The virus is composed of two major capsid proteins of approximately 64 and 7 KDa in size as determined on SDS-PAGE. The HaSV



- 6 -

genome is much later than the largest known nodavirus (another class of RNA viruses) and comprises two ss (+) RNA molecules of approximately 5.3 and 2.4 kb. The genome appears to lack a blockage of unknown structure at the 3' termini that is found in Nodaviridae. The HaSV genome however shares a capped structure and non-polyadenylation with Nodaviridae. HaSV differs significantly from Nodaviridae and Nudaurelia ω virus in terms of its immunological properties. In particular the large capsid protein has different antigenic determinants. Other properties of HaSV are described in the Examples.

10

The host range of HaSV includes Lepidopterans such as from the subfamily Heliothinae. Species known to be hosts are *Helicoverpa (Heliothis) armigera*, *H. punctigera*, *H. zea*, *Heliothis virescens* and other such noctuides as *Spodoptera exigua*. *H. armigera* which is known by the common names corn ear worm, cotton ball worm, tomato grub and tobacco bud worm is a pest of economic significance in most countries. *H. punctigera*, the native bud worm, is a pest of the great economic significance in Australia. Members of the Heliothinae, which include *Helicoverpa* and *Heliothis*, and especially *H. armigera* are among the most important and widespread pests in the world. In the US *Heliothis virescens* and *Helicoverpa zea* are particularly important pests.

20

The first aspect of the invention provides an isolated small RNA virus capable of infecting insects including *Heliothis* species. In a particularly preferred form the invention relates to mutants, variants and derivatives of HaSV. The terms "mutant", "variant" and "derivative" include all naturally occurring and artificially created viruses or viral components which differ from the HaSV isolate as herein described in nucleotide content or sequence, amino acid content or sequence, immunological reactivity, non-glycosylation or glycosylation pattern and/or infectivity but generally retain insecticidal activity. Specifically the terms "mutant", "variant" and "derivative" of HaSV covers small RNA viruses which have one or more functional characteristic of HaSV described herein. Examples of mutants, variants or derivatives of HaSV include small RNA

25

30



- 7 -

viruses that have different nucleic or amino acid sequences from HaSV but retain one of more functional features of HaSV. These may include strains with genetically silent substitutions, strains carrying replication and encapsidation sequences and signals that are functionally related to HaSV, or
5 strains that carry functionally related protein domains.

In a preferred aspect the invention relates to mutants, variants or derivatives of HaSV which encode replication or encapsidation sequences, structures or signals with 60%, preferably 70%, more preferably 80%, still more preferably
10 90% and even more preferably 95% nucleotide sequence identity to the nucleotide sequences HaSV.

In another preferred aspect the invention relates to mutants, variants or derivatives of HaSV which encode proteins with at least 50%, preferably 60%,
15 preferably 70%, more preferably 80%, still more preferably 90% and even more preferably 95% amino acid sequence identity to proteins or polypeptides of HaSV.

In another preferred aspect the invention relates to mutants, variants or
20 derivatives of HaSV with 50%, more preferably 60%, still more preferably 70%, more preferably 80%, still more preferably 90 or 95% nucleotide sequence identity to the following biologically active domains encoded by the HaSV genome:

RNA 1 (SEQ ID No: 39) - amino acid residues 401 to 600 or the other
25 domains in the replicase

RNA 2 (SEQ ID No: 47) (in the capsid protein)

- amino acid residues 273 to 435
- amino acid residues 50 to 272
- amino acid residues 436 to the COOH terminus

30

Preferably the viral isolate of the present invention is biologically pure which means a preparation of the virus comprising at least 20% relative to other



- 8 -

components as determined by weight, viral activity or any other convenient means. More preferably the isolates are 50% pure, still more preferably it is 60%, even more preferably it is 70% pure, still more preferably it is 80% pure and even more preferably it is 90% or more, pure.

5

In a second aspect the present invention relates to a nucleotide sequence or sequences hybridizable with those of HaSV. The term nucleotide sequence used herein includes RNA, DNA, cDNA and nucleotide sequences complementary thereto. Such nucleotide sequences also include single or
10 double stranded nucleic acid molecules and linear and covalently closed circular molecules. The nucleic acid sequences may be the same as the HaSV sequences as herein described or may contain single or multiple nucleotide substitutions and/or deletions and/or additions thereto. The term nucleotide sequence also includes sequences with sufficient homology to hybridize with
15 the nucleotide sequence under low, preferably medium and most preferably high stringency conditions (Sambrook J, Fritsch, E.F. & Maniatis T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbour Laboratories Press) and to nucleotide sequences encoding functionally equivalent sequences. In still a more preferred embodiment the invention
20 comprises the nucleotide sequences of genome components 1 and 2 (SEQ ID Nos: 39 and 47) as represented by Figures 1 and 2 hereinafter or parts thereof, or mutants, variants, or derivatives thereof. The terms "mutants", "variants" or "derivatives" of nucleotide genome components 1 and 2 (SEQ ID Nos: 39 and 47) has the same meaning, when applied to nucleotide sequences as that given
25 above and includes parts of genome components 1 and 2 (SEQ ID Nos: 39 and 47).

The second aspect of the invention also relates to nucleotide signals, sequences or structures which enable the nucleic acid on which they are present to be
30 replicated by HaSV replicase. Furthermore the invention relates to the nucleotide signals, sequences or structures which enable nucleic acids on which they are present to be encapsidated.

AMENDED SHEET
IPE/AU

- 9 -

In a particularly preferred embodiment of the second aspect, the invention comprises nucleotide sequences which are mutants of the capsid gene having the following sequences:

- 5 ATG GGC GAT GCC GGC GTC GCGT TCA CAG (SEQ ID No: 2)
 ATG GAG GAT GCT GGA GTG GCG TCA CAG (SEQ ID No: 3)
 ATG AGC GAG GCC GGC GTC GCG TCA CAG (SEQ ID No: 4)

10 In a preferred aspect the invention relates to nucleotide sequences of HaSV encoding insecticidal activity including the capsid protein gene and P17 (SEQ ID No: 48) and mutants, variants and derivatives thereof.

In another preferred aspect the invention comprises nucleotide sequences including the following ribozyme oligonucleotides:

- 15 5'CCATCGATGCCGGACTGGTATCCCAGGGGG (called "HVR1Cla" herein) (SEQ ID No: 5)

5' CCATCGATGCCGGACTGGTATCCCGAGGGAC (called "5'HVR2Cla" herein) (SEQ ID No: 6)

- 20 5' CCATCGATGATCCAGCCTCCTCGCGGCGCCGGATGGGCA (called "RZHDV1" herein) (SEQ ID No: 7)

25 5' GCTCTAGATCCATTCGCCATCCGAAGATGCCCATCCGGC (called "RZHDV2" herein) (SEQ ID No: 8)

5' CCATCGATTTATGCCGAGAAGGTAACCAGAGAAACACAC (called "RZHC1" herein) (SEQ ID No: 9)

- 30 5' GCTCTAGACCAGGTAATATACCACAACGTGTGTTTCTCT (called "RZHC2" herein) (SEQ ID No: 10)



- 10 -

Ribozyme sequences are useful for obtaining translation, replication and encapsidation of the transcript. It is therefore desirable to cleave the transcript downstream of its t-RNA-like structure or poly A tail prior to translation, replication or encapsidation of the transcript.

5

The present invention also further extends to oligonucleotide primers for the above sequences, antisense sequences and nucleotide probes for the above sequences and homologues and analogues of said primers, antisense sequences and probes. Such primers and probes are useful in the identification, isolation and/or cloning of genes encoding insecticidally effective proteins or proteins required for viral activity, from HaSV or another virus (whether related or unrelated) carrying a similar gene or similar RNA sequence. They are also useful in screening for HaSV or other viruses in the field or in identifying HaSV or other viruses in insects, especially in order to identify related viruses capable of causing pathogenicity similar to HaSV.

Any pair of oligonucleotide primers derived from either RNA 1 or RNA 2 (SEQ ID Nos: 39 and 47) and located between ca 300 and 1500 bp apart can be used as primers. The following pairs of primer sequences exemplify particularly preferred embodiments of the present invention: Specifically for RNA 1 (SEQ ID No: 39):

1. HVR1B5' (SEQ ID No: 38) (described below) and the primer complementary to nucleotides 1192-1212 of Figure 1.
2. The primer corresponding to nucleotides 4084 and 4100 of Fig. 13 and the primer HVRI3p (SEQ ID No: 12) described below

Specifically for RNA 2 (SEQ ID No: 47):

1. The primer corresponding to nucleotides 459 to 476 of Fig. 2 and the primer complementary to nucleotides 1653 to 1669 of Fig. 2 (this would include the central variable domain)
2. R2cdha5 and the primer complementary to nucleotides 1156 to 1172 of Fig. 2



- 11 -

3. The primer corresponding to nucleotides 1178 to 1194 and the primer complementary to nucleotides 2072 to 2091 (of Fig. 2).

Other combinations giving shorter fragments are also possible.

5 Further preferred primers include:

5' GGGGGGAATTTCATTTAGGTG[^]CACTATAGTTCTGCCTCCCCGGAC
(called "HvR1SP5p" herein) (SEQ ID No: 11)

10 5' GGGGGGATCCTG[^]GTATCCCAGGGGGGC (called "HvR13p" herein)
(SEQ ID No: 12)

5' CCGGAAGCTTGTTTTCTTTCTTTACCA (called "Hr2cdna5" herein)
(SEQ ID No: 13)

15

5' GGGGGATCCGATGGTATCCCGAGGGACGC
TCAGCAGGTGGCATAGG (called "HvR23p" herein) (SEQ ID No: 14)

AAATAATTTTGTTACTTTAGAAGGAGA[^]ATACATATGAGCGAGCGA
20 GCACAC (called "HVPET65N" herein) (SEQ ID No: 15)

AAATAATTTTGTTTAACCTTAAGAAGGAGATCTACATATGCTGGAGT
GGCGTCAC (called "HVPET63N" herein) (SEQ ID No: 16)

25 GGAGATCTACATATGGGAGATGCTGGAGTG (called "HVPET64N"
herein) (SEQ ID No: 17)

GTAGCGAACGTCGAGAA (called "HVRNA2F3" herein) (SEQ ID No: 18)

30 GGGGGATCCTCAGTTGTCAGTGGCGGGGTAG (called "HVP65C"
herein) (SEQ ID No: 19)



- 12 -

GGGGATCCCTAATTGGCACGAGCGGCGC (called "HVP6C2" herein)
 (SEQ ID No: 20)

AATTACATATGGCGGCCGCGTTTCTGCC (called "HVP6MA" herein)
 5 (SEQ ID No: 21)

AATTACATATGTTCGCGGCCGCGTTTCT (called "HVP6MF" herein)
 (SEQ ID No: 22)

10 The invention also relates to vectors encoding the nucleotide sequence described above and to host cells including the same. Preferably these vectors are capable of expression in animal, plant or bacterial cell or are capable of transferring the sequences of the present invention to the genome of other organisms such as plants. More preferably they are capable of expression in
 15 insect and crop plant cells.

In a preferred aspect the invention relates to the vectors pDHVR1, pDHVR1RZ, pDHVR2, pDHVR2RZ, p17V71, p17E71, pPH, pV71, p17V64, p17E64, pP64, pV64, pBacHVR1, pBacHVR1RZ, pBacHUR2, pBacHVR2RZ,
 20 pHSPR1, pHSPR1RZ, pHSPR2, pHSPR2RZ, pSR1(E3)A, pSR1(E3)B, pSR2A, pSR2B, pSX2P70, pSXR2P70, pSRP2B, pBHVR1B, pBHVR2B, pT7T2P64, pSR2P70, pT7T2P65, pT7T2P70, pT7T2-P71, pBSKSE3, pBSR15, pBSR25p, pSR25, pHR236P70, pHR235P65, pGemp63N, pGemp64N, pGemp65N, pP64N, pP65H, pTP6MA, pTP6MF, pTP17, pTP17delBB, pP656
 25 or p70G as described hereinafter.

In a third aspect the invention relates to polypeptides or proteins encoded by HaSV and to homologues and analogues thereof. This aspect of the invention also relates to derivatives and variants of the polypeptides and proteins of
 30 HaSV. Such derivatives and variants include substitutions and/or deletions of one or more amino acids, and amino and carboxy terminal fusions with other polypeptides or proteins. In a preferred aspect the invention relates to the



- 13 -

proteins P7, P16, P17 (SEQ ID No: 48), P64, P70 (SEQ ID No: 52), P71 (SEQ ID No: 50), P11a (SEQ ID No: 42), P11b (SEQ ID No: 44), P14 (SEQ ID No: 46) and P187 (SEQ ID No: 40) described herein and to homologues and analogues thereof, including fusion proteins particularly of P71 (SEQ ID No: 50) such as P70 (SEQ ID No: 52) described herein. In a most preferred aspect the invention relates to polypeptides or proteins from HaSV which have insecticidal activity themselves or provide target specificity for insecticidal agents. In particular the invention relates to polypeptides or fragments thereof with insect gut binding specificity, particularly to the variable domains thereof as herein described. In addition, homologues and analogues with said insecticidal activity of the polypeptides and proteins are also included within the scope of the invention. In addition the invention also relates to antibodies (such as monoclonal or polyclonal antibodies or chimeric antibodies including phage antibodies produced in bacteria) specific for said polypeptide and protein sequences. Such antibodies are useful in detecting HaSV and related viruses or the protein products thereof.

In a fourth aspect the invention provides an infectious, recombinant insect virus including a vector, an expressible nucleic acid sequence comprising all of, or a portion of the HaSV genome, including an insecticidally effective portion of the genome and optionally, material derived from another insect virus species or isolate(s).

Insect virus vectors suitable for the invention according to this aspect, include baculoviruses, entomopoxviruses and cytoplasmic polyhedrosis viruses. Most preferably, the insect virus vector is selected from the group comprising the baculovirus genera of nuclear polyhedrosis viruses (NPV's) and granulosis viruses (GV's). In this aspect of the invention the vector acts as a carrier for the HaSV genes encoding insecticidal activity. The recombinant insect virus vector may be grown by either established procedures Shieh, (1989), Vlcek (in press) or any other suitable procedure and the virus disseminated as needed.



- 14 -

The insect virus vectors may be those described in copending International application No. PCT/AU92/00413.

- The nucleic acid sequence or sequences incorporated into the recombinant vector may be a cDNA, DNA or RNA sequence and may comprise the genome or portion thereof of a DNA or RNA of HaSV or another species. The term "material derived from another insect virus species or isolate" includes any nucleic acid sequence, or protein sequence or parts thereof which are useful in exerting an insecticidal effect when incorporated in the recombinant vector of the invention. Suitable nucleic acid sequences for incorporation into the recombinant vector include insecticidally effective agents such as a neurotoxin from the mite *Pyemotes tritici* (Tomalski, M.D. & Miller, L.K. Nature 352, 82-85 (1991) a toxin component of the venom of the North African scorpion *Androctonus australis* Maeda, S. et al. Virology 184-777-780 (1991) Stewart, L.M.D. et al., Nature 352, 85-88 (1991), Conotoxins from the venom of *Conus* spp. (Olivera B.M. et al., Science 249, 257-263 (1990); Woodward S.R. et al., EMBO J. 9, 1015-1020 (1990); Olivera B.M. et al., Eur. J. Biochem. 202, 589-595 (1991).
- The exogenous nucleic acid sequence may be operably placed into the insect virus vector between a viral or cellular promoter and a polyadenylation signal. Upon infection of an insect cell, the vector virus will cause the production of either infectious virus genomic RNA or infectious encapsidated viral particles.
- The promoters may be constitutively expressed or inducible. These include tissue specific promoters, temperature sensitive promoters or promoters which are activated when the insect feeds on a metabolite in the plant that it is desired to protect.
- Recombinant insect virus vectors according to the present invention may include nucleic acid sequences comprising all or an infectious or insecticidally



- 15 -

effective portion of genome the HaSV and optionally another insect virus species or isolate.

In a particularly preferred embodiment of the present invention there is
5 provided assembled capsids comprising one or more of the capsid proteins of
the present invention, or derivatives or variants thereof as contemplated or
described herein. These assembled virus capsids are useful as vectors for
insecticidal agents. As such the assembled viral capsids may be used to
administer insecticidal agents such as various nucleotide sequences with
10 insecticidal activity or various toxins to an insect. Nucleotide sequences in the
form of RNA or DNA which can be used include those of the HaSV genome
or other insect viruses. Toxins which can be used advantageously include those
which are active intracellularly and may also include neurotoxins with an
appropriate transportation mechanism to reach the insect neurones.

15

The efficacy or insecticidal activity of infectious genomic RNA or viral
particles produced by insect cells infected with insect vectors according to this
aspect of the invention, may be enhanced as described below. Moreover the
virus vector itself may include within a non-essential region(s), one or more
20 nucleic acid sequences encoding substances that are deleterious to insects such
as the insecticidally effective agents described above. Alternatively an extra
genome component may be added to the HaSV genome either by insertion
into one of the HaSV genes or by adding it to the ends of the genome.

25 In a particularly preferred embodiment there is provided a recombinant
baculovirus vector comprising HaSV or part thereof having insecticidal
properties.

Other modifications which may be made to the infectious recombinant insect
30 virus according to the fourth aspect include:



- 16 -

- i) splitting the exogenous HaSV nucleic acid molecules comprising the genome and cloning the fragments into the insect vector so that they cannot rejoin. One component, preferably the virus RNA replicase, could be expressed from a separately-transcribed fragment, the transcripts of which would not be replicated by the replicase they encode. The remainder of the genome (having insecticidal activity or encoding the capsid protein or a separate toxin m-RNA) could be encoded by, for example, a second separately-transcribed fragment, the transcripts of which are capable of being amplified by the replicase. Consequently, whilst the transcripts from the second or other fragment would effect their insecticidal activity upon the infected insect cell, they would not be able to infect another insect cell, (even if encapsidated) - because the replicase or replicase-encoding transcripts would be absent;
- This modification would allow an inherent biological containment to be built into the insecticidal vectors, which, when used in conjunction with the use of non-persistent DNA virus vectors such as those described in the above mentioned copending application, would allow a new level of environmental safety greatly extending earlier approaches based on baculovirus vectors.
- ii) Manipulation of encapsidation signals or sequences essential for replicase binding or production of sub-genomic mRNA's including expression of exogenous insect control factors as RNAs dependent on the virus for replication. This involves determination of RNA sequences and signals important for replication and encapsidation of virus RNAs, such as by analysis of replication of deletion mutants carrying reporter genes in appropriate cells, followed by studies on the transmission of the reporter gene to larvae by feeding of virus. These deletion mutants can be used to carry genes for insect control factors/toxins to larvae after replacing the reporter gene by a suitable toxin gene such as shown in Fig. 12;



- 17 -

- iii) using an insect promoter responsive to virus infection and, for example, placing copies of the viral replicase gene under the control of two promoters, one which is constitutive or expressed at early stages of vector infection, and the other being a cellular promoter turned on by the ensuing RNA viral infection. This system would then make more copies of the replicase mRNA available as the amount of its template increased. Such a promoter may be isolated using techniques analogous to enhancer trapping, that is, transforming insect cells with a suitable reporter gene and looking for induction of the reporter upon virus infection of a population of transformed cells.

In a fifth aspect the invention relates to a method of controlling insect attack in plants by genetically manipulating plants to express HaSV or parts thereof which can confer insecticidal activity optionally in combination with other insecticidally effective agents. Such plants are referred to as transgenic plants.

The term "express" should be understood as referring to the process of transcribing the genome or portion thereof into RNA or, alternatively, the process of transcribing the genome or portion thereof into RNA and then, in turn, translating the RNA into a protein or peptide.

In a sixth aspect the invention relates to the transgenic plants per se as described above. Transgenic plants according to the invention may be prepared for example by introducing a DNA construct including a cDNA or DNA fragment encoding all or a desired infectious portion of HaSV, into the genome of a plant. The cDNA or DNA fragment may, preferably, be operably placed between a plant promoter and a polyadenylation signal. Promoters may cause constitutive or inducible expression of the sequences under their control. Furthermore they may be specific to certain tissues, such as the leaves of a plant where insect attack occurs but not to other parts of the plant such as that used for food. The inducible promoters may be induced by stimuli such as disturbance of wind or insect movement on the plant's tissues, or may be



- 18 -

specifically turned on by insect damage to plant tissues. Heat may also be a stimulus for promoter induction such as in spring where temperatures increase and likelihood of insect attack also increases. Other stimuli such as spraying by a chemical (for instance a harmless chemical) may induce the promoter.

5

The cDNA or DNA fragment may encode all or a desired infectious portion of the wild-type, recombinant or otherwise mutated HaSV. For example, deletion mutants could be used which lack segments of the viral genome which are non-essential for replication or perhaps pathogenicity.

10

The nucleotide sequences of the invention can be inserted into a plant genome by already established techniques, for example by an *Agrobacterium* transfer system or by electroporation.

15 Plants which may be used in this aspect of the invention include plants of both economic and scientific interest. Such plants may be those in general which need protection against the insect pests discussed herein and in particular include tomato, potato, corn, cotton, field pea and tobacco.

20 To enhance the efficacy of infectious genomic RNA or viral particles expressed by transgenic plants according to the invention, the DNA construct introduced into the plants' genome may be engineered to include one or more exogenous nucleic acid sequences encoding substances that are deleterious to insects. Such substances include, for example, *Bacillus thuringiensis* δ -toxin, insect
25 neurohormones, insecticidal compounds from wasp or scorpion venom or of heterologous origin, or factors designed to attack and kill infected cells in such a way so as to cause pathogenesis in the infected tissue (for example, a ribozyme targeted against an essential cellular function).

30 DNA constructs may also be provided which include:



- 19 -

- i) mechanisms for regulating pathogen expression (for example, mechanisms which restrict the expression of ribozymes to the insect cells) by tying for example, expression to abundant virus replication, production of minus-strand RNA or sub-genomic mRNA's; and/or
- 5 ii) mechanisms similar to, or analogous to, those described in copending International patent application number PCT/AU92/00413 so as to achieve a limited-spread system (such as control of replication).
- 10 Transgenic plants according to the present invention may also be capable of expressing all or an infectious or insecticidal portion of genomes from HaSV and one or more species or isolates of insect viruses.
- In a seventh aspect of the invention HaSV, or insecticidally effective parts
- 15 thereof, or the infectious recombinant virus vectors of the fourth aspect of the present invention may be applied directly to the plant to control insect attack. HaSV or the recombinant virus vectors may be produced either in whole or in part in either whole insects or in culture cells of insects or in bacteria or in yeast or in some other expression system. HaSV or the recombinant virus
- 20 forms may be applied in a crude form, semi purified or purified form optionally in admixture with agriculturally acceptable carrier to the crop, in need of protection. HaSV may also be applied as a facilitator of infection where existing insect populations already infected with another agent, such as one or more other viruses whereby HaSV is able to act synergistically to bring
- 25 about an insecticidal effect. Alternatively HaSV and another agent such as one or more viruses may be applied together to plants to control insects feeding thereon.

A deposit of HaSV No. 18.4 was made on August 5th 1992 at the Australian

30 Government Analytical Laboratories. The deposit was given accession No. N92/35575.



AMENDED SHEET

- 20 -

EXAMPLE 1

TAXANOMIC, PHYSIOCHEMICAL
AND BIOCHEMICAL CHARACTERISATION
OF AN INSECT VIRUS: HaSV

5 Materials and Methods

- A Animals and virus production. H. Armigara larvae were raised as described in Teakle R.E. and Jensen J.M. (1985) *Heliothis punctiger* in Singh P and Moore R.F. (eds) Handbook of Insect Rearing Vol 2., Elsevier, Amsterdam pp 313-322. Larvae were infected for virus production by feeding five day old larvae on 10mg pieces of diet to which 0.064 OD₂₆₀ units of HaSV had been applied. After 24 hours the larvae were then transferred to covered 12-well plates (BioScientific, Sydney, Australia) that contained sufficient diet and grown for eight days after which they were collected and frozen at -80 °C until further processed. Frozen larvae were weighed to 100g, placed into 200ml of 50mM Tris buffer (pH 7.4), homogenized, and filtered through four layers of muslin. This homogenate was centrifuged in a Sorvall SS-34 rotor at 10,000 x g for 30 minutes whereupon the supernatant was transferred to fresh tubes and recentrifuged in Beckman SW-28 rotor at 100K xg for 3 hours. The resultant band was collected and repelleted in 50 mM pH 7.2 Tris buffer in a Beckman SW-28 tube by centrifugation at 100K xg for 3 hours. The pelleted virus was resuspended overnight in 1ml of buffer at 4 °C then layered onto a discontinuous CsCl gradient containing equal volumes of 60% and 30% CsCl (w/v) in a Beckman SW-41 tube and centrifuged at 12 h at 200 xg. The resultant pellet was suspended in 100µl of buffer and frozen for further use.
- B Particle characterization. Staining with acridine orange was as described in Mayor H.D. and Hill N.O. (1961) Virology 14: p264. Buoyant density was estimated in CsCl gradients according to Scotti P.D., Longworth J.F., Plus N, Crozier G. and Reiganum C. (1981) Advances in Virus Research 26: 117-143.



- 21 -

- C Immunological procedure. Rabbit anti-sera to HaSV was produced by standard immunological procedures. Rabbit antisera to the Nudaurelia virus in addition to the virus itself was provided by Don Hendry (Rhodes University, Grahamstown, South Africa). Rabbit antisera to the Nudaurelia b virus was supplied by the late Carl Reinganum (Plant Research Institute, Burnley, Vic, Australia). The immunological relationship to the Nudaurelia ω virus was determined by the standard reciprocal double diffusion technique. Immunoblotting was performed according to Towbin H., Staehelin T. and Gordon J. (1979) PNAS.
- 5 Antibodies monospecific for the major 65 kDa capsid protein were prepared by incubating polyclonal antisera with sections of nitrocellulose blotted with the 65 kDa protein. After extensive washing in Tris buffered saline, the bound antibodies were eluted in 50mM citric buffer, pH 8.0 after a 5 minute incubation.
- 15
- D Protein characterization. Polyacrylamide gel electrophoresis in the presence of SDS followed the procedure of Laemmli UK 1970 Nature 227: 680-685 and was done with 12.5% gels unless otherwise noted with low and high molecular weight standards from BioRad. Staining was done with a colloidal preparation of Coomassie Blue G-250 (Gradipore Ltd, Pyrmont, New South Wales, Australia). Determination of the M_r of the smallest protein was done with a 16% gel and standards of 3.4 kDa, 12.5 kDa and 21.5 kDa (Boehringer Mannheim). Glycosylation of the viral proteins was determined by a general glycan staining procedure with reagents supplied by Boehringer Mannheim; the positive control was fetuin. N-termini of proteins were sequenced using procedures described by Matsudaira (1989) Purification of Proteins and Peptides by SDS-PAGE in A Practical Guide to Protein and Peptide Purification for Microsequencing ed Matsudaira P.T. Academic Press, San Diego pp
- 20
- 25
- 30 52-72 on an Applied Biosystems 477A gas phase sequencer.



- 22 -

E Nucleic acid characterization. RNA was removed from capsids by twice vortexing a virus suspension with equal volumes of neutralized phenol then with phenol/chloroform (50:50). RNA was then precipitated from the aqueous phase in the presence of 300 mM sodium acetate and 2.5
5 volumes of ethanol. Digestions of the HaSV nucleic acid with RNase A and DNase I (Boehringer Mannheim) were done with pBSSK(-) phagemid ssDNA and dsDNA (Stratagene) and RNA controls (BRL). Denaturing agarose gel electrophoresis in the presence of formaldehyde was performed according to Sambrook et al (1989). The state of
10 polyadenylation of the viral RNA was determined by two methods. The first method was to compare the binding of identical amounts (20 µg) of viral RNA and poly(A)-selected RNA from *Helicoverpa virescens* to a 1ml slurry of 5mg of oligo-d(T) cellulose (Pharmacia) in a binding buffer consisting of 20 mM Tris pH 7.8, 500 mM NaCl, 1 mM EDTA
15 and 0.04% SDS. The second method was to observe specific priming of viral RNA and viral RNA polyadenylated with poly(A) polymerase (Pharmacia) with d(T)₁₆A/C/G primers in RNA sequencing reactions using reverse transcriptase (US Biochemical) and a protocol provided by the supplier. The 5' cap structure of the genomic RNA and HaSV
20 was determined by observing the ability of polynucleotide kinase to phosphorylate viral RNA with and without preincubation with tobacco acid pyrophosphatase and alkaline phosphatase (Promega) under conditions described by the supplier.

25

F *In vitro* translation of HaSV RNA. *In vitro* translation of HaSV RNA was performed with lysates of both rabbit reticulocytes and wheat germ (Promega) as directed by the supplier. Reactions were conducted in 10 µl volumes with 1.0 µg of RNA in the presence of five uCi ³⁵S-methionine. The labelled proteins were resolved on 10% and 14%
30 SDS-PAGE gels as described above then visualised by autoradiography of the dried gels. The two viral RNAs were separated by a "freeze and



- 23 -

5 squeeze" method after resolution on nordenaturing low-melting-point agarose gels in TAE (Sambrook, et al. 1989). Briefly, agarose slices containing the RNA were melted at 65 ° C in a volume of TAE buffer equal to six times the agarose volume. The solution was allowed to gel on ice before freezing at -80 ° C for 30 minutes. The frozen solution was thawed on ice then centrifuged at 14,500xg for 10 minutes after which the supernatant was withdrawn and precipitated by the addition of ethanol.

10

G Bioassay of virus-induced pathogenesis

Known amounts of virus isolate, as shown in Figure 3, were fed to larvae at the growth stages indicated by admixture to standard diet. At the time point shown, the larvae were weighed and the mean and SD calculated. Growth of infected larvae was compared to those of uninfected control populations from the same hatching batch in every experiment.

15

Results

20 i) Characteristics and taxonomy of HaSV

The virus particles are isometric and are approximately 36 - 38 nm in diameter. They are composed of two major capsid proteins, of 65 kDa and 6kD is size. The virions contain two single-stranded (+) RNA species of 5.3 kb and 2.4 kb length. The virus bears a similarity in these respects to the Nudaurelia ω virus, which has been tentatively regarded as a member of the Tetraviridae; these two viruses differ however, in the above respects from other viruses in this group and are likely to form a new virus family, sharing chiefly their capsid structure (T=4) with the Tetraviridae.

25

30 ii) Particle characterization and serology.

The buoyant density of HaSV was calculated to be 1.296g/ml in CsCl at pH 7.2. The A_{260}/A_{280} ratio of HaSV viral particles was 1.22 indicating a nucleic



- 24 -

acid content of approximately 7% (Gibbs and Harrison, (1976) Plant Virology: The Principles London: Edward Arnold. Reciprocal immuno-double diffusion comparisons between HaSV and the Nudaurelia ω virus showed no serological relationship. The more sensitive technique of immunoblotting also showed a
 5 complete lack of any antigenic relationship. In addition, HaSV did not react with antisera to the Nudaurelia β virus in a immuno-diffusion test or when immunoblotted. However, no Nudaurelia β virus was available as a positive control in these latter two immunological experiments. When HaSV was stained with acridine orange then irradiated with 310nm UV light, the particles
 10 fluoresced red which indicated a single stranded genome.

iii) Protein characterization.

Examination of the capsid proteins of HaSV with polyacrylamide gel electrophoresis in the presence of SDS showed variable results depending on
 15 the quantity of protein present. At low protein loadings, two proteins in major abundance were evident that had M_r 's of 65,000 and 6,000 along with a protein in minor abundance with M_r of 72,000 (data not shown). When more protein was present on the gels, however, at least 12 more distinct bands with M_r 's ranging between 15,000 and 62,000 became evident. Probing the resolved and
 20 blotted proteins with antibodies monospecific for the major 65 kDa capsid protein showed all but two of the proteins shared common antigens with the major 65 kDa protein. The major 6 kDa capsid protein and a minor band migrating at M_r = 16,000 failed to react with both the monospecific antibodies and untreated antisera.

25

The capsid proteins were shown to be non-glycosylated as they failed to react with a hydrazine analog after oxidation with periodic acid. The N-terminus of the 65 kDa protein appeared to be blocked in some manner as two efforts to
 30 conduct an Edman degradation failed. After the second attempt, the sample was treated with n-chlorosuccinimide and shown to be in a quantity normally adequate for sequencing. The N-terminus of the 6 kDa protein, however, was



- 25 -

not blocked as an unambiguous 16-residue sequence was readily obtained. The sequence of the N-terminus of the 6 kDa capsid protein and those of a cyanogen bromide cleaved fragment of the 65 kDa protein are as follows:

- 5 6 kDa protein:
PheAlaAlaAlaValSerAlaPheAlaAlaAsnMetLeuSerSerValLeuLysSer
(SEQ ID No: 23)
65 kDa protein:
ProThrLeuValAspGlnGlyPheTrpIleGlyGlyGlnTyrAlaLeuThrProThrSer
10 (SEQ ID No: 24)

- Detailed sequence analysis of the RNA genome carried out in Example 3 showed that RNA 1 (SEQ ID No: 39) encodes a protein of molecular weight 186,980 hereinafter referred to as P187 (SEQ ID No: 40) and RNA 2 (SEQ ID No: 47) encodes proteins with molecular weight 16,522 (called P17 (SEQ ID No: 48)) and 70,670 (called P71 (SEQ ID No: 50)). P71 (SEQ ID No: 50) is processed into two proteins of molecular weight 63,378 (called P64) and 7,309 (called P7).

iv) Nucleic acid characterization

- The extracted nucleic acid from HaSV was readily hydrolysed by RNase A but
20 not by DNase I. Denaturing agarose gel electrophoresis of the extracted RNA genome of HaSV indicated two strands that migrated at 5.5 kb and 2.4 kb. The RNA strands were shown not to have extensive regions of polyadenylation as only 24% of the viral RNA bound to the oligo-d(T) cellulose matrix as opposed to 82% of poly(A)-selected RNA. Further evidence for the non-
25 polyadenylation of the viral genome was provided by the observation that the oligo primer, d(T)₁₆G, gave a clear sequencing ladder using reverse transcriptase only after *in vitro* polyadenylation of the viral strands with poly(A)-polymerase.
- 30 The demonstration that the strands could be modified with poly(A)-polymerase also showed the lack of any 3' modification. The 5' termini of the viral strands were shown to be capped, most likely with m⁷G(5')ppp(5')G, as they could not



- 26 -

be labelled with polynucleotide kinase unless pretreated with tobacco acid pyrophosphatase and alkaline phosphatase.

v) *In vitro* translation.

- 5 *In vitro* translation of the viral RNA yielded different results in the two translation systems used (data not shown). The 5.5 kb RNA translated very poorly in the reticulocyte system whereas it produced in the wheatgerm system more than 20 proteins ranging in size from $M_r = 195,000$ to $M_r = 12,000$. The 2.4 kb viral RNA strand yielded a major protein with an $M_r = 24,000$ in both
- 10 systems in addition to a minor protein at $M_r = 70$ kDa. A time course of the translation reaction with the 5.5 kb RNA strand showed all labelled proteins were produced at similar rates indicating that the smaller products did not arise through processing of the larger ones. However when a time course experiment was done with translation of the smaller 2.4 kb RNA strand, the 24
- 15 kDa protein appeared before the 70 kDa protein.

vi) Presence of another form of HaSV

- Frequently, during purification of HaSV virions, a minor band appeared in varying amounts on the CsCl gradient that had a buoyant density of 1.3 g/ml.
- 20 On four occasions, when particles from this minor band were used to infect *H. armiger* larvae that were then processed as before for purification of HaSV virions, the HaSV band with a density of 1.296g/ml was again recovered in vast excess to a varying minor amount of the more dense band. No virions of either type were recovered from uninfected control larvae. Proteins extracted
- 25 from the more dense particles appeared identical to those from the less dense particles when examined by SDS-PAGE and immunoblotting with antibodies specific for the 65 kDa capsid protein of HaSV. Extraction and examination of the RNA genome with denaturing agarose gel electrophoresis also showed the same 5.5 and 2.4 kb bands. When particles from the more dense band
- 30 were examined by electron microscopy as before, they appeared to have a larger diameter 45nm but otherwise highly similar to the 38nm particles.



- 27 -

The molar ratio of the two RNA strands was determined by quantitative densitometry of fluorograms of the resolved strands. The ratio derived from an average of four measurements of various loadings on denaturing gels proved to be 1.7:1 (5.5 kb strand: 2.4 kb strand) which is somewhat lower than
5 the expected ratio of 2.3:1 for equimolar amounts of each strand.

The genome of HaSV has major differences that make it distinct from those of the nodaviruses, the only other group of bipartite small RNA viruses pathogenic to animals. Although HaSV shares the characteristic of a bipartite
10 genome with the only animal viruses having such a divided genome, the nodaviridae, it differs in virtually every other aspect from this group. Both segments of its genome are considerably larger than the corresponding nodaviral RNAs (Hendry D.A., (1991) Nodaviridae of Invertebrates. in (ed. E. Kurstak) Viruses of Invertebrates. Marcel Dekker, New York, pp. 227-276).
15 However, the division of genetic labour is similar with the larger component carrying the replicase gene and the smaller one encoding the capsid proteins. Direct comparison of the sequences shows little homology between these viruses, at either RNA or protein level. The Nodaviruses, have the already mentioned unusual 3'blockage (probably a protein), whereas the HaSV RNAs
20 terminate in a distinctive secondary structure resembling a tRNA.

vii) Bioassays of virus isolates on larvae

The original constructs made to express the capsid proteins (precursor and processed forms) in *E. coli* for bioassay started at the first AUG (nts 284 to
25 286). Production of full-length, immuno-reactive protein from these was due to these clones being the 5C sequence version with the extra C residue. Bioassays of these proteins have been difficult due to problems with obtaining suitable *Heliothis* larvae for the tests.

30 Purified native HaSV was used to conduct bioassays in non-noctuid insect species. The native HaSV was orally administered, the larvae scored for symptoms of infection and growth was measured. Dot blotting for HaSV RNA



RECEIVED 09 SEP 1994

- 27/1 -

was also conducted. Based on these experiments native HaSV does not appear to infect the following larvae.

Species	Order	Family
<i>Galleria mellonella</i>	Lepidoptera	Pyradidae
5 <i>Tineola bissellia</i>	Lepidoptera	Tineidae
<i>Epiphyas postvittana</i>	Lepidoptera	Tortricidae
<i>Lucilia cuprina</i>	Diptera	Calliphoridae
<i>Dacus tyronii</i>	Diptera	Tephritidae
<i>Antitroglus parvulus</i>	Coleoptera	Scarabaediae
10 <i>Lepidiota picticollis</i>	Coleoptera	Scarabaediae
<i>Sericesthis germinata</i>	Coleoptera	Scarabaediae

The above experiment conducted with the larvae of *Spodoptera exigua* and *S. litura* showed that native HaSV infects these species but not to the same

15 degree as seen in *Heliothis armigera*.



AMENDED SHEET
IPEA/AU

- 28 -

EXAMPLE 2

OTHER VIRUS ISOLATES

Materials and Methods

A Virus isolation

- 5 Apparently infected (*viz* diseased) larvae of *Helicoverpa* *sp* were collected in February 1993 at Mullaley (NSW), Narrabri (NSW) and Toowoomba (QLD) (Australia). Referring to Fig. 10 the samples in wells 2A-2D were from parasitised *H. armigera* larvae collected from sorghum at Mullaley; the sample in 6C was collected from sunflower at Toowoomba; the sample in 7D was
- 10 collected from cotton at the Narrabri Research Station. The latter two larvae may have been either *H. armigera* or *H. punctigera*, which are both easily infected with HaSV.

B Virus RNA Extraction

- 15 Larvae collected were ground up and RNA extracted. RNA extraction and purification were as per Example 1.

C Dot-Blot Northern Hybridization

- Extracts of viral RNA was analysed by Northern dot-blot hybridisation using a
- 20 probe made from cloned HaSV sequences derived from 3'-terminal 1000 units of RNA 1 and RNA 2 by random priming in a Boehringer Mannheim kit according to the supplier's instructions were employed. RNA extracts were transferred to Zeta-Probe (BioRad) for probing. Hybridization under high stringency washing conditions were as specified by BioRad. Hybridizations
- 25 were carried out in the following solution:
- 1 mM EDTA, 500 mM NaH_2PO_4 , pH 7.2, 7% SDS, at 65 °C in a rotating Hybaid hybridization chamber. After completion of hybridization and removal of the solution containing the probe, the filters were washed twice in 1 mM EDTA, 40 mM NaH_2PO_4 ,
- 30 pH 7.2, 5% SDS, at 65 °C (1 h each), followed by 2 washes in 1 mM EDTA, 40 mM NaH_2PO_4 , pH 7.2 1% SDS, at 65 °C (1 h each), before autoradiography.



- 29 -

RESULTS

- Referring to Fig. 10, samples 9A, 9B, 10A, 10B and 10C contain HaSV infected positive control lab-raised larvae; 9C-H contain healthy (HaSV-free) negative control lab-raised larvae; All other wells (beginning 1-8) contain
- 5 extract from field-collected larvae. Numbers 2A-D, 6C and 7D gave positive signals indicating that these isolates are either the same as HaSV or derivatives or variants thereof. Election microscopy employing (-) staining confirmed that the samples which gave positive signals contained abundant icosohedral virus particles of approximately 36nm in size.
- 10
- The presence of HaSV in larvae which had tested positive in the Northern hybridization dot-blot was confirmed by Western blotting of crude extracts from such infected larvae, using the polyclonal antibody to the HaSV capsid protein. For routine screening of such extracts in order to identify further
- 15 isolates of HaSV or to confirm the presence of the virus, use of a monoclonal antibody or its equivalent is preferable, in order to achieve (i) higher sensitivity of detection and (ii) greater specificity of detection.

20

EXAMPLE 3**IDENTIFICATION, ISOLATION AND CHARACTERISATION OF INSECT
VIRUS GENES.****Materials and Methods**

- 25 **A Animals and virus production.**

H. armigera larvae were raised as described in Example 1.

- B Protein characterization**

Was conducted as described in Example 1.

30

- C Nucleic acid characterization**

Was conducted as in Example 1.



- 30 -

D Fractionation of virus RNA

The two viral RNAs were separated by a "freeze and squeeze" method after resolution on nondenaturing low melting point agarose gels in TAE (Sambrook, et al, 1989). Briefly, agarose slices containing the RNA were
5 melted at 65° C in a volume of TAE buffer equal to six times the agarose volume. The solution was allowed to gel on ice before freezing it at -80° C for 30 minutes. The frozen solution was thawed on ice then centrifuged at 14,500g for 5 minutes after which the supernatant was withdrawn and precipitated by addition of ethanol.

10

E *In vitro* translation of HaSV RNA

Was as in Example 1.

F cDNA synthesis and cloning of virus genome

15 The virus RNAs were reverse transcribed into cDNA using the Superscript RTase (a modified form of the Moloney murine leukaemia virus (MMLV) RTase, produced by Life Technologies Inc). Oligo(dT) was used as a primer on RNA which had been polyadenylated *in vitro*. After size selection of DNA fragments over 1 kbp in length, the cDNA was then blunt-end ligated using T4
20 DNA ligase (Boehringer Mannheim or Promega, under conditions described by the suppliers) into vector pBSSK(-) (Stratagene) which had been cut with EcoRV and dephosphorylated with calf intestinal alkaline phosphatase (Boehringer Mannheim). E.coli strain JM109 or JPA101 were electroporated with the ligation mixture and white colonies selected on colour-indicator plates
25 Sambrook et al, 1989.

For some clones of RNA2 (SEQ ID No: 47), cDNA was synthesised using the RTase of AMV (Promega) and a specific primer complementary to nucleotide sequence 2285 - 2301 of RNA 2 (SEQ ID No: 47). The same buffer and
30 conditions were used for the Superscript RTase (above). The AMV RTase was found not to make cDNA form a primer annealing to the terminal 18



- 31 -

nucleotide sequence (see below), nor to be able to reach the 5'-end of the RNA with the primer here described.

G Sequencing of DNA and RNA

- 5 The cDNA clones were separated as single-stranded or double-stranded DNA, using the deaza-dGTP and deaza-dITP nucleotide analogues (Pharmacia) in the deaza T7 sequencing kit as recommended by this supplier. Synthetic oligonucleotides were used as primers. The 5' terminal sequences of the two RNAs were determined using reverse transcriptase to sequence the RNA
- 10 template directly, from specific oligonucleotide primers located about 200 nucleotides downstream from the termini. Such RNA sequencing was performed using the reverse transcriptase sequencing kit from Promega, under the conditions described by the manufacturer.
- 15 The sequence of the 20 or so nucleotides at the 5' terminus of each RNA was checked using direct RNase digestion of 5'-labelled RNA under conditions designed to confer sequence-specificity. Direct RNA sequence using RNases was performed with the RNase sequencing kit from US Biochemicals, following the protocols provided by the manufacturer. This also confirmed
- 20 that the sequence of the most abundant RNA is consistent with that of the RNA analysed using the specific primer and RTase.

- All transcription of plasmids linearized as described were performed as recommended by the suppliers of SP6 RNA polymerase, in the presence of
- 25 1mM cap analogue, 0.2mM GTP, and 0.5mM of the other NTPs.

H Subcloning and expression

PCR amplification

- The polymerase chain reaction (PCR) was used to obtain sequences covering
- 30 virus genes in a form suitable for cloning into expression vectors. The reaction was performed with Taq DNA polymerase (Promega) as described by the supplier, in a rapid cycling thermal sequencer manufactured by Corbett



- 32 -

Research (Sydney, Australia). A typical reaction involved 1 cycle of 1 min at 90 °C, 25 cycles of 95 °C (10 sec), 50 °C (20 sec), 72 °C (1.5 min), followed by one cycle of 72 °C for 5 min. Templates were generally cDNA or cDNA clones derived from HaSV RNAs, made as described below. Primers were as
5 described below for the relevant constructs.

Upon termination of the PCR reaction, the product's ends were made blunt by treatment with E.coli DNA polymerase I (Klenow fragment) at ambient temperature for 15 minutes. After heating at 65 ° C for 10 minutes, the
10 reaction was cooled on ice and the reaction mix made 1mM in ATP. The product then 5'-phosphorylated using 5 units of T4 polynucleotide kinase at 37 ° C for 30 minutes. After heating at 65 ° C for 10 minutes, the product was run on a 1% low-melting agarose gel and purified as described for RNA in section E above.

15

ligations: Vectors and restriction fragments cut with the enzymes described were run on 1% low-melting-point agarose gels and excised as slices. These slices were then melted at 65 ° C for 5 minutes, before cooling to 37 ° C. Fragment and vectors were then ligated in 10ul total volume at 14 ° C
20 overnight using T4DNA ligase (BRL, Boehringer Mannheim or Promega), in the buffers supplied by the manufacturers.

expression: Expression plasmids containing viral genes (e.g. for the capsid protein) were transformed into *E. coli* strain BL21 (DE3) or HMS174 (DE3)
25 (supplied by Novagen). After growth as specified by the supplier, protein expression was induced by the addition of isopropyl β -D-thiogalactopyranoside (IPTG), at 0.4 mM to the growing culture for a period of 3h. Expressed proteins were analysed by SDS-polyacrylamide gel electrophoresis of bacterial extracts (Laemmli, 1970).

30



- 33 -

Results**i) Mapping cDNA clones of HaSV**

The template for cDNA synthesis was virus RNA which had been polyadenylated in vitro. Oligo(dT) was used as a primer for the Superscript reverse transcriptase (RTase; a modified form of the Moloney murine leukaemia virus (MMLV) RTase, produced by Life Technologies Inc). The cDNA was cloned into vector pBSSK(-) as described earlier. The larger clones were selected for further analysis by restriction mapping and Northern hybridization. All the probes tested hybridized either to RNA 1 or to RNA 2, suggesting that there are no regions of extensive sequence homology between the two RNA's. Furthermore, screening of a number of other clones excluded the theoretical possibility that either RNA band may actually contain more than one species.

ii) RNA 1 (SEQ ID No: 39) clones

Three large RNA 1 (SEQ ID No: 39) clones (B11U, B11O and B35) obtained for the first round of clones were further analysed by restriction mapping and shown to form an overlap spanning over 3 kbp (this was later confirmed by sequencing). The second round of cloning then yielded E3 of 5.3 kbp, representing 99.7% of RNA 1 (SEQ ID No: 39). A complete restriction map of clone E3 showed it to align with that previously determined for three overlapping clones. On the basis of this alignment, the 5' end of the insert in B11U was placed about 300 nucleotides downstream from the 5' end of the RNA.

25

Once clones covering a contiguous block had been identified, the orientation relative to the RNA was determined.

iii) RNA 2 (SEQ ID No: 47) clones

Three significant cDNA clones were isolated for RNA 2 (SEQ ID No: 47) (Fig. 2). One, hr236, contains about 88% of RNA 2 (SEQ ID No: 47) (2470 bp total length), and runs from the 3' end to 240 bp from the 5' end. The



- 34 -

other clones, hr247 and hr 249 are 3' coterminal subgenomic fragments of 1520 bp and 760 bp, respectively. Orientation of clone hr236 was determined by strand specific hybridization. While a much stronger signal was seen with a probe for one orientation, the probe specific for the other orientation also
5 yielded a signal, indicating that there are extensive regions of reverse complementarity within the positive strand sequence. Such sequences are likely to form extensive short and long-range secondary structure.

The clones contain the 3' sequence of HaSV RNA 2 (SEQ ID No: 47) as they
10 all have the same 3' sequence adjacent to the poly (A) stretch added *in vitro* before cDNA priming. The remaining 5' sequence of RNA 2 has been obtained by direct RNA sequencing using two reverse transcriptases as described above.

15 iv) Sequencing of virus genome

The clones mapped in section (i) were selected for further analysis by sequencing.

The cDNA clones were completely sequenced as single-stranded DNA in both
20 orientations, using the deaza-dGTP and deaza-dITP nucleotide analogues (Pharmacia) and synthetic oligonucleotides as primers.

v) Sequence of genome component 1 (SEQ ID No: 39) (see Figure 1)

The 5310 nucleotides of RNA 1 (SEQ ID No: 39) encode a protein of
25 molecular weight 187,000 which is regarded as the RNA-dependent RNA polymerase (replicase) in view of its amino acid sequence similarity in certain limited regions to replicases of other RNA viruses. The apparent molecular weight of this protein upon *in vitro* translation of virus RNA and SDS-PAGE is 195,000.

30

Sequence analysis of RNA 1 (SEQ ID No: 39) was concentrated on clone E3 which extends from the 3' end of RNA 1 to 18 nucleotides from the 5' end



- 35 -

(Figure 1). The complete sequence has been confirmed by sequencing in both directions. An ORF of 1750 amino acids and spanning virtually the complete RNA (5310 nucleotides in length) has been detected. This ORF begins with the first AUG on the sequence at position 34 and terminates at nucleotide

5 5290 and is thought to encode the RNA-dependent RNA polymerase (replicase)(referred to as P187 (SEQ ID No: 40) in Fig. 1) required for virus replication, since it contains the Gly-Asp-Asp conserved triplet and surrounding sequences identified in these enzymes, which are usually large (over 100 kDa), in addition to further homology with the polymerase encoded

10 by tobacco mosaic virus and other plus-stranded RNA viruses.

Referring to Fig. 1 the sequence is presented as the upper strand of the cDNA sequence. This strand is therefore in the same sense as the viral (positive-sense) RNA. The sequence of the protein encoded by the major open reading

15 frame, encoding the putative RNA-dependent RNA replicase, is shown, as are those of the small open reading frames at the 3' end, corresponding to the proteins P11a (SEQ ID No: 42), P11b (SEQ ID No: 44) and P14 (SEQ ID No: 46).

20 Clone E3 was inserted downstream of the SP6 promoter for *in vitro* transcription. As mentioned above, the transcript of this clone can be translated in the wheat germ system to yield the 195 kDa protein observed upon translation of fractionated RNA 1 (SEQ ID No: 39) from the virus. The latter yields more lower molecular weight products, presumably due to being

25 contaminated with nicked and degraded RNA. The products derived from the *in vitro* transcript can therefore be regarded as defining the coding capacity of the complete RNA 1 (SEQ ID No: 39) of HaSV.

vi) **Sequence of genome component 2 (see Figure 2)**

30 The 2470 nucleotides encode a protein of molecular weight 71,000 which contains the peptide sequences corresponding to those determined from the two virus capsid proteins. This protein is therefore the precursor of these



- 36 -

capsid proteins. The protein is a major product of *in vitro* translation of this RNA obtained either from virus particles or by *in vitro* transcription of a full-length cDNA clone; in addition, another major translation product of apparent molecular weight 24,000 is obtained. This protein is derived from a
5 molecular weight 17,000 reading frame overlapping the slab of the capsid protein gene.

Clones hr236 and hr247 were completely sequenced as the first step in RNA 2 sequencing. These sequences were then extensively compared to that obtained
10 by direct RNA sequencing using AMV reverse transcriptase.

Comparison of the cloned sequence with that by direct RNA sequencing showed both clones lacked 50 nucleotide present in the RNA (at around nucleotide 1500). The sequence of this stretch was obtained by direct RNA
15 sequencing using the AMV RTase. The MMLV "Superscript" RTase, which was used to make all the cDNA clones, was found to simply by-pass this region in sequencing reactions. These 50 nucleotides contain a very stable GC-rich hairpin flanked by a 6 bp direct repeat, and the MMLV RTase skips from the first repeat to the second.

20

The sequence of RNA 2 (SEQ ID No: 47) was then completed using plasmids pSR2A and pSR2P70 constructed as described below. The plasmids contain a segment of cDNA derived for the AMV RTase, as well as the sequence corresponding to the 5' 240 nucleotides of RNA 2 (SEQ ID No: 47) which are
25 not present on phr236 (Fig. 2). The sequence of RNA in Fig. 2 is presented as the upper strand of the cDNA sequence. This strand is therefore in the same sense as the viral (positive-sense) RNA. The sequences of the proteins encoded by the major open reading frames, encoding the capsid protein precursor P71 (SEQ ID No: 50), and P17 (SEQ ID No: 48).

30

The sequence of RNA 2 (SEQ ID No: 47) encodes a major ORF running from a methionine initiation codon at nucleotides 366 to 368 to a termination codon



- 37 -

at nucleotides 2307 to 2309. This protein encoded by this ORF has a theoretical molecular weight of 71,000 (SEQ ID No: 50). This initiation codon is in a good context (AGGatgG), suggesting that it will be well recognized by scanning ribosomes. The size of the product is close to that of the residual
5 putative precursor protein identified in purified virus, and to the size of the *in vitro* translation product obtained from RNA 2 (SEQ ID No: 47).

The approach adopted to identify the gene encoding the capsid protein was to obtain amino acid sequence information from the two abundant capsid
10 proteins and then locate these on the protein encoded by the sequence of the virus RNA's. CNBr cleaved products of the capsid protein were therefore sequenced. These fragments gave a clear and unambiguous sequence shown in Example 1. These sequences determined were then located on the large ORF of RNA 2 (SEQ ID No: 47). (Figure 2)

15

In the case of the small capsid protein, the clear and unambiguous sequence, obtained is located near the carboxy terminus of the major ORF on RNA 2 (SEQ ID No: 47). Starting at the point corresponding to the amino-terminal residue of the sequence determined for the 6 kDa protein, and continuing to
20 the carboxy-terminus of the complete reading frame, the protein encoded by the sequence 7.2 kDa and has a hydrophobic N-terminal region and an arginine rich (basic) C-terminal region. It is an extremely basic protein with a pI of 12.6.

25 The two abundant capsid proteins are derived from a single precursor, which is processed at a specific site. This is presumably immediately amino-terminal to the sequence FAAVS.... (SEQ ID No: 25)

RNA 2 (SEQ ID No: 47) appears to be a bicistronic mRNA (see Figs. 2 and
30 5). The first methionine codon is encoded on the sequence of RNA at nucleotides 283 to 285. This ATG is in a poor context (TTTatgA), making it a weaker initiation codon. It initiates a reading frame of 157 amino acids,



- 38 -

- encoding a protein of molecular weight 17,000 (SEQ ID No: 48). (The second AUG [nts 366 to 368] initiates the 71 kDa (SEQ ID No: 50) precursor of the capsid protein). Since the first AUG is in a poor context, abundant expression of the capsid precursor would be expected. In fact, *in vitro* translation of a full length RNA 2 (SEQ ID No: 47) transcribed from a reconstructed cDNA clone yields two major protein products of relative mobility 71,000 (SEQ ID No: 50) and 24,000, similar to those already observed upon translation of viral RNA 2 (SEQ ID No: 47). The protein of Mr 24,000 appears to correspond to the 157 amino acid protein, despite the significant anomaly in apparent size. The 24,000 Mr product was also observed upon translation of an *in vitro* transcript covering only nucleotides 220 to 1200 of RNA 2 (SEQ ID No: 47). This region contains no open reading frame other than those already mentioned and cannot encode a protein longer than 157 amino acids.
- The protein of Mr 24,000 seen upon *in vitro* translation appears to correspond to P17 (SEQ ID No: 48), with the anomaly in apparent size probably being due to the high content of proline (P), glutamate (E), serine (S) and threonine (T). These amino acids cause the protein run more slowly on a gel thereby giving it an apparent size of Mr 24,000.
- The Mr 24,000 protein (hereinafter referred to as P17 (SEQ ID No: 48)) may have a function in modifying or manipulating the growth characteristics or cell cycle of HaSV-infected cells. Although a protein of 16kDa (identified in Example 1) is found in small amounts in the capsid, it does not react with antiserum against the virus particles this is unlikely to correspond to P17 (SEQ ID No: 48), since a preparation of the latter proteins migrates with a molecular weight of 24,000 on SDS gels.

- Sequence analysis of the Region from nucleotide 500 to 600 of RNA 2 (SEQ ID No: 47) showed that it has the sequence shown in Fig. 2, as do the plasmids pSR2A, pSR2P70, pSR2B and pSXR2P70. However, plasmids pT7T72P65 and pT7T2P70 have an extra C residue at nucleotide 570. The RNA sequence from



- 39 -

which they are derived from is shown in Fig. 2 (the "5C" version). In this sequence the first ATG (nucleotides 283 to 285) is in the same reading frame as most of the capsid protein gene. The resultant fusion protein is called "P70" (SEQ ID No: 52) and its carboxyterminal-truncated version (a variant of the native P64) is "P65". In view of these clones it was considered important to resolve whether any virus RNA carrying the extra C residue was present in the viral RNA population first isolated for investigation.

Direct sequencing of the virus RNA using reverse transcriptase confirmed that the 4C version lacking the extra residue was the abundant form of the RNA. In order to exclude the possibility of a small amount of the RNA having the extra residue, a sensitive PCR assay was designed. This showed that the extra C residue was not present on any RNA in the viral population, and had been introduced into some clones as a PCR artefact. These clones were however retained and used in bacterial expression experiments (below) because of the high level expression obtained of the P65 and P70 (SEQ ID No: 52) fusion proteins.

vii) Comparison with the sequence of the *Nudaulrelia* ω capsid gene

The sequence of most of the RNA2 of the *Nudaulrelia* ω virus has recently been published by Agrawal D.K. and Johnson J.E. (Virology 190 806-814, 1992). From the published sequence it has been determined that this sequence shows 63% homology to that of HaSV RNA2 (SEQ ID No: 47) at the nucleotide level and 66% at the overall amino acid level. A detailed comparison of the capsid proteins of these two viruses shows the amino-terminal 45 residues to be variable, the next 220 residues to be highly conserved, the next 180 residues to be variable and the c-terminal 200 residues covering the small protein P7 to be highly conserved. A more detailed comparison is discussed below.



- 40 -

The published report did not find a complete reading frame corresponding to the 157 amino acid protein (P17 (SEQ ID No: 48)) gene reported above. The AUG is however present, as is a reading frame - starting upstream of the start of the capsid gene - showing considerable amino acid homology to P17 (SEQ ID No: 48) of HaSV. In vitro translation of purified *Nudaurelia* ω virus RNA 2 and a re-examination of the nucleotide sequencing data for this RNA may help to resolve the question of whether the *Nudaurelia* ω virus also encodes a protein homologous to the HaSV P17.

10 More interestingly, antisera against these two viruses, which are similar at a nucleotide sequence level, do not show any cross-reactivity.

viii) Construction of full-length clones

15 RNA 1 (SEQ ID No: 39)

cDNA clone E3, described above contains all but the 5'-18 nucleotides of RNA 1 (SEQ ID No: 39) and included the complete ORF present on the sequence. The first full-length clone of RNA 1 (SEQ ID No: 39) is therefore based on E3. The 4.9 kbp XbaI-ClaI fragment from clone E3 was recloned into

20 pBSKS(-) (Stratagene) cut with XbaI and ClaI, giving pBSKSE3.

The full-length clone of RNA 1 (SEQ ID No: 39) was completed using PCR. The primer defining the 5' end of the RNA carried an EcoRI site, the promoter for the SP6 RNA polymerase and a sequence corresponding to the 5' 17 nucleotides of RNA 1, as shown in Figure 1. The sequence of this primer was:

HvR1SP5p:

5'-GGGGGGAATTCATTTAGGTGACACTATAGTTCTGCCTCCCCGGAC
(SEQ ID No: 11) (The G which initiates transcription is underlined)

30 Using an oligonucleotide complementary to nucleotides 1192 - 1212, a PCR product of 1240 bp was efficiently made. The template was cDNA synthesised using the MMLV RTase and the same oligonucleotide complementary to



- 41 -

nucleotides 1192 - 1212 was the primer. Upon termination of the PCR reaction, the product's ends were made blunt and then 5'-phosphorylated as described below. The purified PCR fragment was then cleaved with restriction endonuclease XbaI and the 450 bp subfragment corresponding to the 5' end of
5 RNA 1 (SEQ ID No: 39) cloned into the plasmid pBSSK(-)(Stragene) cut with EcoRV and XbaI, to give pBSR15.

To assemble the full-length of RNA 1 (SEQ ID No: 39), pBSKSE3 (above) was cut with XbaI and ScaI giving fragments of 1.2 kbp and 6.8 kbp. pBSR15
10 was cut with the same enzymes, giving fragments of 2 and 1.8 kbp. Ligation of the 6.8 kbp fragment for pBSKSE3 and the 1.8 kbp fragment for mpBSR15 yielded pSR1(E3)A. Upon linearization at ClaI and *in vitro* transcription with the SP6 RNA polymerase, and RNA corresponding to RNA 1 (SEQ ID No: 39), and terminating in a poly(A) stretch of about 50 nucleotides, is obtained.

15

Since the natural RNA 1 (SEQ ID No: 39) does not have a poly (A) tail, an alternative plasmid was constructed which carries a BamHI restriction site immediately downstream of the 3'end of RNA 1 (SEQ ID No: 39). Again this terminal fragment was made using PCR as above. The sequence of the primer
20 was as follows:

HvR13p: 5'-GGGGGGATCCTGGTATCCCAGGGGCGC (SEQ ID No: 12)
(the nucleotide complementary to that which was determined as the 3' one, based on its adjacency to the poly(A) stretch, is underlined; RNA terminating at the BamHI site will have the sequence GCGCCCCCUGGGAUACCaggauc
25 (SEQ ID No: 26)).

The template was clone E3 and an oligonucleotide corresponding to nucleotides 4084 - 4100 was the other primer. The 1220 bp product was blunt-ended, kinased and gel-purified as described above, before cleavage with
30 HindIII. The resulting 420 bp subfragment corresponding to the 3' end of RNA 1 (SEQ ID No: 39) cloned into plasmid pSR1(E3)A cut with ClaI, end-filled with Klenow and then cut with HindIII. The resulting plasmid is



- 42 -

pSR1(E3)B. Upon linearization at BamHI and *in vitro* transcription with the SP6 RNA polymerase, and RNA corresponding to RNA 1 (SEQ ID No: 39), and terminating as described immediately above is obtained.

5 ix) RNA 2 (SEQ ID No: 47)

In constructing the full-length cDNA clone to enable *in vitro* transcription of this RNA hr236 described above was used as a basis. Two separate PCR products, one corresponding to the 5' portion of RNA 2 (SEQ ID No: 47), which is missing from this clone altogether, and another covering the region
10 where clone hr236 lacks the hairpin-forming sequence described above, were required.

The primer defining the 5' end of the RNA carried a HindIII site and a sequence corresponding to the 5' 18 nucleotides of RNA 2 (SEQ ID No: 47),
15 as shown in Figure 2. The sequence of this primer was:

Hr2cdna5: 5'-CCGGAAGCTTGTTTTTCTTTCTTTACCA (SEQ ID No: 13)
(The nucleotide underlined corresponds to that identified as the first nucleotide of RNA 2. (SEQ ID No: 47))

Using an oligonucleotide complementary to nucleotides 1653 - 1669, a PCR
20 product of 1.67 kbp was made. The template was cDNA synthesised using the MMLV RTase and an oligonucleotide complementary to the 18 nucleotides at the 3' end of RNA 2 (SEQ ID No: 47) as the primer. Upon termination of the PCR reaction, the product was blunt-ended, kinased and gel-purified as described above, before cleavage with PstI. The resulting 1.3 kbp subfragment
25 corresponding to the 5' half of RNA 2 (SEQ ID No: 47) was cloned into plasmid pBSSK(-) (Stragene) cut with EcoRV and PstI, giving plasmid pBSR25p. In order to place this subfragment corresponding to the 5' half of RNA 2 (SEQ ID No: 47) downstream of the SP6 promoter for *in vitro* transcription, a 1.3 kbp HindIII - BamHI fragment was excised from pBSR25p
30 and ligated into HindIII - BamHI cut pGEM-1 (Promega), giving plasmid pSR25.



- 43 -

The second PCR product, covering the region where clone hr236 lacks the hairpin-forming sequence described above, was synthesised using as primers oligonucleotides corresponding to nucleotide sequence 873 to 889 of RNA 2 (SEQ ID No: 47) and to the complement of nucleotide sequence 2290 - 2309.

- 5 Upon termination of the PCR reaction, the product was blunt-ended, kinased and gel-purified as described above, before cleavage with AatII. The resulting 1.1 kbp subfragment covering the required region was cloned into plasmid phr236 cut with HindIII, end-filled with Klenow and cut with AatII, giving plasmid phr236P70.

10

The two segments were joined covering the first 230 nucleotides of RNA 2 (SEQ ID No: 47) together. Plasmid phr236P70 was cut at the SacI site in the vector adjacent to the 5' end of the insert and this made blunt-ended using Klenow in the absence of dNTPs. After heat-inactivation of the Klenow, the

- 15 plasmid was cut with EcoRI, yielding fragments of 4.5 kbp and 380 bp. Plasmid pSR25 was cut with NheI, blunt-ended by end-filling with Klenow and cut with EcoRI, yielding fragments of 2.8 kbp, 900 bp and 750 bp. The 4.5 kbp fragment of phr236P70 and the 900 bp fragment of pSR25 were ligated to give pSR2P70. This clone covers all of RNA 2 (SEQ ID No: 47) except for the 3'
- 20 169 nucleotides.

To complete the full-length clone of RNA 2 (SEQ ID No: 47), it was necessary to insert a fragment covering the 3' end. As with RNA 1 (SEQ ID No: 39), two versions were made. One, called pSR2A, used the 3' end as present in

25 phr236, together with the poly(A) tail present in this version. The other pSR2B, used a PCR fragment carrying a BamHI site immediately downstream of the 3' nucleotide, as in pSR1(E3)B above. To construct pSR2A, a 350 bp NotI-ClaI fragment was excised from phr236 and cloned into pSR2P70 cut with the same endonucleases. Linearization at the unique ClaI site allows *in vitro*

30 transcription of the complete RNA 2 (SEQ ID No: 47) and a poly(A) tail of about 50 nucleotides in length.



- 44 -

To make pSR2B, an appropriate PCR product was made using as primers an oligonucleotide corresponding to nucleotide sequence 1178 to 1194 and to the 3' terminal 18 nucleotides of RNA 2 (SEQ ID No: 47). The latter primer carried a BamHII site attached, giving it the sequence:

- 5 HvR23p: 5'-GGGGGATCCGATGGTATCCCGAGGGACGC
(SEQ ID No: 14)

The template used was a plasmid phr236. Upon termination of the PCR reaction, the product was blunt-ended, kinased and gel-purified as described
10 above, before cleavage with NotI. The resulting 400 bp subfragment covering the required region was cloned into plasmid pSR2P70 cut with ClaI, end-filled with Klenow and cut with NotI, giving plasmid pSRP2B. Linearization at the unique BamHI site allows *in vitro* transcription of the complete RNA 2 (SEQ ID No: 47), terminating with the sequence ACCaggatc.

15

x) Construction of pSXR2P70

This plasmid was made to determine where p24 starts. A 2.1 kbp XhoI-BamHI fragment was cut from clone pSR2P70 and ligated into the vector pGEM-1 (Promega) which had been cut with SalI and BamHI. In vitro
20 transcription of the resulting plasmid after linearization at the unique BamHI site yielded an RNA covering about 70 nucleotides upstream of the first ATG at nucleotides 283 to 286, plus a short sequence derived from the vector.

In *vitro* translation of the RNA from pSXR2P70 yielded both proteins (P70
25 (SEQ ID No: 52) + P24).

xi) Description of virus-induced pathology

The virus induces a rapid anti-feeding effect in *Helicoverpa* larvae as determined by experiments with larvae the results of which are shown in Fig.
30 3. Fig. 3 shows: A. neonate larvae (less than 24 h old) were fed the designated concentrations of isolated virus (in particles per ml [of diet] added to solid diet). They were weighed on following days and the mean of a statistically



- 45 -

significant number (24) of larvae shown. Where necessary, mortality was recorded for the higher concentrations. The vertical axis shows the fold-increase in weight from the hatching weight of 0.1 mg per larvae. This scale therefore also corresponds to weight in units of 0.1 mg (ie 300 is equivalent to 5 30 mg). B. As for A, but the larvae were 5 days old at the start of the virus feeding. The vertical scale is in mg weight.

No weight gain at all was detectable with neonates which had been fed the doses of virus over 10^8 particles per ml (virus added to diet). In addition, 10 100% mortality was evident after four days at the highest doses. Virus doses as low as 10^6 particles per ml (virus added to diet) still cause significant stunting. The five day old larvae showed a cessation of feeding after 48 hours and significant stunting at 4 dpi, but no mortality at comparable virus doses (Figure 3). Neonates are therefore very sensitive indeed to this virus. Virus 15 particles accumulate specifically in the midgut. This potent anti-feeding effect may be due to the capsid protein or another protein encoded by the virus, or to the effect of any combination of such proteins.

xii) Expression of virus-encoded proteins in bacteria.

20 The vectors

The expression system used initially was derived from the pET-11 system (Novagen). Trimmed down versions of pET-11b and c were constructed and used to compare expression of the capsid proteins. However, due to difficulties experienced with this system substantial modification of the original vectors 25 was carried out in order to achieve much higher yields. These results are described in iii-b) below.

The initial trimmed-down vectors discussed above were made as follows: pGEM-2 (Promega) which carries T7 promoter adjacent to a poly-linker 30 sequence, but has no sequences corresponding to the lac operon, was cut at the unique XbaI (34) and ScaI (1651) sites, giving fragments of 1.61 and 1.25 kbp. The plasmids pET-11b and c were cut with the same enzymes, giving fragments



- 46 -

of 4.77 and 0.91 kbp. The 1.61 kbp fragment of pGEM-2, carrying the c-terminal portion of the ampicillin-resistance gene, the origin of replication and the T7 promoter, was then ligated to the 0.91 kbp fragment of the pET vector, which carries a sequence covering the Shine-Dalgarno sequence, the ATG (in
5 a NdeI site), the terminator for the T7 polymerase and the N-terminal portion of the ampicillin-resistance gene. The resulting plasmids of approximately 2.53 kbp, called pT7T2-b and c, therefore carry a complete T7 transcription unit, which may be used as an expression system in a manner similar to the original pET-11 plasmids, but are repressor-neutral within the cell; they neither titrate
10 away repressor by carrying a binding site, nor do they carry the gene producing the repressor. They were found to grow very well in *E.coli* strains JM109 and BL21 (DE3), and to be very efficient expression vectors. The repressor present in the cells was found to be sufficient to keep the genomic T7 polymerase gene uninduced and therefore the foreign gene unexpressed in the
15 absence of IPTG.

xiii-a) Construction of plasmids for expression of capsid proteins

In this section, all proteins expressed from segments of HaSV RNA 2 (SEQ ID No: 47) are referred to by the size of their gene, as defined in Fig. 4 and in
20 section vi) of this example. The following plasmids were constructed by PCR, using the abovementioned full-length clone of RNA 2 (SEQ ID No: 47), plasmid pSR2A as the template, except where mentioned otherwise.

Groups of plasmids expressed protein starting at each of the first three
25 methionine initiation codons found on the sequence of HaSV RNA 2 (SEQ ID No: 47). For those proteins initiating at the first methionine initiation codon found on the sequence of HaSV RNA 2 (SEQ ID No: 47) (which initiates the P17 (SEQ ID No: 48) gene; oligonucleotide primer HVPET65N (SEQ ID No: 15)), an extra group of plasmids was made by PCR using as a template the
30 version of the RNA 2 sequence carrying an extra C residue inserted at residue 570 (SEQ ID No: 51) (as depicted in Figure 2). Expression constructs initiating at the third methionine initiation codon found on the sequence of



- 47 -

- HaSV RNA 2 (which is located within the P17 gene; oligonucleotide primer HVPET63N (SEQ ID No: 16)) were made by PCR using as a template only the version of the RNA 2 sequence carrying an extra C residue inserted at residue 570 (SEQ ID No: 51). For these latter expression constructs, as well
- 5 as those designed to initiate expression from the second methionine initiation codon found on the sequence of HaSV RNA 2 (SEQ ID No: 47) (which initiates the P71 gene; oligonucleotide primer HVPET64N (SEQ ID No: 17)), two versions were constructed.
- 10 One version terminated at a point corresponding to the c-terminus of the processed (P64) form of the capsid protein and was made using oligonucleotide primer HVP65C (SEQ ID No: 19). The other version terminated at a point corresponding to the c-terminus of the precursor (P71 (SEQ ID No: 50)) form of the capsid protein and was made using
- 15 oligonucleotide primer HVP6C2 (SEQ ID No: 20).

- The sequence encoding P64 (or the precursor, P71 (SEQ ID No: 50)) was synthesised in two segments using PCR. The amino-terminal half of the gene was obtained using as primers oligonucleotides incorporating one of the three
- 20 ATG possible initiation codons for the ORF, in addition to an oligonucleotide with the sequence TCAGCAGGTGGCATAGG (SEQ ID No: 27); complementary to nucleotides 1653 to 1669 of the sequence shown in Fig. 2. The forward primers were as follows:

HVPET65N:

- 25 AAATAATTTTGTCTTACTTTAGAAAGGAGATATACATATGAGCGAGCGA
GCACAC (SEQ ID No: 15)

(the underlined sequence corresponds to nucleotides 283 to 296 of the sequence shown in Figure 2)

- 30 HVPET63N
AAATAATTTTGTCTTAACCTTAAGAAGGAGATCTACATATGCTGGAGTG
GCGTCAC (SEQ ID No: 16)



- 48 -

(the underlined sequence corresponds to nucleotides 373 to 390 of the sequence shown in Figure 2; the AflII (CTTAAG) and BglII (AGATCT) sites introduced into the sequence by single nucleotide changes (shown in *italics*) in the oligonucleotide are shown in bold).

5

HVPET64N

GGAGATCTACATATGGGAGATGCTGGAGTG (SEQ ID No: 17)

(the underlined sequence corresponds to nucleotides 366 to 383 of the sequence shown in Figure 2; the BglII site introduced into the sequence by a single nucleotide change in the oligonucleotide is shown in bold).

10

The PCR products obtained from each combination of one of these primers with the abovementioned one were treated with the Klenow fragment of *E. coli* DNA polymerase, and then with T4 polynucleotide kinase in the presence of 1 mM ATP, before purification by agarose gel electrophoresis as described above. Each product was then cleaved with AatII to yield fragments of 0.95 and 0.4 kbp, and each resulting fragment of about .95 kbp cloned into vector pGEM-2 (Promega) cut with HincII and AatII, giving plasmids pGEMP63N (in which the insert commenced with oligonucleotide HVPET63N (SEQ ID No: 16)), pGEMP64N (in which the insert commenced with oligonucleotide HVPET64N (SEQ ID No: 17)) and pGemP65N (in which the insert commenced with oligonucleotide HVPET65N (SEQ ID No: 15)). The fragment covering portion of the HaSV capsid gene was then excised with enzymes AatII and XbaI.

25

Two versions of plasmid pGemP65N were made, using different templates as described above. pGemP65N was derived from the sequence of the viral RNA, as in plasmid pSF2A; plasmid pGemP65Nc was derived from the sequence carrying an extra C residue, as shown in Fig. 2 (see "5C version").

30

In parallel, the carboxy-terminal halves of the major capsid protein variant, whether terminating as for P64 or for P71 (SEQ ID No: 50), were also



- 49 -

produced using PCR. An oligonucleotide primer, HVRNA2F3, with the sequence GTAGCGAACGTCTGAGAA (SEQ ID No: 18) (corresponding to nucleotides 873 to 889 of the sequence shown in Figure 2) was used in conjunction with each of the two primers following:

5

HVP65C

GGGGGATCCTCAGTTCCTCAGTGGCGGGGTAG (SEQ ID No: 19)

(the underlined sequence is complementary to nucleotides 2072 to 2091 of the sequence shown in Figure 2).

10

HVP6C2

GGGGATCCCTAATTGGCAGGAGCGGCGC (SEQ ID No: 20)

(the underlined sequence is complementary to nucleotides 2290 to 2309 of the sequence shown in Figure 2).

15

The PCR products obtained from each combination of one of these primers with the above mentioned one (HVRNA2F3 (SEQ ID No: 18)) were treated with the Klenow fragment of *E.coli* DNA polymerase, and then with T4 polynucleotide kinase in the presence of 1 mM ATP, before purification by agarose gel electrophoresis as described above. Each product was then cleaved with AatII to yield fragments of 0.9 kbp (in the case of HVP65C (SEQ ID No: 19)) or 1.1 kbp (in the case of HVP6C2 (SEQ ID No: 20)) and 0.4 kbp, and each resulting fragment of about .9 or 1.1 kbp cloned into plasmid phr236 cut with HindIII, treated with Klenow and AatII, giving plasmids phr236P65C and phr236P70 (which has already been described above), respectively. The fragment covering the c-terminus of the capsid protein gene was then excised with enzymes AatII and BamHI.

To assemble plasmids for expression in suitable strains of *E. coli*, the excised XbaI-AatII fragments of 0.95 kbp covering the amino-terminal half of the gene and the excised AatII - BamHI fragments of 0.9 or 1.1 kbp covering the carboxy-terminal half of the gene were simultaneously ligated into the vector



- 50 -

pT7T2 cut with XbaI and BamHI. Initial transformation was of *E. coli* strain JM109. Recombinant plasmids carrying the correct insert were then transformed into strain BL21(DE3) for expression as described above.

- 5 The plasmid obtained by ligating the aminoterminal fragment commencing with oligonucleotide primer HVPET63N (SEQ ID No: 16) to the c-terminal fragment ending at oligonucleotide primer HVP65C (SEQ ID No: 19) in the expression vector pT7T2b was called pP65G.
- 10 In the case of plasmid pP64N, containing an insert from HVPET64N (SEQ ID No: 17) to HVP65C (SEQ ID No: 19), the fragment covering the amino-terminal half of the oligonucleotide was excised by BglII and ScaI from the plasmid pGemP64N and the fragment covering the remainder of the gene was excised with ScaI and EcoRI from plasmid pT7T2-P65. These two fragments
- 15 were then ligated simultaneously into pP65G which had been cut with BglII and EcoRI.

- The resulting construct carrying the complete P71 (SEQ ID No: 50) precursor gene was called pT7T2-P71 and that carrying the P64 form of the gene was
- 20 called pT7T2-P64. In the case of plasmids derived from pGemP65N and pGemP65Nc, carrying inserts commencing as defined by primer HVPET65N, the expression plasmid derived from pGemP65N which is based on PCR products made using as the template the sequence of the viral RNA, as in plasmid pSR2A, was called pTP17; a truncated form of this plasmid, which
- 25 expresses P17 (SEQ ID No: 48), was made by cutting at the unique BglII and BamHI sites, removing the intervening fragment (which corresponds to the c-terminal part of the insert) and religating the compatible cohesive ends, to give pTP17delBB. The expression plasmids derived from plasmid pGemP65Nc (which was derived from the sequence carrying an extra C residue, were
- 30 called pT7T2-P65 (carrying an insert terminating at the primer HVP65C (SEQ ID No: 19)) and pT7T2-P70 (carrying an insert terminating at the primer HVP6C2 (SEQ ID No: 20)).



- 51 -

Expression of P6

Two forms of this protein, which arises through processing of the large capsid protein variant precursor P70 (SEQ ID No: 52) and therefore lacks its own initiation codon, were made. One form (protein MA) replaced the

5 phenylalanine at the start of this protein with methionine, giving it the amino-terminal sequence MAA...; the other carries an additional methionine residue, giving the amino-terminal sequence MFAA... The oligonucleotides used for PCR-amplified products covering the p6 coding sequence carried a NdeI site (bold) at the ATG codon, for direct ligation into the pET-11 vectors. The

10 primers used were:

HVP6MA: AATTACATATGGCGGCCGCGTTTCTGCC (SEQ ID No: 21)

HVP6MF: AATTACATATGTTTCGCGGCCGCGTTTCT (SEQ ID No: 22)

15

Each of these primers was used in conjunction with primer HVP6C2 (SEQ ID No: 20) to generate a PCR product of 0.2 kbp. These products were blunt-end ligated into vector pBSSK(-) which had been cut with EcoRV and dephosphorylated. The insert corresponding to the p6 gene was excised with

20 NdeI and BamHI (using the BamHI site in the primer HVP6C2 (SEQ ID No: 20)) and ligated into the expression vector pET-11b, which had been cut with the same enzymes. For expression at higher levels, the insert was transferred to PT7T2 as a XbaI - BamHI fragment, yielding plasmids pTP6MA and pTP6MF.

25

IPTG induction of bacteria containing plasmids pTP6MA or pTP6MF were used produce p6 for bioassay.

xiii-b) Expression of viral genes in *E. coli* and bioassay in larvae30 **Expression of P64**

IPTG induction of bacteria containing plasmid pT7T2-P65, which contains an insert running from the location of primer HVPET65N (SEQ ID No: 15) to



- 52 -

that of primer HVP65C (SEQ ID No: 19), yielded a protein of molecular weight 68 000. This was 3 000 molecular weight greater than the size of the authentic coat protein, as expected. Expression of pP65G, which contains an insert running from HVPET63N (SEQ ID No: 16) to HVP65C (SEQ ID No: 19), yielded a protein of 65 000 molecular weight.

The authentic capsid protein (P64) was expressed poorly from plasmid pT7T2-P64. Recloning this insert as a NdeI-BamHI fragment back into the other form of the vector (PT7T2b) did not alter this.

10

Expression of P70

IPTG induction of bacteria containing plasmid pT7T2-P70, which contains an insert running from the location of primer HVPET65N (SEQ ID No: 15) to that of primer HVP6C2 (SEQ ID No: 20), yielded a protein of molecular weight 73 000. This was 3 000 molecular weight larger than the size of the precursor of the coat protein, as expected.

The authentic capsid protein precursor (P71 (SEQ ID No: 50)) was expressed poorly from plasmid pT7T2-P71. Recloning this insert as a NdeI-BamHI fragment back into the other form of the vector (pT7T2b) did not alter this.

25

Due to the observation mentioned in vi) above, plasmids designed to express all forms of the capsid proteins from several possible ATG's at the start of the open reading frame were constructed.

It was found that both authentic P64 and P71 (SEQ ID No: 50) were expressed poorly in bacteria. In contrast, P17 (SEQ ID No: 48) and the forms of the capsid protein commencing at the P17 ATG were expressed very well. The extra C residue present in the latter two constructs resulted in a fusion protein being made from these expression plasmid. The sequence of the fusion proteins can be derived from Fig. 2 by including an extra C at position 570. The fusion caused the first 67 residues of the HaSV capsid protein to be

30



- 53 -

replaced by the first 95 residues of P17 (SEQ ID No: 48). Good expression of the large capsid precursor and protein was achieved, but the size of these proteins were above 3 kDa larger than the authentic forms. Notwithstanding this the expression products of the vectors containing the 5C variant of RNA 2 (SEQ ID No: 51) are still useful because the resulting product, a P70 (SEQ ID No: 52) variant, is only modified at the NH₂ terminus. Since this terminus is thought to be embedded in the capsid structure and therefore not to participate in the initial interaction with the larval midgut cell, the variant is still useful.

10

In order to produce constructs which ensure that the expressed proteins possessed the native amino terminus, new plasmids carrying the correct sequence were then cloned into the expression vector (pT7T2). It was found these plasmids to express proteins of the correct size.

15

The P6 has not yet been to expressed from the new constructs. No evidence has been found for processing of P70 to yield the mature proteins in bacteria, nor upon *in vitro* translation of synthetic full-length RNA 2 (SEQ ID No: 47).

20 The P17 (SEQ ID No: 48) gene has also been cloned into the same vectors for expression and bio-assay. This protein accumulates well in bacteria upon induction, and electron microscopy analysis has shown it form spectacular honeycomb-like structures under the bacterial cell wall, completely surrounding the cell interior (results not shown). The properties of this protein including
25 its amino acid composition and ability to form tube-like structures when expressed in bacteria suggest that it may be an homolog of a gap junction protein. The latter is involved in forming the channels linking the cytoplasm of adjacent epithelial cells in the insect gut. P17 could then play a role in enlarging or forming these channels, thereby enabling cell-to-cell movement of
30 the virus in the insect gut, analogous to the movement or spreading proteins encoded by plant RNA viruses.



- 54 -

In order to ensure that the expressed proteins carried the native amino terminus the correct sequence has also been cloned into the expression vector (pT7T2). The vector had been very slightly modified to that described above to introduce two novel restriction sites (for AflIII and BglIII) flanking the Shine-Dalgarno sequence. The resulting constructs have been found to be poor producers of the capsid proteins. The complete coding regions (which have been completely checked by re-sequencing) have therefore been recloned into the more satisfactory vectors. Results using these constructs suggest that the amino-terminus of the capsid protein presents inherent difficulties in expression. These difficulties may be imposed by either the nucleotide sequence encoding the amino terminus, or the actual amino acid sequence itself. To discriminate between these possibilities, two types of mutants were made in the sequence encoding the amino terminal 5 residues of the HaSV capsid protein. These amino-terminal mutants are as follows:

15

HVP71GJ Y

CCCATATG GGC GAT GCC GGC GTC GCG TCA CAG (SEQ ID No: 28)

Met Gly Asp Ala Gly Val Ala Ser Gln (SEQ ID No: 29)

20 HVP71SER:

CCCATATG AGC GAG GCC GGC GTC GCG TCA CAG (SEQ ID No: 30)

Met Ser Glu Ala Gly Val Ala Ser Gln (SEQ ID No: 31)

Native HaSV seq:

25 ATG GGA GAT GCT GGA GTG GCG TCA CAG (SEQ ID No: 32)

Met Gly Asp Ala Gly Val Ala Ser Gln (SEQ ID No: 33)

30



- 55 -

EXAMPLE 4

EXPRESSION IN BACULOVIRUS VECTORS AND BIOASSAY ON LARVAE

Materials and Methods

5 A(i) Cloning of HaSV capsid protein gene.

The capsid protein gene was amplified by PCR using the following primers:

5' primers:

HV17V71:

5' GGGGGATCCCGCGGATTTATGAGCGAG (SEQ ID No: 34)

10 HV17E71:

5' GGGGGATCCCGCGGAGACATGAGCGAGCACAC (SEQ ID No: 35)

HVP71:

5' GGGGGATCCAGCGACATGAGAGATGCTGGAGTGG
(SEQ ID No: 36)

15 HVV71:

5' GGGGGATCCAGCGACATGAGAGATGCTGGAGTGG
(SEQ ID No: 37)

The ATG triplets initiating P17 (SEQ ID No: 48) (in HV17V71 (SEQ ID No: 34) and HV17E71 (SEQ ID No: 35)) or P71 (SEQ ID No: 50) (in HVP71 and
20 HVV71) are underlined)

3' primers:

Primers HVP65C (SEQ ID No: 19) and HVP6C2 (SEQ ID No: 20), described
in Example 3. Results section Xiiia, were used. These constructs were made
25 using one of the four 5' primers and HVP6C2 (SEQ ID No: 20). Plasmids
constructed from PCR products made using one of the four 5' primers and
HVP65C (SEQ ID No: 19) are called 17V64 (made using 5' primer 17E71
(SEQ ID No: 35)), P64 (made using 5' primer P71 (SEQ ID No: 36)) and V64
(made using 5' primer V71 (SEQ ID No: 37)). These plasmids allow
30 expression of P64.



- 56 -

A(ii) Cloning a full length cDNA of HaSV RNA 1 (SEQ ID No: 39).

For expression of an RNA transcript corresponding to full length HaSV RNA 1 (SEQ ID No: 39), in insect cells by baculovirus infection or plasmid transfection, PCR was used to generate a fragment of cDNA linking the 5' end of RNA 1 (SEQ ID No: 39) to a Bam HI site.

The primers were:

HVR1B5'

5' GGGGGATCCCTTCTGCCTCCCCGGAC (SEQ ID No: 38)

(where the underlined nucleotide represents the start of natural RNA 1 (SEQ ID No: 39)), and an oligonucleotide complementary to nucleotides 1192=1212 of RNA 1 (SEQ ID No: 39).

The template was plasmid pSR1(E3)B described in Example 3 above.

A segment of the 1240 bp PCR fragment corresponding to the 5' 320 nucleotides of RNA 1 (SEQ ID No: 39) was excised with Bam HI and ASC II and cloned into the Bam HI site of pBSSK(-)[Stratagene] together with the 5 kbp ASCII - Bam HI fragment of pSR1(E3)B, giving plasmid pBHVR1B, which carries the complete cDNA to HaSV RNA 1 (SEQ ID No: 39), flanked by Bam HI sites.

20

A(iii) Cloning a full length CDNA of HaSV RNA 2 (SEQ ID No: 47).

For expression of an RNA transcript corresponding to full length RNA 2 (SEQ ID No: 47) in insect cells by baculovirus infection or plasmid transfection, plasmid pB+NR2B was made by inserting a fragment carrying Hind III and Bam HI sites from the multiple cloning site of vector pBSSK(-) [Stratagene] into plasmid pSR2B described above. The resulting plasmid, called pBHVR2B, carried the cDNA corresponding to full length HaSV RNA 2 (SEQ ID No: 47), flanked by Bam HI sites.

A(iv) Baculovirus transfer plasmids.

Bam HI fragments of 5.3 and 2.5 kbp corresponding to HaSV RNA's 1 and 2 (SEQ ID Nos: 39 and 47) respectively, were excised from pBHVR1B and



- 57 -

pBHVR2B respectively and inserted into the baculovirus transfer vectors described below, which had been linearised with Bam HI.

B. Baculovirus Expression of Proteins.

- 5 Baculovirus transfer vectors and engineered AcMNPV virus were transfected into *Spodoptera frugiperda* (SF9) cells as described by the supplier (Clontech) and as described in the following references:
- Vlak, J.M. & Kens, R.J.A. (1990) in 'Viral Vaccines', Wiley-Liss Inc., NY, pp.92-128; Kitts, P.A. et al (1990) Nucleic Acids Research 18: 5667-5672; Kitts, 10 P.A. and Possee, R.P. (in preparation); Possee, R.D. (1986) Virus Research, 5: 43-59.

C. Western Blotting

As in Example 1

15 D. Oligonucleotides

The following Ribozyme Oligonucleotides were produced according to standard methods.

HVR1Cla

5' CCATCGATGCCGGACTGGTATCCCAGGGGG (SEQ ID No: 5)

20

5' HVR2Cla

5' CCATCGATGCCGGACTGGTATCCCGAGGGAC (SEQ ID No: 6)

RZHDV1

- 25 5' CCATCGATGATCCAGCCTCCTCGCGGCGCCGGATGGGCA (SEQ ID No: 7)

RZHDV2

- 30 5' GCTCTAGATCCATTCGCCATCCGAAGATGCCCATCCGGC (SEQ ID No: 8)



- 58 -

RZHC1

5' CCATCGATTTATGCCGAGAAGGTAACCAGAGAAACACAC (SEQ ID
No: 9)

5 RZHC2

5' GCTCTAGACCAGGTAATATACCACAACGTGTGTTTCTCT (SEQ ID
No: 10)

Results

- 10 A series of recombinant baculoviruses has been constructed, based on the pVL941 transfer vector (PharMingen) or pBakPak8 (Clontech) and the AcMNPV. These are designed to express the correct forms of the precursor and processed HaSV capsid proteins (P64 and P71 (SEQ ID No: 50)) as well as the smaller capsid protein P6, and P17 (SEQ ID No: 48). In all systems
- 15 where replicatable RNA encoding the nucleotide sequences of the present invention are to be used, such as eukaryotic systems, in order to get efficient replication, translation or encapsidation of the RNA it is necessary to excise structures downstream of the t-RNA like structure such as the 3' extension or poly A tail on the RNA. In order to carry out such an excision, ribozymes or
- 20 other suitable mechanisms may be employed. This self cleavage activity of the ribozyme containing transcript should proceed at such a rate that most of the transcript is transported into the cytoplasm of the cell before the regeneration of a replicatable 3' end occurs. Such ribozyme systems are more fully explained in Example 7. In the results presented here highly efficient
- 25 production of P64 and P71 (SEQ ID No: 50) has been achieved. Electron microscopy and density gradient analysis have confirmed that empty particles ("capsoids") are being produced in infected cells that efficiently express the P71 precursor gene. P17 (SEQ ID No: 48) placed in the context of the *H. virescens* juvenile hormone esterase (JHE) gene (Hanzlik T.N., et al, J. Biol. Chem. 264,
- 30 12419-25 (1989)) is produced, but not in large amounts. The latter construct results in a reduction of expression of the capsid protein from the same



- 59 -

recombinant, presumably due to a reduction in the number of ribosomes reaching the AUG for the capsid gene.

SF9 cells infected with recombinant baculovirus have been shown to contain
5 large amounts of icosahedral virus particles by electron microscopy (data not shown). These particles contained no RNA, and were empty inside. This observation shows that signals on the viral RNA required for encapsidation of RNA must be located in either the 5' 270 nucleotides or the 3' 170 nucleotides, or both, since these sequences were missing from the RNA
10 transcripts made using recombinant baculovirus. Expression of HaSV proteins was confirmed by Western blotting of total protein extracts from infected insect cells.

In addition, the pAcUW31 vector (Clontech), which carries two promoters, is
15 being used to simultaneously express p6 and p64 as separate proteins. In order to bioassay the capsid protein produced in baculovirus infected cells, it is first necessary to purify it from the baculovirus expression vector. Preliminary attempts have made use of density gradients, based on the observation that empty virus particles ("assembled capsids") are in fact
20 produced in infected cells.

As outlined earlier, the HaSV genome or portion thereof is a particularly effective insecticidal agent for insertion into baculovirus vectors. Such a vector is constructed by insertion of the complete virus genome or portion thereof
25 (preferably the replicase gene) into the baculovirus genome as shown in Fig. 13. Preferably the virus genome or replicase is transcribed from a promoter active constitutively in insect cells or active at early stages upon baculovirus infection. An example of such a promoter is the heat shock promoter described in Example 7. Heat shock promoters are also activated in stressed
30 cells, for example cells stressed by baculovirus infection. An even more preferable use of such a baculovirus construct is to use the HSP promoter to drive the HaSV replicase and another gene for a toxin (as exemplified



- 60 -

elsewhere in the specification) where the RNA expressing the toxin gene is capable of being replicated by the HaSV replicase. Such recombinant baculoviruses carrying the HaSV genome or portions thereof for expression in larvae at early or other stages of the baculovirus infection cycle are particularly
5 effective biological insecticides.

EXAMPLE 5

EFFECT OF HaSV GENES AND THEIR PRODUCTS ON PLANTS

10 Materials and Methods

A. Electroporation of protoplasts.

Protoplasts of *Nicotiana tobacum*, *N. plumbaginifolia* and *Triticum aestivum* and oats were produced and electroporated with either HaSV or HaSV RNA as described in Matsunaga et al (1992) J.Gen. Virol. 73: 763-766.

15

B. Northern blot analysis - RNA extraction from protoplasts after harvest

The protoplasts are subjected to 3 cycles of freezing and thawing, and then an equal volume of 2x extraction buffer (100 mM Tris-HCl, pH 7.5, 25 mM

20 EDTA, 1% SDS, made in DEPC treated water) is added, followed by 1 volume of phenol (equilibrated in 10 mM Tris-HCl pH 8.0) heated to 65 °C.

The samples are mixed by vortexing and incubated at 65 °C for 15 min, vortexing every 5 min. After phase separation by centrifugation at room

temperature for 5 min, the aqueous phase is re-extracted with phenol, re

25 separated by centrifugation and re-extracted with chloroform/isoamyl alcohol.

To the aqueous phase are then added 0.1 volume of DEPC-treated sodium acetate (pH 5.0) and 2 volumes of ethanol. The RNA is recovered by precipitation at -70 °C, followed by centrifugation at 4 °C for 15 min. The samples were then analysed by agarose gel electrophoresis as described in

30 example 1.



- 61 -

After blotting to Zeta-Probe membrane (BioRad), the hybridization protocols were as above for Example 2.

C. Total protein from HaSV - electroporated protoplasts.

- 5 Protoplasts were analysed by SDS-polyacrylamide gel electrophoresis and Western blotting as described in Example 1.

Results

- 10 i) Use of complete (replication-competent) RNA virus genome in protoplasts

a) HaSV replication in protoplasts

The nodavirus FHV has previously been shown to replicate in barley protoplasts (Selling H.H., Allison, R. F. and Kaesberg, P. Proc. Natl. Acad. Sci. USA 87,434-8 (1990). To determine whether HaSV virus RNA can replicate
15 in plants protoplasts, when introduced by electroporation, experiments using protoplasts from *Nicotiana plumbaginifolia* and wheat have been conducted. (These are all species for which protoplasts are regularly available in the Division of Plant industry). Assays for replication including RNA (Northern) blots using probes derived from cloned fragments of cDNA to RNAs 1 and 2
20 (SEQ ID Nos: 39 and 47), and Western blots, using the antiserum to purified HaSV particles. Initial experiments showed that both HaSV virus and RNA electroporated into protoplasts of *N. plumbaginifolia* resulted in HaSV replication as studied using and verified by northern blots and ELISA. As a positive control TMV RNA was electroporated and was replication observed.

25

b) Bioassays

Protoplasts into which HaSV RNA had been introduced by electroporation were harvested after 6 or 7 days post electroporation and used in bioassays on neonate larvae by addition to normal diet. The results showed significant
30 stunting of test larvae in comparison to control larvae (see Table 1 below). Protoplasts lacking HaSV RNAs had no effect on the larvae, confirming the result of control experiments. This result confirms that HaSV RNA, when



- 62 -

expressed or replicated in plant cells, is able to cause the formation of infectious virus particles able to control insect larvae feeding on the plant material.

- 5 Northern blotting has been used to confirm that RNA electroporation into protoplasts leads to RNA replication.

Table 1: Results of Bioassay from a typical experiment with Nicotiana and oat protoplasts (oat results are shown in brackets) [see over]

10

	Treatment	Number	Escapes	Number stunted
	1. diet only	12 (12)	2 (3)	0/10 (0/9)
	2. diet + protoplasts	12 (12)	0 (1)	0/12 (0/11)
	3. HaSV + diet	12 (12)	0 (1)	12/12 (11/11)
15	4. diet + HaSV/protoplasts	12 (n.d.)	0 (n.d.)	12/12 (n.d.)
	5. diet + RNA/protoplasts	12 (12)	0 (0)	11/12 (10*/12)
	* HaSV replication in the larvae was confirmed except for two larvae which were dead. The letters "n.d." mean the experiment was not done.			

- 20 The above results demonstrate assembly of HaSV particles from electroporated RNA in protoplasts of both monocot and dicot plant species.

c) Plasmids to test replication of cloned and engineered forms of HaSV

- (1) Plasmids allowing in vitro transcription of HaSV RNAs 1 and 2 (SEQ ID Nos: 39 and 47) for electroporation into protoplasts have already been described above.
- (2) Plasmids for transient expression of individual HaSV RNAs (1 or 2) (SEQ ID Nos: 39 and 47) in protoplasts. Full-length cDNAs for the two viral RNAs have been inserted into expression plasmids pDH51 (with the CaMV 35 S promoter. Pietrzak M., et al (1986) Nucl. Acids Res. 14, 5857-68) for dicots and pActI.cas (with the rice actin promoter) for monocots (McElroy et al (1990) The Plant Cell 2: 163-171). As with the vectors for expression in insect



- 63 -

cells, these expression plasmids are being modified to include a cis-acting ribozyme for generation of authentic ends. The non-ribozyme plasmids gave no virus replication.

ii) Expression of capsid protein in plants

- 5 In view of the present inventors' observation that empty particles ("assembled capsids") are being produced in baculovirus-infected cells that efficiently express the P71 precursor gene, expression of the coding region for the capsid protein in tobacco plants was investigated. The vector chosen for this purpose is based on pDH51 which carries the CaMV 35S promoter and polyadenylation
- 10 signal. If necessary for improved expression, this vector can be modified by the addition of a translation enhancer sequence from e.g. TMV. Although certain groups have constructed transgenic plants expressing the capsid proteins of plant viruses, there has been only one recent report of assembly of empty capsids in such plants (Bertioli et al.,(1991) J. gen. Virol. 72: 1801-9).
- 15 Bertioli et al point out that the protein-protein interactions in most icosohedral plant RNA viruses may be too weak to allow assembly of such capsids. In addition to the present inventors' observation of empty HaSV capsids, it has been found these capsids are very tough, showing great resilience to e.g. repeated cycles of freezing and thawing, so that it is expected to see assembly
- 20 of empty HaSV capsids ("assembled capsids") in transgenic plants.

EXAMPLE 6

IDENTIFICATION OF MIDGUT BINDING DOMAINS

Materials & Methods

25 A. Plasmid construction

Was as described in Examples 3 and 4.

B. Western blotting

Was as described in Examples 1 and 3.

30

C. Invitro translation



- 64 -

In vitro transcripts of cloned CDNA of HaSV RNA's was translated in vitro as in Examples 1 and 3.

D. Preparation of Brush Border Membrane Vesicles.

- 5 Brush Border Membrane Vesicles were prepared from freshly isolated larvae midguts of *H. Armigera* by the method of M. Wolfersberger et al (1987) Comp. Biochem. Physiol. 86A: 301-308, as modified by S.F. Garczynski et.al. (1991) Applied Environ. Micro-biol 57: 1816-2820. Brush Border Membrane Vesicles binding assays using invitro labelled protein or ¹²⁵I-labelled protein were as
10 described in Garczynski et.al. (1991) or in H.M. Horton and Burand, J.P. (1993) J. Virol. 67: 1860-1868.

Results

i) Determination of epitopes on the capsid surface

- 15 Comparison of the recently published sequence of the *Nudaurelia* ω virus (*N ω V*) capsid protein with that of HaSV shown that these proteins are closely related and fall into four distinct domains, which are alternatively variable and highly conserved. These domains are summarised as follows:

20 Residues:	HaSV 1-49	50-272	273-435	437-647
	<i>NωV</i> : 1-46	47-269	270-430	431-645
% identity:	37	81	34	81

- Comparison of this observation with the alignment by Agrawal and Johnson
25 (1992) between the *N ω V* and the nodavirus BBV (whose crystal structure is known: Hosur et al (1987) Proteins: Structure, Function & Genetics 2: 167-176) showed that the variable region coincided with a region forming the most prominent surface protrusion on the BBV capsid. Both HaSV and *N ω V* carry large insertions at this point relative to BBV, and these insertions are largely
30 different in sequence. Assuming that the alignment by Agrawal and Johnson (1992) is correct, then this means that HaSV and *N ω V* have a more prominent pyramid-like structures as a surface protrusion than do the nodaviruses, and



- 65 -

the pyramid-like structures are different. As already noted, there is no immunological cross-reactivity between the two viruses, despite the high degree of identity. There is thus a strong implication of the variable domain as a surface protrusion which functions as the sole antigenic region.

5

To confirm this a 400 bp NarI fragment spanning the variable region was deleted from the capsid gene in the expression vector. With end-filling of these sites the deletion is in-frame, so that a truncated protein of ca. 57 KDa is produced in bacteria upon induction. This protein was recognized only poorly
10 on Western blots by the antiserum against intact HaSV particles made in rabbits. The central variable domain was recognized well by the antiserum when expressed in isolation from the rest of the capsid gene.

As shown in the table above the region of HaSV capsid protein comprising
15 residues 273-439 shows great divergence from the corresponding region of the N ω V capsid protein, compared to its immediate flanking regions. Within this region an especially divergent domain is found from residue 351 to residue 411, which shows only 25% identity to the corresponding region of the N ω V capsid protein. This region is flanked by the sequences corresponding to the
20 β -sheet structural features β -E(residues 339-349) and β -F(residues 424-431) of the HaSV capsid protein, based on the alignment the N ω V and nodavirus capsid proteins by Agrawal and Johnson (1992), and is therefore likely to form the loop of the most prominent surface protrusion on the HaSV capsid. This is based on comparison and correspondence to the nodavirus capsid protein
25 structure and capsid structure as described by Wery J.-P. and Johnson, J.E. (1989) Analytical Chemistry 61, 1341A-1350A and Kaesberg, P., et al. (1990) J. Mol. Biol. 214, 423-435. This loop is thought to contain important epitopes. It is significant that this exterior loop on the nodavirus capsid protein is one of the most variable regions when capsid proteins sequences from a number of
30 nodaviruses are compared (Kaesberg et al. 1990).



- 66 -

Finally, the present inventors have observed a significant level of immunological cross-reaction on Western blots, between antisera against the CryIA(c) Bt toxin and HaSV capsid protein, whether obtained from virus or expressed in bacteria. Initial data from the NarI deletion mutant described above suggest that this binding is not to the central variable domain, but to other regions of the capsid protein. The only other region of the proteins which shows extensive sequence variability, the amino terminus, cannot be responsible for the binding, since both authentic capsid protein and the protein with an altered amino terminus expressed in bacteria are recognized by the anti Bt antisera.

ii) In-Vitro binding assays

The full-length clones for *in vitro* translation yielding highly ^{35}S or ^3H labelled proteins were constructed by replacing the bacterial translation interaction signal in the T7 plasmids above by the more active eucaryotic context sequence from the JHE gene. The labelled capsid protein made by *in vitro* translation of the *in vitro* transcripts may be tested for binding to brush border membrane vesicles (BBMV's). Conditions are optimised by testing different procedures. The deletion mutant lacking approximately 125 amino acids in the central region, and containing the variable domain, as well as others derived from it are also tested.

iii) Fusion proteins comprising virus capsid midgut binding domains and other proteins

The idea behind these tests is to fuse the binding domain from the HaSV capsid protein to either large proteins (preferably indigestible, causing protein to aggregate in or on the midgut cells) or toxin domains from other proteins with suitable properties but normally different binding specificities (e.g. Bt). In initial experiments, the gene for the complete capsid protein has been fused to the GUS gene, as has a deletion mutant containing essentially only the central portion of the capsid gene. The resulting fusion proteins are being expressed



- 67 -

in bacteria and tested for GUS activity, and makes them sensitive probes for binding experiments on midgut tissue.

iv) Mapping binding sites using Bt/HaSV fusion proteins

- 5 Analysis of deletion mutants of the CryIA(c) Bt toxin has identified domains which may be involved in determining the host-specificity of this Bt by acting as receptor-binding sites (Schnepf et al (1990) J. Biol. Chem. 265: 20923-20930; Li et al (1991), Nature 353: 815-21. The present inventors have obtained a clone of this toxin gene. Deletion mutants corresponding to those identified by
- 10 Schnepf et al are constructed. Segments of the HaSV capsid protein gene can then be inserted into these mutants, the protein expressed in bacteria and their insecticidal function assayed.

EXAMPLE 7

15 VIRAL GROWTH IN CELL CULTURE

Materials & Methods

A. Cell Lines

The following cultured insect cell lines were tested for infection by HaSV:

- 20 *Drosophila melanogaster*, *Helicoverpa armigera* (ovarian derived), *Heliothis zea* (ovarian derived), *Plutella xylostella*, *Spodoptera frugiperda* (SF9).

- All lines were grown under standard conditions. Upon reaching confluence, the culture medium was removed and all mono-layers covered with 1.5 ml of cell culture medium into which HaSV had been diluted; the average multiplicity of infection (M.O.I.) was 10^4 . After adsorption at 26 °C for 2h, the
- 25 inoculum was removed, the cells carefully washed twice with phosphate buffered saline (pH 7.0) and incubation continued with 5 ml of 10% Foetal calf serum in TC199 culture medium (Cyto Systems).

B. Northern Blotting Analysis.

- 30 Virus replication in all the above cell lines was confirmed by northern blotting analysis. Total RNA was extracted from infected cells by the method of Chomczynski and Sacchi (1987). Anal. Biochem. 162: 156-159. The cells were



AMENDED SHEET
IPEA/AU

- 68 -

lysed in 1 ml of lysis solution (4M guanidinium thiocyanate, 25mM sodium citrate, pH 7, 0.5% sarcosyl, 0.1M 2-mercaptoethanol). In order, 0.1 ml of 2M sodium acetate, pH 4, 1 ml of phenol (0.2M sodium acetate equilibrated), and 0.2 ml of chloroform-isoamyl alcohol mixture (49:1) were added with thorough
5 mixing between reagents. This was then vortexed for 10 s and cooled on ice for 15 min. Tubes were centrifuged in an Eppendorf centrifuge at 14k for 15 min at 4 °C for at least 15 min to allow RNA precipitation. RNA was pelleted by centrifugation at 14k for 15 min, washed with 0.6 ml of ice-cold 70% ethanol, pelleted once again (10K, 10 min), air dried at room temperature and
10 resuspended in DEPC (Sigma) treated millipore water. RNA was subject to denaturing agarose gel electrophoresis in the presence of formaldehyde according to Sambrook et.al. (1989). The gel was Northern transferred to a zeta-probe membrane (Biorad) as described by Sambrook et.al. (1989). The probe was prepared by random-priming the 3' sequences of the HaSV genome
15 using DNA and cDNA clones pSHVR15GB and pT7T2p71SR-1 as per manufacturer's instructions (Boehringer-Mannheim). Hybridization was carried out as described for the standard DNA probe protocol contained within the literature for the zeta-probe membrane (Biorad).

20 C. Vectors

Vectors as described below.

Results

It has been found that HaSV will replicate in several continuous cell lines, of
25 which the best is the *Spodoptera frugiperda* line SF9. Time course assays by Northern blotting in SF9 cells have shown that RNA 1 (SEQ ID No: 39) replication is clearly detectable within a few hours of infection. RNA 2 (SEQ ID No: 47) is present only in very small amounts early in infection and accumulates much more slowly than RNA 1 (SEQ ID No: 39) does. This
30 observation is consistent with one made earlier in HaSV-infected larvae, where RNA 2 (SEQ ID No: 47) replication was not observed until 3 days after infection.



- 69 -

Some apparent replication was also observed in *Drosophila* cells (DL2), but with the difference that more RNA 2 (SEQ ID No: 47) replication was observed at the early time points compared to the lepidopteran cell lines above.

5

Plasmids that express the HaSV genome as RNA transcripts from full length cDNA clones have been constructed and tested. These clones, constructed by PCR and carefully checked, have restriction sites immediately adjacent to the ends of the sequence. Transcription is driven from a specially-re-engineered *Drosophila* HSP70 promoter.

10

i) Constructs for expression in insect cells

The constructs are based on vectors carrying the *Drosophila* HSP or actin promoters and suitable polyadenylation signals from *Drosophila* (Corces & Pellicer (1984) J. Biol. Chem. 259: 14812-14817) or SV40 (Angelichio et al (1991) Nucl. Acids. Res. 18: 5037-5043). Since transcription from such plasmids generates viral RNAs carrying long 3' terminal extensions derived from sequences in the polyadenylation signal fragment, it is necessary to achieve cleavage of the transcript immediately after the 3' sequence of the viral RNA. These plasmids gave no virus replication, presumably because of the 3' terminal extension. The method of choice for obtaining authentic 3' termini is based on introduction of DNA sequences encoding a cis-acting ribozyme into the constructs. With suitable engineering, such a ribozyme will cleave immediately 3' to the viral sequences within the transcript. Suitable ribozymes, based on the hepatitis delta virus (Been M.D., Perrotta, A. T. & Rosenstein, S.P. Biochemistry 31, 11843-52 (1992) or the hairpin cassette ribozyme (Altschuler, M., Tritz R. & Hampel, A. Gene 122, 85-90 (1992) have been designed. This involves synthesis of overlapping oligonucleotides, which are then annealed and end-filled with the Klenow fragment of DNA polymerase, to create short DNA fragments encoding the desired ribozyme. These fragments carry restriction sites at their termini allowing them to be ligated

20

25

30



- 70 -

into plasmids between the viral RNA cDNA (which has a 3' restriction site added by PCR) and the restriction fragment carrying the poladenylation signal.

EXAMPLE 8

5 SHEDDING OF INFECTED CELLS

Materials & Methods

A. Confocal Laser Scanning Microscopy. (CLSM)

CLSM enables the visualisation and analysis of three-dimensional cell and tissue structures at the macro and molecular levels. The Leica CLSM used in
10 this example is based on an MC 68020/68881 VME bus (20MHz) with standard 2Mbyte framestore and 4Mbyte RAM and OS9 operating system with programmes written in C code. It incorporates a Leica Diaplan research microscope and using X10/0.45, X25/0.75, X40/1.30 and X63/1.30 Fluotar objectives has a claimed optical efficiency better than 90%. The confocal
15 pinhole is software controlled over the range of 20 to 200 μ m. Excitation at 488 and 514 nm is provided by a 2 to 50 mW argon-ion laser.

B. Immunocytochemistry (ICC).

For whole mount ICC, tissues were dissected under saline and fixed in fresh 4% formaldehyde in phosphate buffered saline (PBS) for at least 15 mins.
20 After multiple washes in PBS they were permeabilized either by 60 mins incubation in PBT (PLBS with 0.1% Triton X-100 plus 0.2% bovine serum albumin). After 30 mins blocking in PBT+N (5% normal goat serum) tissue was incubated in primary antibody diluted (1:40) in PBT+N for at least 2 hrs at room temperature then at 4°C overnight. After extensive washing in PBT
25 and 30 mins blocking in PBT+N the FITC conjugated secondary antibody diluted (1:60) in PBT+N was incubated for 2 hrs at room temperature plus overnight at 4°C. After multiple washes in PBT and PBS the tissue was cleared in 70% glycerol and mounted in 0.01%w/v p-phenylenediamine (Sigma#P1519) dissolved in 70% glycerol. All processing was at room
30 temperature unless otherwise stated.



- 71 -

Results

The inventors' current model for the effect of HaSV involves the detection by the insect midgut of infected cells, their identification as infected and their subsequent shedding in numbers sufficient to cause irreparable damage to the insect midgut. The evidence for this is based on the above and on the following direct observation of the fate of infected cells in midgut tissue over 1-3 days post infection. These results in repeat experiments were complicated by the discovery that another unrelated virus was present in the larval population being tested. Preliminary findings indicated that HaSV infection activates or facilitates pathogenesis of the unrelated virus and together these cause severe disruption of the larval gut cells. Thus these two agents appear to act synergistically in causing gut cell disruption.

Midguts from larvae infected with HaSV were treated with the antiserum to purified HaSV particles (above) and examined under the Laser confocal microscope (described above). This established that some midgut cells were sufficiently infected with HaSV to give strong fluorescence signals. Such cells were moreover clearly separating from the surrounding tissue, a sign that they were in the process of being shed.

Similar observation have been made with other insect viruses (Flipsen et al (1992) Society for Invertebrate Pathology Abstract #96) although in these cases the effect is too localised and weak to cause any anti-feeding effect apparently only the small RNA virus of the tetraviridae which are localised to the gut and cause more-or-less severe anti-feeding effects in their hosts (Moore, N.F. in Kurstak E. (Ed) (1991) Viruses of Invertebrates. Marcel Dekker, New York pp277-285) are capable of such an effect to an extent sufficient for pest control.

Following on from the immune-fluorescence work, *in situ* hybridization can be carried out to detect RNA replication in infected cells. Furthermore, larvae infected with a recombinant HaSV expressing a foreign gene at early stages



- 72 -

(by insertion of that gene into RNA 1 in place of the N-terminal portion of the replicase gene) can be studied. A correlation between virus replication and cell rejection can be confirmed by histochemical analysis of the midgut cells of the infected larvae. Thus the cell-shedding phenomenon offers a direct and rapid

5 assay for early events in HaSV-infected gut tissue. Extracts of baculo-vector infected insect cells carrying empty HaSV particles can be fed to larvae directly and the midgut examined by toluidine blue staining and immune-fluorescence at intervals after infection. This will allow direct determination of whether the particles can bind the brush border membranes in intact gut, and

10 whether such binding can induce the massive disruption evident in normally infected larvae. Control experiments using extracts from cells infected with the baculovector alone can be conducted to observe and distinguish effects due to the vector. The immune-fluorescence assay on midgut tissue allows analysis of binding to midgut brushborder membranes. Once determined for wild-type

15 capsid protein expressed from a baculo-vector, deletion or replacement mutants can be inserted into the baculovectors. Suitable cell extracts from these can be used to infect larvae.

20

EXAMPLE 9

ENGINEERED VIRUS AND USES

Materials & Methods

(as indicated in earlier Examples)

25 i) Engineered virus as a vector for other toxin genes

This involves placing suitable genes under control of HaSV replication and encapsidation signals. Genes which may be suitable include intracellular insect toxins such as ricin, neurotoxins, gelonin and diphtheria toxins. The toxin gene may be placed in the viral gene such that it is a silent (downstream) cistron on

30 a polycistronic RNA, or in a minus strand orientation, requiring replication by the viral polymerase to be expressed. Standard techniques in molecular biology can be used to engineer these vectors.



- 73 -

A discussion of two recombinant HaSV vectors which have been designed is given below:

for RNA 1 (SEQ ID No: 39):

The reporter gene (or one of the toxin genes mentioned above) is inserted in place of the amino-terminal portion of the putative replicase gene, such that the initiation codon used for the replicase (ie that at nucleotides 37-39 of the sequence) is now used to commence reporter gene translation. The fusion is achieved by the use of artificial NcoI restriction sites common to both sequences.

10 The short 36 nucleotide 5'-untranslated leader of RNA 1 (SEQ ID No: 39) (shown in upper case) is synthesised as the following sequence:
gggatccacaGTTCTGCCTCCCCCGGACGGTAAATATAGGGGAACCATG
Gtctagagg, (SEQ ID No: 53)

using two overlapping oligonucleotides comprising the first 40 nucleotides and the complement of the last 40 nucleotides respectively. These primers are annealed and end-filled by Klenow. The resulting fragment is then cut with BamHI and XbaI (sites underlined) and cloned with plasmid vector pBSIISK(-).

The GUS gene carrying a NcoI site at the ATG codon was obtained as a NcoI-SacI fragment from plasmid pRAJ275 (Jefferson, RAJ Plant Mol. Biol. Rep 5, 3387-405 (1987)). This SacI site is located just downstream from the coding sequence for the GUS gene.

The 5' leader of RNA is excised as a BamHI-NcoI fragment from the above vector, and is ligated together with the NcoI-SacI fragment carrying the GUS gene into plasmid pHSPRIRZ or pDHVRIRZ carrying the full-length cDNA insert of RNA 1 (see above) which has been cut with BamHI and SacI. The resulting plasmid then carries a complete form of RNA 1 (SEQ ID No: 39) but with the amino-terminal portion of the replicase gene substituted by the GUS gene. It is desirable to produce a construct with approximately the same size as RNA 1 (SEQ ID No: 39) for encapsidation purposes.



- 74 -

Similar approaches are adopted for RNA 2 (SEQ ID No: 47), with the foreign, reporter or toxin gene fused to the initiation codon of either P17 or P71. In either case the context sequence of the introduced gene is modified to give the necessary expression level of that protein. The foreign gene is introduced into
5 plasmids pHSPR2RZ or pDHVR2RZ.

The above recombinants have been described specifically as insertions of a reporter gene (GUS). The toxin genes to be inserted are described on page 14 of the specification. These preferably further require a signal peptide
10 sequence added at the amino-terminus of the protein.

ii) Capsid technology

Identification of encapsidation (and replication) signals on virus RNA allows
15 design of RNAs which can be encapsidated in HaSV particles during assembly of virus in a suitable production system. The virus capsids then carry the RNA of choice into the insects midgut cells where the RNA can perform its intended function. Examples of RNAs which may be encapsidated in this manner include RNAs for specific toxins such as intracellular toxins, such as
20 ricin, gelonin, diphtheria toxins or neurotoxins. This strategy is based on the resistance of the virus particle to the harsh gut environment.

iii) Other uses of the capsid particle

The capsid particles can be used as vectors for protein toxins. Knowledge of
25 icosahedral particle structure elucidated by the inventors suggests that the amino and especially the C-termini are present within the capsid interior. It is possible to replace or modify the amino acid sequence corresponding to P7 such that it encodes a suitable protein toxin which is cleaved off the bulk of the capsid protein during capsid maturation. As with toxin-encoding mRNAs,
30 the HaSV capsid delivers it to the midgut cell of the feeding insect, where it exerts the desired toxic effect.



- 75 -

iv) Use of HaSV in plants

The use of HaSV in the production of insect-resistant transgenic plants are shown in Fig. 12. These inventions are based on the use of either the complete HaSV genome, or of the replicase gene as a tool for the

5 amplification of suitable amplifiable mRNAs (e.g. encoding toxin) or of the capsid protein as a means to deliver insecticidal agents. These strategies are now described in some detail.

a) Use of the complete HaSV genome

10 Fragments of cDNA corresponding to the full-length HaSV genome components RNAs 1 and 2 (SEQ ID Nos: 39 and 47) are placed in a suitable vector for plant transformation under the control of either a constitutive plant promoter (e.g. the CaMV 35S promoter mentioned above) or an inducible promoter or a tissue specific (e.g. leaf-specific) promoter. The cDNAs are

15 followed by a cis-cleaving ribozyme and a suitable plant polyadenylation signal. Transcription and translation of these genes in transgenic plant tissues and cells leads to assembly of fully infectious virus particles to infect and kill feeding larvae.

20 The following experiments were conducted. The plasmids for expression used the CaMV 35S promoter to generate transcripts commencing at the first nucleotide of the HaSV RNAs 1 and 2 (SEQ ID Nos: 39 and 47). The vector pDH51 (M. Pietrzak, R. Shilito, T. Hohn and I. Potrykus (1986). Nucleic Acids Research 14, 5857) which carries the CaMV 35S promoter followed by a

25 multiple cloning site and the CaMV polyadenylation fragment was modified to make a suitable vector, pDH51Stu, carrying a StuI site at the immediate 3' end of the CaMV 35S promoter. The promoter thereby terminates in the sequence GAGAGGCCT, with the underlined residue being that at which transcription would start. (Similar vectors have been described by Mori *et al*, J. General

30 Virology 72, 243-246 (1991) and Dessens and Lomonosoff, *ibid* 74, 889-892 (1993).) The StuI site (AGG/CCT) is followed by a BamHI site (GGATCC). Cleavage of this vector with StuI and BamHI generates a vector DNA



- 75/1 -

molecule with one blunt end (from *Stu*I cleavage) and one sticky *Bam*HI end. This allows ligation of cDNA molecules corresponding to the full-length HaSV genomic RNAs, and carrying a blunt end at the 5' end of the full-length cDNA and a *Bam*HI site after the 3'-end of the full-length cDNA.

5

Suitable cDNA fragments carrying a blunt end corresponding to the 5'-terminal nucleotide of either RNA 1 or 2 (SEQ ID Nos: 39 and 47) were generated using PCR and an oligonucleotide primer corresponding to the 5'-terminal first 18 nucleotides of the sequence of either RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47). The cDNA sequence corresponding to the 3' terminal sequences of either RNA 1 (SEQ ID No. 39) or RNA 2 (SEQ ID No 47) were followed on these DNA fragments by sequences corresponding to one of the ribozymes whose sequences are shown in Fig. 8 and whose construction is described in Example 7. The 3'-terminal sequence corresponding to an *Xba*I site (TCTAGA) shown in these ribozyme sequences was followed on the suitable DNA fragments by a *Bam*HI site, which upon cleavage with this enzyme yielded a sticky end capable of being ligated into the *Bam*HI end of the vector cleaved as described above. There were therefore a total of four suitable DNA fragments for insertion into the vector:

20

RNA 1 (SEQ ID No: 39) followed by the hairpin cassette (HC) ribozyme
RNA 1 (SEQ ID No: 39) followed by the hepatitis delta virus (HDV) ribozyme

RNA 2 (SEQ ID No: 47) followed by the hairpin (HC) ribozyme

25 RNA 2 (SEQ ID No: 47) followed by the hepatitis delta virus (HDV) ribozyme.

These four fragments were individually ligated into the vector pDH51*Stu*I cleaved with *Stu*I and *Bam*HI to generate four distinct plasmids as follows:

pDH*Stu*R1HC

30 pDH*Stu*R1HDV

pDH*Stu*R2HC

pDH*Stu*R2HDV



- 75/2 -

Transcription from the 35S promoter in these plasmids results in RNAs commencing at the first nucleotide of either the RNA 1 sequence (SEQ ID No: 39) or RNA 2 sequence (SEQ ID No: 47) and terminating in the CaMV polyadenylation fragment. Self-cleavage at the locations shown in Fig. 8 by the

5 cis-acting ribozymes obtained within these transcripts generates RNA molecules with the 3'-termini corresponding to the natural virus termini.

After amplification and purification on CsCl gradients, thirty µg of each of these four plasmids was transfected by electroporation into aliquots of two

10 million *N. plumbaginifolia* protoplasts (as described in Example 5) either individually or in the combinations listed below:

pDHStuR1HC + pDHStuR2HC

pDHStuR1HDV + pDHStuR2HDV

15 The production of infectious HaSV particles within transfected protoplasts was then demonstrated by bioassay on heliothis larvae. After incubation at 25 °C for 3-5 days, the protoplasts were recovered by low speed centrifugation and applied directly to standard heliothis diet as surface contamination for bioassay as described in Example 1. Stunting was only observed when plasmids

20 expressing HaSV RNA 1 (SEQ ID No: 39) and RNA 2 (SEQ ID No: 47) were co-transfected, and then only in the case of those carrying the hairpin ribozyme to generate the viral 3' ends (see Table 2). In contrast, constructs carrying the HDV ribozyme at the 3' end were not infectious. The reasons for this have not been determined. As expected, expression of RNA 1 or 2 (SEQ ID Nos:

25 39 and 47) alone in protoplasts did not lead to the assembly of infectious particles. Western blot analysis of protoplasts transfected with the RNA 2 (SEQ ID No: 47) constructs did show production of limited amounts of the capsid protein.

30 Suitable control experiments confirmed that larval stunting was due to HaSV particles generated *de novo* in the protoplasts. As shown in the Table 2,



-75/3 -

neither the protoplasts alone nor protoplasts mixed with plasmid DNA were capable of initiating stunting.

Table 2

5	Treatment	No. of larvae escapes	No. stunted
	1. diet alone	24	0
	2. diet + HaSV	24	24
	3. diet + protoplasts	24	0
10	4. diet + pDHStuR1HC	24	0
	5. diet + pDHStuR1HDV	24	0
	6. diet + pDHStuR2HC	24	0
	7. diet + pDHStuR2HDV	24	0
	8. diet + pDHStuR1HC		
15	+ pDHStuR2HC	24	22
	9. diet + pDHStuR1HDV		
	+ pDHStuR2HDV	24	0
	10. pDHStuR1HC + pDHStuR2HC (but mixed with protoplasts)	24	0

20

HaSV infection of stunted larvae was confirmed by dot-blotting of RNA using HaSV specific probes. After weighing, larva were sacrificed and total RNA extracted as follows. Each larva was homogenised in the presence of 260 µl deionised water, 24 µl 2M sodium acetate pH 4.0 and 200 µl phenol

25 equilibrated with 2M sodium acetate pH 4.0. After centrifugation at 14 000 rpm for 15 min at 4°C, the supernatant (about 200 µl) was removed and extracted once with an equal volume of chloroform. After centrifugation, the supernatant (about 200 µl) was mixed with 20 µl of sodium acetate and 400 µl of absolute ethanol. The precipitate after centrifugation was vacuum dried and
30 redissolved in 5-10 µl of sterile, DEPC-treated water. For dot-blotting, the RNA was mixed with 70 µl of DEPC-treated water and 30 µl of 10 mM EDTA, 30 mM NaOH. HaSV RNA was determined and quantified by dot



- 75/4 -

- blotting (as described in Example 2) using a probe random primed DNA from clones corresponding to the terminal 1000 nucleotides of RNA 1 and 2. All larvae recorded as stunted in the bioassays were found to carry HaSV and give signals comparable to those of the larvae fed purified HaSV particles (Table
- 5 2). To confirm that the larvae were infected with HaSV, ten aliquots of protoplasts were electroporated with plasmids pDHStuR1HC + pDHStuR2HC and the protoplasts fed (after incubation) to 150 heliothis larvae. The larvae were allowed to grow for one week, upon which significant stunting was observed in 50% of the larvae, and virus was then purified from these stunted
- 10 larvae as described in Example 1. Analysis on CsCl gradients showed the production of distinct bands characteristic of HaSV; analysis of the virus particles found in these bands by SDS-PAGE and Western blotting with HaSV antiserum confirmed their identity as authentic HaSV.
- 15 These results have therefore demonstrated that DNA plasmids capable of expressing the HaSV genome in plant cells have been constructed. Once introduced into the cells, the plasmids are transcribed to yield HaSV genomic RNA which can drive the assembly of particles able to infect heliothis larvae by the normal oral route. These constructs have therefore been shown to be
- 20 suitable for use in engineering transgenic plants expressing HaSV.

- A variation on this strategy is to remove from the cDNA for RNA 2 (SEQ ID No: 47) the fragments encoding RNA encapsidation and/or replication signals. This results in either the assembly in the plant cells of
- 25 HaSV particles carrying only RNA 1 (SEQ ID No: 39), or of HaSV particles carrying RNA 1 (SEQ ID No: 39) and a form of RNA 2 (SEQ ID No: 47) which cannot be replicated in the infected insect cell.

- A further variation on this strategy is to modify the plant transgene
- 30 encoding RNA 2 (SEQ ID No: 47) so that it is still replicatable and encapsidatable, but no longer express functional capsid protein. HaSV capsids made in such plant cells will then be capable of making both the replicase and



- 75/5 -

P17 (SEQ ID No: 48) in infected insect cells, but not of assembling progeny virus particles therein (such as shown in Fig. 12(d)). These measures confer



- 76 -

inherent biological safety in the form of containment on the use of such transgenic plant material.

(b) Use of portions of HaSV genome to deliver toxins to insect cells

- 5 This approach makes use of any of the systems described in (a) above. Plant cells contain an additional transgene encoding a suitable insect-specific, intracellular toxin (as described above). Such a toxin gene is expressed by plant RNA polymerase in either a positive or a negative sense (the latter is preferred) and in such a form that the RNA can be encapsidated by HaSV
- 10 capsid protein and/or replicated by the HaSV replicase in infected insect cells (see Figs. 12a and 12b)

- Transgenic plants would contain two different transgenes, making either unmodified capsid protein precursor or a modified form in which most of the
- 15 carboxyterminal protein P7 is replaced by a suitable insect-specific toxin or one which is inactive as part of a fusion protein. (Gelonin or other ribosome-inactivating proteins, insect gut toxins or neurotoxins may be suitable here.) Expression from these two transgenes would be regulated so that only the required amounts of the modified and unmodified forms are made in the plant
- 20 cell, and assembled in such proportions into the capsoids. One way to modulate the production of capsotxin fusion proteins is to make translation of the carboxyterminal toxin reading frame dependent on a translational frameshift or read-through of a termination codon. With an appropriate low frequency of frame-shifting (eg 0.1 - 2%), it could even be sufficient to use a
- 25 single transgene, if it were possible to synthesise the P7 portion and the toxin portion as overlapping genes. Upon assembly (which we have demonstrated in insect cells using the baculovirus vectors) and maturation, the protein precursors are cleaved and release the mature P7 and the toxin, which remain within the capsoids. These proteins are not released until capsoid disassembly
- 30 occurs in insect gut cells. The processed form of the toxin is then able to kill the pest.



- 77 -

(c) HaSV particles devoid of nucleic acid carrying one or more suitable protein toxins and/or their mRNA

A protein toxin (or toxins) is expressed as a fusion with the capsid protein. The fusion protein then assembles into capsid carrying the toxin(s).

- 5 These capsids present in the plant tissue exert an antifeeding effect on insects attaching the plant.

EXAMPLE 10

EXPRESSION OF Hasv IN OTHER DELIVERY VECTORS

10 Materials & Methods

(as indicated in earlier Examples)

- Constructs similar to those for plant expression are introduced into yeast or bacteria by standard techniques. Virus particles are assembled for either fully infectious virus or any of the modified or biologically contained forms
- 15 described in Example 9. Microbes produced in suitable fermentation or culture facilities and carrying such forms of the virus are then delivered to the crop by spraying. The microbial cell wall provides extra protection for the virus particles produced within the microbe.
- 20 Well established techniques exist for culture and transformation of yeast (Ausubel, F.M. *et al.* (eds) Current Protocols in Molecular Biology. J. Wiley & Sons, NY, 1989). An example of a yeast expression vector is pBM272, which contains the URA3 selectable marker (Johnston, M. & Davies, R.W. Mol. Cell. Biol. 4, 1440-8, (1984); Stone, D. & Craig, E. Mol. Cell. Biol. 10, 1622-32
- 25 (1990). Another example of an expression vector is pRJ28, carrying the Trp1 and Leu2 selectable markers.

- Yeast has recently been shown to support replication of RNA replicons derived from a plant RNA virus, brome mosaic virus (Janda, M. & Ahlquist, P. Cell 72, 961-70 (1993). Since the BMV replicase is distantly related to that of
- 30 HaSV, and the two viruses are likely to replicate by similar strategies within



- 78 -

cells, yeast cells probably contain all the cellular factors required for HaS \checkmark to generate infectious virus.

For bacteria, suitable expression vectors have been described above.



AMENDED SHEET
IPEA/AU

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Commonwealth Scientific and Industrial Research Organisation and Pacific Seeds Pty. Ltd.
- (ia) INVENTORS: P. D. CHRISTIAN, K. H. J. GORDON and T. N. HANZLIK
- (ii) TITLE OF INVENTION: INSECT VIRUSES AND THEIR USES IN PROTECTING PLANTS
- (iii) NUMBER OF SEQUENCES: 53
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: DAVIES COLLISON CAVE
 - (B) STREET: 1 LITTLE COLLINS STREET
 - (C) CITY: MELBOURNE
 - (D) STATE: VICTORIA
 - (E) COUNTRY: AUSTRALIA
 - (F) ZIP: 3000
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE: 13 AUGUST 1993
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: JOHN M. SLATTERY
 - (B) REGISTRATION NUMBER: NA
 - (C) REFERENCE/DOCKET NUMBER: 1613611
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (613) 254 2777



80

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GGATCCACAG NNN

13

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGGGCGATG CCGGCGTCGC GTTCACAG

28

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGGAGGATG CTGGAGTGGC GTCACAG

27

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGAGCGAGG CCGGCGTCGC GTCACAG

27



81

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CCATCGATGC CGGACTGGTA TCCCAGGGGG

30

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCATCGATGC CGGACTGGTA TCCCGAGGGA C

31

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CCATCGATGA TCCAGCCTCC TCGCGGCGCC GGATGGGCA

39

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCTCTAGATC CATTGCCCAT CCGAAGATGC CCATCCGGC

39



82

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCATCGATT ATGCCGAGAA GGTAACCAGA GAAACACAC

39

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCTCTAGACC AGGTAATATA CCACAACGTG TGTTCCTCT

39

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGGGGGAATT CATTAGGTG ACACTATAGT TCTGCCTCCC CGGAC

45

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGGGGGATCC TGGTATCCCA GGGGGGC

27



AMENDED SHEET
 IPEA/AU

83

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 28 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCGGAAGCTT GTTTTCTTT CTTTACCA

28

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 46 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGGGGATCCG ATGGTATCCC GAGGGACGCT CAGCAGGTGC CATAGG

46

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 52 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AAATAATTTT GTTACTTTAG AAGGAGATAT ACATATGAGC GAGCGAGCAC AC

52

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 55 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AAATAATTTT GTTTAACCTT AAGAAGGAGA TCTACATATG CTGGAGTGGC GTCAC

55



84

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GGAGATCTAC ATATGGGAGA TGCTGGAGTG

30

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 17 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GTAGCGAACG TCGAGAA

17

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGGGGATCCT CAGTTGTCAG TGGCGGGGTA G

31

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 28 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGGGATCCCT AATTGGCACG AGCGGGCGC

28



85

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AATTACATAT GGCGGCCGCC GTTCTGCC

29

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

AATTACATAT GTTCGCGGCC GCCGTTTCT

29

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein - N terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Phe Ala Ala Ala Val Ser Ala Phe Ala Ala Asn Met Leu Ser Ser Val
1 5 10 15

Leu Lys Ser



86

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein - internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
- Pro Ser Leu Val Asp Gln Gly Phe Trp Ile Gly Gly Gln Tyr Ala Leu
1 5 10 15
- Thr Pro Thr Ser
 20

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein - internal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Phe Ala Ala Ala Val Ser
1 5

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: RNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCGCCCCCUG GGAUACCAGG AUC

23



87

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TCAGCAGGTG GCATAGG

17

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 6..32

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CCCAT ATG GGC GAT GCC GGC GTC GCG TCA CAG
Met Gly Asp Ala Gly Val Ala Ser Gln
1 5

32

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein - N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Gly Asp Ala Gly Val Ala Ser Gln
1 5



88

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 6..32
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CCCAT ATG AGC GAG GCC GGC GTC GCG TCA CAG
Met Ser Glu Ala Gly Val Ala Ser Gln
1 5

32

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 9 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein - N-terminal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met Ser Glu Ala Gly Val Ala Ser Gln
1 5

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..27
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ATG GGA GAT GCT GGA GTG GCG TCA CAG
Met Gly Asp Ala Gly Val Ala Ser Gln
1 5

27



89

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein - N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met Gly Asp Ala Gly Val Ala Ser Gln
1 5

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GGGGGATCCC GCGGATTTAT GAGCGAG

27

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGGGGATCCC GCGGAGACAT GAGCGAGCAC AC

32

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GGGGGATCCA GCGACATGAG AGATGCTGGA GTGG

34



(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GGGGGATCCA GCGACATGAG AGATGCTGGA GTGG

34

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GGGGGATCCG TTCTGCCTCC CCGGAC

26

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5312 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 37..5145

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GTCTGCTC CCGGACGG TAAATATAGG GGAACA ATG TAC GCG AAA GCG ACA
 Met Tyr Ala Lys Ala Thr
 1 5

54

GAC GTG GCG CGT GTC TAC GCC GCG GCA GAT GTC GCC TAC GCG AAC GTA
 Asp Val Ala Arg Val Tyr Ala Ala Ala Asp Val Ala Tyr Ala Asn Val
 10 15 20

102

CTG CAG CAG AGA GCA GTC AAG TTG GAC TTC GCC CCG CCA CTG AAG GCA
 Leu Gln Gln Arg Ala Val Lys Leu Asp Phe Ala Pro Pro Lcu Lys Ala
 25 30 35

150

CTA GAA ACC CTC CAC AGA CTG TAC TAT CCG CTG CGC TTC AAA GGG GGC
 Leu Glu Thr Leu His Arg Leu Tyr Tyr Pro Leu Arg Phe Lys Gly Gly
 40 45 50

198



91

ACT Thr 55	TTA Leu	CCC Pro	CCG Pro	ACA Thr	CAA Gln 60	CAC His	CCG Pro	ATC Ile	CTG Leu	GCC Ala 65	GGG Gly	CAC His	CAA Gln	CGT Arg	GTC Val 70	246
GCA Ala	GAA Glu	GAG Glu	GTT Val	CTG Leu 75	CAC His	AAT Asn	TTC Phe	GCC Ala	AGG Arg 80	GGA Gly	CGT Arg	AGC Ser	ACA Thr	GTG Val 85	CTC Leu	294
GAG Glu	ATA Ile	GGG Gly	CCG Pro 90	TCT Ser	CTG Leu	CAC His	AGC Ser	GCA Ala 95	CTT Leu	AAG Lys	CTA Leu	CAT His	GGG Gly 100	GCA Ala	CCG Pro	342
AAC Asn	GCC Ala	CCC Pro 105	GTC Val	GCA Ala	GAC Asp	TAT Tyr	CAC His 110	GGG Gly	TGC Cys	ACC Thr	AAG Lys	TAC Tyr 115	GGC Gly	ACC Thr	CGC Arg	390
GAC Asp	GGC Gly 120	TCG Ser	CGA Arg	CAC His	ATT Ile	ACG Thr 125	GCC Ala	TTA Leu	GAG Glu	TCT Ser	AGA Arg 130	TCC Ser	GTC Val	GCC Ala	ACA Thr	438
GGC Gly 135	CGG Arg	CCC Pro	GAG Glu	TTC Phe	AAG Lys 140	GCC Ala	GAC Asp	GCC Ala	TCA Ser	CTG Leu 145	CTC Leu	GCC Ala	AAC Asn	GGC Gly	ATT Ile 150	486
CCC Ala	TCC Ser	CGC Arg	ACC Thr	TTC Phe 155	TGC Cys	GTC Val	GAC Asp	GGA Gly	GTC Val 160	GGC Gly	TCT Ser	TGC Cys	GCG Ala	TTC Phe 165	AAA Lys	534
TCG Ser	CGC Arg	GTT Val	GGA Gly 170	ATT Ile	GCC Ala	AAT Asn	CAC His	TCC Ser 175	CTC Leu	TAT Tyr	GAC Asp	GTG Val	ACC Thr 180	CTA Leu	GAG Glu	582
GAG Glu	CTG Leu	GCC Ala 185	AAT Asn	GCG Ala	TTT Phe	GAG Glu	AAC Asn 190	CAC His	GGA Gly	CTT Leu	CAC His	ATG Met 195	GTC Val	CGC Arg	GCG Ala	630
TTC Phe	ATG Met 200	CAC His	ATG Met	CCA Pro	GAA Glu	GAG Glu 205	CTG Leu	CTC Leu	TAC Tyr	ATG Met	GAC Asp 210	AAC Asn	GTG Val	GTT Val	AAT Asn	678
GCC Ala 215	GAG Glu	CTC Leu	GGC Gly	TAC Tyr	CGC Arg 220	TTC Phe	CAC His	GTT Val	ATT Ile	GAA Glu 225	GAG Glu	CCT Pro	ATG Met	GCT Ala	GTG Val 230	726
AAG Lys	GAC Asp	TGC Cys	GCA Ala	TTC Phe 235	CAG Gln	GGG Gly	GGG Gly	GAC Asp	CTC Leu 240	CGT Arg	CTC Leu	CAC His	TTC Phe	CCT Pro 245	GAG Glu	774
TTG Leu	GAC Asp	TTC Phe	ATC Ile 250	AAC Asn	GAG Glu	AGC Ser	CAA Gln	GAG Glu 255	CGG Arg	CGC Arg	ATC Ile	GAG Glu	AGG Arg 260	CTG Leu	GCC Ala	822
GCC Ala	CGC Arg	GGC Gly 265	TCC Ser	TAC Tyr	TCC Ser	AGA Arg	CGC Arg 270	GCC Ala	GTC Val	ATT Ile	TTC Phe	TCC Ser 275	GGC Gly	GAC Asp	GAC Asp	870
GAC Asp	TGG Trp 280	GGT Gly	GAT Asp	GCG Ala	TAC Tyr	TTA Leu 285	CAC His	GAC Asp	TTC Phe	CAC His	ACA Thr 290	TGG Trp	CTC Leu	GCC Ala	TAC Tyr	918
CTA Leu 295	CTG Leu	GTG Val	AGG Arg	AAC Asn	TAC Tyr 300	CCC Pro	ACT Thr	CCG Pro	TTT Phe	GGT Gly 305	TTC Phe	TCA Ser	CTC Leu	CAT His	ATA Ile 310	966



92

GAA	GTC	CAG	AGG	CGC	CAC	GGC	TCC	AGC	ATT	GAG	CTG	CGC	ATC	ACT	CGC	1014
Glu	Val	Gln	Arg	Arg	His	Gly	Ser	Ser	Ile	Glu	Leu	Arg	Ile	Thr	Arg	
				315					320					325		
CGC	CCA	CCT	GGA	GAC	CGC	ATG	CTG	GCC	GTC	GTC	CCA	AGG	ACG	TCC	CAA	1062
Ala	Pro	Pro	Gly	Asp	Arg	Met	Leu	Ala	Val	Val	Pro	Arg	Thr	Ser	Gln	
			330					335					340			
GGC	CTC	TGC	AGA	ATC	CCA	AAC	ATC	TTT	TAT	TAC	GCC	GAC	GGC	TGC	GGC	1110
Gly	Leu	Cys	Arg	Ile	Pro	Asn	Ile	Phe	Tyr	Tyr	Ala	Asp	Ala	Ser	Gly	
		345					350					355				
ACT	GAG	CAT	AAG	ACC	ATC	CTT	ACG	TCA	CAG	CAC	AAA	GTC	AAC	ATG	CTG	1158
Thr	Glu	His	Lys	Thr	Ile	Leu	Thr	Ser	Gln	His	Lys	Val	Asn	Met	Leu	
	360					365					370					
CTC	AAT	TTT	ATG	CAA	ACG	CGT	CCT	GAG	AAG	GAA	CTA	GTC	GAC	ATG	ACC	1206
Leu	Asn	Phe	Met	Gln	Thr	Arg	Pro	Glu	Lys	Glu	Leu	Val	Asp	Met	Thr	
	375				380					385					390	
GTC	TTG	ATG	TCG	TTC	GGC	CGC	GCT	AGG	CTG	CGC	GCG	ATC	GTG	GTC	GCC	1254
Val	Leu	Met	Ser	Phe	Ala	Arg	Ala	Arg	Leu	Arg	Ala	Ile	Val	Val	Ala	
				395					400					405		
TCA	GAA	GTC	ACC	GAG	AGC	TCC	TGG	AAC	ATC	TCA	GCG	GCT	GAC	CTG	GTC	1302
Ser	Glu	Val	Thr	Glu	Ser	Ser	Trp	Asn	Ile	Ser	Pro	Ala	Asp	Leu	Val	
			410					415					420			
CGC	ACT	GTC	GTG	TCT	CTT	TAC	GTC	CTC	CAC	ATC	ATC	GAG	CGC	CGA	AGG	1350
Arg	Thr	Val	Val	Ser	Leu	Tyr	Val	Leu	His	Ile	Ile	Glu	Arg	Arg	Arg	
		425					430					435				
GCT	GCG	GTC	GCT	GTC	AAG	ACC	GCC	AAG	GAC	GAC	GTC	TTT	GGA	GAC	ACT	1398
Ala	Ala	Val	Ala	Val	Lys	Thr	Ala	Lys	Asp	Asp	Val	Phe	Gly	Glu	Thr	
	440					445					450					
TCG	TTC	TGG	GAG	AGT	CTC	AAG	CAC	GTC	TTG	GGC	TCC	TGT	TGC	GGT	CTG	1446
Ser	Phe	Trp	Glu	Ser	Leu	Lys	His	Val	Leu	Gly	Ser	Cys	Cys	Gly	Leu	
	455				460					465					470	
CGC	AAC	CTC	AAA	GGC	ACC	GAC	GTC	GTC	TTT	ACT	AAG	CGC	GTC	GTC	GAT	1494
Arg	Asn	Leu	Lys	Gly	Thr	Asp	Val	Val	Phe	Thr	Lys	Arg	Val	Val	Asp	
				475					480					485		
AAG	TAC	CGA	GTC	CAC	TCG	CTC	GGA	GAC	ATA	ATC	TGC	GAC	GTC	CGC	CTG	1542
Lys	Tyr	Arg	Val	His	Ser	Leu	Gly	Asp	Ile	Ile	Cys	Asp	Val	Arg	Leu	
			490					495					500			
TCC	CCT	GAA	CAG	GTC	GGC	TTC	CTG	CCG	TCC	CGC	GTA	CCA	CCT	GCC	CGC	1590
Ser	Pro	Glu	Gln	Val	Gly	Phe	Leu	Pro	Ser	Arg	Val	Pro	Pro	Ala	Arg	
		505					510					515				
GTC	TTT	CAC	GAC	AGG	GAA	GAG	CTT	GAG	GTC	CTT	CGC	GAA	CCT	GGC	TGC	1638
Val	Phe	His	Asp	Arg	Glu	Glu	Leu	Glu	Val	Leu	Arg	Glu	Ala	Gly	Cys	
	520					525					530					
TAC	AAC	GAA	CGT	CCG	GTA	CCT	TCC	ACT	CCT	CCT	GTG	GAG	GAG	CCC	CAA	1686
Tyr	Asn	Glu	Arg	Pro	Val	Pro	Ser	Thr	Pro	Pro	Val	Glu	Glu	Pro	Gln	
	535				540					545					550	



AMENDED SHEET
 IPEA/AU

93

GGT	TTC	GAC	GCC	GAC	TTG	TGG	CAC	GCG	ACC	GCG	GCC	TCA	CTC	CCC	GAG	1734
Gly	Phe	Asp	Ala	Asp	Leu	Trp	His	Ala	Thr	Ala	Ala	Ser	Leu	Pro	Glu	
				555					560					565		
TAC	CGC	GCC	ACC	TTG	CAG	GCA	GGT	CTC	AAC	ACC	GAC	GTC	AAG	CAG	CTC	1782
Tyr	Arg	Ala	Thr	Leu	Gln	Ala	Gly	Leu	Asn	Thr	Asp	Val	Lys	Gln	Leu	
			570					575					580			
AAG	ATC	ACC	CTC	GAG	AAC	GCC	CTC	AAG	ACC	ATC	GAC	GGG	CTC	ACC	CTC	1830
Lys	Ile	Thr	Leu	Glu	Asn	Ala	Leu	Lys	Thr	Ile	Asp	Gly	Leu	Thr	Leu	
		585					590					595				
TCC	CCA	GTC	AGA	GCC	CTC	GAG	ATG	TAC	GAG	GCG	CCG	CCA	GCG	AGC	GCG	1876
Ser	Pro	Val	Arg	Gly	Leu	Glu	Met	Tyr	Glu	Gly	Pro	Pro	Gly	Ser	Gly	
	600					605					610					
AAG	ACG	GCG	ACC	CTC	ATC	GCC	GCC	CTT	GAG	GCC	GCG	GCG	GCT	AAA	GCA	1926
Lys	Thr	Gly	Thr	Leu	Ile	Ala	Ala	Leu	Glu	Ala	Ala	Gly	Gly	Lys	Ala	
615					620					625					630	
CTT	TAC	GTG	GCA	CCC	ACC	AGA	GAA	CTG	AGA	GAG	GCT	ATG	GAC	CGG	CGG	1974
Leu	Tyr	Val	Ala	Pro	Thr	Arg	Glu	Leu	Arg	Glu	Ala	Met	Asp	Arg	Arg	
				635					640					645		
ATC	AAA	CCG	CCG	TCC	GCC	TCG	GCT	ACG	CAA	CAT	GTC	GCC	CTT	GCG	ATT	2022
Ile	Lys	Pro	Pro	Ser	Ala	Ser	Ala	Thr	Gln	His	Val	Ala	Leu	Ala	Ile	
			650					655					660			
CTC	CGT	CGT	GCC	ACC	GCC	GAG	GCG	GCC	CCT	TTC	GCT	ACC	GTG	GTT	ATC	2070
Leu	Arg	Arg	Ala	Thr	Ala	Glu	Gly	Ala	Pro	Phe	Ala	Thr	Val	Val	Ile	
		665					670					675				
GAC	GAG	TGC	TTC	ATG	TTC	CCG	CTC	GTG	TAC	GTC	GCG	ATC	GTG	CAC	GCC	2118
Asp	Glu	Cys	Phe	Met	Phe	Pro	Leu	Val	Tyr	Val	Ala	Ile	Val	His	Ala	
	680					685					690					
TTG	TCC	CCG	AGC	TCA	CGA	ATA	GTC	CTT	GTA	GGG	GAC	GTC	CAC	CAA	ATC	2166
Leu	Ser	Pro	Ser	Ser	Arg	Ile	Val	Leu	Val	Gly	Asp	Val	His	Gln	Ile	
695					700					705					710	
GGG	TTT	ATA	GAC	TTC	CAA	GCG	ACA	AGC	GCG	AAC	ATG	CCG	CTC	GTT	CGC	2214
Gly	Phe	Ile	Asp	Phe	Gln	Gly	Thr	Ser	Ala	Asn	Met	Pro	Leu	Val	Arg	
				715					720					725		
GAC	GTC	GTT	AAG	CAG	TGC	CGT	CGG	CGC	ACT	TTC	AAC	CAA	ACC	AAG	CGC	2262
Asp	Val	Val	Lys	Gln	Cys	Arg	Arg	Arg	Thr	Phe	Asn	Gln	Thr	Lys	Arg	
			730					735					740			
TGT	CCG	GCC	GAC	GTC	GTT	GCC	ACC	ACG	TTT	TTC	CAG	AGC	TTG	TAC	CCC	2310
Cys	Pro	Ala	Asp	Val	Val	Ala	Thr	Thr	Phe	Phe	Gln	Ser	Leu	Tyr	Pro	
		745					750					755				
GGG	TGC	ACA	ACC	ACC	TCA	GGG	TGC	GTC	GCA	TCC	ATC	AGC	CAC	GTC	GCC	2358
Gly	Cys	Thr	Thr	Thr	Ser	Gly	Cys	Val	Ala	Ser	Ile	Ser	His	Val	Ala	
	760					765					770					
CCA	GAC	TAC	CGC	AAC	AGC	CAG	GCG	CAA	ACG	CTC	TGC	TTC	ACG	CAG	GAG	2406
Pro	Asp	Tyr	Arg	Asn	Ser	Gln	Ala	Gln	Thr	Leu	Cys	Phe	Thr	Gln	Glu	
					780					785					790	



94

GAA AAG TCG CGC CAC GGG GCT GAG GGC GCG ATG ACT GTG CAC GAA GCG Glu Lys Ser Arg His Gly Ala Glu Gly Ala Met Thr Val His Glu Ala 795 800 805	2454
CAG GGA CGC ACT TTT GCG TCT GTC ATT CTG CAT TAC AAC GGC TCC ACA Gln Gly Arg Thr Phe Ala Ser Val Ile Leu His Tyr Asn Gly Ser Thr 810 815 820	2502
GCA GAG CAG AAG CTC CTC GCT GAG AAG TCG CAC CTT CTA GTC GGC ATC Ala Glu Gln Lys Leu Leu Ala Glu Lys Ser His Leu Leu Val Gly Ile 825 830 835	2550
ACG CGC CAC ACC AAC CAC CTG TAC ATC CGC GAC CCG ACA GGT GAC ATT Thr Arg His Thr Asn His Leu Tyr Ile Arg Asp Pro Thr Gly Asp Ile 840 845 850	2598
GAG AGA CAA CTC AAC CAT AGC GCG AAA GCC GAG GTG TTT ACA GAC ATC Glu Arg Gln Leu Asn His Ser Ala Lys Ala Glu Val Phe Thr Asp Ile 855 860 865 870	2646
CCT GCA CCC CTG GAG ATC ACG ACT GTC AAA CCG AGT GAA GAG GTG CAG Pro Ala Pro Leu Glu Ile Thr Thr Val Lys Pro Ser Glu Glu Val Gln 875 880 885	2694
CGC AAC GAA GTG ATG GCA ACG ATA CCC CCG CAG AGT GCC ACG CCG CAC Arg Asn Glu Val Met Ala Thr Ile Pro Pro Gln Ser Ala Thr Pro His 890 895 900	2742
GGA GCA ATC CAT CTG CTC CGC AAG AAC TTC GGG GAC CAA CCC GAC TGT Gly Ala Ile His Leu Leu Arg Lys Asn Phe Gly Asp Gln Pro Asp Cys 905 910 915	2790
GGC TGT GTC GCT TTG GCG AAG ACC GGC TAC GAG GTG TTT GGC GGT CGT Gly Cys Val Ala Leu Ala Lys Thr Gly Tyr Glu Val Phe Gly Gly Arg 920 925 930	2838
GCC AAA ATC AAC GTA GAG CTT GCC GAA CCC GAC GCG ACC CCG AAG CCG Ala Lys Ile Asn Val Glu Leu Ala Glu Pro Asp Ala Thr Pro Lys Pro 935 940 945 950	2886
CAT AGG GCG TTC CAG GAA GGG GTA CAG TGG GTC AAG GTC ACC AAC GCG His Arg Ala Phe Gln Glu Gly Val Gln Trp Val Lys Val Thr Asn Ala 955 960 965	2934
TCT AAC AAA CAC CAG GCG CTC CAG ACG CTG TTG TCC CGC TAC ACC AAG Ser Asn Lys His Gln Ala Leu Gln Thr Leu Leu Ser Arg Tyr Thr Lys 970 975 980	2982
CGA AGC GCT GAC CTG CCG CTA CAC GAA GCT AAG GAG GAC GTC AAA CGC Arg Ser Ala Asp Leu Pro Leu His Glu Ala Lys Glu Asp Val Lys Arg 985 990 995	3030
ATG CTA AAC TCG CTT GAC CGA CAT TGG GAC TGG ACT GTC ACT GAA GAC Met Leu Asn Ser Leu Asp Arg His Trp Asp Trp Thr Val Thr Glu Asp 1000 1005 1010	3078
GCC CGT GAC CGA GCT GTC TTC GAG ACC CAG CTC AAG TTC ACC CAA CGC Ala Arg Asp Arg Ala Val Phe Glu Thr Gln Leu Lys Phe Thr Gln Arg 1015 1020 1025 1030	3126



GGC GGC ACC GTC GAA GAC CTG CTG GAG CCA GAC GAC CCC TAC ATC CGT Gly Gly Thr Val Glu Asp Leu Leu Glu Pro Asp Asp Pro Tyr Ile Arg 1035 1040 1045	3174
GAC ATA GAC TTC CTT ATG AAG ACT CAG CAG AAA GTG TCG CCC AAG CCG Asp Ile Asp Phe Leu Met Lys Thr Gln Gln Lys Val Ser Pro Lys Pro 1050 1055 1060	3222
ATC AAT ACG GGC AAG GTC GGG CAG GGG ATC GCC GCT CAC TCA AAG TCT Ile Asn Thr Gly Lys Val Gly Gln Gly Ile Ala Ala His Ser Lys Ser 1065 1070 1075	3270
CTC AAC TTC GTC CTC GCC GCT TGG ATA CGC ATA CTC GAG GAG ATA CTC Leu Asn Phe Val Leu Ala Ala Trp Ile Arg Ile Leu Glu Glu Ile Leu 1080 1085 1090	3318
CGT ACC GGC AGC CGC ACG GTC CGG TAC AGC AAC GGT CTC CCC GAC GAA Arg Thr Gly Ser Arg Thr Val Arg Tyr Ser Asn Gly Leu Pro Asp Glu 1095 1100 1105 1110	3366
GAA GAG GCC ATG CTG CTC GAA GCG AAG ATC AAT CAA GTC CCA CAC GCC Glu Glu Ala Met Leu Leu Glu Ala Lys Ile Asn Gln Val Pro His Ala 1115 1120 1125	3414
ACG TTC GTC TCG GCG GAC TGG ACC GAG TTT GAC ACC GCC CAC AAT AAC Thr Phe Val Ser Ala Asp Trp Thr Glu Phe Asp Thr Ala His Asn Asn 1130 1135 1140	3462
ACG AGT GAG CTG CTC TTC GCC GCC CTT TTA GAG CGC ATC GGC ACG CCT Thr Ser Glu Leu Leu Phe Ala Ala Leu Leu Glu Arg Ile Gly Thr Pro 1145 1150 1155	3510
GCA GCT GCC GTT AAT CTA TTC AGA GAA CGG TGT GGG AAA CGC ACC TTG Ala Ala Ala Val Asn Leu Phe Arg Glu Arg Cys Gly Lys Arg Thr Leu 1160 1165 1170	3558
CGA GCG AAG GGT CTA GGC TCC GTT GAA GTC GAC GGT CTG CTC GAC TCC Arg Ala Lys Gly Leu Gly Ser Val Glu Val Asp Gly Leu Leu Asp Ser 1175 1180 1185 1190	3606
GGC GCA GCT TGG ACG CCT TGC CGC AAC ACC ATC TTC TCT GCC GCC GTC Gly Ala Ala Trp Thr Pro Cys Arg Asn Thr Ile Phe Ser Ala Ala Val 1195 1200 1205	3654
ATG CTC ACG CTC TTC CGC GGC GTC AAG TTC GCA GCT TTC AAA GGC GAC Met Leu Thr Leu Phe Arg Gly Val Lys Phe Ala Ala Phe Lys Gly Asp 1210 1215 1220	3702
GAC TCG CTC CTC TGT GGT AGC CAT TAC CTC CGT TTC GAC GCT AGC CGC Asp Ser Leu Leu Cys Gly Ser His Tyr Leu Arg Phe Asp Ala Ser Arg 1225 1230 1235	3750
CTT CAC ATG GGC GAA CGT TAC AAG ACC AAA CAT TTG AAG GTC GAG GTG Leu His Met Gly Glu Arg Tyr Lys Thr Lys His Leu Lys Val Glu Val 1240 1245 1250	3798
CAG AAA ATC GTG CCG TAC ATC GGA CTC CTC GTC TCC GCT GAG CAG GTC Gln Lys Ile Val Pro Tyr Ile Gly Leu Leu Val Ser Ala Glu Gln Val 1255 1260 1265 1270	3846



GTC	CTC	GAC	CCT	GTC	AGG	AGC	GCT	CTC	AAG	ATA	TTT	GGG	CGC	TGC	TAC	3894
Val	Leu	Asp	Pro	Val	Arg	Ser	Ala	Leu	Lys	Ile	Phe	Gly	Arg	Cys	Tyr	
				1275					1280					1285		
ACA	AGC	GAA	CTC	CTT	TAC	TCC	AAG	TAC	GTG	GAG	GCT	GTG	AGA	GAC	ATC	3942
Thr	Ser	Glu	Leu	Leu	Tyr	Ser	Lys	Tyr	Val	Glu	Ala	Val	Arg	Asp	Ile	
			1290					1295					1300			
ACC	AAG	GCC	TGG	AGT	GAC	GCC	CGC	TAC	CAC	AGC	CTC	CTG	TGC	CAC	ATG	3990
Thr	Lys	Gly	Trp	Ser	Asp	Ala	Arg	Tyr	His	Ser	Leu	Leu	Cys	His	Met	
		1305					1310					1315				
TCA	GCA	TGC	TAC	TAC	AAT	TAC	GCG	CCG	GAG	TCT	GCG	GCG	TAC	ATC	ATC	4038
Ser	Ala	Cys	Tyr	Tyr	Asn	Tyr	Ala	Pro	Glu	Ser	Ala	Ala	Tyr	Ile	Ile	
	1320					1325					1330					
GAC	GCT	GTT	GTT	CGC	TTT	GGG	CGC	GGC	GAC	TTC	CCG	TTT	GAA	CAA	CTG	4086
Asp	Ala	Val	Val	Arg	Phe	Gly	Arg	Gly	Asp	Phe	Pro	Phe	Glu	Gln	Leu	
1335				1340					1345						1350	
CGC	GTG	GTG	CGT	GCC	CAT	GTG	CAG	GCA	CCC	GAC	GCT	TAC	AGC	AGC	ACG	4134
Arg	Val	Val	Arg	Ala	His	Val	Gln	Ala	Pro	Asp	Ala	Tyr	Ser	Ser	Thr	
				1355				1360						1365		
TAT	CCG	GCT	AAC	GTG	CGC	GCA	TCG	TGC	CTT	GAC	CAC	GTG	TTC	GAG	CCC	4182
Tyr	Pro	Ala	Asn	Val	Arg	Ala	Ser	Cys	Leu	Asp	His	Val	Phe	Glu	Pro	
			1370					1375					1380			
CGC	CAG	GCC	GCC	GCC	CCG	GCA	GGT	TTC	GTT	GCG	ACA	TGT	GCG	AAG	CCG	4230
Arg	Gln	Ala	Ala	Ala	Pro	Ala	Gly	Phe	Val	Ala	Thr	Cys	Ala	Lys	Pro	
		1385					1390					1395				
GAA	ACG	CCT	TCT	TCA	CTT	ACC	GCG	AAA	GCT	GGT	GTT	TCT	GCG	ACT	ACA	4278
Glu	Thr	Pro	Ser	Ser	Leu	Thr	Ala	Lys	Ala	Gly	Val	Ser	Ala	Thr	Thr	
	1400					1405					1410					
AGC	CAC	GTT	GCG	ACT	GGG	ACT	GCG	CCC	CCG	GAG	TCT	CCA	TGG	GAT	GCA	4326
Ser	His	Val	Ala	Thr	Gly	Thr	Ala	Pro	Pro	Glu	Ser	Pro	Trp	Asp	Ala	
1415					1420					1425					1430	
CCT	GCA	GCC	AAC	AGC	TTT	TCG	GAG	TTA	TTG	ACA	CCG	GAG	ACC	CCG	TCC	4374
Pro	Ala	Ala	Asn	Ser	Phe	Ser	Glu	Leu	Leu	Thr	Pro	Glu	Thr	Pro	Ser	
			1435					1440						1445		
ACA	TCA	TCC	TCG	CCG	TCA	TCG	TCT	TCA	TCG	GAC	TCC	TCT	ACA	TCG	TGT	4422
Thr	Ser	Ser	Ser	Pro	Ser	Ser	Ser	Ser	Ser	Asp	Ser	Ser	Ser	Ser	Cys	
			1450					1455					1460			
GGA	AGG	TCG	CTC	AGT	GGT	GGA	GAC	ACC	GCA	AGG	ACC	ACA	GAA	GAC	TTG	4470
Gly	Arg	Ser	Leu	Ser	Gly	Gly	Asp	Thr	Ala	Arg	Thr	Thr	Glu	Asp	Leu	
		1465					1470					1475				
AAC	AGC	AGA	AAG	CCG	CCT	TCG	CAA	GAC	AGG	CAA	TCA	CGC	TCG	TCT	GAA	4518
Asn	Ser	Arg	Lys	Pro	Pro	Ser	Gln	Asp	Arg	Gln	Ser	Arg	Ser	Ser	Glu	
	1480					1485					1490					
TGT	CTG	GAC	AGA	AGC	GGA	GAA	AGG	ACA	GGC	AGT	TCG	TTA	ACT	GCC	CCC	4566
Cys	Leu	Asp	Arg	Ser	Gly	Glu	Arg	Thr	Gly	Ser	Ser	Leu	Thr	Ala	Pro	
1495					1500					1505					1510	



97

ACT GCT CCG AGC CCC TCA TTC TCA TTT TCG GAA AGA GCT CGA CTG GCG	4614
Thr Ala Pro Ser Pro Ser Phe Ser Phe Ser Glu Arg Ala Arg Leu Ala	
1515 1520 1525	
ACC GGG CCG ACT GTC GCC GCT GCG ACA TCA CCT GCA ACC CCA TCC	4662
Thr Gly Pro Thr Val Ala Ala Ala Thr Ser Pro Ser Ala Thr Pro Ser	
1530 1535 1540	
TGC GCC ACG GAC CAG GTT GCC GCG AGG ACC ACG CCG GAC TTT GCG CCT	4710
Cys Ala Thr Asp Gln Val Ala Ala Arg Thr Thr Pro Asp Phe Ala Pro	
1545 1550 1555	
TTC CTG GGT TCC CAG TCT GCC CGT GCT GTC TCG AAG CCG TAC CGG CCC	4758
Phe Leu Gly Ser Gln Ser Ala Arg Ala Val Ser Lys Pro Tyr Arg Pro	
1560 1565 1570	
CCC ACG ACT GCC CGT TGG AAA GAA GTC ACC CCG CTC CAC GCG TGG AAG	4806
Pro Thr Thr Ala Arg Trp Lys Glu Val Thr Pro Leu His Ala Trp Lys	
1575 1580 1585 1590	
GGC GTG ACC GGA GAC CGA CCG GAA GTC AGG GAG GAC CCG GAG ACA GCG	4854
Gly Val Thr Gly Asp Arg Pro Glu Val Arg Glu Asp Pro Glu Thr Ala	
1595 1600 1605	
CGC GTC GTC CAG GCT CTG ATC AGC GGC CGT TAT CCT CAG AAG ACG AAG	4902
Ala Val Val Gln Ala Leu Ile Ser Gly Arg Tyr Pro Gln Lys Thr Lys	
1610 1615 1620	
CTT TCC TCC GAC GCA TCC AAA GGC TAC TCA AGA ACT AAG GGA TGC TCA	4950
Leu Ser Ser Asp Ala Ser Lys Gly Tyr Ser Arg Thr Lys Gly Cys Ser	
1625 1630 1635	
CAA TCC ACC TCT TTT CCT GCC CCG AGT GCG GAT TAC CAG GCC CGC GAC	4998
Gln Ser Thr Ser Phe Pro Ala Pro Ser Ala Asp Tyr Gln Ala Arg Asp	
1640 1645 1650	
TGC CAG ACA GTC CGA GTC TGC CCG GCC GCT GCA GAG ATG GCG CGC TCA	5046
Cys Gln Thr Val Arg Val Cys Arg Ala Ala Ala Glu Met Ala Arg Ser	
1655 1660 1665 1670	
TGT ATT CAC GAG CCG TTG GCT TCA TCT GCC GCC AGT GCC GAC TTG AAG	5094
Cys Ile His Glu Pro Leu Ala Ser Ser Ala Ala Ser Ala Asp Leu Lys	
1675 1680 1685	
CGC ATA CGC TCT ACC TCG GAC TCT GTT CCC GAT GTA AAG ATC AGC AAG	5142
Arg Ile Arg Ser Thr Ser Asp Ser Val Pro Asp Val Lys Ile Ser Lys	
1690 1695 1700	
AGC GCA TGAAGGAACA AAATTAGTTT CCTTGTTCTG AAACAAGGTG GTCCCTCCCA	5198
Ser Ala	
TTGAGGTAAA GACTCTGGTG AGTCCTCAAC GTTACTCGTT GACTCTGCTG CGGTTTCGATT	5258
CGATTCCCAA GCAGCAAAGG GTGCGCAACT AGTACGGCGC CCCCTGGGAT ACCA	5312



(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1704 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```

Met Tyr Ala Lys Ala Thr Asp Val Ala Arg Val Tyr Ala Ala Ala Asp
 1           5           10           15
Val Ala Tyr Ala Asn Val Leu Gln Gln Arg Ala Val Lys Leu Asp Phe
          20           25           30
Ala Pro Pro Leu Lys Ala Leu Glu Thr Leu His Arg Leu Tyr Tyr Pro
      35           40           45
Leu Arg Phe Lys Gly Gly Thr Leu Pro Pro Thr Gln His Pro Ile Leu
 50           55           60
Ala Gly His Gln Arg Val Ala Glu Glu Val Leu His Asn Phe Ala Arg
 65           70           75           80
Gly Arg Ser Thr Val Leu Glu Ile Gly Pro Ser Leu His Ser Ala Leu
          85           90           95
Lys Leu His Gly Ala Pro Asn Ala Pro Val Ala Asp Tyr His Gly Cys
      100           105           110
Thr Lys Tyr Gly Thr Arg Asp Gly Ser Arg His Ile Thr Ala Leu Glu
      115           120           125
Ser Arg Ser Val Ala Thr Gly Arg Pro Glu Phe Lys Ala Asp Ala Ser
      130           135           140
Leu Leu Ala Asn Gly Ile Ala Ser Arg Thr Phe Cys Val Asp Gly Val
      145           150           155           160
Gly Ser Cys Ala Phe Lys Ser Arg Val Gly Ile Ala Asn His Ser Leu
          165           170           175
Tyr Asp Val Thr Leu Glu Glu Leu Ala Asn Ala Phe Glu Asn His Gly
      180           185           190
Leu His Met Val Arg Ala Phe Met His Met Pro Glu Glu Leu Leu Tyr
      195           200           205
Met Asp Asn Val Val Asn Ala Glu Leu Gly Tyr Arg Phe His Val Ile
      210           215           220
Glu Glu Pro Met Ala Val Lys Asp Cys Ala Phe Gln Gly Gly Asp Leu
      225           230           235           240
Arg Leu His Phe Pro Glu Leu Asp Phe Ile Asn Glu Ser Gln Glu Arg
          245           250           255
Arg Ile Glu Arg Leu Ala Ala Arg Gly Ser Tyr Ser Arg Arg Ala Val
      260           265           270

```



Ile Phe Ser Gly Asp Asp Asp Trp Gly Asp Ala Tyr Leu His Asp Phe
 275 280 285

His Thr Trp Leu Ala Tyr Leu Leu Val Arg Asn Tyr Pro Thr Pro Phe
 290 295 300

Gly Phe Ser Leu His Ile Glu Val Gln Arg Arg His Gly Ser Ser Ile
 305 310 315 320

Glu Leu Arg Ile Thr Arg Ala Pro Pro Gly Asp Arg Met Leu Ala Val
 325 330 335

Val Pro Arg Thr Ser Gln Gly Leu Cys Arg Ile Pro Asn Ile Phe Tyr
 340 345 350

Tyr Ala Asp Ala Ser Gly Thr Glu His Lys Thr Ile Leu Thr Ser Gln
 355 360 365

His Lys Val Asn Met Leu Leu Asn ho Met Gln Thr Arg Pro Glu Lys
 370 375 380

Glu Leu Val Asp Met Thr Val Leu Met Ser Phe Ala Arg Ala Arg Leu
 385 390 395 400

Arg Ala Ile Val Val Ala Ser Glu Val Thr Glu Ser Ser Trp Asn Ile
 405 410 415

Ser Pro Ala Asp Leu Val Arg Thr Val Val Ser Leu Tyr Val Leu His
 420 425 430

Ile Ile Glu Arg Arg Arg Ala Ala Val Ala Val Lys Thr Ala Lys Asp
 435 440 445

Asp Val Phe Gly Glu Thr Ser Phe Trp Glu Ser Leu Lys His Val Leu
 450 455 460

Gly Ser Cys Cys Gly Leu Arg Asn Leu Lys Gly Thr Asp Val Val Phe
 465 470 475 480

Thr Lys Arg Val Val Asp Lys Tyr Arg Val His Ser Leu Gly Asp Ile
 485 490 495

Ile Cys Asp Val Arg Leu Ser Pro Glu Gln Val Gly Phe Leu Pro Ser
 500 505 510

Arg Val Pro Pro Ala Arg Val Phe His Asp Arg Glu Glu Leu Glu Val
 515 520 525

Leu Arg Glu Ala Gly Cys Tyr Asn Glu Arg Pro Val Pro Ser Thr Pro
 530 535 540

Pro Val Glu Glu Pro Gln Gly Phe Asp Ala Asp Leu Trp His Ala Thr
 545 550 555 560

Ala Ala Ser Leu Pro Glu Tyr Arg Ala Thr Leu Gln Ala Gly Leu Asn
 565 570 575

Thr Asp Val Lys Gln Leu Lys Ile Thr Leu Glu Asn Ala Leu Lys Thr
 580 585 590



100

Ile Asp Gly Leu Thr Leu Ser Pro Val Arg Gly Leu Glu Met Tyr Glu
595 600 605

Gly Pro Pro Gly Ser Gly Lys Thr Gly Thr Leu Ile Ala Ala Leu Glu
610 615 620

Ala Ala Gly Gly Lys Ala Leu Tyr Val Ala Pro Thr Arg Glu Leu Arg
625 630 635 640

Glu Ala Met Asp Arg Arg Ile Lys Pro Pro Ser Ala Ser Ala Thr Gin
645 650 655

His Val Ala Leu Ala Ile Leu Arg Arg Ala Thr Ala Glu Gly Ala Pro
660 665 670

Phe Ala Thr Val Val Ile Asp Glu Cys Phe Met Phe Pro Leu Val Tyr
675 680 685

Val Ala Ile Val His Ala Leu Ser Pro Ser Ser Arg Ile Val Leu Val
690 695 700

Gly Asp Val His Gln Ile Gly Phe Ile Asp Phe Gln Gly Thr Ser Ala
705 710 715 720

Asn Met Pro Leu Val Arg Asp Val Val Lys Gln Cys Arg Arg Arg Thr
725 730 735

Phe Asn Gln Thr Lys Arg Cys Pro Ala Asp Val Val Ala Thr Thr Phe
740 745 750

Phe Gln Ser Leu Tyr Pro Gly Cys Thr Thr Thr Ser Gly Cys Val Ala
755 760 765

Ser Ile Ser His Val Ala Pro Asp Tyr Arg Asn Ser Gln Ala Gln Thr
770 775 780

Leu Cys Phe Thr Gln Gln Glu Lys Ser Arg His Gly Ala Glu Gly Ala
785 790 795 800

Met Thr Val His Glu Ala Gln Gly Arg Thr Phe Ala Ser Val Ile Leu
805 810 815

His Tyr Asn Gly Ser Thr Ala Glu Gln Lys Leu Leu Ala Glu Lys Ser
820 825 830

His Leu Leu Val Gly Ile Thr Arg His Thr Asn His Leu Tyr Ile Arg
835 840 845

Asp Pro Thr Gly Asp Ile Glu Arg Gln Leu Asn His Ser Ala Lys Ala
850 855 860

Glu Val Phe Thr Asp Ile Pro Ala Pro Leu Glu Ile Thr Thr Val Lys
865 870 875 880

Pro Ser Glu Glu Val Gln Arg Asn Glu Val Met Ala Thr Ile Pro Pro
885 890 895

Gln Ser Ala Thr Pro His Gly Ala Ile His Leu Leu Arg Lys Asn Phe
900 905 910



101

Gly Asp Gln Pro Asp Cys Gly Cys Val Ala Leu Ala Lys Thr Gly Tyr
 915 920 925

Glu Val Phe Gly Gly Arg Ala Lys Ile Asn Val Glu Leu Ala Glu Pro
 930 935 940

Asp Ala Thr Pro Lys Pro His Arg Ala Phe Gln Glu Gly Val Gln Trp
 945 950 955 960

Val Lys Val Thr Asn Ala Ser Asn Lys His Gln Ala Leu Gln Thr Leu
 965 970 975

Leu Ser Arg Tyr Thr Lys Arg Ser Ala Asp Leu Pro Leu His Glu Ala
 980 985 990

Lys Glu Asp Val Lys Arg Met Leu Asn Ser Leu Asp Arg His Trp Asp
 995 1000 1005

Trp Thr Val Thr Glu Asp Ala Arg Asp Arg Ala Val Phe Glu Thr Gln
 1010 1015 1020

Leu Lys Phe Thr Gln Arg Gly Gly Thr Val Glu Asp Leu Leu Glu Pro
 1025 1030 1035 1040

Asp Asp Pro Tyr Ile Arg Asp Ile Asp Phe Leu Met Lys Thr Gln Gln
 1045 1050 1055

Lys Val Ser Pro Lys Pro Ile Asn Thr Gly Lys Val Gly Gln Gly Ile
 1060 1065 1070

Ala Ala His Ser Lys Ser Leu Asn Phe Val Leu Ala Ala Trp Ile Arg
 1075 1080 1085

Ile Leu Glu Glu Ile Leu Arg Thr Gly Ser Arg Thr Val Arg Tyr Ser
 1090 1095 1100

Asn Gly Leu Pro Asp Glu Glu Glu Ala Met Leu Leu Glu Ala Lys Ile
 1105 1110 1115 1120

Asn Gln Val Pro His Ala Thr Phe Val Ser Ala Asp Trp Thr Glu Phe
 1125 1130 1135

Asp Thr Ala His Asn Asn Thr Ser Glu Leu Leu Phe Ala Ala Leu Leu
 1140 1145 1150

Glu Arg Ile Gly Thr Pro Ala Ala Ala Val Asn Leu Phe Arg Glu Arg
 1155 1160 1165

Cys Gly Lys Arg Thr Leu Arg Ala Lys Gly Leu Gly Ser Val Glu Val
 1170 1175 1180

Asp Gly Leu Leu Asp Ser Gly Ala Ala Trp Thr Pro Cys Arg Asn Thr
 1185 1190 1195 1200

Ile Phe Ser Ala Ala Val Met Leu Thr Leu Phe Arg Gly Val Lys Phe
 1205 1210 1215

Ala Ala Phe Lys Gly Asp Asp Ser Leu Leu Cys Gly Ser His Tyr Leu
 1220 1225 1230



Arg Phe Asp Ala Ser Arg Leu His Met Gly Glu Arg Tyr Lys Thr Lys
 1235 1240 1245

His Leu Lys Val Glu Val Gln Lys Ile Val Pro Tyr Ile Gly Leu Leu
 1250 1255 1260

Val Ser Ala Glu Gln Val Val Leu Asp Pro Val Arg Ser Ala Leu Lys
 1265 1270 1275 1280

Ile Phe Gly Arg Cys Tyr Thr Ser Glu Leu Leu Tyr Ser Lys Tyr Val
 1285 1290 1295

Glu Ala Val Arg Asp Ile Thr Lys Gly Trp Ser Asp Ala Arg Tyr His
 1300 1305 1310

Ser Leu Leu Cys His Met Ser Ala Cys Tyr Tyr Asn Tyr Ala Pro Glu
 1315 1320 1325

Ser Ala Ala Tyr Ile Ile Asp Ala Val Val Arg Phe Gly Arg Gly Asp
 1330 1335 1340

Phe Pro Phe Glu Gln Leu Arg Val Val Arg Ala His Val Gln Ala Pro
 1345 1350 1355 1360

Asp Ala Tyr Ser Ser Thr Tyr Pro Ala Asn Val Arg Ala Ser Cys Leu
 1365 1370 1375

Asp His Val Phe Glu Pro Arg Gln Ala Ala Ala Pro Ala Gly Phe Val
 1380 1385 1390

Ala Thr Cys Ala Lys Pro Glu Thr Pro Ser Ser Leu Thr Ala Lys Ala
 1395 1400 1405

Gly Val Ser Ala Thr Thr Ser His Val Ala Thr Gly Thr Ala Pro Pro
 1410 1415 1420

Glu Ser Pro Trp Asp Ala Pro Ala Ala Asn Ser Phe Ser Glu Leu Leu
 1425 1430 1435 1440

Thr Pro Glu Thr Pro Ser Thr Ser Ser Ser Pro Ser Ser Ser Ser
 1445 1450 1455

Asp Ser Ser Thr Ser Cys Gly Arg Ser Leu Ser Gly Gly Asp Thr Ala
 1460 1465 1470

Arg Thr Thr Glu Asp Leu Asn Ser Arg Lys Pro Pro Ser Gln Asp Arg
 1475 1480 1485

Gln Ser Arg Ser Ser Glu Cys Leu Asp Arg Ser Gly Glu Arg Thr Gly
 1490 1495 1500

Ser Ser Leu Thr Ala Pro Thr Ala Pro Ser Pro Ser Phe Ser Phe Ser
 1505 1510 1515 1520

Glu Arg Ala Arg Leu Ala Thr Gly Pro Thr Val Ala Ala Ala Thr Ser
 1525 1530 1535

Pro Ser Ala Thr Pro Ser Cys Ala Thr Asp Gln Val Ala Ala Arg Thr
 1540 1545 1550



103

Thr Pro Asp Phe Ala Pro Phe Leu Gly Ser Gln Ser Ala Arg Ala Val
 1555 1560 1565
 Ser Lys Pro Tyr Arg Pro Pro Thr Thr Ala Arg Trp Lys Glu Val Thr
 1570 1575 1580
 Pro Leu His Ala Trp Lys Gly Val Thr Gly Asp Arg Pro Glu Val Arg
 1585 1590 1595 1600
 Glu Asp Pro Glu Thr Ala Ala Val Val Gln Ala Leu Ile Ser Gly Arg
 1605 1610 1615
 Tyr Pro Gln Lys Thr Lys Leu Ser Ser Asp Ala Ser Lys Gly Tyr Ser
 1620 1625 1630
 Arg Thr Lys Gly Cys Ser Gln Ser Thr Ser Phe Pro Ala Pro Ser Ala
 1635 1640 1645
 Asp Tyr Gln Ala Arg Asp Cys Gln Thr Val Arg Val Cys Arg Ala Ala
 1650 1655 1660
 Ala Glu Met Ala Arg Ser Cys Ile His Glu Pro Leu Ala Ser Ser Ala
 1665 1670 1675 1680
 Ala Ser Ala Asp Leu Lys Arg Ile Arg Ser Thr Ser Asp Ser Val Pro
 1685 1690 1695
 Asp Val Lys Ile Ser Lys Ser Ala
 1700

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5312 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 4218..4512

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GTTCTGCCTC CCCGGGACGG TAAATATAGG GGAACAATGT ACGCGAAAGC GACAGACGTG	60
GCGCGTGTCT ACGCCGCGGC AGATGTCGCC TACGCGAAGC TACTGCAGCA GAGAGCAGTC	120
AAGTTGGACT TCGCCCCGCC ACTGAAGGCA CTAGAAACCC TCCACAGACT GTACTATCCG	180
CTGCGCTTCA AAGGGGGCAG TTTACCCCGG ACACAACACC CGATCCTGGC CGGGCACCAA	240
CGTGTCCGAG AAGAGGTTCT GCACAATTTC GCCAGGGGAC GTAGCACAGT GCTCGAGATA	300
GGGCGGTCTC TGCACAGCGC ACTTAAGCTA CATGGGGCAC CGAACGCCCC CGTCCGAGAC	360



AMENDED SHEET

IPRA/ALL

TATCACGGGT GCACCAAGTA CGGCACCCGC GACGGCTCGC GACACATTAC GGCCTTAGAG	420
TCTAGATCCG TCGCCACAGG CCGGCCCGAG TTCAAGGCCG ACGCCTCACT GCTCGCCAAC	480
GGCATTGCCT CCCGCACCTT CTGCGTCGAC GGAGTCGGCT CTTGCGCGTT CAAATCGCGC	540
GTTGGAATTG CCAATCACTC CCTCTATGAC GTGACCCTAG AGGAGCTGGC CAATGCGTTT	600
GAGAACCACG GACTTCACAT GGTCCGCGCG TTCATGCACA TGCCAGAAGA GCTGCTCTAC	660
ATGGACAACG TGGTTAATGC CGAGCTCGGC TACCGCTTCC ACGTTATTGA AGAGCCTATG	720
GCTGTGAAGG ACTGCGCATT CCAGGGGGGG GACCTCCGTC TCCACTTCCC TGAGTTGGAC	780
TTCATCAACG AGAGCCAAGA GCGGCGCATC GAGAGGCTGG CCGCCCGCGG CTCCTACTCC	840
AGACGGCCCG TCATTTTCTC CGGCGACGAC GACTGGGGTG ATGCGTACTT ACACGACTTC	900
CACACATGGC TCGCCTACCT ACTGGTGAGG AACTACCCCA CTCGGTTTGG TTTCTCACTC	960
CATATAGAAG TCCAGAGGCG CCACGGCTCC AGCATTGAGC TGCGCATCAC TCGCGGCCA	1020
CCTGGAGACC GCATGCTGGC CGTCTGCCA AGGACGTCCC AAGGCCTCTG CAGAATCCCA	1080
AACATCTTTT ATTACGCCGA CGCGTCGGGC ACTGAGCATA AGACCATCCT TACGTCACAG	1140
CACAAAGTCA ACATGCTGCT CAATTTTATG CAAACGCGTC CTGAGAAAGA ACTAGTCGAC	1200
ATGACCGTCT TGATGTCGTT CGCGCGCGCT AGGCTGCGCG CGATCGTGGT CGCCTCAGAA	1260
GTCACCGAGA GCTCCTGGAA CATCTCACCG GCTGACCTGG TCCGCACTGT CGTGTCTCTT	1320
TACGTCCTCC ACATCATCGA GCGCCGAAGG GCTGCGGTGG CTGTCAAGAC CGCCAAGGAC	1380
GACGTCTTTG GAGAGACTTC GTTCTGGGAG AGTCTCAAGC ACGTCTTGGG CTCCTGTTGC	1440
GGTCTGCGCA ACCTCAAAGG CACCGACGTC GTCTTTACTA AGCGCGTCGT CGATAAGTAC	1500
CGAGTCCACT CGCTCGGAGA CATAATCTGC GACGTCCGCC TGTCCCTGA ACAGGTCGGC	1560
TTCCTGCCGT CCCGCGTACC ACCTGCCCCG GTCTTTCACG ACAGGGAAGA GCTTGAGGTC	1620
CTTCGCGAAG CTGGCTGCTA CAACGAACGT CCGGTACCTT CCACTCCTCC TGTGGAGGAG	1680
CCCCAAGGTT TCGACGCCGA CTTGTGGCAC GCGACGCGG CCTCACTCCC CGAGTACCGC	1740
GCCACQTTGC AGGCAGGTCT CAACACCGAC GTCAAGCAGC TCAAGATCAC CCTCGAGAAC	1800
GCCCTCAAGA CCATCGACGG GCTCACCTC TCCCCAGTCA GAGGCCTCGA GATGTACGAG	1860
GGCCCGCCAG GCAGCGGCAA GACGGGCACC CTCATCGCCG CCCTTGAGGC CGCGGGCGGT	1920
AAAGCACTTT ACGTGGCACC CACCAGAGAA CTGAGAGAGG CTATGGACCG GCGGATCAAA	1980
CCGCGTCCG CCTCGGCTAC GCAACATGTC GCCCTTGCGA TTCTCCGTCG TGCCACCGCC	2040
GAGGGCGCCC CTTTCGCTAC CGTGGTTATC GACGAGTGCT TCATGTTCCC GCTCGTGTAC	2100
GTGCGGATCG TGCACGCCTT GTCCCCGAGC TCACGAATAG TCCTGTAGG GGACGTCCAC	2160



CAAATCGGGT TTATAGACTT CCAAGGCACA AGCGCGAACA TGCCGCTCGT TCGCGACGTC	2220
GTTAAGCAGT GCCGTCGGCG CACTTTCAAC CAAACCAAGC GCTGTCCGGC CGACGTCGTT	2280
GCCACCACGT TTTTCCAGAG GTTGTACCCC GGGTGCACAA CCACCTCAGG GTGCGTCGCA	2340
TCCATCAGCC ACGTCGCCCC AGACTACCGC AACAGCCAGG CGCAAACGCT CTGCTTCACG	2400
CAGGAGGAAA AGTCGCGCCA CGGGGCTGAG GGCGCGATGA CTGTGCACGA AGCGCAGGGA	2460
CGCACTTTTG CGTCTGTCTAT TCTGCATTAC AACGGCTCCA CAGCAGAGCA GAAGCTCCTC	2520
GCTGAGAAGT CGCACCTTCT AGTCGGCATC ACGCGCCACA CCAACCACCT GTACATCCGC	2580
GACCCGACAG GTGACATTGA GAGAACTC AACCATAGCG CGAAAGCCGA GGTGTTTACA	2640
GACATCCCTG CACCCCTGGA GATCAGCACT GTCAAACCGA GTGAAGAGGT GCAGCGCAAC	2700
GAAGTGATGG CAACGATACC CCCGACAGGT GCCACGCCGC ACGGAGCAAT CCATCTGCTC	2760
CGCAAGAACT TCGGGGACCA ACCCGACTGT GGCTGTGTCT CTTTGGCGAA GACCGGCTAC	2820
GAGGTGTTTG GCGGTCGTGC CAAAATCAAC GTAGAGCTTG CCGAACCCGA CGCGACCCCG	2880
AAGCCGCATA GGGCGTTCCA GGAAGGGGTA CAGTGGGTCA AGGTCACCAA CGCGTCTAAC	2940
AAACACCAGG CGCTCCAGAC GCTGTTGTCC CGCTACACCA AGCGAAGCGC TGACCTGCCG	3000
CTACACGAAG CTAAGCAGGA CGTCAAACGC ATGCTAACT CGCTTGACCG ACATTGGGAC	3060
TGGACTGTCA CTGAAGACGC CCGTGACCGA GCTGTCTTCG AGACCCAGCT CAAGTTCACC	3120
CAACGCGGCG GCACCGTCGA AGACCTGCTG GAGCCAGACG ACCCCTACAT CCGTGACATA	3180
GACTTCCTTA TGAAGACTCA GCACAAAGTG TCGCCCAAGC CGATCAATAC GGGCAAGGTC	3240
GGGCAGGGGA TCGCCGCTCA CTCAAAGTCT CTCAACTTCG TCCTCGCCGC TTGGATACGC	3300
ATACTCGAGG AGATACTCCG TACCGGGAGC CGCACGGTCC GGTACAGCAA CGGTCTCCCC	3360
GACGAAGAAG AGGCCATGCT GCTCGAAGCG AAGATCAATC AAGTCCACA CGCCACGTTC	3420
GTCTCGGCGG ACTGGACCGA GTTGACACC GCCACAATA ACACGAGTGA GCTGCTCTTC	3480
GCCGCCCTTT TAGAGCGCAT CGGCACGCCT GCAGCTGCCG TTAATCTATT CAGAGAACGG	3540
TGTGGGAAAC GCACCTTGCG AGCGAAGGGT CTAGGCTCCG TTGAAGTCGA CGGTCTGCTC	3600
GACTCCGGCG CAGCTTGAC GCCTTGCCGC AACACCATCT TCTCTGCCGC CGTCATGCTC	3660
ACGCTCTTCC GCGGCGTCAA GTTCGCAGCT TTCAAAGGCG ACGACTCGCT CCTCTGTGGT	3720
AGCCATTACC TCGGTTTCGA CGCTAGCCGC CTTACATGG GCGAACGTTA CAAGACCAAA	3780
CATTTGAAGG TCGAGGTGCA GAAATCGTG CCGTACATCG GACTCCTCGT CTCCGCTGAG	3840
CAGGTGCTCC TCGACCCTGT CAGGAGCGCT CTCAAGATAT TTGGGCGCTG CTACACAAGC	3900
GAACTCCTTT ACTCCAAGTA CGTGGAGGCT GTGAGAGACA TCACCAAGGG CTGGAGTGAC	3960



106

GCCCGCTACC ACAGCCTCCT GTGCCACATG TCAGCATGCT ACTACAATTA CGCGCCGGAG	4020
TCTGCGGGGT ACATCATCGA CGCTGTTGTT CGCTTTGGGC GCGGCGACTT CCCGTTTGAA	4080
CAACTGCGCG TGGTGCCTGC CCATGTGCAG GCACCCGACG CTTACAGCAG CACGTATCCG	4140
GCTAACGTGC GCGCATCGTG CCTTGACCAC GTCTTCGAGC CCCGCCAGGC CGCCGCCCGC	4200
GCAGGTTTCG TTGCGAC ATG TGC GAA GCC GGA AAC GCC TTC TTC ACT TAC	4250
Met Cys Glu Ala Gly Asn Ala Phe Phe Thr Tyr	
1 5 10	
CGC GAA AGC TGG TGT TTC TGC GAC TAC AAG CCA CGT TGC GAC TGG GAC	4298
Arg Glu Ser Trp Cys Phe Cys Asp Tyr Lys Pro Arg Cys Asp Trp Asp	
15 20 25	
TGC GCC CCC GGA GTC TCC ATG GGA TGC ACC TGC AGC CAA CAG CTT TTC	4346
Cys Ala Pro Gly Val Ser Met Gly Cys Thr Cys Ser Gln Gln Leu Phe	
30 35 40	
GGA GTT ATT GAC ACC GGA GAC CCC GTC CAC ATC ATC CTC GCC GTC ATC	4394
Gly Val Ile Asp Thr Gly Asp Pro Val His Ile Ile Leu Ala Val Ile	
45 50 55	
GTC TTC ATC GGA CTC CTC TAC ATC GTG TGG AAG GTC GGT CAG TGG TGG	4442
Val Phe Ile Gly Leu Leu Tyr Ile Val Trp Lys Val Ala Gln Trp Trp	
60 65 70 75	
AGA CAC CGC AAG GAC CAC AGA AGA CTT GAA CAG CAG AAA GCC GCC TTC	4490
Arg His Arg Lys Asp His Arg Arg Leu Glu Gln Gln Lys Ala Ala Phe	
80 85 90	
GCA AGA CAG GCA ATC ACG CTC GTC TGAATGTC TGGACAGAAG CGGAGAAAG	4542
Ala Arg Gln Ala Ile Thr Leu Val	
95	
ACAGGCAGTT CGTTAACTGC CCCCACTGCT CCGAGCCCCT CATTCTCATT TTCGGAAAGA	4602
GCTCGACTGG CGACCGGGCC GACTGTGCCC GCTGCGACAT CACCTTCGGC AACCCCATCC	4662
TGCGCCACGG ACCAGGTTGC CGCGAGGACC ACGCCGGA CTGCGCCTTT CCTGGGTTCC	4722
CAGTCTGCCC GTGCTGTCTC GAAGCCGTAC CGGCCCCCA CCACTGCCCC TTGGAAAGAA	4782
GTCACCCCGC TCCACGCGTG GAAGGGCGTG ACCGGAGACC GACCGGAAGT CAGGGAGGAC	4842
CCGGAGACAG CGGCGGTCGT CCAGGCTCTG ATCAGCGGCC GTTATCCTCA GAAGACGAAG	4902
CTTCTCTCCG ACGCATCCAA AGGCTACTCA AGAACTAAGG GATGCTCACA ATCCACCTCT	4962
TTTCTGCCC CGAGTGCGGA TTACCAGGCC CGCGACTGCC AGACAGTCCG AGTCTGCCGC	5022
GCCGCTGCAG AGATGGCGCG CTCATGTATT CACGAGCCGT TGGCTTCATC TGCCGCCAGT	5082
GCCGACTTGA AGCGCATACG CTCTACCTCG GACTCTGTTC CCGATGTAAA GATCAGCAAG	5142
AGCGCATGAA GGAACAAAAT TAGTTTCCTT GTTCGTAAAC AAGGTGGTCC CTCCCATGTA	5202
GGTAAAGACT CTGGTGAGTC CTCAACGTTA CTCGTTGAGT CTGCTGCGGT TCGATTCCAT	5262
TCCCAAGCAG CAAAGGTGC GCAACTAGTA CGGCGCCCCC TGGGATACCA	5312



107

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 99 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

```

Met Cys Glu Ala Gly Asn Ala Phe Phe Thr Tyr Arg Glu Ser Trp Cys
 1          5          10          15
Phe Cys Asp Tyr Lys Pro Arg Cys Asp Trp Asp Cys Ala Pro Gly Val
          20          25          30
Ser Met Gly Cys Thr Cys Ser Gln Gln Leu Phe Gly Val Ile Asp Thr
          35          40          45
Gly Asp Pro Val His Ile Ile Leu Ala Val Ile Val Phe Ile Gly Leu
          50          55          60
Leu Tyr Ile Val Trp Lys Val Ala Gln Trp Trp Arg His Arg Lys Asp
          65          70          75          80
His Arg Arg Leu Glu Gln Gln Lys Ala Ala Phe Ala Arg Gln Ala Ile
          85          90          95
Thr Leu Val

```

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5312 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 4518..4937

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

```

GTTCTGCCTC CCCCGGACGG TAAATATAGG GGAACAATGT ACGCGAAAGC GACAGACGTG      60
GGCGGTGTCT ACGCCCGCGC AGATGTCGCC TACGCCAAGC TACTGCAGCA GAGAGCAGTC      120
AAGTTGGA CTGCCCCGCC ACTGAAGGCA CTAGAAACCC TCCACAGACT GTACTATCCG      180
CTGCGCTTCA AAGGGGGCAC TTTACCCCGG ACACAACACC CGATCCTGGC CGGGCACCAA      240

```



CGTGTCGCAG AAGAGGTTCT GCACAATTTT GCCAGGGGAC GTAGCACAGT GCTCGAGATA	300
GGGCGGTCTC TGCACAGCGC ACTTAAGCTA CATGGGGCAC CGAACGCCCC CGTCGCAGAC	360
TATCACGGGT GCACCAAGTA CGGCACCCGC GACGGCTCGC GACACATTAC GGCCTTAGAG	420
TCTAGATCCG TCGCCACAGG CCGGCCCCGAG TTCAAGGCCG ACGCCTCACT GCTCGCCAAC	480
GGCATTGCCT CCGGCACCTT CTGCGTCGAC GGAGTCGGCT CTTGCGCGTT CAAATCGCGC	540
GTTGGAATTG CCAATCACTC CCTCTATGAC GTGACCCTAG AGGAGCTGGC CAATGCGTTT	600
GAGAACCACG GACTTCACAT GGTCCGCGCG TTCATGCACA TGCCAGAAGA GCTGCTCTAC	660
ATGGACAACG TGGTTAATGC CGAGCTCGGC TACCGCTTCC ACGTTATTGA AGAGCCTATG	720
GCTGTGAAGC 3CATT CCAGGGGGGG GACCTCCGTC TCCACTTCCG TGAGTTGGAC	780
TTTCATCAACG AAGA GCGGCGCATC GAGAGGCTGG CCGCCCGCGG CTCCTACTCC	840
AGACGCGCGG TCATTTTCTC CGGCGACGAC GACTGGGGTG ATGCGTACTT ACACGACTTC	900
CACACATGGC TCGCTACCT ACTGGTGAGG AACTACCCCA CTCGTTTGG TTTCTCACTC	960
CATATAGAAG TCCAGAGGCG CCACGGCTCC AGCATTGAGC TCGGCATCAC TCGCGGCCA	1020
CCTGGAGACC GCATGCTGGC CGTCGTCCCA AGGACGTCCC AAGGCCTCTG CAGAATCCCA	1080
AACATCTTTT ATTACGCCGA CGCGTCGGGC ACTGAGCATA AGACCATCCT TACGTCACAG	1140
CACAAAGTCA ACATGCTGCT CAATTTTATG CAAACGCGTC CTGAGAAGGA ACTAGTCGAC	1200
ATGACCGTCT TGATGTCGTT CGCGCGCGCT AGGCTGCGCG CGATCGTGGT CGCCTCAGAA	1260
GTCACCGAGA GCTCCTGGAA CATCTCACCG GCTGACCTGG TCCGCACTGT CGTGCTCTT	1320
TACGTCCTCC ACATCATCGA GCGCCGAAGG GCTGCGGTCTG CTGTCAAGAC CGCCAAGGAC	1380
GACGTCTTTG GAGAGACTTC GTTCTGGGAG AGTCTCAAGC ACGTCTTGGG CTCCTGTTGC	1440
GGTCTGCGCA ACCTCAAAGG CACCGACGTC GTCTTTACTA AGCGCGTCGT CGATAAGTAC	1500
CGAGTCCACT CGCTCGGAGA CATAATCTGC GACGTCCGCC TGTCCCCTGA ACAGGTCGGC	1560
TTCTGCGGT CCGCGGTACC ACCTGCCCGG GTCTTTCACG ACAGGGAAGA GCTTGAGGTC	1620
CTTCGCGAAG CTGGCTGCTA CAACGAACGT CCGGTACCTT CCACTCCTCC TGTGGAGGAG	1680
CCCCAAGGTT TCGACGCCGA CTTGTGGCAC GCGACCGCGG CCTCACTCCC CGAGTACCGC	1740
GCCACCTTGC AGGCAGGTCT CAACACCGAC GTCAAGCAGC TCAAGATCAC CCTCGAGAAC	1800
GCCCTCAAGA CCATCGACGG GCTCACCCTC TCCCCAGTCA GAGGCCTCGA GATGTACGAG	1860
GGCCCGCCAG GCAGCGGCAA GACGGGCACC CTCATCGCCG CCCTTGAGGC CGCGGGCGGT	1920
AAAGCACTTT ACGTGGCACC CACCAGAGAA CTGAGAGAGG CTATGGACCG GCGGATCAAA	1980
CGGCGTCCG CCTCGGCTAC GCAACATGTC GCCCTTGCGA TTCTCCGTCTG TGCCACCGCC	2040
GAGGGCGCCC CTTTCGCTAC CGTGGTTATC GACGAGTGCT TCATGTTCCT GCTCGTGTAC	2100
GTCGCGATCG TGCACGCCTT GTCCCCGAGC TCACGAATAG TCCTTGTAGG GGACGTCCAC	2160



CAAATCGGGT TTATAGACTT CCAAGGCACA AGCGCGAACA TGCCGCTCGT TCGCGACGTC	2220
GTTAAGCAGT GCCCTCGGCG CACTTTCAAC CAAACCAAGC GCTGTCCGGC CGACGTCGTT	2280
GCCACCACGT TTTTCCAGAG CTTGTACCCC GGGTGCACAA CCACCTCAGG GTGCGTCGCA	2340
TCCATCAGCC ACGTCGCCCC AGACTACCGC AACAGCCAGG CGCAAACGCT CTGCTTCACG	2400
CAGGAGGAAA AGTCGCGCCA CGGGGCTGAG GCGCGGATGA CTGTGCACGA AGCGCAGGGA	2460
CGCACTTTTG CGTCTGTCAT TCTGCATTAC AACGGCTCCA CAGCAGAGCA GAAGCTCCTC	2520
GCTGAGAAGT CGCACCTTCT AGTCGGCATC ACGCGCCACA CCAACCACCT GTACATCCGC	2580
GACCCGACAG GTGACATTGA GAGACAACTC AACCATAGCG CGAAAGCCGA GGTGTTTACA	2640
GACATCCCTG CACCCCTGGA GATCAGCACT GTCAAACCGA GTGAAGAGGT GCAGCGCAAC	2700
GAAGTGATGG CAACGATACC CCCGCAGAGT GCCACGCCGC ACGGAGCAAT CCATCTGCTC	2760
CGCAAGAACT TCGGGGACCA ACCCGACTGT GGCTGTGTCT CTTTGGCGAA GACCGGCTAC	2820
GAGGTGTTTG GCGGTCGTGC CAAAATCAAC GTAGAGCTTG CCGAACCCGA CGCGACCCCG	2880
AAGCCGCATA GGGCGTTCCA GGAAGGGGTA CAGTGGGTCA AGGTCACCAA CGCGTCTAAC	2940
AAACACCAGG CGCTCCAGAC GCTGTTGTCC CGCTACACCA AGCGAAGCGC TGACCTGCCG	3000
CTACACGAAG CTAAGGAGGA CGTCAAACGC ATGCTAAACT CGCTTGACCG ACATTGGGAC	3060
TGGACTGTCA CTGAAGACGC CCGTGACCGA GCTGTCTTCG AGACCCAGCT CAAGTTCACC	3120
CAACGCGGCG GCACCGTCGA AGACCTGCTG GAGCCAGACG ACCCCTACAT CCGTGACATA	3180
GACTTCCTTA TGAAGACTCA GCAGAAAGTG TCGCCCAAGC CGATCAATAC GGGCAAGGTC	3240
GGGCAGGGGA TCGCCGCTCA CTCAAAGTCT CTCAACTTCG TCCTCGCCGC TTGGATACGC	3300
ATACTCGAGG AGATACTCCG TACCGGGAGC CGCACGGTCC GGTACAGCAA CCGTCTCCCC	3360
GACGAAGAAG AGGCCATGCT GTCGAAGCG AAGATCAATC AAGTCCACAA CGCCACGTTT	3420
GTCTCGGCGG ACTGGACCGA GTTGACACC GCCACAATA ACACGAGTGA GCTGCTCTTC	3480
GCCGCCCTTT TAGAGCGCAT CGGCACGCCCT GCAGCTGCCG TTAATCTATT CAGAGAACGG	3540
TGTGGGAAAC GCACCTTGCG AGCGAAGGGT CTAGGCTCCG TTGAAGTCGA CCGTCTGCTC	3600
GACTCCGGCG CAGCTTGAC GCCTTGCCGC AACACCATCT TCTCTGCCGC CGTCATGCTC	3660
ACGCTCTTCC GCGGCGTCAA GTTCGCAGCT TTCAAAGGCG ACGACTCGCT CCTCTGTGGT	3720
AGCCATTACC TCCGTTTCGA CGCTAGCCGC CTTACATGG CGGAACGTTA CAAGACCAAA	3780
CATTGAAGG TCGAGGTGCA GAAAATCGTG CCGTACATCG GACTCCTCGT CTCCGCTGAG	3840
CAGGTGCTCC TCGACCTGT CAGGAGCGCT CTCAAGATAT TTGGGCGCTG CTACACAAGC	3900
GAATCCTTT ACTCCAAGTA CGTGCAGGCT GTGAGAGACA TCACCAAGGG CTGGAGTGAC	3960



110

GCCCGCTACC ACAGCCTCCT GTGCCACATG TCAGCATGCT ACTACAATTA CGCGCCGGAG	4020
TCTGCGGCGT ACATCATCGA CGCTGTTGTT CGCTTTGGGC GCGGCGACTT CCCGTTTGAA	4080
CAACTGCGCG TGGTGCCTGC CCATGTGCAG GCACCCGACG CTTACAGCAG CACGTATCCG	4140
GCTAACGTGC GCGCATCGTG CCTTGACCAC GTCTTCGAGC CCGGCCAGGC CGCGCCCCCG	4200
GCAGGTTTCG TTGCGACATG TGCGAAGCCG GAAACGCCCTT CTCACTTAC CGCGAAAGCT	4260
GGTGTTCCTG CGACTACAAG CCACGTTGCG ACTGGGACTG CGCCCCCGGA GTCTCCATG	4320
GATGCACCTG CAGCCAACAG CTTTTCGGAG TTATTGACAC CGGAGACCCC GTCCACATCA	4380
TCCTCGCCGT CATCGTCTTC ATCGGACTCC TCTACATCGT GTCGAAGGTC GCTCAGTGGT	4440
GGAGACACCG CAAGGACCAC AGAAGACTTG AACAGCAGAA ACCCGCCTTC GCAAGACAGG	4500
CAATCAGGCT CGTCTGA ATG TCT GGA CAG AAG CGG AGA AAG GAC AGG CAG	4550
Met Ser Gly Gln Lys Arg Arg Lys Asp Arg Gln	
1 5 10	
TTC GTT AAC TGC CCC CAC TGC TCC GAG CCC CTC ATT CTC ATT TTC GGA	4598
Phe Val Asn Cys Pro His Cys Ser Glu Pro Leu Ile Leu Ile Phe Gly	
15 20 25	
AAG AGC TCG ACT GGC GAC CGG GCC GAC TGT CGC CGC TGC GAC ATC ACC	4646
Lys Ser Ser Thr Gly Asp Arg Ala Asp Cys Arg Arg Cys Asp Ile Thr	
30 35 40	
TTC GGC AAC CCC ATC CTG CGC CAC GGA CCA GGT TGC CGC GAG GAC CAC	4694
Phe Gly Asn Pro Ile Leu Arg His Gly Pro Gly Cys Arg Glu Asp His	
45 50 55	
GCC GGA CTT TGC GCC TTT CCT GGG TTC CCA GTC TGC CCG TGC TGT CTC	4742
Ala Gly Leu Cys Ala Phe Pro Gly Phe Pro Val Cys Pro Cys Cys Leu	
60 65 70 75	
GAA GCC GTA CCG GCC CCC CAC GAC TGC CCG TTG GAA AGA AGT CAC CCC	4790
Glu Ala Val Pro Ala Pro His Asp Cys Pro Leu Glu Arg Ser His Pro	
80 85 90	
GCT CCA CGC GTG GAA GGG CGT GAC CGG AGA CCG AOC GGA AGT CAG GGA	4838
Ala Pro Arg Val Glu Gly Arg Asp Arg Arg Pro Thr Gly Ser Gln Gly	
95 100 105	
GGA CCC GGA GAC AGC GGC GGT CGT CCA GGC TCT GAT CAG CGG CCG TTA	4886
Gly Pro Gly Asp Ser Gly Gly Arg Pro Gly Ser Asp Gln Arg Pro Leu	
110 115 120	
TCC TCA GAA GAC GAA GCT TTC CTC CGA CGC ATC CAA AGG CTA CTC AAG	4934
Ser Ser Glu Asp Glu Ala Phe Leu Arg Arg Ile Gln Arg Leu Leu Lys	
125 130 135	
AAC TAAGGGATGC TCACAATCCA CCTCTTTTCC TGCCCCGAGT GCGGATTACC	4987
Asn	
140	
AGGCCCCGGA CTGCCAGACA GTCCGAGTCT GCCGCGCCGC TGCAGAGATG GCGCGCTCAT	5047
GTATTCACGA GCCGTTGGCT TCATCTGCCG CCAGTGCCGA CTTGAAGCGC ATACGCTCTA	5107



111

CCTCGGACTC TGTTCCTCGAT GTAAAGATCA GCAAGAGG3C ATGAAGGAAC AAAATTAGTT	5167
TCCTTGTTTCG TAAACAACCT GGTCCCTCCC ATTGAGGTAA AGACTCTGGT GAGTCCTCAA	5227
CGTTACTCGT TGAGTCTGCT GCGGTTTCGAT TCCATTCCCA A3CAGCAAAG GGTGCGCAAC	5287
TAGTACGGCG CCCCTGGGA TACCA	5312

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 140 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Met	Ser	Gly	Gln	Lys	Arg	Arg	Lys	Asp	Arg	Gln	Phe	Val	Asn	Cys	Pro	1	5	10	15
His	Cys	Ser	Glu	Pro	Leu	Ile	Leu	Ile	Phe	Gly	Lys	Ser	Ser	Thr	Gly	20	25	30	
Asp	Arg	Ala	Asp	Cys	Arg	Arg	Cys	Asp	Ile	Thr	Phe	Gly	Asn	Pro	Ile	35	40	45	
Leu	Arg	His	Gly	Pro	Gly	Cys	Arg	Glu	Asp	His	Ala	Gly	Leu	Cys	Ala	50	55	60	
Phe	Pro	Gly	Phe	Pro	Val	Cys	Pro	Cys	Cys	Leu	Glu	Ala	Val	Pro	Ala	65	70	75	80
Pro	His	Asp	Cys	Pro	Leu	Glu	Arg	Ser	His	Pro	Ala	Pro	Arg	Val	Glu	85	90	95	
Gly	Arg	Asp	Arg	Arg	Pro	Thr	Gly	Ser	Gln	Gly	Gly	Pro	Gly	Asp	Ser	100	105	110	
Gly	Gly	Arg	Pro	Gly	Ser	Asp	Gln	Arg	Pro	Leu	Ser	Ser	Glu	Asp	Glu	115	120	125	
Ala	Phe	Leu	Arg	Arg	Ile	Gln	Arg	Leu	Leu	Lys	Asn					130	135	140	



112

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5312 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 4944..5162

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GTTCTGCCTC CCCCCGACGG TAAATATAGG GGAACAATGT ACGCGAAAGC GACAGACGTG	60
GCGCGTGTCT ACGCCGCGGC AGATGTCGCC TACGCGAACG TACTGCAGCA GAGAGCAGTC	120
AAGTTGGACT TCGCCCCGCC ACTGAAGGCA CTAGAAACCC TCCACAGACT GTACTATCCG	180
CTGCGCTTCA AAGGGGGCAC TTTACCCCCG ACACAACACC CGATCCTGGC CGGGCACCAA	240
CGTGTGCGAG AAGAGGTTCT GCACAATTTC GCCAGGGGAC GTAGCACAGT GCTCGAGATA	300
GGGCGGTCTC TGCACAGCGC ACTTAAGCTA CATGGGGCAC CGAACGCCCC CGTCGCAGAC	360
TATCACGGGT GCACCAAGTA CGGCACCCGC GACGGCTCGC GACACATTAC GGCCTTAGAG	420
TCTAGATCCG TCGCCACAGG CCGGCCCGAG TTCAAGGCCG ACGCCTCACT GCTCGCCAAC	480
GGCATTGCCT CCCGCACCTT CTGCGTCGAC GGAGTCGGCT CTTGCGCGTT CAAATCGCGC	540
GTTGGAATTG CCAATCACTC CCTCTATGAC GTGACCCTAG AGGAGCTGGC CAATGCGTTT	600
GAGAACCACG GACTTCACAT GGTCCGCGCG TTCATGCACA TGCLAGAAGA GCTGCTCTAC	660
ATGGACAACG TGGTTAATGC CGAGCTCGGC TACCGCTTCC ACGTTATTGA AGAGCCTATG	720
GCTGTGAAGG ACTGCGCATT CCAGGGGGGG GACCTCCGTC TCCACTTCCC TGAGTGGAC	780
TTCATCAACG AGAGCCAAGA GCGGCGCATC GAGAGGCTGG CCGCCCCGGG CTCCTACTCC	840
AGAGCGCGCG TCATTTTCTC CGGCGACGAC GACTGGGGTG ATGCGTACTT ACACGACTTC	900
CACACATGGC TCGCCTACCT ACTGGTGAGG AACTACCCCA CTCCGTTTGG TTTCTCACTC	960
CATATAGAAG TCCAGAGGCG CCACGGCTCC AGCATTGAGC TGCGCATCAC TCGCGCGCCA	1020
CCTGGAGACC GCATGCTGGC CGTCGTCCCA AGGACGTCCC AAGGCCTCTG CAGAATCCCA	1080
AACATCTTTT ATTACGCCGA CGCGTCGGGC ACTGAGCATA AGACCATCCT TACGTCACAG	1140
CACAAAGTCA ACATGCTGCT CAATTTTATG CAAACGCGTC CTGAGAAGGA ACTAGTCGAC	1200
ATGACCGTCT TGATGTCGTT CGCGCGCGCT AGGCTGCGCG CGATCGTGGT CGCCTCAGAA	1260



AMENDED SHEET
 IPEA/AU

GTCACCGAGA GCTCCTGGAA CATCTCACCG GCTGACCTGG TCCGCACTGT CGTGTCTCTT 1320
TACGTCTCTC ACATCATCGA GCGCCGAAGG GCTGCGGTGG CTGTCAAGAC CGCCAAGGAC 1380
GACGTCTTTG GAGAGACTTC GTTCTGGGAG AGTCTCAAGC ACGTCTTGGG CTCCTGTTGC 1440
GGTCTGCGCA ACCTCAAAGG CACCGACGTC GTCTTTACTA AGCGCGTCGT CGATAAGTAC 1500
CGAGTCCACT CGCTCGGAGA CATAATCTGC GACGTCCGCC TGTC TCTGA ACAGGTCCGGC 1560
TTCCTGCCGT CCCGCTACC ACCTGCCCCG GTCTTTCACG ACAGGGAAGA GCTTGAGGTC 1620
CTTCGCGAAG CTGGCTGCTA CAACGAACGT CCGGTACCTT CCACTCCTCC TGTGGAGGAG 1680
CCCCAAGGTT TCGACGCCGA CTTGTGGCAC GCGACCGCGG CCTCACTCCC CGAGTACCGC 1740
GCCACCTTGC AGGCAGGTCT CAACACCGAC GTCAAGCAGC TCAAGATCAC CCTCGAGAAC 1800
GCCCTCAAGA CGATCGACGG GCTCACCTC TCCCAGTCA GAGGCCTCGA GATGTACGAG 1860
GGCCCCGCG GCGACGGCAA GACGGGCACC CTCATCGCCG CCCTTGAGGC CGCGGGCGGT 1920
AAAGCACTTT ACGTGGCACC CACCAGAGAA CTGAGAGAGG CTATGGACCG GCGGATCAAA 1980
JCGCCGTCCG CCTCGGCTAC GCAACATGTC GCCCTTGCGA TTCTCCGTCC TGCCACCGCC 2040
GAGGGCGCCC CTTTCGCTAC CGTGGTTATC GACGAGTGCT TCATGTTCCC GCTCGTGTAC 2100
GTCGCGATCG TGCACGCCTT GTCCCCGAGC TCACGAATAG TCCTGTAGG GGACGTCCAC 2160
CAAATCGGGT TTATAGACTT CCAAGGCACA AGCGGGAACA TGCCGCTCGT TCGCGACGTC 2220
GTTAAGCAGT GCCGTGGCG CACTTTCAAC CAAACCAAGC GCTGTCCGGC CGACGTGCTT 2280
GJCACCACGT TTTTCAGAG CTTGTACCCC GGGTGCACAA CCACCTCAGG GTGCGTCGCA 2340
TCCATCAGCC ACGTGGCCCC AGACTACCGC AACAGCCAGG CGCAAACGCT CTGCTTCAGG 2400
CAGGAGGAAA AGTCGGGCCA CGGGGCTGAG GCGCGCATGA CTGTGCACGA AGCGCAGGGA 2460
CGCACTTTTG CGTCTGTCTAT TCTGCATTAC AACGGCTCCA CAGCAGAGCA GAAGCTCCTC 2520
GCTGAGAAGT CGCACCTTCT AGTCGGCATC ACGCGCCACA CCAACGACCT GTACATCCGC 2580
GACCCGACAG GTGACATTGA GAGACAATC AACCATAGCG CGAAAGCCGA GGTGTTTACA 2640
GACATCCCTG CACCCCTGGA GATCAGGACT GTCAAACCGA GTGAAGAGGT GCAGCGCAAC 2700
GAAGTGATGG CAACGATACC CCCGCAGAGT GCCAGCCCGC ACGGAGCAAT CCATCTGCTC 2760
CGCAAGAACT TCGGGGACCA ACCCGACTGT GGCTGTGTCC CTTTGGCGAA GACCGGCTAC 2820
GAGGTGTTTG GCGGTGCTGC CAAAATCAAC GTAGAGCTTG CCGAACCCGA CGCGACCCCG 2880
AAGCCGCATA GGGCGTTCCA GGAAGGGGTA CAGTGGGTCA AGGTCACCAA CGCGTCTAAC 2940
AAACACCAGC CGCTCCAGAC GCTGTTGTCC CGCTACACCA AGCGAAGCGC TGACCTGCGC 3000
CTACACGAAG CTAAGGAGGA CGTCAAACGC ATGCTAAACT CGCTTGACCG ACATTGGGAC 3060



114

TGGAAGACGC CCGTGACCG/ GCTGTCTTCG AGACCCAGCT CAAGTTCACC	3120
CAACGCGGCG GCACCGTCGA AGACCTGCTG GAGCCAGACG ACCCTTACAT CCGTGACATA	3180
GACTTCCTTA TGAAGACTCA GCACAAAGTG TCGCCCAAGC CGATCAATAC GGGCAAGGTC	3240
GGGCASGGGA TCGCGGTCA CTCAAAGTCT CTCAACTTCG TCCTCGCCGC TTGGATACG	3300
ATACTCGAGG AGATACTCCG TACCGGGAGC CGCACGGTCC GGTACAGCAA CGCTCTCCCC	3360
GACGAAGAAG AGGCCATGCT GCTCGAAGCG AAGATCAATC AAGTCCCACA CGCCACGTTC	3420
GTCTCGGCGG ACTGGACCGA GTTTGACACC GCCCACAATA ACACGAGTGA GCTGCCTTC	3480
GCCGCCCTTT TAGAGCGCAT CGGCACGCCT GCAGCTGCCG TTAATCTATT CAGAGAACGG	3540
TGTGGGAAAC GCACCTTGGC AGCGAAGGCT CTAGGCTCCG CTGAAGTCTGA CGGTCTGCTC	3600
GACTCGGCGG CAGCTTGGAC GCCTTGCCGC AACACCATCT TCTCTGCCGC CGTCATGCTC	3660
ACGCTCTTCC GCGGCGTCAA GTTCGCAGCT TTCAAAGGCG ACGACTCGCT CCTCTGTGGT	3720
AGCCATTACC TCCGTTTCGA CGCTAGCCGC CTTACATGG CGGAACGTA CAAGACCAAA	3780
CATTTGAAGG TCGAGGTGCA GAAAATCGTG CCGTACATCG GACTCCTCGT CTCCGCTGAG	3840
CAGGTGCTCU TCGACCCCTG CAGGAGCGCT CTCAAGATAT TTGGGCGCTG CTACACAAGC	3900
GAACTCCTTT ACTCCAAGTA GGTGGAGGCT GTGAGAGACA TCACCAAGGG GTGGAGTGAC	3960
GCCCGCTACC ACAGCCTCCT GTGCCACATG TCAGCATGCT ACTACAATTA CGCGCCGGAG	4020
TCTGCGGCGT ACATCATCGA CGCTGTTGTT CGCTTTGGGC GCGGCGACTT CCGGTTTGAA	4080
CAACTGCGCG TGGTGGCTGC CCATGTGCAG GCACCCGACG CTTACAGCAG CACGTATCCG	4140
GCTAACGTGC GCGCATCGTG CCTTGACCAC GTCTTCGAGC CCGGCGAGG CGCGCCCCG	4200
GCAGGTTTCG TTGACATG TCGGAAGCCG GAAACGCCTT CTTCACTTAC CGCGAAAGCT	4260
GGTGTTCCTG CGACTACAAG CCACGTTGCG ACTGGGACTG CGCCCGCGGA GTCTCCATGG	4320
GATGCACCTG CAGCCAACAG CTTTTCGGAG TTAATTGACAC CGGAGACCCC GTCCACATCA	4380
TCCTCGGCGT CATCGTCTTC ATCGGACTCC TCTACATCGT GTGGAAGGTC GCTCAGTGGT	4440
GGAGACACCG CAAGGACCAC AGAAGACTTG AACAGCAGAA AGCCGCCTTC GCAAGACAGG	4500
CAATCACGCT CGTCTGAATG TCTGGACAGA AGCGGAGAAA GGACAGGCAG TTCGTTAACT	4560
GCCCCCACTG CTCCGAGCCC CTCATTCTCA TTTTCGGAAA GAGCTCGACT GCGGACCGGG	4620
CCGACTGTGC CCGCTGCGAC ATCACCTTCG GCAACCCCAT CCTGCGCCAC GGACCAGGTT	4680
GCCGCGAGGA CCACGCCGGA CTTTGCGCCT TTCTGGGTT CCCAGCTGC CCGTGCTGTC	4740
TCGAAGCCGT ACCGGCCCCC CACGACTGCC CGTTGGAAAG AAGTCACCCC GCTCCACGGG	4800
TGGAAGGCGG TGACCGGAGA CCGACCGGAA GTCAAGGAGG ACCCGGAGAC AGCGGCGGTC	4860
GTCCAGGCTC TGATCAGCGG CCGTTATCCT CAGAAGACGA AGCTTTCCTC CGACGCATCC	4920



115

AAAGGCTACT CAAGAACTAA GGG ATG CTC ACA ATC CAC CTC TTT TCC TGC	4970
Met Leu Thr Ile His Leu Phe Ser Cys	
1 5	
CCC GAG TGC GGA TTA CCA GGC CCG CGA CTG CCA GAC AGT CCG AGT CTG	5018
Pro Glu Cys Gly Leu Pro Gly Pro Arg Leu Pro Asp Ser Pro Ser Leu	
10 15 20 25	
CCG CGC CGC TGC AGA GAT GGC GCG CTC ATG TAT TCA CGA GCC GTT GGC	5066
Pro Arg Arg Cys Arg Asp Gly Ala Leu Met Tyr Ser Arg Ala Val Gly	
30 35 40	
TTC ATC TGC CGC CAG TGC CGA CTT GAA GCG CAT ACG CTC TAC CTC GGA	5114
Phe Ile Cys Arg Gln Cys Arg Leu Glu Ala His Thr Leu Tyr Leu Gly	
45 50 55	
CTC TGT TCC CGA TGT AAA GAT CAG CAA GAG GGC ATG AAG GAA CAA AAT	5162
Leu Cys Ser Arg Cys Lys Asp Gln Gln Glu Arg Met Lys Glu Gln Asn	
60 65 70	
TAGTTTCCTT GTTCGTAAAC AAGGTGGTCC CTCCCATTGA GGTAAGACT CTGGTGAGTC	5222
CTCAACGTTA CTCGTTGAGT CTGCTGCGGT TCGATTCCAT TCCCAAGCAG CAAAGGGTGC	5282
GCAACTAGTA CGGCGCCCCC TGGGATACCA	5312

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Met	Leu	Thr	Ile	His	Leu	Phe	Ser	Cys	Pro	Glu	Cys	Gly	Leu	Pro	Gly
1				5					10					15	
Pro	Arg	Leu	Pro	Asp	Ser	Pro	Ser	Leu	Pro	Arg	Arg	Cys	Arg	Asp	Gly
			20					25					30		
Ala	Leu	Met	Tyr	Ser	Arg	Ala	Val	Gly	Phe	Ile	Cys	Arg	Gln	Cys	Arg
		35					40					45			
Leu	Glu	Ala	His	Thr	Leu	Tyr	Leu	Gly	Leu	Cys	Ser	Arg	Cys	Lys	Asp
	50					55					60				
Gln	Gln	Glu	Arg	Met	Lys	Glu	Gln	Asn							
65					70										



116

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2478 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 283..753

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GTTTTTCTTT CTTTACCAAG TGTGGTAAAA TTAAACAAA GAAGAAAACC AGGACCGTAA	60
CCCGGCCCTT ACACACCTCG AGTCCGTGAC CACCGGATTA TACGTGCCCC ACCACACGGC	120
GCCTTTTCCG ACCACTCTCG AGAGTCGTTG GGAGTTTCGT CCGTGACCAC CCGGTTGGCA	180
GTCGACAGAC GCTTCCGGAC CACTAGAACC TCCTCGAGCG AAGGACACAC AGCACACACA	240
CCGCCTTAGC TGCACCTACG GCAGCGTTGA TAGCGCGGAT TT ATG AGC GAG CAC	294
Met Ser Glu His	
1	
ACC ATC GCC CAC TCC ATC ACA TTA CCA CCC GGT TAC ACC CTT GCC CTA	342
Thr Ile Ala His Ser Ile Thr Leu Pro Pro Gly Tyr Thr Leu Ala Leu	
5 10 15 20	
ATA CCC CCT GAA CCT GAA GCA GGA TGG GAG ATG CTG GAG TGG CGT CAC	390
Ile Pro Pro Glu Pro Glu Ala Gly Trp Glu Met Leu Glu Trp Arg His	
25 30 35	
AGC GAC CTC ACA ACC GTC GCG GAA CCC GTA ACG TTC GGG TCA GCG CCA	438
Ser Asp Leu Thr Thr Val Ala Glu Pro Val Thr Phe Gly Ser Ala Pro	
40 45 50	
ACA CCG TCA CCG TCA ATG GTA GAA GAA ACC AAC GGC GTC GGA CCG GAA	486
Thr Pro Ser Pro Ser Met Val Glu Glu Thr Asn Gly Val Gly Pro Glu	
55 60 65	
GGC AAG TTT CTC CCC CTG ACA ATT TCA CCG CTG CTG CAC AAG ACC TCG	534
Gly Lys Phe Leu Pro Leu Thr Ile Ser Pro Leu Leu His Lys Thr Ser	
70 75 80	
CGC AAA GCC TTG ACG CCA ACA CCG TCA CTT TCC CCG CTA ACA TCT CTA	582
Arg Lys Ala Leu Thr Pro Thr Pro Ser Leu Ser Pro Leu Thr Ser Leu	
85 90 95 100	
GCA TGC CCG AAT TCC GGA ATT GGG CCA AGG GAA AGA TCG ACC TCG ACT	630
Ala Cys Pro Asn Ser Gly Ile Gly Pro Arg Glu Arg Ser Thr Ser Thr	
105 110 115	



117

CCG ATT CCA TCG GCT GGT ACT TCA AGT ACC TTG ACC CAG CGG GTG CTA	678
Pro Ile Pro Ser Ala Gly Thr Ser Ser Thr Leu Thr Gln Arg Val Leu	
120 125 130	
CAG AGT CTG CGC GCG CCG TCG GCG AGT ACT CGA AGA TCC CTG ACG GCC	726
Gln Ser Leu Arg Ala Pro Ser Ala Ser Thr Arg Arg Ser Leu Thr Ala	
135 140 145	
TCG TCA AGT TCT CCG TCG ACG CAG AGA TAAGAGAGAT CTATAACGAG	773
Ser Ser Ser Ser Pro Ser Thr Gln Arg	
150 155	
GAGTGGCCCG TCGTCACTGA CGTGTCGGTC CCCCTCGACG GCCGCCAGTG GAGCCTCTCG	833
ATTTTCTCCT TTCCGATGTT CAGAACCGCC TACGTCGCCG TAGCGAACGT CGAGAACAAG	893
GAGATGTCCG TCGACGTTGT CAACGACCTC ATCGAGTGGC TCAACAATCT CGCCGACTGG	953
CGTTATGTCC TTGACTCTGA ACAGTGGATT AACTTCACCA ATGACACCAC GTACTACGTG	1013
CGCATCCGGG TTCTACGTCC AACCTACGAC GTTCCAGACC CCACAGAGGG CCTTGTTCCG	1073
ACAGTCTCAG ACTACCGCCT CACTTATAAG GCGATAACAT GTGAAGCCAA CATGCCAACA	1133
CTCGTCGACC AAGGCTTTTG GATCGGCGGC CAGTACGCTC TCA JCCGAC TAGCCTACCG	1193
CAGTACGACG TCAGCGAGGC CTACGCTCTG CACACTTTGA CCTTCGCCAG ACCATCCAGC	1253
GCCGCTGCAC TCGCGTTTGT GTGGGCAGGT TTGCCACAGG GTGGCACTGC GCCTGCAGGC	1313
ACTCCAGCCT GGGAGCAGGC ATCCTCGGGT GGCTACCTCA CCTGGGCCCA CAACGGTACT	1373
ACTTTCCCAG CTGGCTCCGT TAGCTACJTT CTCCCTGAGG GTTTCGCCCT TGAGCGCTAC	1433
GACCCGAACG ACGGCTCTTG GACCGACTTC GCTTCGCAG GAGACACCGT CACTTTCGG	1493
CAGGTGCGCG TCGACGAGGT CGTTGTGACC AACAACCCCG CCGGCGGCGG CAGCGCCCC	1553
ACCTTACCG TGAGAGTGCC CCCTTCAAAC GCTTACACCA ACACCGTGTT TAGGAACAGC	1613
CTCTTAGAGA CTCGACCCTC CTCTCGTAGG CTCGAACTCC CTATGCCACC TGCTGACTTT	1673
GGACAGACGG TCGCCAACAA CCCGAAGATC GAGCAGTCGC TTCTTAAAGA AACACTTGGC	1733
TGCTATTTGG TCCACTCCAA AATGCGAAAC CCCGTTTTCC AGCTCACGCC AGCCAGCTCC	1793
TTTGCGCGCG TTTCTTCAA CAATCCGGGT TATGAGCGCA CACGCGACCT CCCGGACTAC	1853
ACTGGCATCC GTGACTCATT CGACCAGAAC ATGTCCACCG CTGTGGCCCA CTTCCGCTCA	1913
CTCTCCCACT CCTGCAGTAT CGTACTAAG ACCTACCAGG GTTGGGAAGG CGTCACGAAC	1973
GTCAACAAGC CTTTCGGCCA ATTCGCGCAC GCGGGCCTCC TCAAGAATGA GGAGATCCTC	2033
TGCCTGCGCG ACGACCTGGC CACCCGTCTC ACAGGTGTCT ACCCGGCCAC TGACAACTTC	2093
GCGGCGCGCG TTTCTGCCTT CGCCGCGAAC ATGCTGTCTT CCGTGCTGAA GTCGGAGGCA	2153
ACGTCTCTCA TCATCAAGTC CGTTGGCGAG ACTGCCGTCT GCGCGGCTCA GTCCGGCCTC	2213



118

GCGAAGCTAC CCGGACTGCT AATGAGTGTA CCAGGGAAGA TTGCCGCGCG TGTCCGCGCG	2273
CGCCGAGCGC GCCGCCGCCG CGCTCGTGCC AATTAGTTTG CTCGCTCCTG TTTCCGCCGT	2333
TCGTAAAACG GCGTGGTCCC GCACATTACG CGTACCCTAA AGACTCTGGT GAGTCCCCGT	2393
CGTTACACGA CGGGTCTGCC GCGGTTTCGAT TCCATTCCCA AGCGGCAAGA AGGACGTAGT	2453
TAGCTCTGCG TCCCTCGGGA TACCA	2478

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 157 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Met	Ser	Glu	His	Thr	Ile	Ala	His	Ser	Ile	Thr	Leu	Pro	Pro	Gly	Tyr
1				5					10					15	
Thr	Leu	Ala	Leu	Ile	Pro	Pro	Glu	Pro	Glu	Ala	Gly	Trp	Glu	Met	Leu
			20					25					30		
Glu	Trp	Arg	His	Ser	Asp	Leu	Thr	Thr	Val	Ala	Glu	Pro	Val	Thr	Phe
		35				40						45			
Gly	Ser	Ala	Pro	Thr	Pro	Ser	Pro	Ser	Met	Val	Glu	Glu	Thr	Asn	Gly
	50					55					60				
Val	Gly	Pro	Glu	Gly	Lys	Phe	Leu	Pro	Leu	Thr	Ile	Ser	Pro	Leu	Leu
	65				70				75						80
His	Lys	Thr	Ser	Arg	Lys	Ala	Leu	Thr	Pro	Thr	Pro	Ser	Leu	Ser	Pro
				85					90					95	
Leu	Thr	Ser	Leu	Ala	Cys	Pro	Asn	Ser	Gly	Ile	Gly	Pro	Arg	Glu	Arg
			100					105					110		
Ser	Thr	Ser	Thr	Pro	Ile	Pro	Ser	Ala	Gly	Thr	Ser	Ser	Thr	Leu	Thr
		115					120					125			
Gln	Arg	Val	Leu	Gln	Ser	Leu	Arg	Ala	Pro	Ser	Ala	Ser	Thr	Arg	Arg
	130					135					140				
Ser	Leu	Thr	Ala	Ser	Ser	Ser	Ser	Pro	Ser	Thr	Gln	Arg			
145					150					155					



119

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2478 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 366..2306

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

```

GTTTTCTTT CTTTACCAAG TGTGCTAAAA TTAAACAAA GAAGAAAACC AGGACCGTAA      60
CCCGGCCCTT ACACACCTCG AGTCCGTGAC CACCGGATTA TACGTCGCCC ACCACACGGC      120
GCCTTTTCCG ACCACTCTCG AGAGTCGTTG GGAGTTTCGT CCGTGACCAC CCGGTTGGCA      180
GTGACAGAC GCTTCCGAC CACTAGAACC TCCTCGAGCG ACGCACACAC AGCACACACA      240
CCGCCTTAGC TGCACCTACG GCACGGTTGA TAGCGCGGAT TTATGAGCGA GCACACCATC      300
GCCCCTTCCA TCACATTACC ACCCGGTTAC ACCCTTGCCC TAATACCCCC TGAACCTGAA      360
GCAGG ATG GGA GAT GCT CGA GTG GCG TCA CAG CGA CCT CAC AAC CGT      407
      Met Gly Asp Ala Gly Val Ala Ser Gln Arg Pro His Asn Arg
      1           5           10

CGC GGA ACC CGT AAC GTT CGG GTC AGC GCC AAC ACC GTC ACC GTC AAT      455
Arg Gly Thr Arg Asn Val Arg Val Ser Ala Asn Thr Val Thr Val Asn
      15           20           25           30

GGT AGA AGA AAC CAA CGG CGT CGG ACC GGA AGG CAA GTT TCT CCC CCT      503
Gly Arg Arg Asn Gln Arg Arg Arg Thr Gly Arg Gln Val Ser Pro Pro
      35           40           45

GAC AAT TTC ACC GCT GCT GCA CAA GAC CTC GCG CAA AGC CTT GAC GCC      551
Asp Asn Phe Thr Ala Ala Ala Gln Asp Leu Ala Gln Ser Leu Asp Ala
      50           55           60

AAC ACC GTC ACT TTC CCC GCT AAC ATC TCT AGC ATG CCC GAA TTC CGG      599
Asn Thr Val Thr Phe Pro Ala Asn Ile Ser Ser Met Pro Glu Phe Arg
      65           70           75

AAT TGG GCC AAG GGA AAG ATC GAC CTC GAC TCC GAT TCC ATC GGC TGG      647
Asn Trp Ala Lys Gly Lys Ile Asp Leu Asp Ser Asp Ser Ile Gly Trp
      80           85           90

TAC TTC AAG TAC CTT GAC CCA GCG GGT GCT ACA GAG TCT GCG CGC GCC      695
Tyr Phe Lys Tyr Leu Asp Pro Ala Gly Ala Thr Glu Ser Ala Arg Ala
      95           100           105           110

GTC GGC GAG TAC TCG AAG ATC CCT GAC GGC CTC GTC AAG TTC TCC GTC      743
Val Gly Glu Tyr Ser Lys Ile Pro Asp Gly Leu Val Lys Phe Ser Val
      115           120           125

```



120

GAC Asp	GCA Ala	GAG Glu	ATA Ile	AGA Arg	GAG Glu	ATC Ile	TAT Tyr	AAC Asn	GAG Glu	GAG Glu	TGC Cys	CCC Pro	GTC Val	GTC Val	ACT Thr	791
			130					135					140			
GAC Asp	GTG Val	TCC Ser	GTC Val	CCC Pro	CTC Leu	GAC Asp	GGC Gly	CGC Arg	CAG Gln	TGG Trp	AGC Ser	CTC Leu	TCG Ser	ATT Ile	TTC Phe	839
		145					150					155				
TCC Ser	TTT Phe	CCG Pro	ATC Met	TTC Phe	AGA Arg	ACC Thr	GCC Ala	TAC Tyr	GTC Val	GCC Ala	GTA Val	GCG Ala	AAC Asn	GTC Val	GAG Glu	887
	160					165					170					
AAC Asn	AAG Lys	GAG Glu	ATG Met	TCG Ser	CTC Leu	GAC Asp	GTT Val	GTC Val	AAC Asn	GAC Asp	CTC Leu	ATC Ile	CAG Glu	TGG Trp	CTC Leu	935
	175				180					185					190	
AAC Asn	AAT Asn	CTC Leu	GCC Ala	GAC Asp	TGG Trp	CGT Arg	TAT Tyr	GTC Val	GTT Val	GAC Asp	TCT Ser	GAA Glu	CAG Gln	TGG Trp	ATT Ile	983
				195					200					205		
AAC Asn	TTC Phe	ACC Thr	AAT Asn	GAC Asp	ACC Thr	ACG Thr	TAC Tyr	TAC Tyr	GTC Val	CGC Arg	ATC Ile	CGC Arg	GTT Val	CTA Leu	CGT Arg	1031
			210					215					220			
CCA Pro	ACC Thr	TAC Tyr	GAC Asp	GTT Val	CCA Pro	GAC Asp	CCC Pro	ACA Thr	GAG Glu	GCC Gly	CTT Leu	GTT Val	CGC Arg	ACA Thr	GTC Val	1079
		225					230					235				
TCA Ser	GAC Asp	TAC Tyr	CGC Arg	CTC Leu	ACT Thr	TAT Tyr	AAG Lys	GCG Ala	ATA Ile	ACA Thr	TGT Cys	GAA Glu	GCC Ala	AAC Asn	ATG Met	1127
	240					245					250					
CCA Pro	ACA Thr	CTC Leu	GTC Val	GAC Asp	CAA Gln	GGC Gly	TTT Phe	TGG Trp	ATC Ile	GGC Gly	GGC Gly	CAG Gln	TAC Tyr	GCT Ala	CTC Leu	1175
	255				260					265					270	
ACC Thr	CCG Pro	ACT Thr	AGC Ser	CTA Leu	CCG Pro	CAG Gln	TAC Tyr	GAC Asp	GTC Val	AGC Ser	GAG Glu	GCC Ala	TAC Tyr	GCT Ala	CTG Leu	1223
				275					280					285		
CAC His	ACT Thr	TTG Leu	ACC Thr	TTC Phe	GCC Ala	AGA Arg	CCA Pro	TCC Ser	AGC Ser	GCC Ala	GCT Ala	GCA Ala	CTC Leu	GCG Ala	TTT Phe	1271
			290					295					300			
GTG Val	TGG Trp	GCA Ala	GGT Gly	TTG Leu	CCA Pro	CAG Gln	GGT Gly	GGC Gly	ACT Thr	GCG Ala	CCT Pro	GCA Ala	GGC Gly	ACT Thr	CCA Pro	1319
		305					310					315				
GCC Ala	TGG Trp	GAG Glu	CAG Gln	GCA Ala	TCC Ser	TCC Ser	GGT Gly	GGC Gly	TAC Tyr	CTC Leu	ACC Thr	TGG Trp	CGC Arg	CAC His	AAC Asn	1367
	320					325					330					
GGT Gly	ACT Thr	ACT Thr	TTC Phe	CCA Pro	GCT Ala	GGC Gly	TCC Ser	GTT Val	AGC Ser	TAC Tyr	GTT Val	CTC Leu	CCT Pro	GAG Glu	GGT Gly	1415
	335				340					345					350	
TTC Phe	GCC Ala	CTT Leu	GAG Glu	CGC Arg	TAC Tyr	GAC Asp	CCG Pro	AAC Asn	GAC Asp	GGC Gly	TCT Ser	TGG Trp	ACC Thr	GAC Asp	TTC Phe	1463
				355					360					365		



121

GCT	TCC	GCA	GGA	GAC	ACC	GTC	ACT	TTC	CGG	CAG	GTC	GCC	GTC	GAC	GAG	1511
Ala	Ser	Ala	Gly	Asp	Thr	Val	Thr	Phe	Arg	Gln	Val	Ala	Val	Asp	Glu	
			370					375					380			
GTC	GTT	GTG	ACC	AAC	AAC	CCC	GCC	GGC	GGC	GGC	AGC	GCC	CCC	ACC	TTC	1559
Val	Val	Val	Thr	Asn	Asn	Pro	Ala	Gly	Gly	Gly	Ser	Ala	Pro	Thr	Phe	
			385				390					395				
ACC	GTG	AGA	GTG	CCC	CCT	TCA	AAC	GCT	TAC	ACC	AAC	ACC	GTG	TTT	AGG	1607
Thr	Val	Arg	Val	Pro	Pro	Ser	Asn	Ala	Tyr	Thr	Asn	Thr	Val	Phe	Arg	
	400					405					410					
AAC	ACG	CTC	TTA	GAG	ACT	CGA	CCC	TCC	TCT	CGT	AGG	CTC	GAA	CTC	CCT	1655
Asn	Thr	Leu	Leu	Glu	Thr	Arg	Pro	Ser	Ser	Arg	Arg	Leu	Glu	Leu	Pro	
	415				420					425					430	
ATG	CCA	CCT	GCT	GAC	TTT	GGA	CAG	ACG	GTC	GCC	AAC	AAC	CCG	AAG	ATC	1703
Met	Pro	Pro	Ala	Asp	Phe	Gly	Gln	Thr	Val	Ala	Asn	Asn	Pro	Lys	Ile	
				435					440					445		
GAG	CAG	TCG	CTT	CTT	AAA	GAA	ACA	CTT	GGC	TGC	TAT	TTG	GTC	CAC	TCC	1751
Glu	Gln	Ser	Leu	Leu	Lys	Glu	Thr	Leu	Gly	Cys	Tyr	Leu	Val	His	Ser	
			450					455					460			
AAA	ATG	CGA	AAC	CCC	GTT	TTC	CAG	CTC	ACG	CCA	GCC	AGC	TCC	TTT	GGC	1799
Lys	Met	Arg	Asn	Pro	Val	Phe	Gln	Leu	Thr	Pro	Ala	Ser	Ser	Phe	Gly	
		465					470					475				
GCC	GTT	TCC	TTC	AAC	AAT	CCG	GGT	TAT	GAG	CGC	ACA	CGC	GAC	CTC	CCG	1847
Ala	Val	Ser	Phe	Asn	Asn	Pro	Gly	Tyr	Glu	Arg	Thr	Arg	Asp	Leu	Pro	
	480					485					490					
GAC	TAC	ACT	GGC	ATC	CGT	GAC	TCA	TTC	GAC	CAG	AAC	ATG	TCC	ACC	GCT	1895
Asp	Tyr	Thr	Gly	Ile	Arg	Asp	Ser	Phe	Asp	Gln	Asn	Met	Ser	Thr	Ala	
					500					505					510	
GTG	GCC	CAC	TTC	CGC	TCA	CTC	TCC	CAC	TCC	TGC	AGT	ATC	GTC	ACT	AAG	1943
Val	Ala	His	Phe	Arg	Ser	Leu	Ser	His	Ser	Cys	Ser	Ile	Val	Thr	Lys	
				515					520					525		
ACC	TAC	CAG	GGT	TGG	GAA	GGC	GTC	ACG	AAC	GTC	AAC	ACG	CCT	TTC	GGC	1991
Thr	Tyr	Gln	Gly	Trp	Glu	Gly	Val	Thr	Asn	Val	Asn	Thr	Pro	Phe	Gly	
			530					535					540			
CAA	TTC	GGG	CAC	GCG	GGC	CTC	CTC	AAG	AAT	GAG	GAG	ATC	CTC	TGC	CTC	2039
Gln	Phe	Ala	His	Ala	Gly	Leu	Leu	Lys	Asn	Glu	Glu	Ile	Leu	Cys	Leu	
			545				550					555				
GCC	GAC	GAC	CTG	GCC	ACC	CGT	CTC	ACA	GGT	GTC	TAC	CCC	GCC	ACT	GAC	2087
Ala	Asp	Asp	Leu	Ala	Thr	Arg	Leu	Thr	Gly	Val	Tyr	Pro	Ala	Thr	Asp	
			560			565					570					
AAC	TTC	GGG	GCC	GCC	GTT	TCT	GCC	TTC	GCC	GGG	AAC	ATG	CTG	TCC	TCC	2135
Asn	Phe	Ala	Ala	Ala	Val	Ser	Ala	Phe	Ala	Ala	Asn	Met	Leu	Ser	Ser	
					580					585					590	
GTG	CTG	AAG	TCG	GAG	GCA	ACG	TCC	TCC	ATC	ATC	AAG	TCC	GTT	GGC	GAG	2183
Val	Leu	Lys	Ser	Glu	Ala	Thr	Ser	Ser	Ile	Ile	Lys	Ser	Val	Gly	Glu	
				595					600					605		



122

ACT GCC GTC GGC GCG GCT CAG TCC GGC CTC GCG AAG CTA CCC GGA CTG	2231
Thr Ala Val Gly Ala Ala Gln Ser Gly Leu Ala Lys Leu Pro Gly Leu	
610 615 620	
CTA ATG AGT GTA CCA GGG AAG ATT GCC GCG CGT GTC CGC GCG CGC CGA	2279
Leu Met Ser Val Pro Gly Lys Ile Ala Ala Arg Val Arg Ala Arg Arg	
625 630 635	
GCG CGC CGC CGC GGC GCT CGT GCC AAT TAGTTTGCTC GCTCCTGTTT	2326
Ala Arg Arg Arg Ala Ala Arg Ala Asn	
640 645	
CGCCGTTTCG TAAACGGCG TGGTCCCGCA CATTACGCGT ACCCTAAAGA CTCTGGTGAG	2386
TCCCGTCGT TACACGACGG GTCTGCCGCG GTTCGATTCC ATTCCCAAGC GGCAAGAAGG	2446
ACGTAGTTAG CTCTGCGTCC CTCGGGATAC CA	2478

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 647 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Met	Gly	Asp	Ala	Gly	Val	Ala	Ser	Gln	Arg	Pro	His	Asn	Arg	Arg	Gly
1				5					10					15	
Thr	Arg	Asn	Val	Arg	Val	Ser	Ala	Asn	Thr	Val	Thr	Val	Asn	Gly	Arg
		20						25					30		
Arg	Asn	Gln	Arg	Arg	Arg	Thr	Gly	Arg	Gln	Val	Ser	Pro	Pro	Asp	Asn
		35					40					45			
Phe	Thr	Ala	Ala	Ala	Gln	Asp	Leu	Ala	Gln	Ser	Leu	Asp	Ala	Asn	Thr
	50				55						60				
Val	Thr	Phe	Pro	Ala	Asn	Ile	Ser	Ser	Met	Pro	Glu	Phe	Arg	Asn	Trp
	65				70				75					80	
Ala	Lys	Gly	Lys	Ile	Asp	Leu	Asp	Ser	Asp	Ser	Ile	Gly	Trp	Tyr	Phe
				85				90						95	
Lys	Tyr	Leu	Asp	Pro	Ala	Gly	Ala	Thr	Glu	Ser	Ala	Arg	Ala	Val	Gly
		100						105					110		
Glu	Tyr	Ser	Lys	Ile	Pro	Asp	Gly	Leu	Val	Lys	Phe	Ser	Val	Asp	Ala
		115					120					125			
Glu	Ile	Arg	Glu	Ile	Tyr	Asn	Glu	Glu	Cys	Pro	Val	Val	Thr	Asp	Val
	130					135					140				
Ser	Val	Pro	Leu	Asp	Gly	Arg	Gln	Trp	Ser	Leu	Ser	Ile	Phe	Ser	Phe
	145				150					155					160



123

Pro Met Phe Arg Thr Ala Tyr Val Ala Val Ala Asn Val Glu Asn Lys
 165 170 175

Glu Met Ser Leu Asp Val Val Asn Asp Leu Ile Glu Trp Leu Asn Asn
 180 185 190

Leu Ala Asp Trp Arg Tyr Val Val Asp Ser Glu Gln Trp Ile Asn Phe
 195 200 205

Thr Asn Asp Thr Thr Tyr Tyr Val Arg Ile Arg Val Leu Arg Pro Thr
 210 215 220

Tyr Asp Val Pro Asp Pro Thr Glu Gly Leu Val Arg Thr Val Ser Asp
 225 230 235 240

Tyr Arg Leu Thr Tyr Lys Ala Ile Thr Cys Glu Ala Asn Met Pro Thr
 245 250 255

Leu Val Asp Gln Gly Phe Trp Ile Gly Gly Gln Tyr Ala Leu Thr Pro
 260 265 270

Thr Ser Leu Pro Gln Tyr Asp Val Ser Glu Ala Tyr Ala Leu His Thr
 275 280 285

Leu Thr Phe Ala Arg Pro Ser Ser Ala Ala Ala Leu Ala Phe Val Trp
 290 295 300

Ala Gly Leu Pro Gln Gly Gly Thr Ala Pro Ala Gly Thr Pro Ala Trp
 305 310 315 320

Glu Gln Ala Ser Ser Gly Gly Tyr Leu Thr Trp Arg His Asn Gly Thr
 325 330 335

Thr Phe Pro Ala Gly Ser Val Ser Tyr Val Leu Pro Glu Gly Phe Ala
 340 345 350

Leu Glu Arg Tyr Asp Pro Asn Asp Gly Ser Trp Thr Asp Phe Ala Ser
 355 360 365

Ala Gly Asp Thr Val Thr Phe Arg Gln Val Ala Val Asp Glu Val Val
 370 375 380

Val Thr Asn Asn Pro Ala Gly Gly Gly Ser Ala Pro Thr Phe Thr Val
 385 390 395 400

Arg Val Pro Pro Ser Asn Ala Tyr Thr Asn Thr Val Phe Arg Asn Thr
 405 410 415

Leu Leu Glu Thr Arg Pro Ser Ser Arg Arg Leu Glu Leu Pro Met Pro
 420 425 430

Pro Ala Asp Phe Gly Gln Thr Val Ala Asn Asn Pro Lys Ile Glu Gln
 435 440 445

Ser Leu Leu Lys Glu Thr Leu Gly Cys Tyr Leu Val His Ser Lys Met
 450 455 460

Arg Asn Pro Val Phe Gln Leu Thr Pro Ala Ser Ser Phe Gly Ala Val
 465 470 475 480



AMENDED SHEET
 IPEA/AU

124

Ser Phe Asn Asn Pro Gly Tyr Glu Arg Thr Arg Asp Leu Pro Asp Tyr
485 490 495
Thr Gly Ile Arg Asp Ser Phe Asp Gln Asn Met Ser Thr Ala Val Ala
500 505 510
His Phe Arg Ser Leu Ser His Ser Cys Ser Ile Val Thr Lys Thr Tyr
515 520 525
Gln Gly Trp Glu Gly Val Thr Asn Val Asn Thr Pro Phe Gly Gln Phe
530 535 540
Ala His Ala Gly Leu Leu Lys Asn Glu Glu Ile Leu Cys Leu Ala Asp
545 550 555 560
Asp Leu Ala Thr Arg Leu Thr Gly Val Tyr Pro Ala Thr Asp Asn Phe
565 570 575
Ala Ala Ala Val Ser Ala Phe Ala Ala Asn Met Leu Ser Ser Val Leu
580 585 590
Lys Ser Glu Ala Thr Ser Ser Ile Ile Lys Ser Val Gly Glu Thr Ala
595 600 605
Val Gly Ala Ala Gln Ser Gly Leu Ala Lys Leu Pro Gly Leu Leu Met
610 615 620
Ser Val Pro Gly Lys Ile Ala Ala Arg Val Arg Ala Arg Arg Ala Arg
625 630 635 640
Arg Arg Ala Ala Arg Ala Asn
645

(2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2479 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 283..2307

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GTTTTCTTT CTTTACCAAG TGTGGTAAAA TTAAACAAA GAAGAAACC AGCACCGTAA	60
CCCGGCCCTT ACACACCTCG AGTCCGTGAC CACCGGATTA TACGTGCCC ACCACACGGC	120
GCCTTTTCCG ACCACTCTCG AGAGTCGTTG GGAGTTTCGT CCGTGACCAC CCGGTTGGCA	180
GTGACAGAC GCTTCCGGAC CACTAGAACC TCCTCGAGCG ACGCACACAC AGCACACACA	240



125

CCGCCTTAGC TGCACCTACG GCAGCGTTGA TAGCGCGGAT TT ATG AGC GAG CAC	294
Me ¹ Ser Glu His 1	
ACC ATC GCC CAC TCC ATC ACA TTA CCA CCC GGT TAC ACC CTT GCC CTA	342
Thr Ile Ala His Ser Ile Thr Leu Pro Pro Gly Tyr Thr Leu Ala Leu 5 10 15 20	
ATA CCC CCT GAA CCT GAA GCA GGA TGG GAG ATG CTG GAG TGG CGT CAC	390
Ile Pro Pro Glu Pro Glu Ala Gly Trp Glu Met Leu Glu Trp Arg His 25 30 35	
AGC GAC CTC ACA ACC GTC GCG GAA CCC GTA ACG TTC GGG TCA GCG CCA	438
Ser Asp Leu Thr Val Ala Glu Pro Val Thr Phe Gly Ser Ala Pro 40 45 50	
ACA CCG TCA CCG TCA ATG GTA GAA GAA ACC AAC GGC GTC GGA CCG GAA	486
Thr Pro Ser Pro Ser Met Val Glu Glu Thr Asn Gly Val Gly Pro Glu 55 60 65	
GGC AAG TTT CTC CCC CTG ACA ATT TCA CCG CTG CTG CAC AAG ACC TCG	534
Gly Lys Phe Leu Pro Leu Thr Ile Ser Pro Leu Leu His Lys Thr Ser 70 75 80	
CGC AAA GCC TTG ACG CCA ACA CCG TCA CTT TCC CCC GCT AAC ATC TCT	582
Arg Lys Ala Leu Thr Pro Thr Pro Ser Leu Ser Pro Ala Asn Ile Ser 85 90 95 100	
ACC ATG CCC GAA TTC CCG AAT TGG GCC AAG GGA AAG ATC GAC CTC GAC	630
Ser Met Pro Glu Phe Arg Asn Trp Ala Lys Gly Lys Ile Asp Leu Asp 105 110 115	
TCC GAT TCC ATC GGC TGG TAC TTC AAG TAC CTT GAC CCA GCG GGT GCT	678
Ser Asp Ser Ile Gly Trp Tyr Phe Lys Tyr Leu Asp Pro Ala Gly Ala 120 125 130	
ACA GAG TCT GCG CGC GCC GTC GGC GAG TAC TCG AAG ATC CCT GAC GGC	726
Thr Glu Ser Ala Arg Ala Val Gly Glu Tyr Ser Lys Ile Pro Asp Gly 135 140 145	
CTC GTC AAG TTC TCC GTC GAC GCA GAG ATA AGA GAG ATC TAT AAC GAG	774
Leu Val Lys Phe Ser Val Asp Ala Glu Ile Arg Glu Ile Tyr Asn Glu 150 155 160	
GAG TGC CCC GTC GTC ACT GAC GTG TCC GTC CCC CTC GAC GGC CGC CAG	822
Glu Cys Pro Val Val Thr Asp Val Ser Val Pro Leu Asp Gly Arg Gln 165 170 175 180	
TGG AGC CTC TCG ATT TTC TCC TTT CCG ATG TTC AGA ACC GCC TAC GTC	870
Trp Ser Leu Ser Ile Phe Ser Phe Pro Met Phe Arg Thr Ala Tyr Val 185 190 195	
GCC GTA CCG AAC GTC GAG AAC AAG GAG ATG TCG CTC GAC GTT GTC AAC	918
Ala Val Ala Asn Val Glu Asn Lys Glu Met Ser Leu Asp Val Val Asn 200 205 210	
GAC CTC ATC GAG TGG CTC AAC AAT CTC GCC GAC TGG CGT TAT GTC GTT	966
Asp Leu Ile Glu Trp Leu Asn Asn Leu Ala Asp Trp Arg Tyr Val Val 215 220 225	



126

GAC Asp 230	TCT Ser	GAA Glu	CAG Gln	TGG Trp	ATT Ile	AAC Asn 235	TTC Phe	ACC Thr	AAT Asn	GAC Asp 240	ACC Thr	ACG Thr	TAC Tyr	TAC Tyr	GTC Val	1014
CGC Arg 245	ATC Ile	CGC Arg	GTT Val	CTA Leu	CGT Arg 250	CCA Pro	ACC Thr	TAC Tyr	GAC Asp	GTT Val 255	CCA Pro	GAC Asp	CCC Pro	ACA Thr	GAG Glu 260	1062
GGC Gly	CTT Leu	GTT Val	CGC Arg	ACA Thr 265	GTC Val	TCA Ser	GAC Asp	TAC Tyr	CGC Arg 270	CTC Leu	ACT Thr	TAT Tyr	AAG Lys	GCG Ala 275	ATA Ile	1110
ACA Thr	TGT Cys	GAA Glu	GCC Ala 280	AAC Asn	ATG Met	CCA Pro	ACA Thr	CTC Leu 285	GTC Val	GAC Asp	CAA Gln	GGC Gly	TTT Phe 290	TGG Trp	ATC Ile	1158
GGC Gly	GGC Gly	CAG Gln 295	TAC Tyr	GCT Ala	CTC Leu	ACC Thr	CCG Pro 300	ACT Thr	AGC Ser	CTA Leu	CCG Pro 305	CAG Gln	TAC Tyr	GAC Asp	GTC Val	1206
AGC Ser 310	GAG Glu	GCC Ala	TAC Tyr	GCT Ala	CTG Leu	CAC His 315	ACT Thr	TTG Leu	ACC Thr	TTC Phe 320	GCC Ala	AGA Arg	CCA Pro	TCC Ser	AGC Ser	1254
GCC Ala 325	GCT Ala	GCA Ala	CTC Leu	GCG Ala	TTT Phe 330	GTG Val	TGG Trp	GCA Ala	GGT Gly	TTG Leu 335	CCA Pro	CAG Gln	GGT Gly	GGC Gly	ACT Thr 340	1302
GCG Ala	CCT Pro	GCA Ala	GGC Gly	ACT Thr 345	CCA Pro	GCC Ala	TGG Trp	GAG Glu	CAG Gln 350	GCA Ala	TCC Ser	TCG Ser	GGT Gly	GGC Gly 355	TAC Tyr	1350
CTC Leu	ACC Thr	TGG Trp	CGC Arg 360	CAC His	AAC Asn	GGT Gly	ACT Thr	ACT Thr	TTC Phe	CCA Pro	GCT Ala	GGC Gly	TCC Ser 370	GTT Val	AGC Ser	1398
TAC Tyr	GTT Val	CTC Leu 375	CCT Pro	GAG Glu	GGT Gly	TTC Phe	GCC Ala	CTT Leu	GAG Glu	CGC Arg	TAC Tyr	GAC Asp 385	CCG Pro	AAC Asn	GAC Asp	1446
GGC Gly 390	TCT Ser	TGG Trp	ACC Thr	GAC Asp	TTC Phe	GCT Ala 395	TCC Ser	GCA Ala	GGA Gly	GAC Asp 400	ACC Thr	GTC Val	ACT Thr	TTC Phe	CGG Arg	1494
CAG Gln 405	GTC Val	GCC Ala	GTC Val	GAC Asp	GAG Glu 410	GTC Val	GTT Val	GTG Val	ACC Thr	AAC Asn 415	AAC Asn	CCC Pro	GCC Ala	GGC Gly	GGC Gly 420	1542
GGC Gly	AGC Ser	GCC Ala	CCC Pro	ACC Thr 425	TTC Phe	ACC Thr	GTG Val	AGA Arg	GTG Val 430	CCC Pro	CCT Pro	TCA Ser	AAC Asn	GCT Ala 435	TAC Tyr	1590
ACC Thr	AAC Asn	ACC Thr	GTG Val 440	TTT Phe	AGG Arg	AAC Asn	ACG Thr	CTC Leu 445	TTA Leu	GAG Glu	ACT Thr	CGA Arg	CCC Pro 450	TCC Ser	TCT Ser	1638
CGT Arg	AGG Arg	CTC Leu 455	GAA Glu	CTC Leu	CCT Pro	ATG Met	CCA Pro 460	CCT Pro	GCT Ala	GAC Asp	TTT Phe	GGA Gly 465	CAG Gln	ACG Thr	GTC Val	1686



127

GCC Ala	AAC Asn 470	AAC Asn	CCG Pro	AAG Lys	ATC Ile	GAG Glu 475	CAG Glu	TCG Ser	CTT Leu	CTT Leu	AAA Lys 480	GAA Glu	ACA Thr	CTT Leu	GGC Gly	1734
TGC Cys 485	TAT Tyr	TTG Leu	GTC Val	CAC His	TCC Ser 490	AAA Lys	ATG Met	CGA Arg	AAC Asn	CCC Pro 495	GTT Val	TTC Phe	CAG Gln	CTC Leu	ACG Thr 500	1782
CCA Pro	GCC Ala	AGC Ser	TCC Ser	TTT Phe 505	GGC Gly	GCC Ala	GTT Val	TCC Ser	TTC Phe 510	AAC Asn	AAT Asn	CCG Pro	GGT Gly	TAT Tyr 515	GAG Glu	1830
CGC Arg	ACA Thr	CGC Arg	GAC Asp 520	CTC Leu	CCG Pro	GAG Asp	TAC Tyr	ACT Thr 525	GGC Gly	ATC Ile	CGT Arg	GAC Asp	TCA Ser 530	TTC Phe	GAC Asp	1878
CAG Gln	AAC Asn	ATG Met 535	TCC Ser	ACC Thr	GCT Ala	GTG Val	GCC Ala 540	CAC His	TTC Phe	CGC Arg	TCA Ser	CTC Leu 545	TCC Ser	CAC His	TCC Ser	1926
TGC Cys 550	AGT Ser	ATC Ile	GTC Val	ACT Thr	AAG Lys	ACC Thr 555	TAC Tyr	CAG Gln	GGT Gly	TGG Trp	GAA Glu 560	GGC Gly	GTC Val	ACG Thr	AAC Asn	1974
GTC Val 565	AAC Asn	ACG Thr	CCT Pro	TTC Phe	GGC Gly 570	CAA Gln	TTC Phe	GCG Ala	CAC His	GCG Ala 575	GGC Gly	CTC Leu	CTC Leu	AAG Lys	AAT Asn 580	2022
GAG Glu	GAG Glu	ATC Ile	CTC Leu	TGC Cys 585	CTC Leu	GCC Ala	GAC Asp	GAC Asp	CTG Leu 590	GCC Ala	ACC Thr	CGT Arg	CTC Leu	ACA Thr 595	GGT Gly	2070
GTC Val	TAC Tyr	CCC Pro	GCC Ala 600	ACT Thr	GAC Asp	AAC Asn	TTC Phe	GCG Ala 605	GCC Ala	GCC Ala	GTT Val	TCT Ser	GCC Ala 610	TTC Phe	GCC Ala	2118
GCG Ala	AAC Asn	ATG Met 615	CTG Leu	TCC Ser	TCC Ser	GTG Val	CTG Leu 620	AAG Lys	TCG Ser	GAG Glu	GCA Ala	ACG Thr 625	TCC Ser	TCC Ser	ATC Ile	2166
ATC Ile	AAG Lys 630	TCC Ser	GTT Val	GGC Gly	GAG Glu	ACT Thr 635	GCC Ala	GTC Val	GGC Gly	GCG Ala	GCT Ala 640	CAG Gln	TCC Ser	GGC Gly	CTC Leu	2214
GCG Ala 645	AAG Lys	CTA Leu	CCC Pro	GGA Gly	CTG Leu 650	CTA Leu	ATG Met	AGT Ser	GTA Val	CCA Pro 655	GGG Gly	AAG Lys	ATT Ile	GCC Ala 660	GCG Ala 660	2262
CGT Arg	GTC Val	CGC Arg	GCG Ala	CGC Arg 665	CGA Arg	GCG Ala	CGC Arg	CGC Arg	CGC Arg 670	GCC Ala	GCT Ala	CGT Arg	GCC Ala	AAT Asn 675		2307
TAGTTTGCTC GCTCCTGTTT CGCCGTTTCG TAAAACGGCG TGGTCCCGCA CATTACGCGT																2367
ACCCTAAAGA CTCTGGTGAG TCCCCGTCGT TACACGACGG GTCTGCCGCG GTTCGATTCC																2427
ATTCCCAAGC GGCAAGAAGG ACGTAGTTAG CTCTGCGTCC CTCGGGATAC CA																2479



128

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 675 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

```

Met Ser Glu His Thr Ile Ala His Ser Ile Thr Leu Pro Pro Gly Tyr
 1          5          10          15
Thr Leu Ala Leu Ile Pro Pro Glu Pro Glu Ala Gly Trp Glu Met Leu
          20          25          30
Glu Trp Arg His Asp Leu Thr Thr Val Ala Glu Pro Val Thr Phe
          35          40          45
Gly Ser Ala Pro Thr Pro Ser Pro Ser Met Val Glu Glu Thr Asn Gly
          50          55          60
Val Gly Pro Glu Gly Lys Phe Leu Pro Leu Thr Ile Ser Pro Leu Leu
          65          70          75          80
His Lys Thr Ser Arg Lys Ala Leu Thr Pro Thr Pro Ser Leu Ser Pro
          85          90          95
Ala Asn Ile Ser Ser Met Pro Glu Phe Arg Asn Trp Ala Lys Gly Lys
          100          105          110
Ile Asp Leu Asp Ser Asp Ser Ile Gly p Tyr Phe Lys Tyr Leu Asp
          115          120          125
Pro Ala Gly Ala Thr Glu Ser Ala Arg Ala Val Gly Glu Tyr Ser Lys
          130          135          140
Ile Pro Asp Gly Leu Val Lys Phe Ser Val Asp Ala Glu Ile Arg Glu
          145          150          155          160
Ile Tyr Asn Glu Glu Cys Pro Val Val Thr Asp Val Ser Val Pro Leu
          165          170          175
Asp Gly Arg Gln Trp Ser Leu Ser Ile Phe Ser Phe Pro Met Phe Arg
          180          185          190
Thr Ala Tyr Val Ala Val Ala Asn Val Glu Asn Lys Glu Met Ser Leu
          195          200          205
Asp Val Val Asn Asp Leu Ile Glu Trp Leu Asn Asn Leu Ala Asp Trp
          210          215          220
Arg Tyr Val Val Asp Ser Glu Gln Trp Ile Asn Phe Thr Asn Asp Thr
          225          230          235          240
Thr Tyr Tyr Val Arg Ile Arg Val Leu Arg Pro Thr Tyr Asp Val Pro
          245          250          255
Asp Pro Thr Glu Gly Leu Val Arg Thr Val Ser Asp Tyr Arg Leu Thr
          260          265          270

```



129

Tyr Lys Ala Ile Thr Cys Glu Ala Asn Met Pro Thr Leu Val Asp Gln
 275 280 285

Gly Phe Trp Ile Gly Gly Gln Tyr Ala Leu Thr Pro Thr Ser Leu Pro
 290 295 300

Gln Tyr Asp Val Ser Glu Ala Tyr Ala Leu His Thr Leu Thr Phe Ala
 305 310 315 320

Arg Pro Ser Ser Ala Ala Ala Leu Ala Phe Val Trp Ala Gly Leu Pro
 325 330 335

Gln C Gly Thr Ala Pro Ala Gly Thr Pro Ala Trp Glu Gln Ala Ser
 340 345 350

Ser Gly Gly Tyr Leu Thr Trp Arg His Asn Gly Thr Thr Phe Pro Ala
 355 360 365

Gly Ser Val Ser Tyr Val Leu Pro Glu Gly Phe Ala Leu Glu Arg Tyr
 370 375 380

Asp Pro Asn Asp Gly Ser Trp Thr Asp Phe Ala Ser Ala Gly Asp Thr
 385 390 395 400

Val Thr Phe Arg Gln Val Ala Val Asp Glu Val Val Val Thr Asn Asn
 405 410 415

Pro Ala Gly Gly Gly Ser Ala Pro Thr Phe Thr Val Arg Val Pro Pro
 420 425 430

Ser Asn Ala Tyr Thr Asn Thr Val Phe Arg Asn Thr Leu Leu Glu Thr
 435 440 445

Arg Pro Ser Ser Arg Arg Leu Glu Leu Pro Met Pro Pro Ala Asp Phe
 450 455 460

Gly Gln Thr Val Ala Asn Asn Pro Lys Ile Glu Gln Ser Leu Leu Lys
 465 470 475 480

Glu Thr Leu Gly Cys Tyr Leu Val His Ser Lys Met Arg Asn Pro Val
 485 490 495

Phe Gln Leu Thr Pro Ala Ser Ser Phe Gly Ala Val Ser Phe Asn Asn
 500 505 510

Pro Gly Tyr Glu Arg Thr Arg Asp Leu Pro Asp Tyr Thr Gly Ile Arg
 515 520 525

Asp Ser Phe Asp Gln Asn Met Ser Thr Ala Val Ala His Phe Arg Ser
 530 535 540

Leu Ser His Ser Cys Ser Ile Val Thr Lys Thr Tyr Gln Gly Trp Glu
 545 550 555 560

Gly Val Thr Asn Val Asn Thr Pro Phe Gly Gln Phe Ala His Ala Gly
 565 570 575

Leu Leu Lys Asn Glu Glu Ile Leu Cys Leu Ala Asp Asp Leu Ala Thr
 580 585 590



130

Arg Leu Thr Gly Val Tyr Pro Ala Thr Asp Asn Phe Ala Ala Ala Val
595 600 605
Ser Ala Phe Ala Ala Asn Met Leu Ser Ser Val Leu Lys Ser Glu Ala
610 615 620
Thr Ser Ser Ile Ile Lys Ser Val Gly Glu Thr Ala Val Gly Ala Ala
625 630 635 640
Gln Ser Gly Leu Ala Lys Leu Pro Gly Leu Leu Met Ser Val Pro Gly
645 650 655
Lys Ile Ala Ala Arg Val Arg Ala Arg Arg Ala Arg Arg Ala Ala
660 665 670
Arg Ala Asn
675

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GGGGATCCAC AGTTCTGCCT CCCCCGGACG GTAAATATAG GGGAACCATG GTCTAGAGG

59



THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:-

1. An isolated small RNA virus wherein the virus is up to 40nm in size, is not occluded, infects insect species including Heliothis species, and comprises a genome hybridizable with the nucleotide sequence of RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47).
2. An isolated small RNA virus wherein the virus is up to 40nm in size, is not occluded, infects insect species including Heliothis species, and comprises proteins which are capable of generating antibodies said antibodies being immunologically reactive with the large coat protein of HaSV.
3. The virus of claim 1 or 2 wherein said virus has a particle size of 35 to 38nm and comprises a genome with RNA of substantially 5.3 and 2.4 kb in length.
4. The virus of claim 3 wherein said particle comprises coat proteins of substantially 7 and 64 KDa.
5. The virus of any one of the preceding claims wherein said virus is HaSV or a mutant, variant or derivative thereof as herein described.
6. The virus of claim 5 wherein said virus comprises a nucleic acid sequence which is an encapsidation sequence, structure or signal with at least 50% nucleotide sequence identity to the corresponding nucleotide sequences of HaSV.



7. The virus of claim 5 wherein said virus comprises a nucleic acid sequence which encodes proteins with at least 60% amino acid sequence identity to the corresponding proteins or polypeptides of HaSV.

8. The virus of claim 5 wherein said virus comprises a nucleic acid sequence which has at least 50% nucleotide sequence identity to the portions of the HaSV genome which encode: amino acid residues 401 to 600 of the replicase enzyme or amino acid residues 273 to 435 or 50 to 272 or 436 to the COOH terminus of the capsid protein.

9. An isolated nucleic acid molecule of 15 or more nucleotides in length, said molecule comprising a nucleic acid sequence hybridizable with RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47) under low stringency conditions.

10. The molecule of claim 9 wherein said sequence is hybridizable under medium stringency conditions.

11. The molecule of claim 10 wherein said sequence is hybridizable under high stringency conditions.

12. The molecule of claim 9 wherein said sequence encodes P7, P16, P17 (SEQ ID No: 48), P64, P70 (SEQ ID No: 52), P71 (SEQ ID No: 50), P11a (SEQ ID No: 42), P11b (SEQ ID No: 44), P14 (SEQ ID No: 46) or P187 (SEQ ID No: 40) or a functionally-equivalent mutant, variant or derivative thereof.

13. The molecule of claim 12 wherein said sequence encodes P7, P64, or P71 (SEQ ID No: 50) or a functionally-equivalent mutant, variant or derivative thereof.



14. The molecule of claim 9 in which the molecule is a probe or primer for the nucleic acid sequence of RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47), or functionally-equivalent mutants, variants or derivatives thereof, said molecule comprising nucleic acid sequences suitable for detection of, or replication of, RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47), or portions thereof under appropriate conditions.

15. The molecule of claim 14 in which the molecule is one of a primer pair, wherein said primer is derived from a sequence of RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47) which is located between 300 and 1500 bp from another sequence of RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47) being the sequence of the other primer of said primer pair.

16. The molecule of claim 15 comprising the following sequences

5' GGGGGGAATTCATTTAGGTGACACTATAGTTCTGCCTCCCCGGAC
(called "HvR1SP5p" herein) (SEQ ID No: 11)

5' GGGGGGATCCTGTTATCCCAGGGGGGC (called "HvR13p" herein)
(SEQ ID No: 12)

5' CCGGAAGCTTGTTTTCTTTCTTTACCA (called "Hr2cdna5" herein)
(SEQ ID No: 13)

5' GGGGGATCCGATGGTATCCCGAGGGACGC
TCAGCAGGTGGCATAGC (called "HvR23p" herein) (SEQ ID No: 14)

AAATAATTTTGTACTTTAGAAAGGAGATATACATATGAGCGAGCGAGCA
CAC (called "HVPET65N" herein) (SEQ ID No: 15)



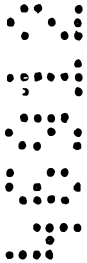
AAATAATTTTGTTTAACCTTAAGAAGGAGATCTACATATGCTGGAGTGGC
GTCAC (called "HVPET63N" herein) (SEQ ID No: 16)

GGAGATCTACATATGGGAGATGCTGGAGTG (called "HVPET64N" herein)
 (SEQ ID No: 17)

GTAGCGAACGTCGAGAA (called "HVRNA2F3" herein) (SEQ ID No: 18)

GGGGGATCCTCAGTTGTCAGTGGCGGGGTAG (called "HVP65C" herein)
 (SEQ ID No: 19)

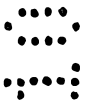
GGGGATCCCTAATTGGCACGAGCGGCGC (called "HVP6C2" herein) (SEQ
 ID No: 20)



AATTACATATGGCGGCCGCCGTTTCTGCC (called "HVP6MA" herein)
 (SEQ ID No: 21)



AATTACATATGTTGCGGCCGCCGTTTCT (called "HVP6MF" herein).
 (SEQ ID No: 22)



17. The molecule of claim 9 additionally comprising a ribozyme sequence.
18. A vector comprising the molecule of claim 9.
19. A vector comprising the molecule of claim 12.
20. A vector comprising the molecule of claim 13.
21. A vector comprising the molecule of claim 14.



22. A vector comprising the molecule of claim 15.
23. A vector comprising the molecule of claim 16.
24. A vector comprising the molecule of claim 9 capable of replication, expression and/or encapsidation in an animal, plant, yeast or bacterial cell.
25. A vector comprising the molecule of claim 9 capable of replication, expression and/or encapsidation in yeast.
26. A vector comprising the molecule of claim 9 capable of transferring said nucleic acid molecule to a plant cell.

27. The vector of any one of claims 24 to 26 which comprises a ribozyme for facilitating replication, expression or encapsidation of the transcript.

28. The vector of any one of claims 24 to 26 wherein having a ribozyme sequence selected from one of the following sequences

5' CCATCGATGCCGGACTGGTATCCCAGGGGG (SEQ ID No: 5)

5' CCATCGATGCCGGACTGGTATCCCGAGGGAC (SEQ ID No: 6)

5' CCATCGATGATCCAGCCTCCTCGCGGCGCCGGATGGGCA (SEQ ID No: 7)

5' GCTCTAGATCCATTTCGCCATCCGAAGATGCCCATCCGGC (SEQ ID No: 8)

5' CCATCGATTATGCCGAGAAGGTAACCAGAGAAACACAC (SEQ ID No: 9)



5' GCTCTAGACCAGGTAATATACCACAACGTGTGTTTCTCT (SEQ ID No: 10)

29. The vector of any one of claims 24 to 26 which comprises a promoter for facilitating expression said promoter selected from the group of the *Drosophila* promoters, heat shock promoters, baculovirus promoters, CMV promoters.

30. A vector according to claim 19 in which the vector is selected from the group consisting of pDHVR1, pDHVR1RZ, pDHVR2, pDHVR2RZ, p17V71, p17E71, pPH, pV71, p17V64, p17E64, pP64, pV64, pBacHVR1, pBacHVR1RZ, pBacHUR2, pBacHVR2RZ, pHSPR1, pHSPR1RZ, pHSPR2, pHSPR2RZ, pSR1(E3)A, pSR1(E3)B, pSR2A, pSR2B, pSX2P70, pSXR2P70, pSRP2B, pBHVR1B, pBHVR2B, pT7T2F64, pSR2P70, pT7T2P65, pT7T2P70, pT7T2-P71, pBSKSE3, pBSR15, pBSR25p, pSR25, phr236P70, phr235P65, pGemP63N, pGemP64N, pGemP65N, pP64N, pP65H, pTP6MA, pTP6MF, pTP17, pTP17delBB, pP656 and p70G.

31. A host cell comprising the vector of any one of claims 18 to 30.

32. The host cell of claim 31 wherein said cell is an insect cell or a plant cell.

33. The host cell of claim 31 wherein said cell is a yeast.

34. A recombinant insect virus vector comprising the nucleic acid molecule of claim 9.



35. The virus vector of claim 34 comprising material derived from baculovirus including NPV and GV, entomopoxvirus, cytoplasmic polyhedrosis virus.

36. The virus vector of claim 34 wherein said vector is capable of infecting insect species including *Heliothis* species.

37. The virus vector of claim 34 comprising one or more nucleic acid sequences which encode substances which are deleterious to insects.

38. A method of producing HaSV or mutants, variants or derivatives thereof as herein described, comprising culturing a yeast host cell according to claim 33 under conditions suitable for expression of said HaSV or mutants, variants or derivatives thereof.

39. A method of controlling insect attack in a plant comprising genetically manipulating said plant so that it is capable of producing HaSV or mutants, variants or derivatives thereof as herein described, or an insecticidally effective portion of HaSV or mutants, variants or derivatives thereof as herein described, such that insects feeding on the plant are deleteriously effected.

40. A method according to claim 39, wherein said plant is genetically manipulated to express an isolated nucleic acid molecule according to any one of claims 9 to 13 or 17.

41. A transgenic plant resistant to insect attack comprising a genome or subgenome capable of expressing the molecule of claim 9 such that the transgenic plant produces HaSV or mutants, variants or derivatives thereof as herein described, or an insecticidally effective portion of HaSV or mutants,



variants or derivatives thereof as herein described, such that insects feeding on the transgenic plant are deleteriously effected.

42. A preparation of HaSV or a mutant, variant or derivative thereof as herein described, or an insecticidally effective portion of said HaSV, or mutant, variant or derivatives thereof as herein described, suitable for application to plants, said preparation capable of imparting an insect protective effect to plants.

43. The plasmid vectors pT7T2b and pT7T2C.

44. A method of identifying HaSV or mutants, variants or derivatives thereof as herein described, using the molecule of claim 9 to detect the presence of said HaSV in a sample.

45. A capsovector for use in controlling insect pests, comprising a capsid protein from an insect small RNA virus of up to 40nm in size which is not occluded and infects insect species including Heliothis species, said capsid protein encapsidating an insecticidal protein toxin such that the protein toxin is protected from inactivation in the gut of an insect following ingestion of the capsovector by the insect.

46. A capsovector as claimed in claim 45 in which the insect small RNA virus is HaSV or a mutant, variant or derivative thereof as described herein.

47. A capsovector as claimed in claim 45 in which the insect small RNA virus is HaSV.

48. A capsovector as claimed in any one of claims 45 to 47 in which the capsid protein is P71 (SEQ ID No. 50)



49. A capsovector as claimed in any one of claims 45 to 48 in which the insecticidal toxin is of plant origin.

50. A capsovector as claimed in any one of claims 45 to 48 in which the insecticidal toxin is Ricin A or diptheria toxin.

51. A capsovector for use in controlling insect pests, comprising a capsid protein of an insect small RNA virus of up to 40nm in size which is not occluded and infects insect species including *Heliothis* species, said capsid protein encapsidating a nucleic acid molecule which is insecticidal or which encodes an insecticidal protein toxin, such that the nucleic acid molecule is protected from inactivation in the gut of an insect following ingestion of the capsovector by the insect.

52. A capsovector as claimed in claim 51 in which the nucleic acid is RNA.

53. A capsovector as claimed in claims 51 or 52 in which the insect small RNA virus is HaSV or a mutant, variant or derivative thereof as described herein.

54. A capsovector as claimed in claim 51 or 52 in which the insect small RNA virus is HaSV.

55. A capsovector as claimed in any one of claims 51 to 54 in which the capsid protein is P71 (SEQ ID No. 50)

56. A capsovector as claimed in any one of claims 51 to 55 in which the insecticidal toxin is of plant origin.



57. A capsovector as claimed in any one of claims 51 to 55 in which the insecticidal toxin is Ricin A or diptheria toxin.

58. An isolated nucleic acid molecule comprising a first sequence encoding at least one capsid protein of an insect small RNA virus of up to 40nm in size which is not occluded and infects insect species including *Heliothis* species, and a second sequence which is insecticidal or which encodes an insecticidal protein toxin.

59. An isolated nucleic acid molecule as claimed in claim 58 in which the nucleic acid is RNA.

60. An isolated nucleic acid molecule as claimed in claim 58 or 59 in which the insect small RNA virus is HaSV.

61. An isolated nucleic acid molecule as claimed in any one of claims 58 to 60 in which the capsid protein is P71 (SEQ ID No. 50)

62. An isolated nucleic acid molecule as claimed in any one of claims 58 to 61 in which the insecticidal toxin is of plant origin.

63. An isolated nucleic acid molecule as claimed in any one of claims 58 to 61 in which the insecticidal toxin is Ricin A or diptheria toxin.

64. A transgenic plant resistant to insect attack comprising a genome or subgenome capable of expressing the nucleic acid molecule as claimed in any one of claims 53 to 63 such that the transgenic plant produces capsid protein in which is encapsidated the nucleic acid molecule.



Dated this 10th day of April 1997.

COMMONWEALTH SCIENTIFIC
AND INDUSTRIAL RESEARCH
ORGANISATION and PACIFIC SEEDS
PTY LTD

Patent Attorneys for the Applicant:

F.B. RICE & CO.

3
3
3
3
3

3
3
3
3

3
3
3



1/45

10 30 50
ETTCTGCCTCCCCCGGACGGTAAATATAGGGGAACAATGTACGCGAAAGCGACAGACGTG
-----+-----+-----+-----+-----+-----+
M Y A K A T D V
replication start

70 90 110
GCGCGTGTCTACGCCGCGGCAGATGTCGCCTACGCGAACGTACTGCAGCAGAGAGCAGTC
-----+-----+-----+-----+-----+-----+
A R V Y A A A D V A Y A N V L Q Q R A V

130 150 170
AAGTTGGACTTCGCCCCGCCACTGAAGGCACTAGAAACCCTCCACAGACTGTACTATCCG
-----+-----+-----+-----+-----+-----+
K L D F A P P L K A L E T L H R L Y Y F

190 210 230
CTGCGCTTCAAAGGGGGCACTTTACCCCCGACACAACACCCGATCCTGGCCGGGCACCAA
-----+-----+-----+-----+-----+-----+
L R F K G G T L P P T Q H P I L A G H Q

250 270 290
CGTGTCGCAGAAGAGGTTCTGCACAATTTTCGCCAGGGGACGTAGCACAGTGCTCGAGATA
-----+-----+-----+-----+-----+-----+
R V A E E V L H N F A R G R S T V L E I

FIG. 1

SUBSTITUTE SHEET

310 330 350
GGGCCGTCTCTGCACAGCGCACTTAAGCTACATGGGGCACCGAACGCCCCCGTCGCAGAC
-----+-----+-----+-----+-----+-----+-----+
G P S L H S A L K L H G A P N A P V A D

370 390 410
TATCACGGGTGCACCAAGTACGGCACCCGCGACGGCTCGCGACACATTACGGCCTTAGAG
-----+-----+-----+-----+-----+-----+-----+
Y H G C T K Y G T R D G S R H I T A L E

430 450 470
TCTAGATCCGTCGCCACAGGCCGCGCCCGAGTTCAAGGCCGACGCCTCACTGCTCGCCAAC
-----+-----+-----+-----+-----+-----+-----+
S R S V A T G R P E F K A D A S L L A N

490 510 530
GGCATTGCCTCCCGCACCTTCTGCGTCGACGGAGTCGGCTCTTGCGCGTTCAAATCGCGC
-----+-----+-----+-----+-----+-----+-----+
G I A S R T F C V D G V G S C A F K S R

550 570 590
GTTGGAATTGCCAATCACTCCCTCTATGACGTGACCCTAGAGGAGCTGGCCAATGCGTTT
-----+-----+-----+-----+-----+-----+-----+
V G I A N H S L Y D V T L E E L A N A F

610 630 650
GAGAACCACGGACTTCACATGGTCCGCGCGTTCATGCACATGCCAGAAGAGCTGCTCTAC
-----+-----+-----+-----+-----+-----+-----+
E N H G L L M V R A F M H M P E E L L Y

670 690 710
ATGGACAACGTGGTTAATGCCGAGCTCGGCTACCGCTTCCACGTTATTGAAGAGCCTATG
-----+-----+-----+-----+-----+-----+-----+
M D N V V N A E L G Y R F H V I E E P M

730 750 770
GCTGTGAAGGACTGCGCATTCCAGGGGGGGGACCTCCGTCTCCACTTCCCTGAGTTGGAC
-----+-----+-----+-----+-----+-----+-----+
A V K D C A F Q G G D L R L H F P E , D

790 810 830
TTCATCAACGAGAGCCAAGAGCGGCGCATCGAGAGGCTGGCCGCCCGCGGGCTCCTACTCC
-----+-----+-----+-----+-----+-----+-----+
F I N E S Q E R R I E R L A A R G S Y S

850 870 890
AGACGCGCCGTCATTTTCTCCGGCGACGACGACTGGGGTGATGCGTACTTACACGACTTC
-----+-----+-----+-----+-----+-----+-----+
R R A V I F S G D D D W G D A Y L H D F

910 930 950
CACACATGGCTCGCCTACCTACTGGTGAGGAACTACCCCACTCCGTTTGGTTTCTCACTC
-----+-----+-----+-----+-----+-----+-----+
H T W L A Y L L V R N Y P T P F G F S L

970 990 1010
CATATAGAAGTCCAGAGGGCGCCACGGCTCCAGCATTGAGCTGCGCATCACTCGCGCGCCA
-----+-----+-----+-----+-----+-----+-----+
H I E V Q R R H G S S I E L R I T R A P

1030 1050 1070
CCTGGAGACCGCATGCTGGCCGTCGTCCCAAGGACGTCCCAAGGCCTCTGCAGAATCCCA
-----+-----+-----+-----+-----+-----+-----+
P G D R M L A V V P R T S Q G L C R I P

1090 1110 1130
AACATCTTTTATTACGCCGACGCGTCGGGCACTGAGCATAAGACCATCCTTACGTCACAG
-----+-----+-----+-----+-----+-----+-----+
N I F Y Y A D A S G T E H K T I L T S Q

1150 1170 1190
CACAAAGTCAACATGCTGCTCAATTTTATGCAAACGCGTCCTGAGAAGGAAGTAGTCGAC
-----+-----+-----+-----+-----+-----+-----+
H K V N M L L N F M Q T R P E K E L V D

SUBSTITUTE SHEET

FIG. 1 Cont'd

SUBSTITUTE SHEET

FIG. 1 Cont'd

1210	1230	1250
ATGACCGTCTTGATGTCGTTTCGCGCGCGCTAGGCTGCGCGCGATCGTGGTCGCCTCAGAA		
-----+-----+-----+-----+-----+-----+-----+		
M	T	V
L	M	S
F	A	R
A	R	L
R	A	I
V	V	A
S	E	

1270	1290	1310
GTCACCGAGAGCTCCTGGAACATCTCACCGGCTGACCTGGTCCGCACTGTCGTGTCTCTT		
-----+-----+-----+-----+-----+-----+-----+		
V	T	E
S	S	W
N	I	S
P	A	D
L	V	R
T	V	V
S	L	

1330	1350	1370
TACGTCCTCCACATCATCGAGCGCCGAAGGGCTGCGGTCGCTGTCAAGACCGCCAAGGAC		
-----+-----+-----+-----+-----+-----+-----+		
Y	V	L
H	I	I
E	R	R
A	A	V
A	V	K
T	A	K
D		

1390	1410	1430
GACGTCTTTGGAGAGACTTCGTTCTGGGAGAGTCTCAAGCACGTCTTGGGCTCCTGTTGC		
-----+-----+-----+-----+-----+-----+-----+		
D	V	F
G	E	T
S	F	W
E	S	L
K	H	V
L	G	S
C	C	

1450	1470	1490
GGTCTGCGCAACCTCAAAGGCACCGACGTCTTTACTAAGCGCGTCGTCGATAAGTAC		
-----+-----+-----+-----+-----+-----+-----+		
G	L	R
N	L	K
G	T	D
V	V	F
T	K	R
V	V	D
K	Y	

1510 1530 1550
CGAGTCCACTCGCTCGGAGACATAATCTGCGACGTCCGCCTGTCCCCTGAACAGGTCGGC
-----+-----+-----+-----+-----+-----+-----+
R V H S L G D I I C D V R L S P E Q V G

1570 1590 1610
TTCCTGCCGTCCCGCGTACCACCTGCCCCGCGTCTTTCACGACAGGGAAGAGCTTGAGGTC
-----+-----+-----+-----+-----+-----+-----+
F L P S R V P P A R V F H D R E E L E V

1630 1650 1670
CTTCGCGAAGCTGGCTGCTACAACGAACGTCCGGTACCTTCCACTCCTCCTGTGGAGGAG
-----+-----+-----+-----+-----+-----+-----+
L R E A G C Y N E R P V P S T P P V E E

1690 1710 1730
CCCCAAGGTTTCGACGCCGACTTGTGGCACGCGACCGCGGCCTCACTCCCCGAGTACCGC
-----+-----+-----+-----+-----+-----+-----+
P Q G F D A D L W H A T A A S L P E Y R

1750 1770 1790
GCCACCTTGCAGGCAGGTCTCAACACCGACGTCAAGCAGCTCAAGATCACCCCTCGAGAAC
-----+-----+-----+-----+-----+-----+-----+
A T L Q A G L N T D V K Q L K I T L E N

1810	1830	1850
GCCCTCAAGACCATCGACGGGCTCACCTCTCCCCAGTCAGAGGCCTCGAGATGTACGAG		
-----+-----+-----+-----+-----+-----+-----+		
A L K T I D G L T L S P V R G L E M Y E		
1870	1890	1910
GGCCCGCCAGGCAGCGGCAAGACGGGCACCCTCATCGCCGCCCTTGAGGCCGCGGGCGGT		
-----+-----+-----+-----+-----+-----+-----+		
G P P G S G K T G T L I A A L E A A G G		
1930	1950	1970
AAAGCACTTTACGTGGCACCCACCAGAGAACTGAGAGAGGCTATGGACCGGCGGATCAAA		
-----+-----+-----+-----+-----+-----+-----+		
K A L Y V A P T R E L R E A M D R R I K		
1990	2010	2030
CCGCCGTCCGCCTCGGCTACGCAACATGTCGCCCTTGCGATTCTCCGTCGTGCCACCGCC		
-----+-----+-----+-----+-----+-----+-----+		
P P S A S A T Q H V A L A I L R R A T A		
2050	2070	2090
GAGGGCGCCCCTTTCGCTACCGTGGTTATCGACGAGTGCTTCATGTTCCCGCTCGTGTAC		
-----+-----+-----+-----+-----+-----+-----+		
E G A P F A T V V I D E C F M F P L V Y		

SUBSTITUTE SHEET

FIG. 1 Cont'd

2110 2130 2150
GTCGCGATCGTGCACGCCTTGTCCCCGAGCTCACGAATAGTCCTTGTAGGGGACGTCCAC
-----+-----+-----+-----+-----+-----+-----+
V A I V H A L S P S S R I V L V G D V H

2170 2190 2210
CAAATCGGGTTTATAGACTTCCAAGGCACAAGCGCGAACATGCCGCTCGTTCGCGACGTC
-----+-----+-----+-----+-----+-----+-----+
Q I G F I D F Q G T S A N M P L V R D V

2230 2250 2270
GTTAAGCAGTGCCGTCGGCGCACTTTCAACCAAACCAAGCGCTGTCCGGCCGACGTCGTT
-----+-----+-----+-----+-----+-----+-----+
V K Q C R R R T F N Q T K R C P A D V V

2290 2310 2330
GCCACCACGTTTTTCCAGAGCTTGTACCCCGGGTGCACAACCACCTCAGGGTGCGTCGCA
-----+-----+-----+-----+-----+-----+-----+
A T T F F Q S L Y P G C T T T S G C V A

2350 2370 2390
TCCATCAGCCACGTCGCCCCAGACTACCGCAACAGCCAGGCGCAAACGCTCTGCTTCACG
-----+-----+-----+-----+-----+-----+-----+
S I S H V A P D Y R N S Q A Q T L C F T

FIG. 1 Cont'd

SUBSTITUTE SHEET

2410 2430 2450
CAGGAGGAAAAGTCGCGCCACGGGGCTGAGGGCGCGATGACTGTGCACGAAGCGCAGGGA
-----+-----+-----+-----+-----+-----+-----+
Q E E K S R H G A E G A M T V H E A Q G

2470 2490 2510
CGCACTTTTGCGTCTGTCATTCTGCATTACAACGGCTCCACAGCAGAGCAGAAGCTCCTC
-----+-----+-----+-----+-----+-----+-----+
R T F A S V I L H Y N G S T A E Q K L L

2530 2550 2570
GCTGAGAAGTCGCACCTTCTAGTCGGCATCACGCGCCACACCAACCACCTGTACATCCGC
-----+-----+-----+-----+-----+-----+-----+
A E K S H L L V G I T R H T N H L Y I R

2590 2610 2630
GACCCGACAGGTGACATTGAGAGACAACCTCAACCATAGCGCGAAAGCCGAGGTGTTTACA
-----+-----+-----+-----+-----+-----+-----+
D P T G D I E R Q L N H S A K A E V F T

2650 2670 2690
GACATCCCTGCACCCCTGGAGATCACGACTGTCAAACCGAGTGAAGAGGTGCAGCGCAAC
-----+-----+-----+-----+-----+-----+-----+
D I P A P L E I T T V K P S E E V Q R N

SUBSTITUTE SHEET

FIG. 1 Cont'd

SUBSTITUTE SHEET

FIG. 1 Cont'd

2710	2730	2750																	
GAAGTGATGGCAACGATACCCCCGCAGAGTGCCACGCCGCACGGAGCAATCCATCTGCTC																			
-----+-----+-----+-----+-----+-----+-----+																			
E	V	M	A	T	I	P	P	Q	S	A	T	P	H	G	A	I	H	L	L
2770			2790			2810													
CGCAAGAACTTCGGGGACCAACCCGACTGTGGCTGTGTCGCTTTGGCGAAGACCGGCTAC																			
-----+-----+-----+-----+-----+-----+-----+																			
R	K	N	F	G	D	Q	P	D	C	G	C	V	A	L	A	K	T	G	Y
2830			2850			2870													
GAGGTGTTTGGCGGTTCGTGCCAAAATCAACGTAGAGCTTGCCGAACCCGACGCGACCCCG																			
-----+-----+-----+-----+-----+-----+-----+																			
E	V	F	G	G	R	A	K	I	N	V	E	L	A	E	P	D	A	T	P
2890			2910			2930													
AAGCCGCATAGGGCGTTCCAGGAAGGGGTACAGTGGGTCAAGGTCACCAACGCGTCTAAC																			
-----+-----+-----+-----+-----+-----+-----+																			
K	P	H	R	A	F	Q	E	G	V	Q	W	V	K	V	T	N	A	S	N
2950			2970			2990													
AAACACCAGGCCCTCCAGACGCTGTTGTCCCGCTACACCAAGCGAAGCGCTGACCTGCCG																			
-----+-----+-----+-----+-----+-----+-----+																			
K	H	Q	A	L	Q	T	L	L	S	R	Y	T	K	R	S	A	D	L	P

3010 3030 3050
CTACACGAAGCTAAGGAGGACGTCAAACGCATGCTAAACTCGCTTGACCGACATTGGGAC
-----+-----+-----+-----+-----+-----+-----+
L H E A K E D V K R M L N S L D R H W D

3070 3090 3110
TGGACTGTCACTGAAGACGCCCCGTGACCGAGCTGTCTTCGAGACCCAGCTCAAGTTCACC
-----+-----+-----+-----+-----+-----+-----+
W T V T E D A R D R A V F E T Q L K F T

3130 3150 3170
CAACGCGGCGGCACCGTCTGAAGACCTGCTGGAGCCAGACGACCCCTACATCCGTGACATA
-----+-----+-----+-----+-----+-----+-----+
Q R G G T V E D L L E P D D P Y I R D I

3190 3210 3230
GACTTCCTTATGAAGACTCAGCAGAAAGTGTCGCCCCAAGCCGATCAATACGGGCAAGGTC
-----+-----+-----+-----+-----+-----+-----+
D F L M K T Q Q K V S P K P I N T G K V

3250 3270 3290
GGGCAGGGGATCGCCGCTCACTCAAAGTCTCTCAACTTCGTCCTCGCCGCTTGGATACGC
-----+-----+-----+-----+-----+-----+-----+
G Q G I A A H S K S L N F V L A A W I R

SUBSTITUTE SHEET

FIG. 1 Cont'd

3310 3330 3350
ATACTCGAGGAGATACTCCGTACCGGGAGCCGCACGGTCCGGTACAGCAACGGTCTCCCC
-----+-----+-----+-----+-----+-----+-----+
I L E E I L R T G S R T V R Y S N G L P

3370 3390 3410
GACGAAGAAGAGGCCATGCTGCTCGAAGCGAAGATCAATCAAGTCCCACACGCCACGTTC
-----+-----+-----+-----+-----+-----+-----+
D E E E A M L L E A K I N Q V P H A T F

3430 3450 3470
GTCTCGGCGGACTGGACCGAGTTTGACACCGCCCACAATAACACGAGTGAGCTGCTCTTC
-----+-----+-----+-----+-----+-----+-----+
V S A D W T E F D T A H N N T S E L L F

3490 3510 3530
GCCGCCCTTTTAGAGCGCATCGGCACGCCTGCAGCTGCCGTTAATCTATTCAGAGAACGG
-----+-----+-----+-----+-----+-----+-----+
A A L L E R I G T P A A A V N L F R E R

3550 3570 3590
TGTGGGAAACGCACCTTGCGAGCGAAGGCTCTAGGCTCCGTTGAAGTCGACGGTCTGCTC
-----+-----+-----+-----+-----+-----+-----+
C G K R T L R A K G L G S V E V D G L L

SUBSTITUTE SHEET

FIG. 1 Cont'd

SUBSTITUTE SHEET

FIG. 1 Cont'd

3610	3630	3650
GACTCCGGCGCAGCTTGGACGCCTTGCCGCAACACCATCTTCTCTGCCGCCGTCATGCTC		
-----+-----+-----+-----+-----+-----+-----+		
D S G A A W T P C R N T I F S A A V M L		
3670	3690	3710
ACGCTCTTCCGCGGCGTCAAGTTCGCAGCTTTCAAAGGCGACGACTCGCTCCTCTGTGGT		
-----+-----+-----+-----+-----+-----+-----+		
T L F R G V K F A A F K G D D S L L C G		
3730	3750	3770
AGCCATTACCTCCGTTTCGACGCTAGCCGCCTTCACATGGGCGAACGTTACAAGACCAAA		
-----+-----+-----+-----+-----+-----+-----+		
S H Y L R F D A S R L H M G E R Y K T K		
3790	3810	3830
CATTTGAAGGTTCGAGGTGCAGAAAATCGTGCCGTACATCGGACTCCTCCTCTCCGCTGAG		
-----+-----+-----+-----+-----+-----+-----+		
H L K V E V Q K I V P Y I G L L V S A E		
3850	3870	3890
CAGGTCGTCCTCGACCCTGTCAGGAGCGCTCTCAAGATATTTGGGCGCTGCTACACAAGC		
-----+-----+-----+-----+-----+-----+-----+		
Q V V L D P V R S A L K I F G R C Y T S		

3910 3930 3950
GAACTCCTTTACTCCAAGTACGTGGAGGCTGTGAGAGACATCACCAAGGGCTGGAGTGAC
-----+-----+-----+-----+-----+-----+-----+
E L L Y S K Y V E A V R D I T K G W S D

3970 3990 4010
GCCCCGTACCACAGCCTCCTGTGCCACATGTCAGCATGCTACTACAATTACGCGCCGGAG
-----+-----+-----+-----+-----+-----+-----+
A R Y H S L L C H M S A C Y Y N Y A P E

4030 4050 4070
TCTGCGGCGTACATCATCGACGCTGTTGTTTCGCTTTGGGCGCGGCGACTTCCCGTTTGAA
-----+-----+-----+-----+-----+-----+-----+
S A A Y I I D A V V R F G R G D F P F E

4090 4110 4130
CAACTGCGCGTGCGTGCGTGCCCATGTGCAGGCACCCGACGCTTACAGCAGCACGTATCCG
-----+-----+-----+-----+-----+-----+-----+
Q L R V V R A H V Q A P D A Y S S T Y P

4150 4170 4190
GCTAACGTGCGCGCATCGTGCTTGACCACGTCTTCGAGCCCCGCCAGGCCGCCGCCCG
-----+-----+-----+-----+-----+-----+-----+
A N V R A S C L D H V F E P R Q A A A P

FIG. 1 Cont'd

SUBSTITUTE SHEET

4210 4230 4250
 GCAGGTTTCGTTGCGACATGTGCGAAGCCGGAAACGCCTTCTTCACTTACCGCGAAAGCT
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 A G F V A T C A K P E T P S S L T A K A
 M C E A G N A F F T Y R E S W
 P11a start

4270 4290 4310
 GGTGTTTCTGCGACTACAAGCCACGTTGCGACTGGGACTGCGCCCCCGGAGTCTCCATGG
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 G V S A T T S H V A T G T A P P E S P W
 C F C D Y K P R C D W D C A P G V S M G

4330 4350 4370
 GATGCACCTGCAGCCAACAGCTTTTCGGAGTTATTGACACCGGAGACCCCGTCCACATCA
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 D A P A A N S F S E L L T P E T P S T S
 C T C S Q Q L F G V I D T G D P V H I I

4390 4410 4430
 TCCTCGCCGTCATCGTCTTCATCGGACTCCTCTACATCGTGTGGAAGGTCGCTCAGTGGT
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 S S P S S S S S D S S T S C G R S L S G
 L A V I V F I G L L Y I V W K V A Q W W

4450	4470	4490
GGAGACACCGCAAGGACCACAGAAGACTTGAACAGCAGAAAGCCGCCTTCGCAAGACAGG		
-----+-----+-----+-----+-----+-----+-----+		
G	D	T
A	R	T
T	T	E
D	L	N
S	R	K
P	P	S
Q	D	R
R	H	R
K	D	H
R	R	L
E	Q	Q
K	A	A
F	A	R
Q	A	

4510	4530	4550
CAATCACGCTCGTCTGAATGTCTGGACAGAAGCGGAGAAAGGACAGGCAGTTCGTAACT		
-----+-----+-----+-----+-----+-----+-----+		
Q	S	R
S	S	E
C	L	D
R	S	G
E	R	T
G	S	S
L	T	
I	T	L
V	*	M
S	G	Q
K	R	R
K	D	R
Q	F	V
N	C	

P11b start

4570	4590	4610
GCCCCCACTGCTCCGAGCCCCTCATTCTCATTTTCGGAAAGAGCTCGACTGGCGACCGGG		
-----+-----+-----+-----+-----+-----+-----+		
A	P	T
A	P	S
P	S	P
S	F	S
F	S	E
R	A	R
L	A	T
G		
P	H	C
S	E	P
L	I	L
I	F	G
K	S	S
T	G	D
R	A	

4630	4650	4670
CCGACTGTCGCGCTGCGACATCACCTTCGGCAACCCCATCCTGCGCCACGGACCAGGTT		
-----+-----+-----+-----+-----+-----+-----+		
P	T	V
A	A	A
T	S	P
S	A	T
P	S	C
A	T	D
Q	V	
D	C	R
R	C	D
I	T	F
G	N	P
I	L	R
H	G	P
G	C	

4690 4710 4730
GCCGCGAGGACCACGCCGGACTTTGCGCCTTTCCTGGGTTCCTCAGTCTGCCCCGTGCTGTC
-----+-----+-----+-----+-----+-----+-----+-----+
A A R T T P D F A P F L G S Q S A R A V
R E D H A G L C A F P G F P V C P C C L

4750 4770 4790
TCGAAGCCGTACCGGCCCCCAGGACTGCCCGTTGGAAAGAAGTCACCCCGCTCCACGCG
-----+-----+-----+-----+-----+-----+-----+-----+
S K P Y R P P T T A R W K E V T P L H A
E A V P A P H D C P L E R S H P A P R V

4810 4830 4850
TGGAAGGGCGTGACCGGAGACCGACCGGAAGTCAGGGAGGACCCGGAGACAGCGGCGGTC
-----+-----+-----+-----+-----+-----+-----+-----+
W K G V T G D R P E V R E D P E T A A V
E G R D R R P T G S Q G G P G D S G G R

4870 4890 4910
GTCCAGGCTCTGATCAGCGGCCGTTATCCTCAGAAGACGAAGCTTTCCTCCGACGCATCC
-----+-----+-----+-----+-----+-----+-----+-----+
V Q A L I S G R Y P Q K T K L S S D A S
P G S D Q R P L S S E D E A F L R R I Q

4930 4950 4970
AAAGGCTACTCAAGAACTAAGGGATGCTCACAATCCACCTCTTTTCCTGCCCCGAGTGCG
-----+-----+-----+-----+-----+-----+-----+
K G Y S R T K G C S Q S T S F P A P S A
R L L K N * M L T I H L F S C P E C G
P14 start

4990 5010 5030
GATTACCAGGCCCGCGACTGCCAGACAGTCCGAGTCTGCCGCGCCGCTGCAGAGATGGCG
-----+-----+-----+-----+-----+-----+-----+
D Y Q A R D C Q T V R V C R A A A E M A
L P G P R L P D S P S L P R R C R D G A

5050 5070 5090
CGCTCATGTATTCACGAGCCGTTGGCTTCATCTGCCGCCAGTGCCGACTTGAAGCGCATA
-----+-----+-----+-----+-----+-----+-----+
R S C I H E P L A S S A A S A D L K R I
L M Y S R A V G F I C R Q C R L E A H T

5110 5130 5150
CGCTCTACCTCGGACTCTGTTCCCGATGTAAAGATCAGCAAGAGCGCATGAAGGAACAAA
-----+-----+-----+-----+-----+-----+-----+
R S T S D S V P D V K I S K S A *
L Y L G L C S R C K D Q Q E R M K E Q N

FIG. 1 Cont'd
SUBSTITUTE SHEET

5170 5190 5210
ATTAGTTTCCTTGTTTCGTAAACAAGGTGGTCCCTCCCATTTGAGGTAAAGACTCTGGTGAG
-----+-----+-----+-----+-----+-----+
*
5230 5250 5270
TCCTCAACGTTACTCGTTGAGTCTGCTGCGGTTCGATTCCATTCCCAAGCAGCAAAGGGT
-----+-----+-----+-----+-----+-----+
5290 5310
GCGCAACTAGTACGGCGCCCCCTGGGATACCA
-----+-----+-----+-----+-----+-----

Fig. 1 cont'd
SUBSTITUTE SHEET

20/45

FIG. 2

SUBSTITUTE SHEET

21/45

* Extra C residue here in "5C Version"

610 630 650
ATTGGGCCAAGGGAAAGATCGACCTCGACTCCGATTCCATCGGCTGGTACTTCAAGTACC
-----+-----+-----+-----+-----+-----+-----+
I G P R E R S T S T P I P S A G T S S T
W A K G K I D L D S D S I G W Y F K Y L

670 690 710
TTGACCCAGCGGGTGCTACAGAGTCTGCGCGCGCCGTCGGCGAGTACTCGAAGATCCCTG
-----+-----+-----+-----+-----+-----+-----+
L T Q R V L Q S L R A P S A S T R R S L
D P A G A T E S A R A V G E Y S K I P D

730 750 770
ACGGCCTCGTCAAGTTCTCCGTCGACGCAGAGATAAGAGAGATCTATAACGAGGAGTGCC
-----+-----+-----+-----+-----+-----+-----+
T A S S S S P S T Q R *
G L V K F S V D A E I R E I Y N E E C P

790 810 830
CCGTCGTCACTGACGTGTCCGTCCCCCTCGACGGCCGCCAGTGGAGCCTCTCGATTTTCT
-----+-----+-----+-----+-----+-----+-----+
V V T D V S V P L D G R Q W S L S I F S

FIG. 2 Cont'd

SUBSTITUTE SHEET

850 870 890
CCTTTCCGATGTTTCAGAACCGCCTACGTCGCCGTAGCGAACGTCGAGAACAAGGAGATGT
-----+-----+-----+-----+-----+-----+-----+
F P M F R T A Y V A V A N V E N K E M S

910 930 950
CGCTCGACGTTGTCAACGACCTCATCGAGTGGCTCAACAATCTCGCCGACTGGCGTTATG
-----+-----+-----+-----+-----+-----+-----+
L D V V N D L I E W L N N L A D W R Y V

970 990 1010
TCGTTGACTCTGAACAGTGGATTAACCTTCACCAATGACACCACGTACTACGTCCGCATCC
-----+-----+-----+-----+-----+-----+-----+
V D S E Q W I N F T N D T T Y Y V R I R

1030 1050 1070
GCGTTCTACGTCCAACCTACGACGTTCCAGACCCACAGAGGGCCTTGTTTCGCACAGTCT
-----+-----+-----+-----+-----+-----+-----+
V L R P T Y D V P D P T E G L V R T V S

1090 1110 1130
CAGACTACCGCCTCACTTATAAGGCGATAACATGTGAAGCCAACATGCCAACACTCGTCG
-----+-----+-----+-----+-----+-----+-----+
D Y R L T Y K A I T C E A N M P T L V D

SUBSTITUTE SHEET

1150	1170	1190																	
ACCAAGGCTTTTGGATCGGCGGCCAGTACGCTCTACCCCGACTAGCCTACCGCAGTACG																			
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+																			
Q	G	F	W	I	G	G	Q	Y	A	L	T	P	T	S	L	P	Q	Y	D
1210	1230	1250																	
ACGTCAGCGAGGCCTACGCTCTGCACACTTTGACCTTCGCCAGACCATCCAGCGCCGCTG																			
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+																			
V	S	E	A	Y	A	L	H	T	L	T	F	A	R	P	S	S	A	A	A
1270	1290	1310																	
CACTCGCGTTTGTGTGGGCAGGTTTGCCACAGGGTGGCACTGCGCCTGCAGGCACTCCAG																			
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+																			
L	A	F	V	W	A	G	L	P	Q	G	G	T	A	P	A	G	T	P	A
1330	1350	1370																	
CCTGGGAGCAGGCATCCTCGGGTGGCTACCTCACCTGGCGCCACAACGGTACTACTTTCC																			
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+																			
W	E	Q	A	S	S	G	G	Y	L	T	W	R	H	N	G	T	T	F	P
1390	1410	1430																	
CAGCTGGCTCCGTTAGCTACGTTCTCCCTGAGGGTTTCGCCCTTGAGCGCTACGACCCGA																			
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+																			
A	G	S	V	S	Y	V	L	P	E	G	F	A	L	E	R	Y	D	P	N

FIG. 2 Cont'd

1450 1470 1490
ACGACGGCTCTTGGACCGACTTCGCTTCCGCAGGAGACACCGTCACTTTCGGGCAGGTCG
-----+-----+-----+-----+-----+-----+-----+
D G S W T D F A S A G D T V T F R Q V A

1510 1530 1550
CCGTCGACGAGGTCGTTGTGACCAACAACCCCGCCGGCGGCAGCGCCCCACCTTCA
-----+-----+-----+-----+-----+-----+-----+
V D E V V V T N N P A G G G S A P T F T

1570 1590 1610
CCGTGAGAGTGCCCCCTTCAAACGCTTACACCAACACCGTGTTTAGGAACACGCTCTTAG
-----+-----+-----+-----+-----+-----+-----+
V R V P P S N A Y T N T V F R N T L L E

1630 1650 1670
AGACTCGACCCTCCTCTCGTAGGCTCGAACTCCCTATGCCACCTGCTGACTTTGGACAGA
-----+-----+-----+-----+-----+-----+-----+
T R P S S R R L E L P M P P A D F G Q T

1690 1710 1730
CGGTCGCCAACAACCCGAAGATCGAGCAGTCGCTTCTTAAAGAAACACTTGGCTGCTATT
-----+-----+-----+-----+-----+-----+-----+
V A N N P K I E Q S L L K E T L G C Y L

FIG. 2 Cont'd

SUBSTITUTE SHEET

1750 1770 1790
TGGTCCACTCCAAAATGCGAAACCCCGTTTTCCAGCTCACGCCAGCCAGCTCCTTTGGCG
-----+-----+-----+-----+-----+-----+-----+
V H S K M R N P V F Q L T P A S S F G A

1810 1830 1850
CCGTTTCCTTCAACAATCCGGGTTATGAGCGCACACGCGACCTCCCGGACTACACTGGCA
-----+-----+-----+-----+-----+-----+-----+
V S F N N P G Y E R T R D L P D Y T G I

1870 1890 1910
TCCGTGACTCATTCGACCAGAACATGTCCACCGCTGTGGCCCACTTCCGCTCACTCTCCC
-----+-----+-----+-----+-----+-----+-----+
R D S F D Q N M S T A V A H F R S L S H

1930 1950 1970
ACTCCTGCAGTATCGTCACTAAGACCTACCAGGGTTGGGAAGGCGTCACGAACGTCAACA
-----+-----+-----+-----+-----+-----+-----+
S C S I V T K T Y Q G W E G V T N V N T

1990 2010 2030
CGCCTTTCGGCCAATTCGCGCACGCGGGCCTCCTCAAGAATGAGGAGATCCTCTGCCTCG
-----+-----+-----+-----+-----+-----+-----+
P F G Q F A H A G L L K N E E I L C L A

FIG. 2 Cont'd

SUBSTITUTE SHEET

2050 2070 2090
CCGACGACCTGGCCACCCGTCTCACAGGTGTCTACCCCGCCACTGACAACTTCGCGGCCG
-----+-----+-----+-----+-----+-----+-----+
D D L A T R L T G V Y P A T D N F A A A

2110 2130 2150
CCGTTTCTGCCCTTCGCCGCGAACATGCTGTCCTCCGTGCTGAAGTCGGAGGCAACGTCCT
-----+-----+-----+-----+-----+-----+-----+
V S A F A A N M L S S V L K S E A T S S

2170 2190 2210
CCATCATCAAGTCCGTTGGCGAGACTGCCGTCGGCGCGGCTCAGTCCGGCCTCGCGAAGC
-----+-----+-----+-----+-----+-----+-----+
I I K S V G E T A V G A A Q S G L A K L

2230 2250 2270
TACCCGGACTGCTAATGAGTGTACCAGGGAAGATTGCCGCGCGTGTCCGCGCGCGCCGAG
-----+-----+-----+-----+-----+-----+-----+
P G L L M S V P G K I A A R V R A R R A

2290 2310 2330
CGCGCCGCGCGCCGCTCGTGCCAATTAGTTTGCTCGCTCCTGTTTCGCCGTTTCGTAAA
-----+-----+-----+-----+-----+-----+-----+
R R R A A R A N *

FIG. 2 Cont'd

SUBSTITUTE SHEET

2350 2370 2390
 ACGGCGTGGTCCCGCACATTACGCGTACCCTAAAGACTCTGGTGAGTCCCGTCGTTACA
 -----+-----+-----+-----+-----+-----+-----+

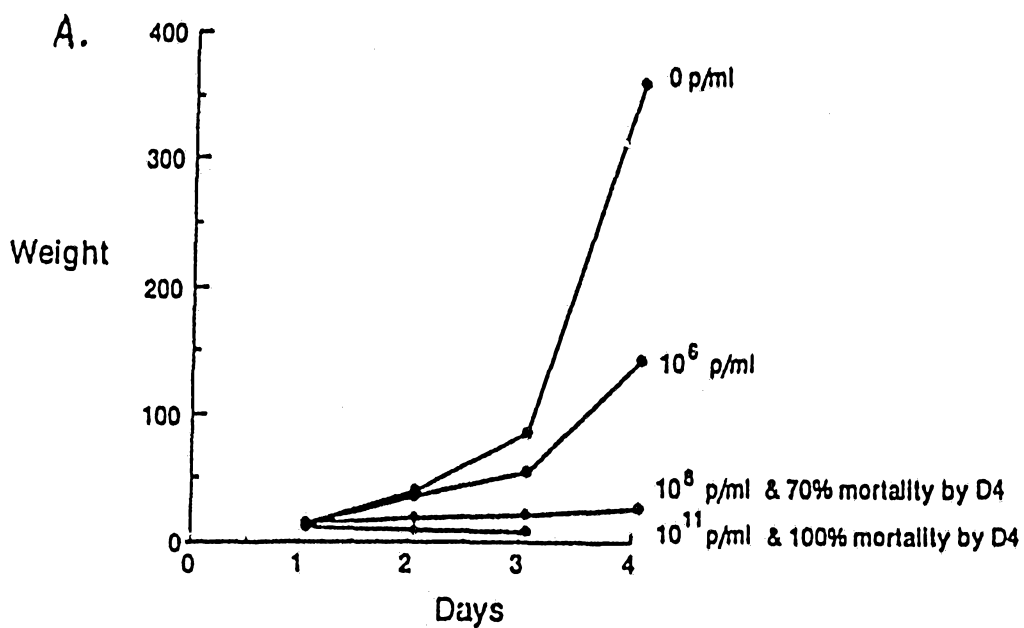
2410 2430 2450
 CGACGGGTCTGCCGCGGTTCGATTCCATTCCCAAGCGGCAAGAAGGACGTAGTTAGCTCT
 -----+-----+-----+-----+-----+-----+-----+

2470
 GCGTCCCTCGGGATACCA
 -----+-----

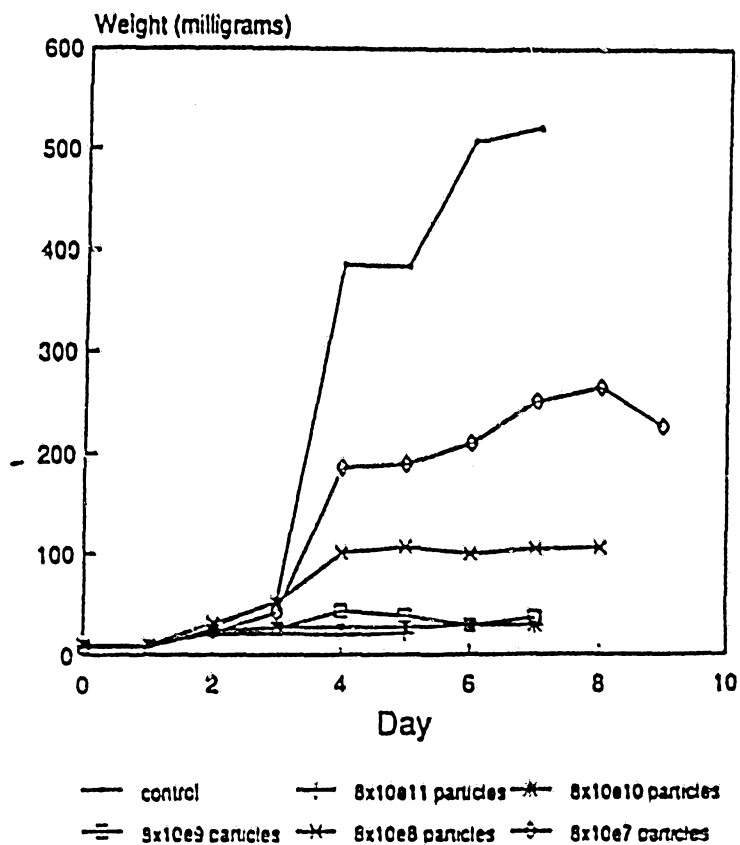
28/45

29/45

FIG. 3



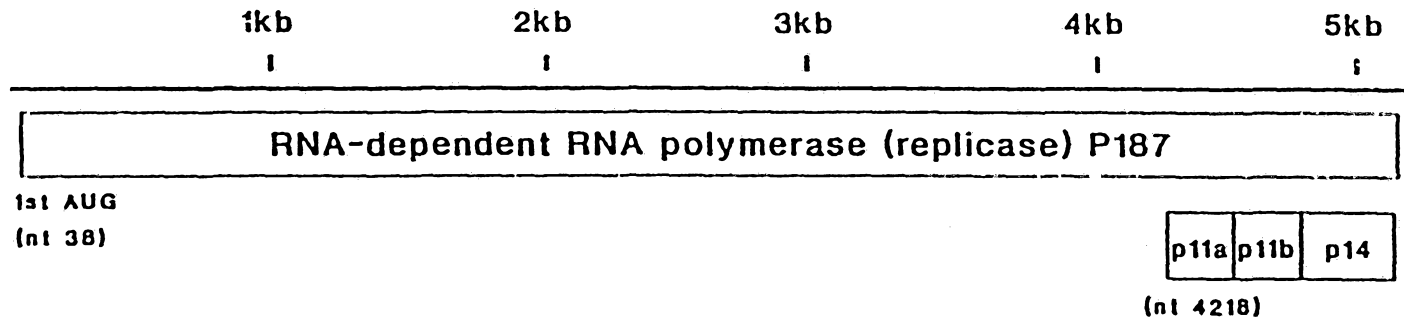
B Weight gain of infected larvae



SUBSTITUTE SHEET

Proteins encoded by the HaSV genome

RNA 1



RNA 2

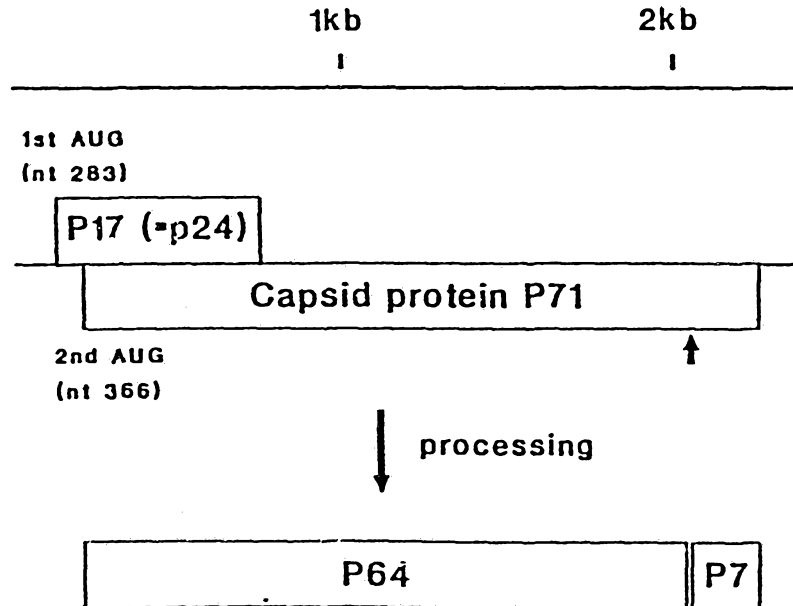


FIG. 4

PROTEINS EXPRESSED FROM HaSV RNA 2

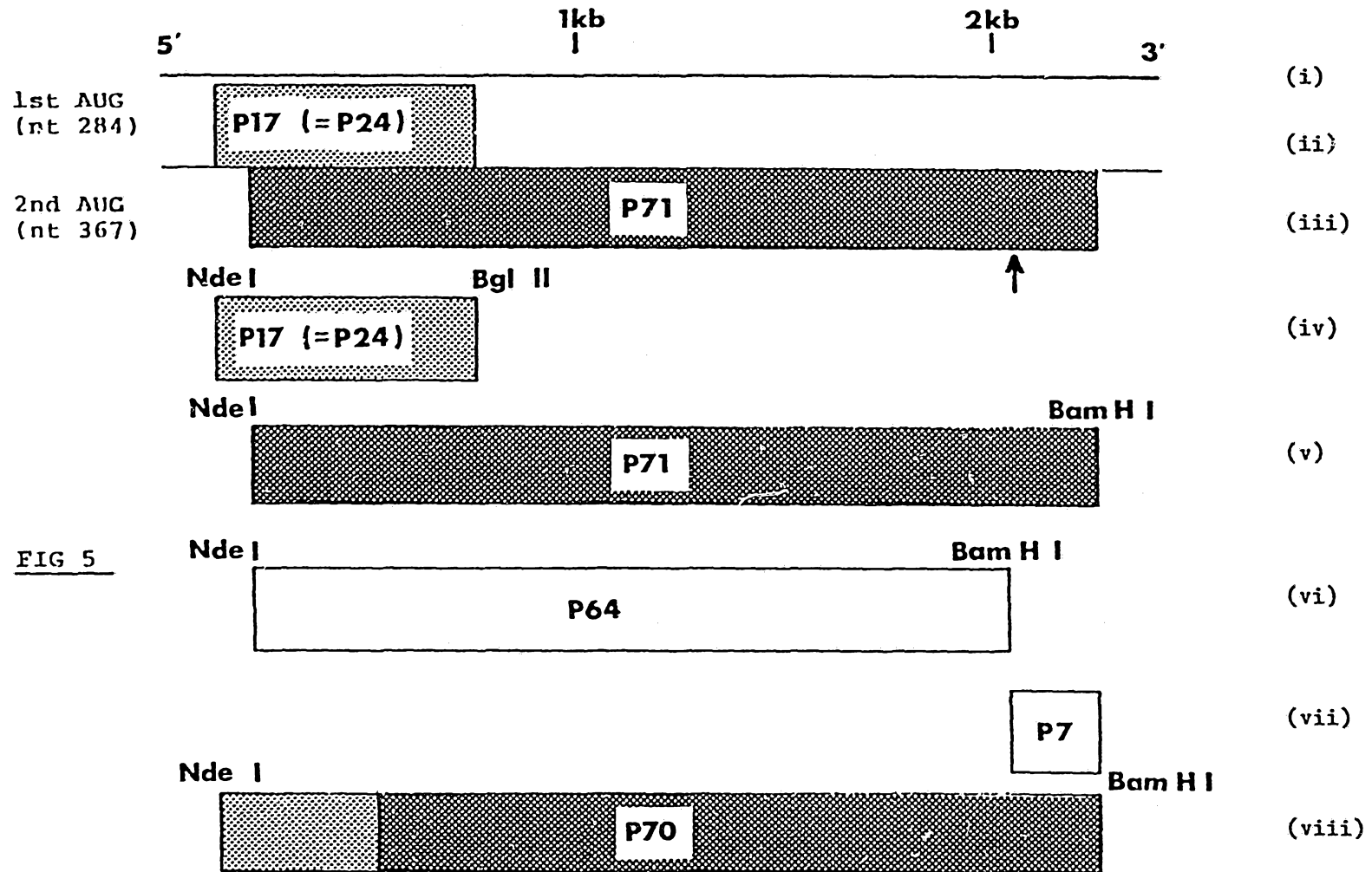


FIG 5

SUBSTITUTE SHEET

RNA 2 (11)

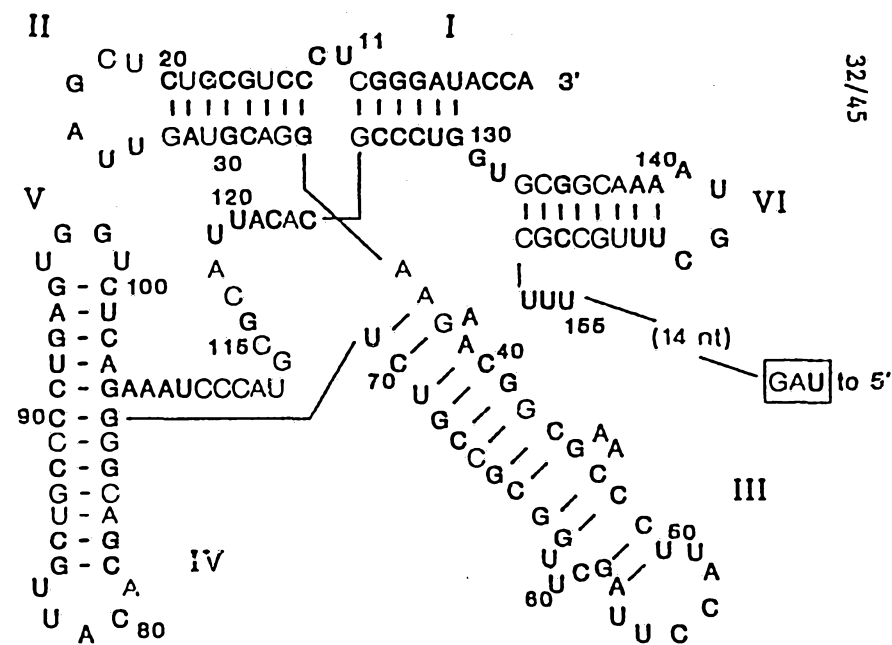
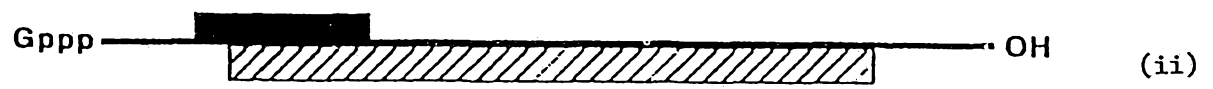


FIG. 6

HaSV RNA1

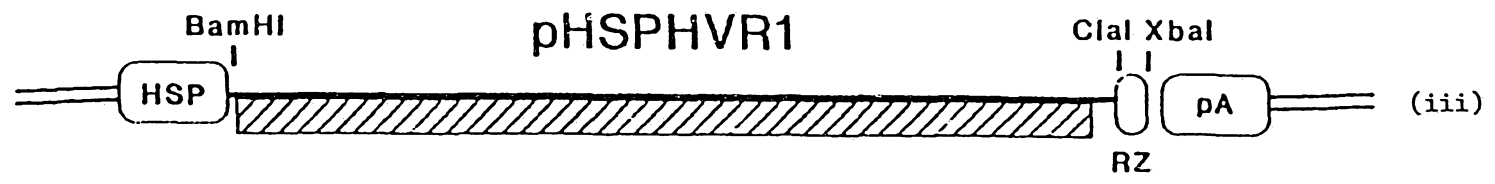


HaSV RNA2



Insect cell expression constructs

pHSPHVR1



pHSPHVR2

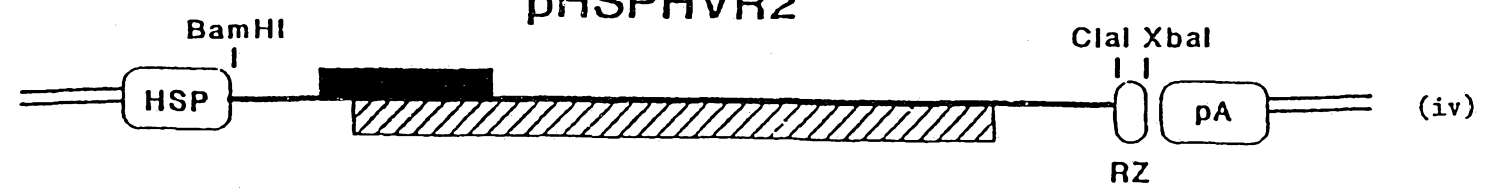
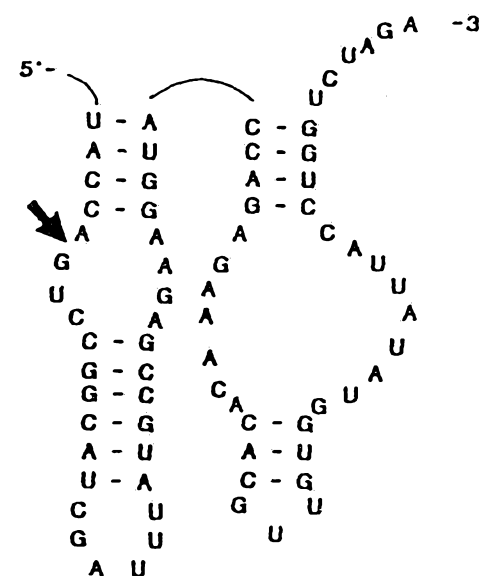


FIG. 7

SUBSTITUTE SHEET

34/45

(ii) **HAIRPIN**



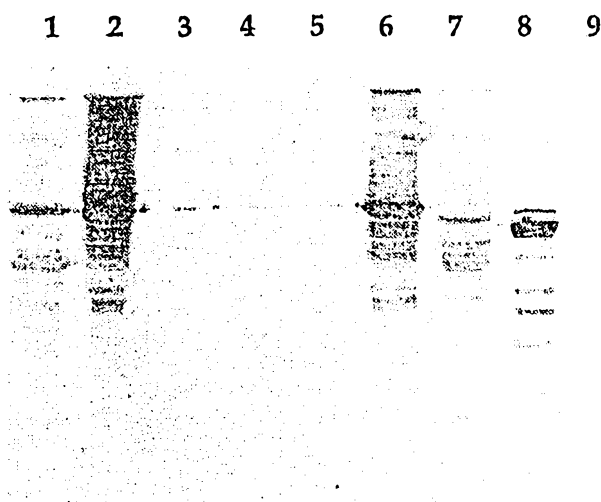
SUBSTITUTE SHEET

FIG. 8

Fig. 9

WESTERN BLOTS OF HaSV CAPSID PROTEIN

A. HaSV ANTISERUM



B. HaSV ANTISERUM



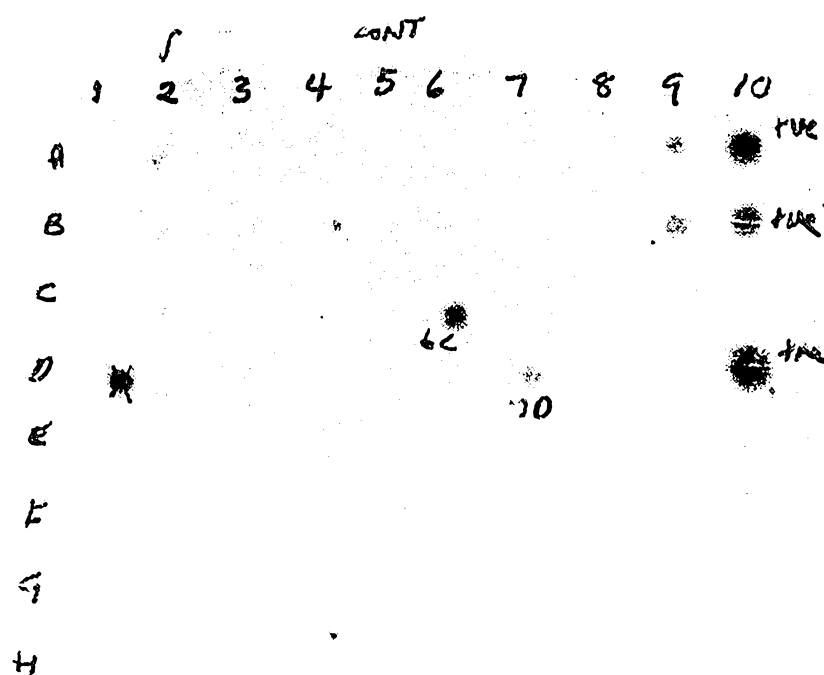
C. Bt ANTISERUM

1 2 3

Bt tain a/s

Fig. 10

DOT-BLOT DETECTION OF HaSV IN FIELD-COLLECTED HELICOVERPA LARVAE



SUBSTITUTE SHEET

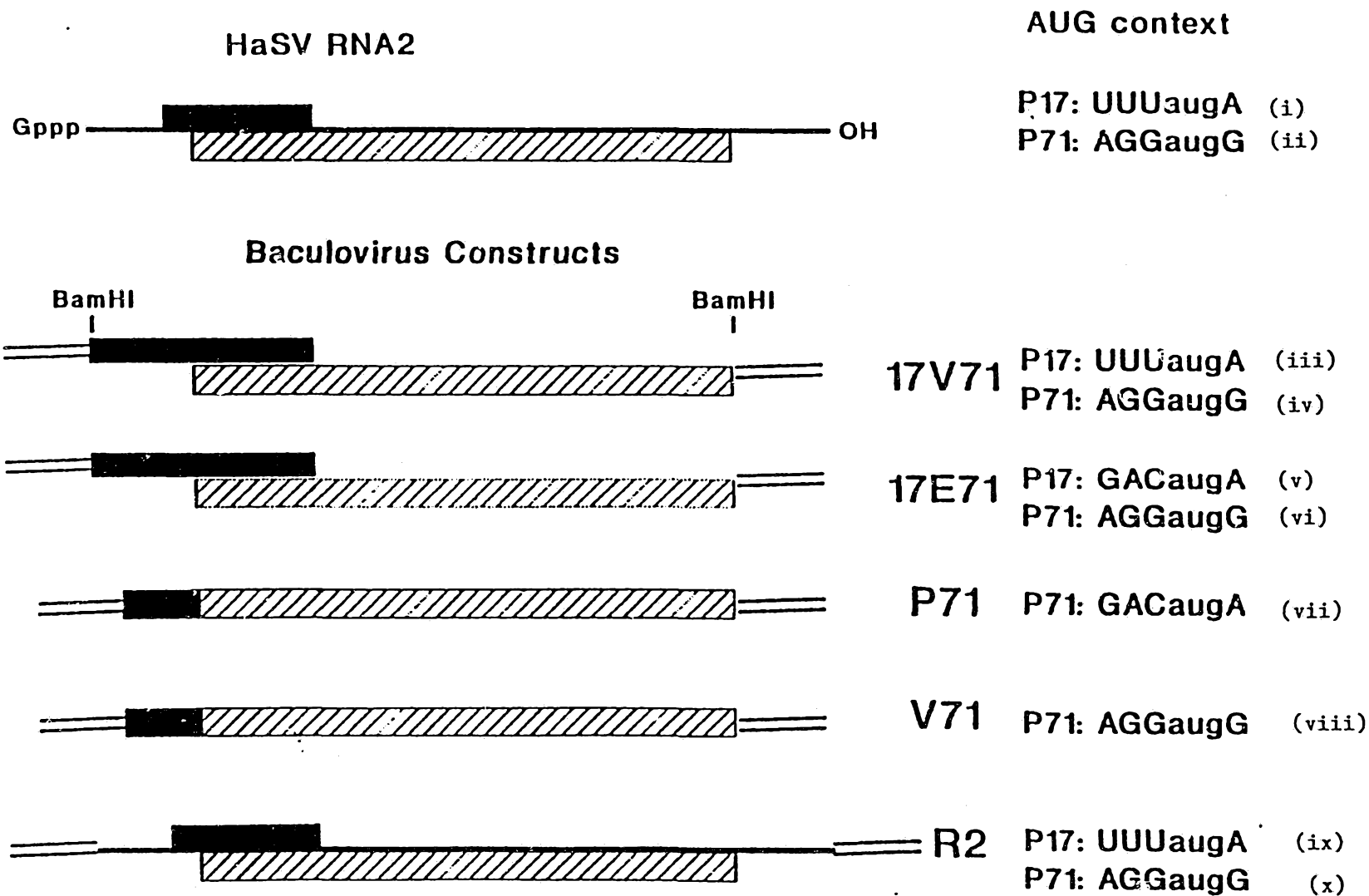


FIG. 11

38/45

virus capsoid strategy:

capsotoxin encapsulation

transgenic plant genome:

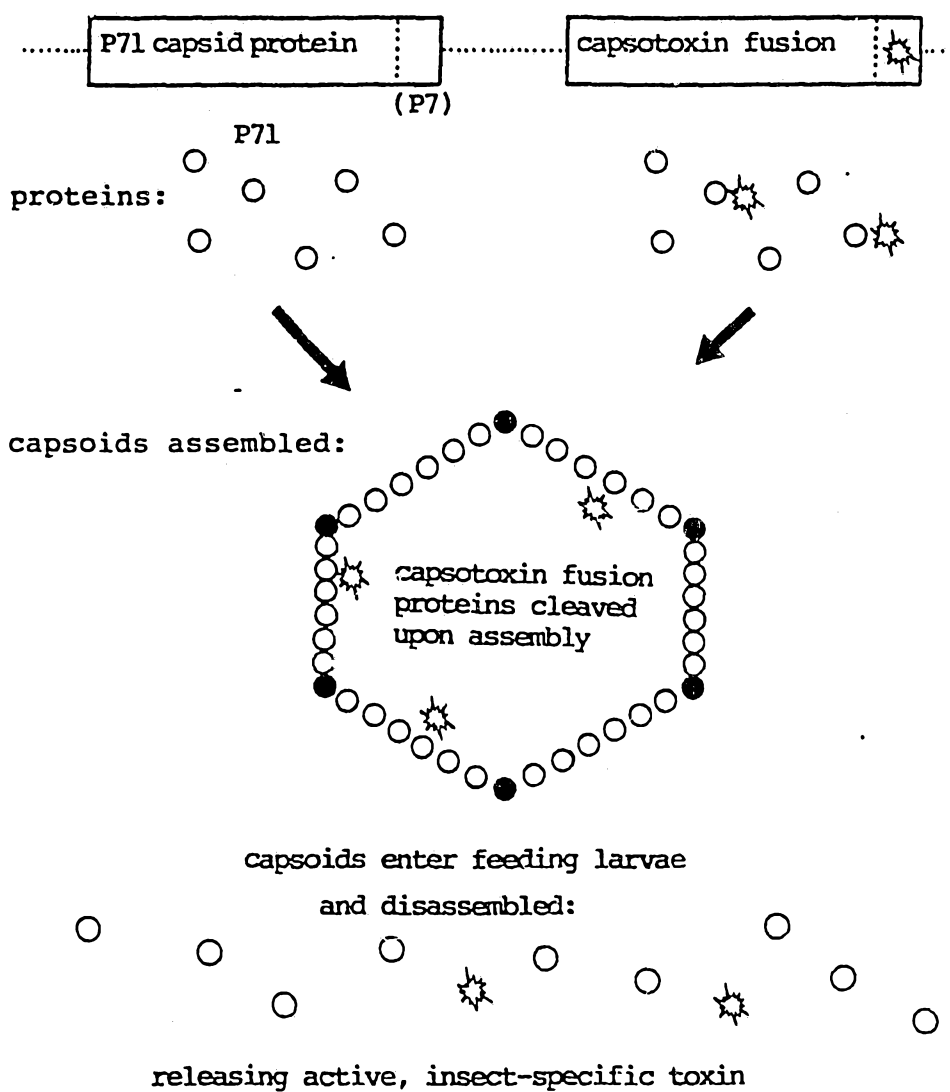


FIGURE 12a

SUBSTITUTE SHEET

39/45

HaSV capsoid strategy:

toxin message encapsulation and mplification

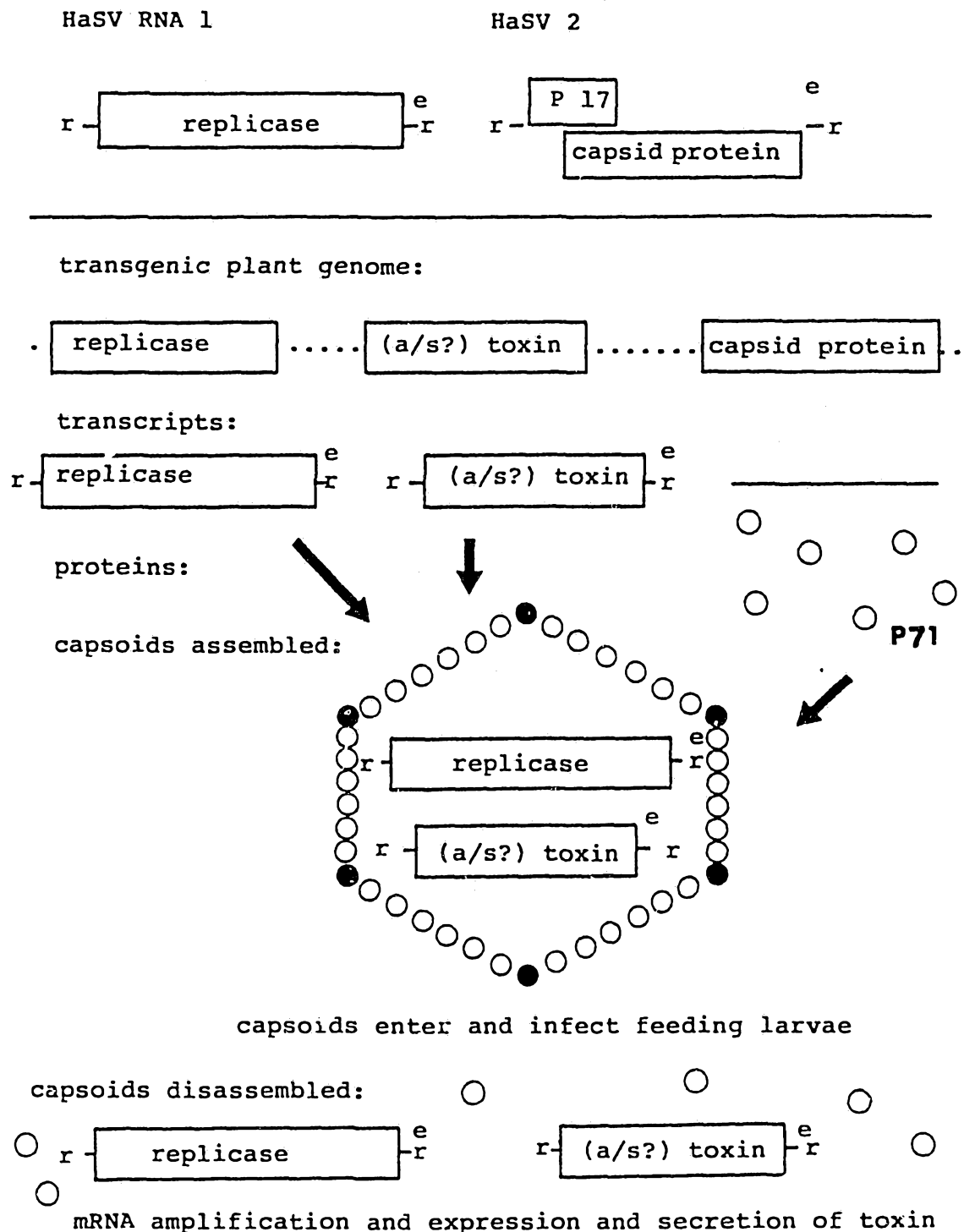


FIGURE 12b

SUBSTITUTE SHEET

40/45

HaSV expression in plants:
the one-way vector

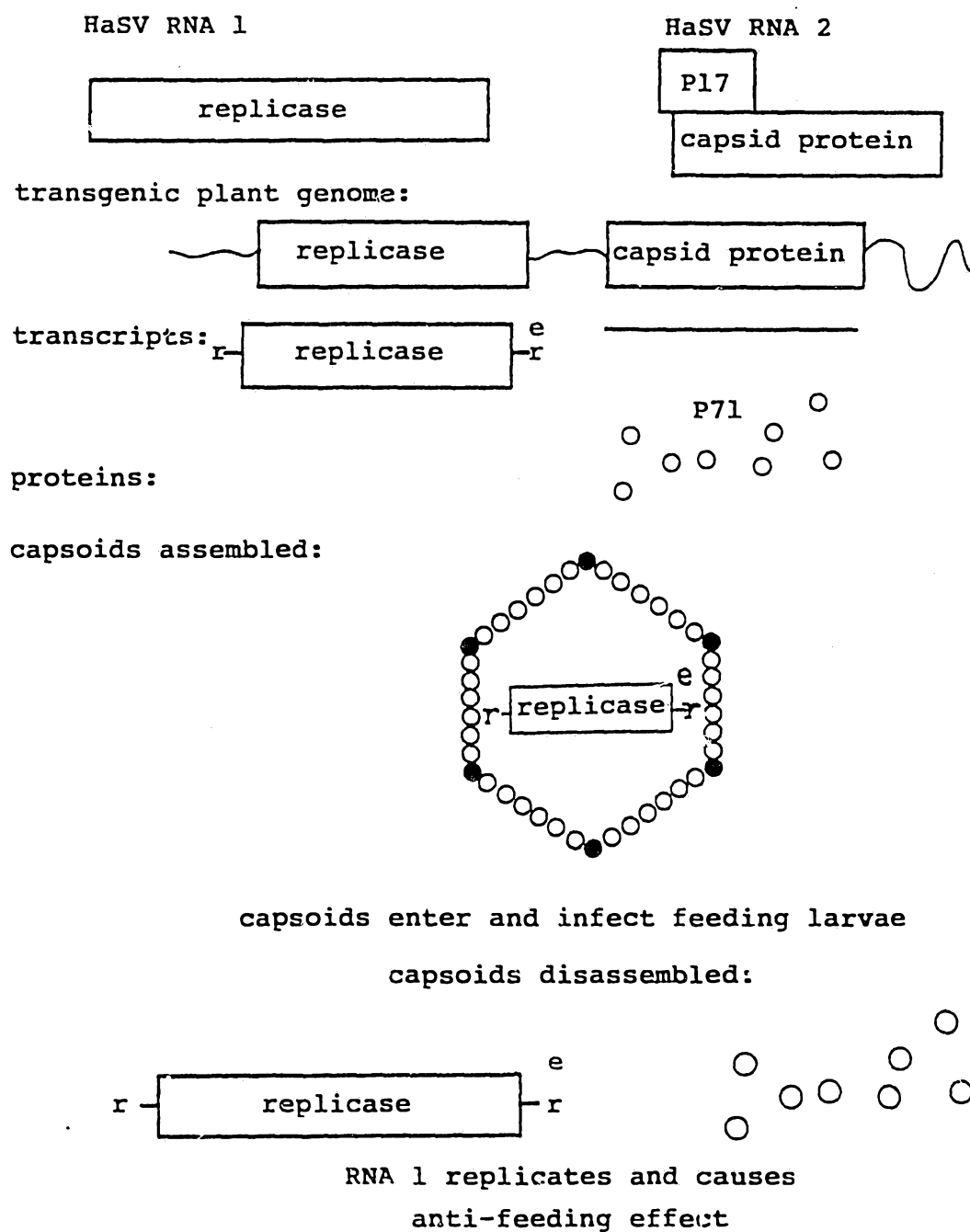


FIGURE 12c

SUBSTITUTE SHEET

41/45

HaSV expression in plants:

the one-way vector

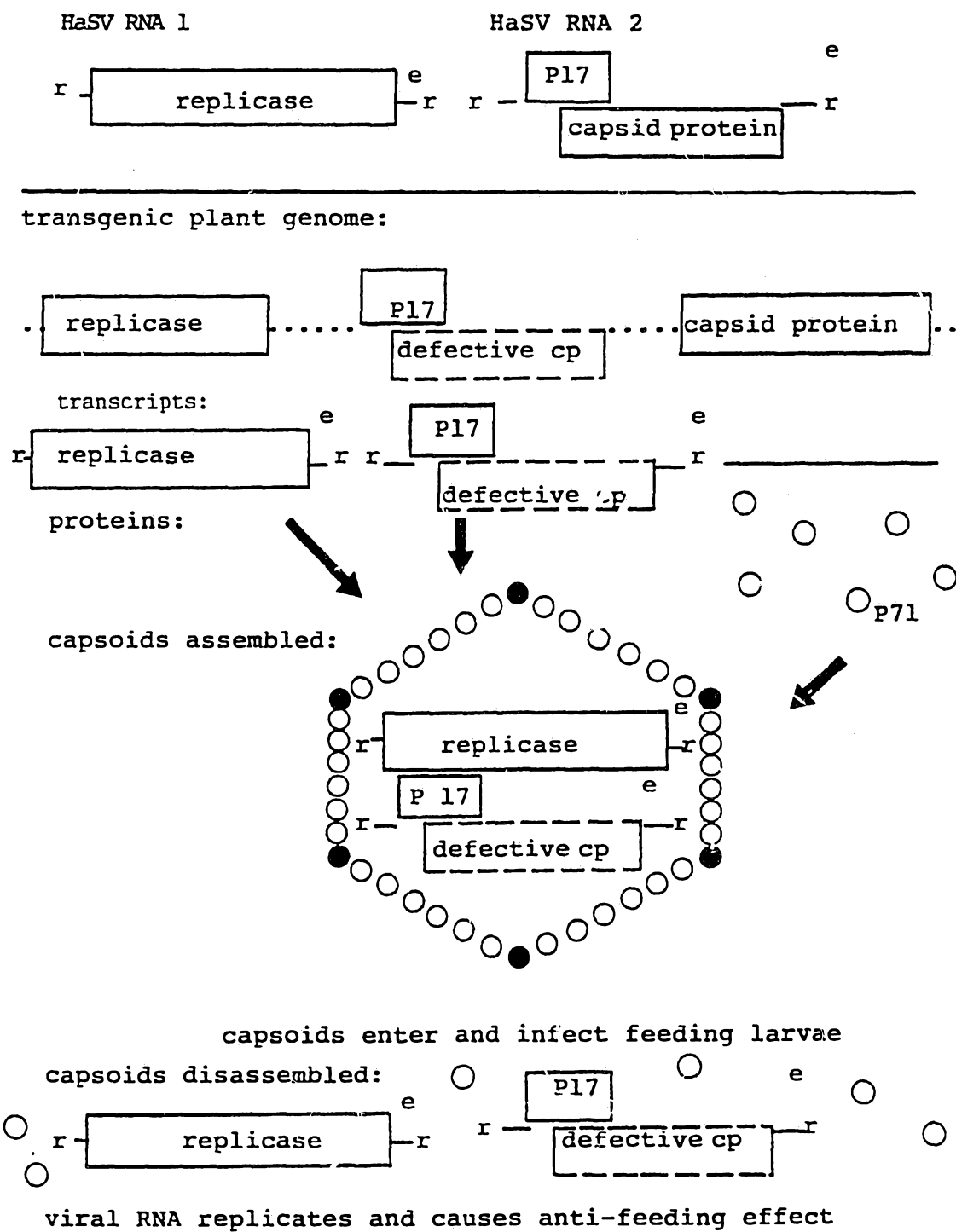


FIGURE 12d

SUBSTITUTE SHEET

42/45

HaSV expression in plants:

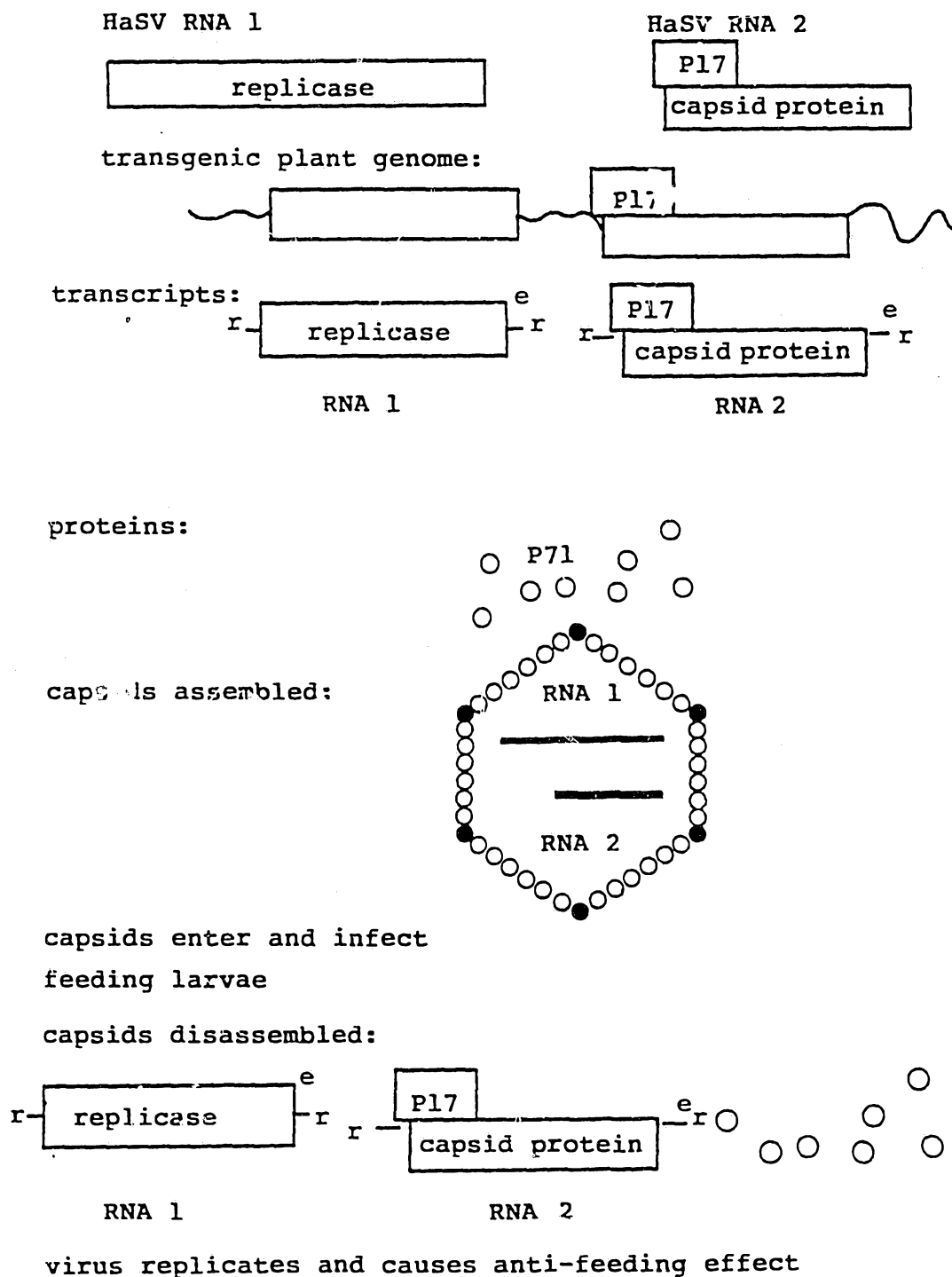


FIGURE 12e

SUBSTITUTE SHEET

43/45

HaSV expression in plants:
the one-way vector for a toxin

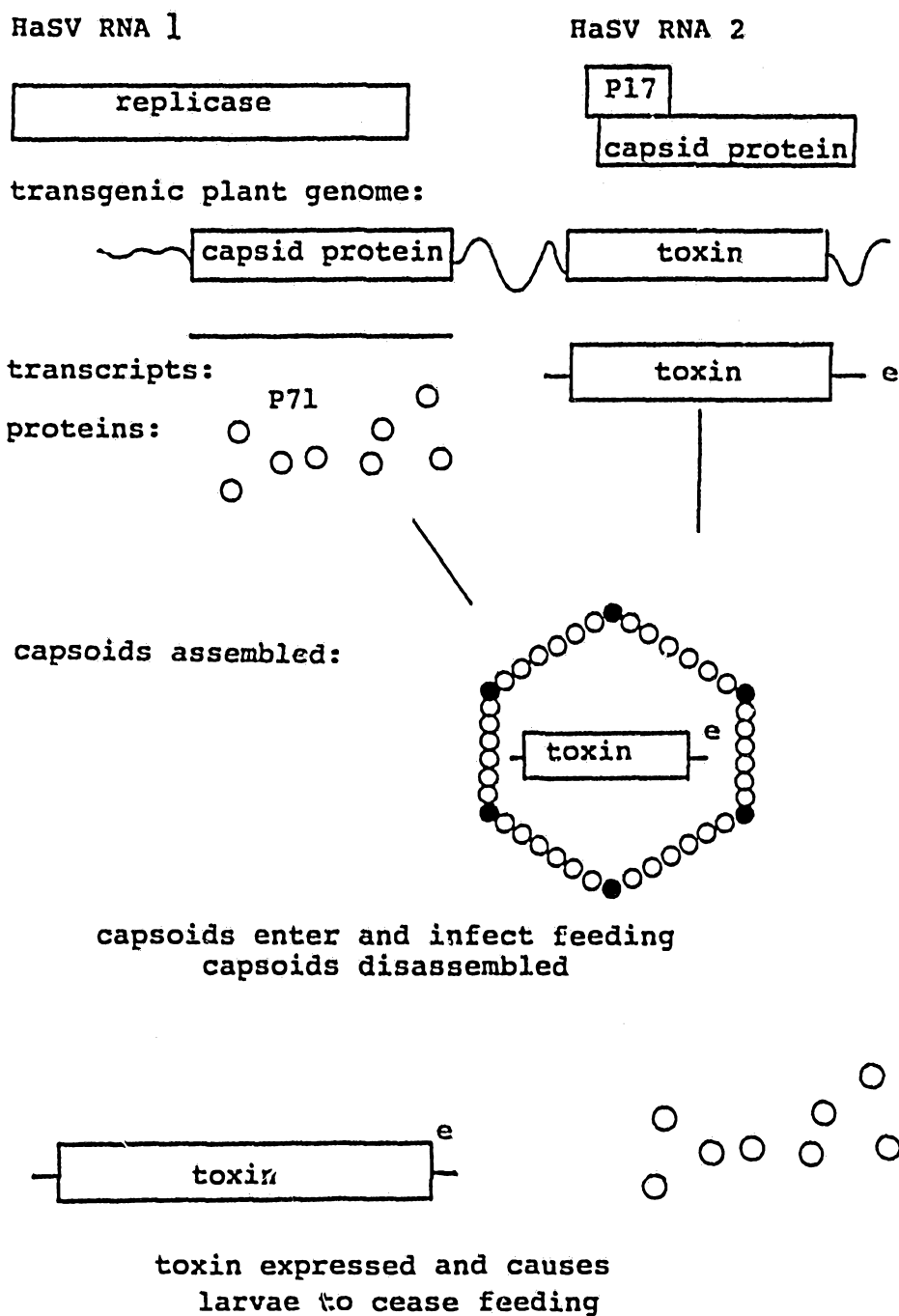
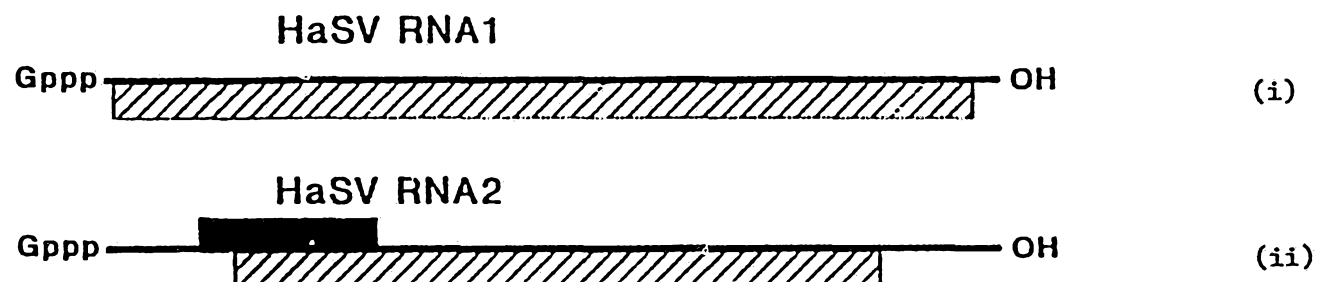


FIGURE 12f

SUBSTITUTE SHEET



baculovirus expression constructs

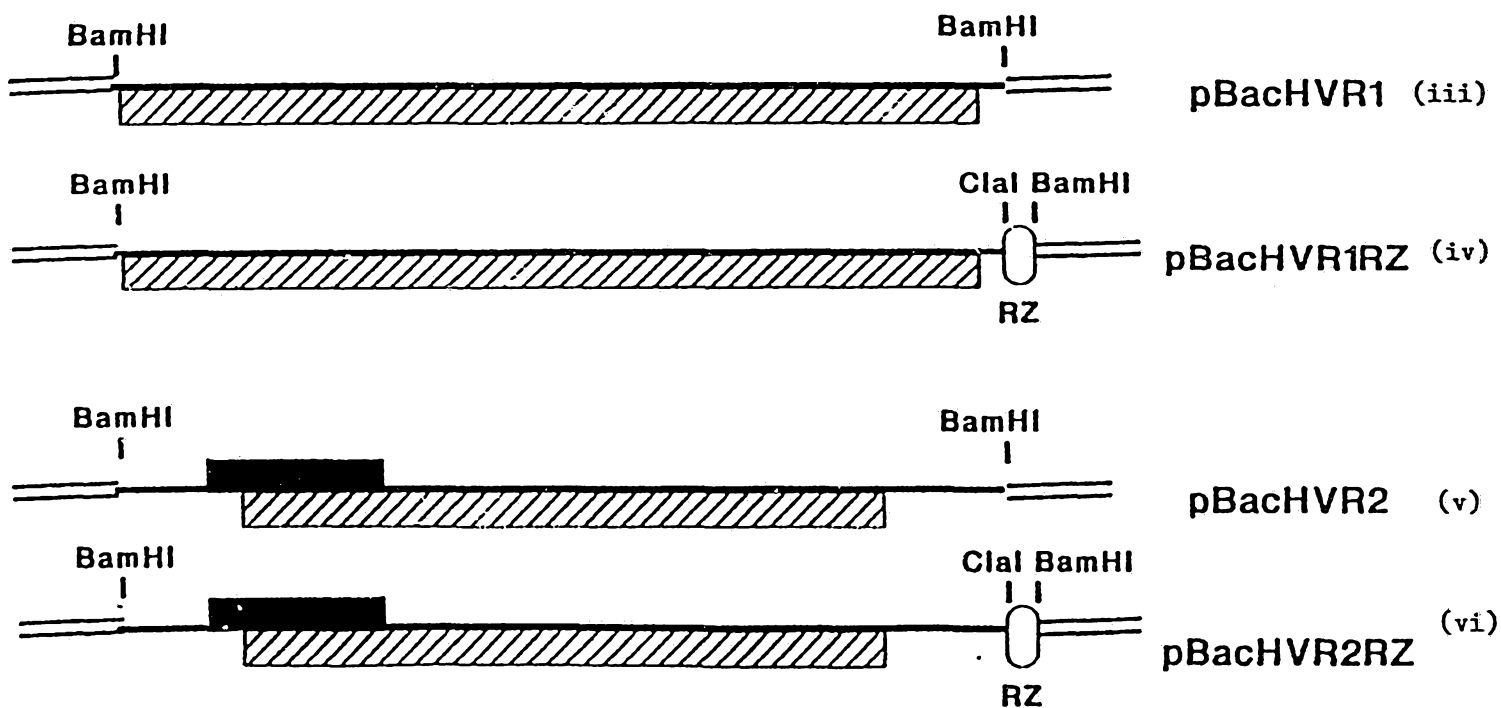


FIG. 13

44/45



Protoplast expression constructs

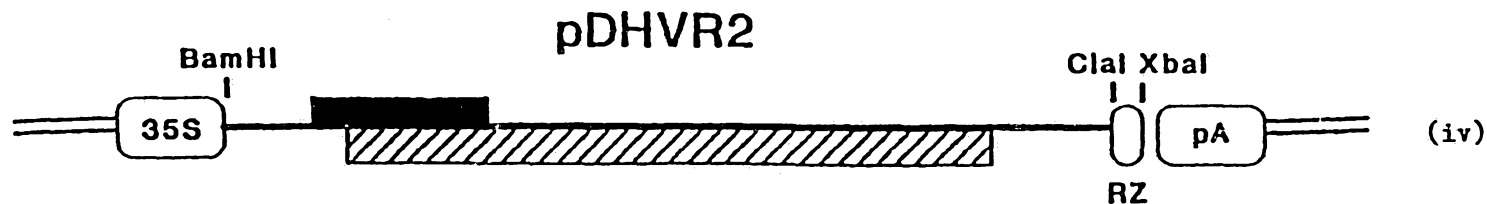


FIG. 14

A. CLASSIFICATION OF SUBJECT MATTER Int. Cl. ⁵ C12N 7/00, 15/11, 15/40, 15/86, C12Q 1/68, C12P 21/08, A01H 5/00, A01N 63/00 According to International Patent Classification (IPC) or to both national classification and IPC				
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC: C12N 7/00, C12N 15/40, C12N 15/86, A01N 63/00, A01H 5/00 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base, and where practicable, search terms used) WPAT) RNA(W) VIRUS#, INSECT(W) VIRUS#, HELIOTHIS(S) ARMIGERA and STN D/B + ORZIT(WPAT) CASA) KEYWORDS: SMALL () RNA () VIRUS ##, PLASMID () PT7T2B, PLASMID () PT7T2C, BIOT)				
C. DOCUMENTS CONSIDERED TO BE RELEVANT				
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.		
X	Applied and Environmental Microbiology, volume 53, No. 1, January 1987, T Manousis and N F Moore; "Cricket Paralysis Virus, a Potential Control Agent for the Olive Fruit-Fly, Dacus Oleae Gmel", pages 142-148 (whole article)	1, 52		
X	WO 88/01833 (Institut Pasteur and Institut Francais de Recherche Scientifique pour le developpement en cooperation), 24 March 1988 (24.03.88) See abstract, page 5 line 25-page 9 line 4	1		
<div style="display: flex; justify-content: space-between;"> <div> <input type="checkbox"/> Further documents are listed in the continuation of Box C. </div> <div> <input type="checkbox"/> See patent family annex. </div> </div>				
<table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top;"> <p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"C" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </td> <td style="width: 50%; vertical-align: top;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&" document member of the same patent family</p> </td> </tr> </table>			<p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"C" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&" document member of the same patent family</p>
<p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"C" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&" document member of the same patent family</p>			
Date of the actual completion of the international search 22 October 1993 (22.10.93)		Date of mailing of the international search report 1 NOV 1993 (1.11.93)		
Name and mailing address of the ISA/AU AUSTRALIAN INDUSTRIAL PROPERTY ORGANISATION PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No. 06 2853929		Authorized officer CARMELA MONGER Telephone No. (06) 2832486		