



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |           |                                                                                                                                                                                                                                                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>(51) International Patent Classification <sup>5</sup> :</b><br><b>C12N 15/40, 15/86, C07K 15/08, C12N 5/10, C12P 21/00</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <b>A2</b> | <b>(11) International Publication Number:</b> <b>WO 94/13812</b><br><b>(43) International Publication Date:</b> 23 June 1994 (23.06.94)                                                                                                                   |
| <b>(21) International Application Number:</b> PCT/US93/11907<br><b>(22) International Filing Date:</b> 7 December 1993 (07.12.93)<br><br><b>(30) Priority Data:</b><br>07/991,867 7 December 1992 (07.12.92) US<br><br><b>(71) Applicant (for all designated States except US):</b> UNIVERSITY OF FLORIDA [US/US]; 186 Grinter Hall, Gainesville, FL 32611 (US).<br><br><b>(72) Inventors; and</b><br><b>(75) Inventors/Applicants (for US only):</b> MOYER, Richard, W. [US/US]; 1225 N.W. 23rd Terrace, Gainesville, FL 32605 (US). HALL, Richard, L. [US/US]; Apartment 178, 3611 S.W. 34th Street, Gainesville, FL 32608 (US). GRUIDL, Michael, E. [US/US]; 211 Alhambra, Columbia, MO 65203 (US).<br><br><b>(74) Agents:</b> SALIWANCHIK, David, R. et al.; Saliwanchik & Saliwanchik, Suite A-1, 2421 N.W. 41st Street, Gainesville, FL 32606 (US). |           | <b>(81) Designated States:</b> AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).<br><br><b>Published</b><br><i>Without international search report and to be republished upon receipt of that report.</i> |
| <b>(54) Title:</b> NOVEL ENTOMOPOXVIRUS GENES, PROTEINS AND METHODS OF USE THEREOF<br><br><b>(57) Abstract</b><br><br>The subject invention pertains to novel Entomopoxvirus (EPV) polynucleotide sequences free from association with other viral sequences with which they are naturally associated, recombinant polynucleotide vectors containing the sequences, recombinant viruses containing the sequences, and host cells infected with the recombinant viruses are provided herein, as well as methods for use thereof in the expression of heterologous proteins in both insect and mammalian host cells.                                                                                                                                                                                                                                        |           |                                                                                                                                                                                                                                                           |

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |                                       |    |                          |
|----|--------------------------|----|---------------------------------------|----|--------------------------|
| AT | Austria                  | GB | United Kingdom                        | MR | Mauritania               |
| AU | Australia                | GE | Georgia                               | MW | Malawi                   |
| BB | Barbados                 | GN | Guinea                                | NE | Niger                    |
| BE | Belgium                  | GR | Greece                                | NL | Netherlands              |
| BF | Burkina Faso             | HU | Hungary                               | NO | Norway                   |
| BG | Bulgaria                 | IE | Ireland                               | NZ | New Zealand              |
| BJ | Benin                    | IT | Italy                                 | PL | Poland                   |
| BR | Brazil                   | JP | Japan                                 | PT | Portugal                 |
| BY | Belarus                  | KE | Kenya                                 | RO | Romania                  |
| CA | Canada                   | KG | Kyrgyzstan                            | RU | Russian Federation       |
| CF | Central African Republic | KP | Democratic People's Republic of Korea | SD | Sudan                    |
| CG | Congo                    | KR | Republic of Korea                     | SE | Sweden                   |
| CH | Switzerland              | KZ | Kazakhstan                            | SI | Slovenia                 |
| CI | Côte d'Ivoire            | LI | Liechtenstein                         | SK | Slovakia                 |
| CM | Cameroon                 | LK | Sri Lanka                             | SN | Senegal                  |
| CN | China                    | LU | Luxembourg                            | TD | Chad                     |
| CS | Czechoslovakia           | LV | Latvia                                | TG | Togo                     |
| CZ | Czech Republic           | MC | Monaco                                | TJ | Tajikistan               |
| DE | Germany                  | MD | Republic of Moldova                   | TT | Trinidad and Tobago      |
| DK | Denmark                  | MG | Madagascar                            | UA | Ukraine                  |
| ES | Spain                    | ML | Mali                                  | US | United States of America |
| FI | Finland                  | MN | Mongolia                              | UZ | Uzbekistan               |
| FR | France                   |    |                                       | VN | Viet Nam                 |
| GA | Gabon                    |    |                                       |    |                          |

DESCRIPTIONNOVEL ENTOMOPOXVIRUS GENES, PROTEINS AND  
METHODS OF USE THEREOF

5

This invention was made with Government support under Grant No. R01 AI15722-12 awarded by the National Institutes of Health and NIH Training Grant T32 AI-07110. The Government has certain rights in this invention.

10

Cross-Reference to a Related Application

This is a continuation-in-part of co-pending PCT application WO/92/14818, which is a continuation-in-part of co-pending U.S. application Serial No. 07/827,685, filed January 30, 1992, which is a continuation-in-part of U.S. application Serial No. 07/657,584, filed February 19, 1991, now abandoned.

15

Field of the Invention

This invention relates generally to the field of recombinantly-produced proteins and specifically to novel, recombinant Entomopoxvirus genes, proteins, protein regulatory sequences and their uses in expressing heterologous genes in transformed hosts.

20

Background of the Invention

Poxviruses are taxonomically classified into the family Chordopoxvirinae, whose members infect vertebrate hosts, e.g., the Orthopoxvirus vaccinia, or into the family Entomopoxvirinae. Very little is known about members of the Entomopoxvirinae family other than the insect host range of individual members. One species of Entomopoxvirus (EPV) is the *Amsacta moorei* Entomopoxvirus (AmEPV), which was first isolated from larvae of the red hairy caterpillar *Amsacta moorei* (Roberts and Granados [1968] *J. Invertebr. Pathol.* 12:141-143). AmEPV is the type species of genus B of EPVs and is one of three known EPVs which will replicate in cultured insect cells (R. R. Granados et al. [1976] "Replication of *Amsacta moorei* Entomopoxvirus and *Autographa californica*

25

30

Nuclear Polyhedrosis Virus in Hemocyte Cell Lines from *Estigmene acrea*,” in *Invertebrate Tissue Culture Applications in Medicine, Biology, and Agriculture*, E. Kurstak and K. Maramorosch (ed.), Academic Press, New York, pp. 379-389; T. Hukuhara et al. [1990] *J. Invertebr. Pathol.* 56:222-232; and Sato, T. [1989] “Establishment of Eight Cell Lines from Neonate Larvae of Tortricids (Lepidoptera), and Their Several Characteristics Including Susceptibility to Insect Viruses,” in *Invertebrate Cell Systems Applications*, J. Mitsuhashi (ed.), Vol. II, CRC Press, Inc., Boca Raton, Florida, pp. 187-198).

AmEPV is one of the few insect poxviruses which can replicate in insect cell culture; AmEPV is unable to replicate in vertebrate cell lines. The AmEPV double-stranded DNA genome is about 225 kb and is unusually A+T rich (18.5% G+C) (W. Langridge, H.R., et al. [1977] *Virology* 76:616-620). Recently, a series of restriction maps for AmEPV were published (Hall, R.L., et al. [1990] *Arch. Virol.* 110:77-90). No DNA homology to vaccinia has been detected (Langridge, W.H. [1983] *J. Invertebr. Pathol.* 42:77-82; Langridge, W.H. [1984] *J. Invertebr. Pathol.* 43:41-46).

The viral replication cycle of AmEPV resembles that of other poxviruses except for the appearance of occluded virus late in infection. For AmEPV, once a cell is infected, both occluded and extracellular virus particles are generated. The mature occlusion body particle, which is responsible for environmentally protecting the virion during infection, consists of virus embedded within a crystalline matrix consisting primarily of a single protein, spheroidin. Spheroidin, the major structural protein of AmEPV, has been reported to be 110 kDa in molecular weight and to consist of a high percentage of charged and sulfur-containing amino acids (Langridge and Roberts [1982] *J. Invertebr. Pathol.* 39:346-353).

Another insect virus is the baculovirus. Like baculoviruses, a characteristic feature of entomopoxviruses is the amalgamation of virions within environmentally stable occlusion bodies. It is this occluded form of the virus that is primarily responsible for dissemination to other insects. While the major protein (polyhedrin) of baculovirus occlusions is quite similar between viruses, it has been reported that the major occlusion body protein (spheroidin) of two group B

entomopoxviruses, *Amsacta moorei* (AmEPV) and *Choristoneura biennis* (CbEPV) is quite different both in terms of amino acid sequence and coding capacity of the corresponding spheroidin genes (115 and 47 kDa for AmEPV and CbEPV, respectively).

5           The entomopoxviruses and the role of occlusion bodies have recently been reviewed by Arif and Kurstak (Arif, B.M., E. Kurstak, E. [1991] "The entomopoxviruses," In *Viruses of Invertebrates* (E. Kurstak, Ed.), pp. 179-195, Marcel Dekker, Inc., New York) and Goodwin *et al.* (Goodwin, R.H., R.J. Milner, C.D. Beaton [1991] "Entomopoxvirinae," In *Atlas of Invertebrate Viruses* (J.R. Adams and J.R. Bonami, Eds.), pp. 259-285, CRC Press, Inc., Boca Raton ). The gene which encodes the AmEPV spheroidin, a 115 kDa protein, has been identified and sequenced (Hall, R.L., R.W. Moyer [1991] *J. Virol.* 65, 6516-6527; Banville, M., F. Dumas, F., S. Trifiro, B. Arif, C. Richardson [1992] *J. Gen. Virol.* 73, 559-566). The AmEPV gene was also mapped and found to be located at the 3' end of a nucleoside triphosphate phosphohydrolase gene (NPH I or NTPase I, Hall and Moyer [1991], *supra*). The spheroidin gene of *Choristoneura biennis* entomopoxvirus (CbEPV) has been reported to be derived from a gene capable of encoding a 47 kDa protein (Yuen, L., J. Dionne, B. Arif, C. Richardson [1990] *Virology* 175:427-433.). A comparison of the sequence of the two spheroidins shows no relationship between the two encoded proteins.

20           We have investigated the spheroidin genes of *Choristoneura biennis*, *Choristoneura fumiferana*, and *Amsacta moorei* viruses. Our results indicate, in contrast to published results, that the initial *Choristoneura* EPV spheroidin assignment is likely incorrect and that the *Choristoneura* spheroidin is instead a highly conserved homolog of the AmEPV spheroidin.

25           The use of viruses and virus proteins in eukaryotic host-vector systems has been the subject of a considerable amount of investigation and speculation. Many existing viral vector systems suffer from significant disadvantages and limitations which diminish their utility. For example, a number of eukaryotic viral vectors are either tumorigenic or oncogenic in mammalian systems, creating the potential for serious health and safety problems associated with resultant gene products and accidental infections. Further, in some eukaryotic host-viral vector systems, the

30

gene product itself exhibits antiviral activity, thereby decreasing the yield of that protein.

In the case of simple viruses, the amount of exogenous DNA which can be packaged into a simple virus is limited. This limitation becomes a particularly acute problem when the genes used are eukaryotic. Because eukaryotic genes usually contain intervening sequences, they are too large to fit into simple viruses. Further, because they have many restriction sites, it is more difficult to insert exogenous DNA into complex viruses at specific locations.

Vaccinia virus has recently been developed as a eukaryotic cloning and expression vector (Mackett, M., et al. [1985] *DNA Cloning*, Vol. II, ed. D. M. Glover, Oxford: IRL Press, pp. 191-212; Panicali, D., et al. [1982] *Proc. Natl. Acad. Sci. USA*, 88:5364-5368). Numerous viral antigens have been expressed using vaccinia virus vectors (Paoletti, E., et al. [1984] *Proc. Natl. Acad. Sci. USA* 81:193-197; Piccine, A., et al. [1986] *BioEssays* 5:248-252) including, among others, HBsAg, rabies G protein and the gp120/gp41 of human immunodeficiency virus (HIV). Regulatory sequences from the spruce budworm EPV have been used previously with vaccinia (Yuen, L., et al. [1990] *Virology* 175:427-433).

Additionally, studies with vaccinia virus have demonstrated that poxviruses have several advantageous features as vaccine vectors. These include the ability of poxvirus-based vaccines to stimulate both cell-mediated and humoral immunity, minimal cost to mass produce vaccine and the stability of the lyophilized vaccine without refrigeration, ease of administration under non-sterile condition, and the ability to insert at least 25,000 base pairs of foreign DNA into an infectious recombinant, thereby permitting the simultaneous expression of many antigens from one recombinant.

There exists a need in the art for additional viral compositions and methods for use in expressing heterologous genes in selected host cells, and in performing other research and production techniques associated therewith.

Brief Summary of the Invention

This invention pertains to novel vectors useful for producing proteins via the expression of a heterologous gene in a novel expression system. More particularly, this invention relates to methods for incorporating a selected heterologous gene (also referred to as exogenous DNA) into a poxvirus genome to produce a recombinant expression vector capable of expression of the selected gene in a host cell.

The expression systems described herein utilize novel structural and/or regulatory DNA elements from Entomopoxvirus genomes. For example, according to the subject invention, the entomopoxvirus spheroidin gene and/or the thymidine kinase gene can be used as the location for the insertion of exogenous DNA. These Entomopoxvirus genes have been discovered to be attractive sites for insertion of heterologous genes because it is possible to transfer the strongly expressed spheroidin gene, or the thymidine kinase gene, as an expression cassette, not only in insect cells, but for use in vertebrate poxviruses such as vaccinia and swinepox virus.

Another aspect of the subject invention pertains to the use of the entomopoxvirus spheroidin or thymidine kinase gene regulatory sequences in other virus vector systems to enhance the performance of those systems. Thus, the subject invention further pertains to the use of regulatory elements from entomopoxvirus to construct novel chimeric vaccines and expression systems which are functional across genera of mammalian poxviruses.

As one aspect, the invention provides novel Entomopoxvirus polynucleotide sequences, free from other viral sequences with which the Entomopoxvirus sequences are associated in nature. Specifically, the subject invention provides nucleotide sequences of Entomopoxvirus spheroidin and thymidine kinase genes, including flanking sequences and regulatory sequences. In particular embodiments, the spheroidin DNA sequence is that which occurs in the *Choristoneura biennis*, *Choristoneura fumiferana*, or *Amsacta moorei* Entomopoxviruses. Also specifically exemplified is the *Amsacta moorei* Entomopoxvirus thymidine kinase nucleotide sequence. As explained more fully herein, fragments and variants of the exemplified sequences are within the scope

of the subject invention. Fragments and variants can be any sequence having substantial homology with the exemplified sequences so long as the fragment or variant retains the utility of the exemplified sequence. One specific type of variant pertains to spheroidin or tk genes from Entomopoxviruses other than those specifically exemplified herein. As described herein, for example, the current inventors have discovered that the spheroidin genes are highly conserved among different species of Entomopoxvirus. Specifically exemplified herein are three different Entomopoxviruse spheroidin genes having a high degree of homology. Other such spheroidin variants or tk variants from other Entomopoxviruses could be readily located and used by the ordinarily skilled artisan having the benefit of the subject application.

As another aspect, the present invention provides recombinant polynucleotide sequences comprising a sequence encoding an Entomopoxvirus spheroidin protein and/or its regulatory sequences, or a variant or fragment of the spheroidin sequence, linked to a second polynucleotide sequence encoding a heterologous gene. One embodiment of such a polynucleotide sequence provides a spheroidin promoter sequence operably linked to a heterologous gene to direct the expression of the heterologous gene in a selected host cell. Another embodiment provides a sequence encoding a spheroidin protein linked to the heterologous gene in a manner permitting expression of a fusion protein. Still another embodiment provides the heterologous gene inserted into a site in a spheroidin gene so that the heterologous gene is flanked on both termini by spheroidin sequences.

Yet a further aspect of the invention provides a recombinant polynucleotide sequence encoding an Entomopoxvirus tk gene and/or its regulatory sequences, or a variant or fragment thereof, linked to a second polynucleotide sequence encoding a heterologous gene. One embodiment of this polynucleotide sequence provides the tk promoter sequence operably linked to the heterologous gene to direct the expression of the heterologous gene in a selected host cell. Another embodiment provides the sequence encoding the tk protein linked to the heterologous gene in a manner permitting expression of a fusion protein. Still another embodiment provides the heterologous gene inserted into

a site in the tk gene so that the heterologous gene is flanked on both termini by tk sequences.

Another aspect of the invention pertains to Entomopoxvirus spheroidin polypeptides, fragments thereof, or analogs thereof, optionally fused to a heterologous protein or peptide. Also provided is an Entomopoxvirus tk polypeptide, fragments thereof, or analogs thereof, optionally linked to a heterologous protein or peptide.

Yet another aspect of the invention is provided by recombinant polynucleotide molecules which comprise one or more of the polynucleotide sequences described above. This molecule may be an expression vector or shuttle vector. The molecule may also contain viral sequences originating from a virus other than the Entomopoxvirus which contributed a spheroidin or tk polynucleotide sequence, e.g., vaccinia.

In another aspect, the present invention provides a recombinant virus comprising a polynucleotide sequence as described above. Also provided are host cells infected with one or more of the described recombinant viruses.

The present invention also provides a method for producing a selected polypeptide comprising culturing a selected host cell infected with a recombinant virus, as described above, and recovering said polypeptide from the culture medium.

As a final aspect, the invention provides a method for screening recombinant host cells for insertion of heterologous genes comprising infecting the cells with a recombinant virus containing a polynucleotide molecule comprising the selected heterologous gene sequence linked to an incomplete spheroidin or tk polynucleotide sequence or inserted into and interrupting the coding sequences thereof so that the heterologous gene is flanked at each termini by an Entomopoxvirus spheroidin or tk polynucleotide sequence. The absence of occlusion bodies formed by the expression of a spheroidin protein in the spheroidin-containing cell indicates the integration of the heterologous gene. Alternatively, the absence of the thymidine kinase function, i.e., resistance to bromodeoxyuridine, formed by the integration of the inactive thymidine kinase sequence indicates the insertion of the heterologous gene.

Other aspects and advantages of the present invention are described further in the following detailed description of embodiments of the present invention.

5

#### Brief Description of the Sequences

SEQ ID NO. 1 is the DNA sequence of the *Amsacta moorei* Entomopoxvirus spheroidin gene and flanking sequences.

SEQ ID NO. 2 is the amino acid sequence encoded by the G1L ORF.

SEQ ID NO. 3 is the amino acid sequence encoded by the G2R ORF.

10

SEQ ID NO. 4 is the amino acid sequence encoded by the G3L ORF.

SEQ ID NO. 5 is the amino acid sequence encoded by the G4R ORF.

SEQ ID NO. 6 is the deduced amino acid sequence of the spheroidin protein.

SEQ ID NO. 7 is the amino acid sequence encoded by the G6L ORF.

15

SEQ ID NO. 8 is the DNA sequence of the *Amsacta moorei* Entomopoxvirus thymidine kinase (tk) gene and flanking sequences.

SEQ ID NO. 9 is a small peptide of 66 amino acids potentially encoded by ORF Q1.

SEQ ID NO. 10 is the amino acid sequence encoded by the Q3 ORF.

20

SEQ ID NO. 11 is the deduced amino acid sequence of the tk protein.

SEQ ID NO. 12 is the synthetic oligonucleotide designated RM58.

SEQ ID NO. 13 is the synthetic oligonucleotide designated RM82.

SEQ ID NO. 14 is the synthetic oligonucleotide designated RM83.

SEQ ID NO. 15 is the synthetic oligonucleotide designated RM92.

25

SEQ ID NO. 16 is the synthetic oligonucleotide designated RM118.

SEQ ID NO. 17 is the synthetic oligonucleotide designated RM165.

SEQ ID NO. 18 is the synthetic oligonucleotide designated RM03.

SEQ ID NO. 19 is the synthetic oligonucleotide designated RM04.

SEQ ID NO. 20 is the synthetic oligonucleotide designated RM129.

30

SEQ ID NO. 21 is the spheroidin gene coding sequence spanning nucleotides #3080 through #6091 of SEQ ID NO. 1.

SEQ ID NO. 22 is a fragment of the spheroidin gene spanning nucleotides #2781 through 3191 of SEQ ID NO. 1 which is likely to contain the promoter sequence.

SEQ ID NO. 23 is the G2R ORF.

5 SEQ ID NO. 24 is the G4R ORF.

SEQ ID NO. 25 is the G1L ORF.

SEQ ID NO. 26 is the G3L ORF.

SEQ ID NO. 27 is the G6L ORF.

10 SEQ ID NO. 28 is the tk gene coding sequence spanning nucleotides #234 through #782 of SEQ ID NO. 8.

SEQ ID NO. 29 is a fragment of the tk gene spanning nucleotides #783 through #851 of SEQ ID NO. 8.

SEQ ID NO. 30 is a fragment spanning nucleotides #750 through #890 of SEQ ID NO. 8 which is likely to contain the promoter sequence.

15 SEQ ID NO. 31 is the Q1 ORF.

SEQ ID NO. 32 is the Q3 OFR.

SEQ ID NO. 33 is a fragment included within the sequence spanning nucleotides #2274 through #6182 of SEQ ID NO. 1 containing the entire spheroidin open reading frame and some flanking sequences.

20 SEQ ID NO. 34 is a polypeptide cleavage product according to the subject invention.

SEQ ID NO. 35 is a polypeptide cleavage product according to the subject invention.

25 SEQ ID NO. 36 is a polypeptide cleavage product according to the subject invention.

SEQ ID NO. 37 is the peptide sequence encoded by the RM58 probe.

SEQ ID NO. 38 is a nucleotide fragment spanning nucleotides #4883 through #4957 of SEQ ID NO. 1.

30 SEQ ID NO. 39 is a nucleotide fragment spanning nucleotides #3962 through #4012 of SEQ ID NO. 1.

SEQ ID NO. 40 is a nucleotide fragment spanning nucleotides #4628 through #4651 of SEQ ID NO. 1.

SEQ ID NO. 41 is the AmEPV NPHI nucleotide sequence shown in Figure 7.

SEQ ID NO. 42 is the AmEPV NPHI amino acid sequence from Figure 7. This sequence is in the order opposite that shown in the Figure.

5        SEQ ID NO. 43 is the CbEPV nucleotide sequence shown in part A of Figure 6.

SEQ ID NO. 44 is the CbEPV amino acid sequence shown in part B of Figure 6.

10       SEQ ID NO. 45 is the CfEPV nucleotide sequence shown in part A of Figure 6.

SEQ ID NO. 46 is the CfEPV amino acid sequence shown in part B of Figure 6.

SEQ ID NO. 47 is the CbEPV amino acid sequence corresponding to amino acids 211 to 221 of AmEPV.

15       SEQ ID NO. 48 is the CbEPV amino acid sequence corresponding to amino acids 682 to 691 of AmEPV.

SEQ ID NO. 49 is the CbEPV amino acid sequence corresponding to amino acids 726-736 of AmEPV.

SEQ ID NO. 50 is the sequence of RM206.

20       SEQ ID NO. 51 is the sequence of RM212.

SEQ ID NO. 52 is the sequence of RM58.

SEQ ID NO. 53 is the sequence of RM75.

SEQ ID NO. 54 is the sequence of RM76.

SEQ ID NO. 55 is the sequence of RM78.

25       SEQ ID NO. 56 is the sequence of RM79.

SEQ ID NO. 57 is the sequence of RM82.

SEQ ID NO. 58 is the sequence of RM83.

SEQ ID NO. 59 is the sequence of RM87.

SEQ ID NO. 60 is the sequence of RM91.

30       SEQ ID NO. 61 is the sequence of RM92.

SEQ ID NO. 62 is the sequence of RM93.

SEQ ID NO. 63 is the sequence of RM95.

SEQ ID NO. 64 is the sequence of RM118.

SEQ ID NO. 65 is the sequence of RM169.

SEQ ID NO. 66 is the sequence of RM170.

5

#### Brief Description of the Drawings

Figure 1 is a physical map of AmEPV illustrating restriction fragments thereof and showing the spheroidin gene just to the right of site #29 in the HindIII-G fragment.

10 Figure 2 provides the AmEPV DNA sequence of the *Amsacta moorei* Entomopoxvirus spheroidin gene and flanking sequences (SEQ ID NO. 1), the deduced amino acid sequences of the spheroidin protein (SEQ ID NO. 6), and five additional open reading frames (ORFs). The complete sequence of the G6 ORF is not shown in this figure but is provided in SEQ ID NO. 1.

15 Figure 3 provides the DNA sequence of the *Amsacta moorei* Entomopoxvirus thymidine kinase (tk) gene and flanking sequences (SEQ ID NO. 8), the deduced amino acid sequences of the tk protein (SEQ ID NO. 11), and two additional ORFs.

Figure 4 is a schematic map of an AmEPV fragment illustrating the orientation of the spheroidin ORF on the physical map and indicating homologies.

20 Figure 5 shows the location of the AmEPV spheroidin specific oligonucleotide primers used in PCR to identify an AmEPV spheroidin-like gene in *Choristoneura* EPV DNAs. The arrowheads over the numbers in the diagram represent oligonucleotide primers. The arrowhead base shows the approximate starting point of the primer, and the 5' to 3' direction is shown by the direction of the arrowhead. Primers from the upper line were paired with various primers from the lower line as shown in Table 2 for PCR reactions. The sequences of the primers are shown in Table 2.

25 Figure 6 shows sequences of corresponding regions of the spheroidin-like gene of CbEPV and CfEPV and the spheroidin of AmEPV. Panel A shows the AmEPV spheroidin sequence and PCR product sequences derived from CbEPV and CfEPV DNA using RM58 as the sequencing primer for the 1 kb PCR products resulting from the AmEPV spheroidin specific primer pair RM58-RM118

30

(Table 2) and either CbEPV or CfEPV DNA. The alignment of the *Choristoneura* EPV sequences with bases 4044-4278 of AmEPV spheroidin is shown. The predicted amino acid sequences from the sequences shown in Panel A are shown in Panel B. Identity is shown by vertical lines and two degrees of conserved changes are indicated by periods and colons. Panel C shows CbEPV spheroidin amino acid sequences derived from protein microsequencing of 3 lys-c endoprotease fragments. Corresponding AmEPV spheroidin sequence and amino acid position numbers are shown.

Figure 7 shows the sequence of the 5' end of the AmEPV NPH I gene including the deduced amino acids. The base numbers represent the extension of the sequence shown in Figure 2, which includes the partial AmEPV NPH I (NTPase I) gene. The sequence in Figure 1 ends at base 6768. The base numbers correspond to those in Figure 8.

Figure 8 shows the spheroidin and NPH I gene orientations in AmEPV, and the locations of primers used in PCR reactions. The arrowheads over the numbers in the figure represent oligonucleotide primers. The arrowhead base shows the approximate starting point of the oligo and 5' to 3' is shown by the direction of the arrowhead. RM83 and RM161 are AmEPV spheroidin specific primers. Primer RM212 is from the 3' end of the CbEPV NPH I gene and RM206 is just downstream of the CbEPV NPH I gene but within the published CbEPV sequence (Yuen *et al.* [1991], *supra*).

#### Detailed Description of the Invention

The present invention provides novel Entomopoxvirus (EPV) polynucleotide sequences free from association with other viral sequences with which they are naturally associated. Recombinant polynucleotide vectors comprising the sequences, recombinant viruses comprising the sequences, and host cells infected with the recombinant viruses are also disclosed herein. These compositions are useful in methods of the invention for the expression of heterologous genes and production of selected proteins in both insect and mammalian host cells.

Novel polynucleotide sequences of the invention encode EPV spheroidin genes and/or flanking sequences, including sequences which provide regulatory signals for the expression of the gene. The invention also provides novel polynucleotide sequences encoding an EPV thymidine kinase (tk) gene and/or its  
5 flanking sequences. The polynucleotide sequences of this invention may be either RNA or DNA sequences. More preferably, the polynucleotide sequences of this invention are DNA sequences.

Specifically disclosed by the present invention are spheroidin polynucleotide sequences obtainable from the *Amsacta moorei* Entomopoxvirus (AmEPV),  
10 *Choristoneura biennis* Entomopoxvirus (CbEPV), and *Choristoneura fumiferana* Entomopoxvirus (CfEPV). Also specifically exemplified is a tk polynucleotide sequence obtained from AmEPV. While these species are exemplified for practice of the methods and compositions of this invention, utilizing the techniques described herein, substantially homologous sequences may be obtained by one of  
15 ordinary skill in the art from other Entomopoxvirus species.

The AmEPV spheroidin DNA sequence, including flanking and regulatory sequences, is reported in Figure 2 as spanning nucleotides # 1 through 6768 (SEQ ID NO. 1). Within this sequence, the spheroidin gene coding sequence spans nucleotides #3080 through #6091 (SEQ ID NO. 21). A fragment which is likely  
20 to contain the promoter sequences spans nucleotides #2781 through #3199 (SEQ ID NO. 22). Other regions of that sequence have also been identified as putative coding regions for other structural or regulatory genes associated with spheroidin. These other fragments of interest include the following sequences: the G2R ORF, which is nucleotides #1472 through #2151 (SEQ ID NO. 23) encoding the amino  
25 acid sequence shown in SEQ ID NO. 3; the G4 ORF, which is nucleotides #2502 through #2987 (SEQ ID NO. 24) encoding the amino acid sequence shown in SEQ ID NO. 5; and the following sequences transcribed left to right on Figure 2: the G1L ORF, which is nucleotides #65 through #1459 (SEQ ID NO. 25) encoding the amino acid sequence shown in SEQ ID NO. 2; the G3L ORF, which  
30 is nucleotides #2239 through #2475 (SEQ ID NO. 26) encoding the amino acid sequence shown in SEQ ID NO. 4; and the G6 ORF, which includes nucleotides #6277 through #6768 (SEQ ID NO. 27) encoding the amino acid sequence shown

in SEQ ID NO. 7. These ORFs are identified in Figure 2. It should be noted that the full length of the G6 ORF extends beyond nucleotide #6768, is shown in SEQ ID NO. 1 and Figure 7, and is discussed more fully below.

5 The AmEPV ORF G4R (SEQ ID NO. 24) which encodes G4R (SEQ ID NO. 5) is immediately upstream of the spheroidin gene has significant homology to the capripoxvirus HM3 ORF. A homolog of the HM3 ORF is found in vaccinia virus just upstream of a truncated version of the cowpox virus ATI gene. Therefore, the microenvironments in this region are similar in the two viruses. Two other ORFs relate to counterparts in vaccinia virus. These ORFs include the  
10 I7 ORF of the vaccinia virus *HindIII*-I fragment (I7) (Schmitt, J.F.C., et al. [1988] *J. Virol.* 62:1889-1897) which relates to the AmEPV G1L ORF (SEQ ID NO. 25) and the NTPase I (NPH I) ORF of the *HindIII*-D fragment which relates to the AmEPV G6L ORF (SEQ ID NO. 27) (Broyles, S.S., et al. [1987] *J. Virol.* 61:1738-1742; and Rodriguez, J.F., et al. [1986] *Proc. Natl. Acad. Sci. USA* 83:9566-9570).  
15 The genomic location of the AmEPV ORFs compared with that of the vaccinia virus ORFs suggests that the arrangement of essential "core genes," which are centrally located and colinear in many, if not all, of the vertebrate poxviruses on a more macroscopic scale, is quite different in the insect virus.

As set out in detail in the accompanying examples below, the spheroidin  
20 gene of AmEPV was identified through direct microsequencing of the protein, and the results were used for the design of oligonucleotide probes. Transcription of the spheroidin gene is inhibited by cycloheximide, suggesting it is a late gene. Consistent with this prediction are the observations that spheroidin transcripts were initiated within a TAAATG motif (See Figure 2, nucleotide #3077-3082) and  
25 that there is a 5' poly(A) sequence, both characteristic of late transcripts.

The isolation and sequencing of the CbEPV and CfEPV spheroidin genes are also described in detail below.

The AmEPV thymidine kinase (tk) DNA sequence, including flanking and regulatory sequences, is reported in Figure 3, as spanning nucleotides #1 through  
30 #1511 (SEQ ID NO. 8). Within this sequence, the tk gene coding sequence spans nucleotides #234 through #782 (SEQ ID NO. 28) (transcribed right to left on Figure 3). Another fragment of interest may include nucleotides #783 through

#851 (SEQ ID NO. 29) of that sequence or fragments thereof. A fragment likely to contain the promoter regions spans nucleotides #750 through #890 (SEQ ID NO. 30). Other regions of that sequence have also been identified as putative coding regions for other structural or regulatory genes associated with tk. These other fragments of interest include the following sequences (transcribed left to right on Figure 3: ORF Q1, which is nucleotides #18 through #218 (SEQ ID NO. 31) encoding the amino acid sequence shown in SEQ ID NO. 10); and ORF Q3, which is nucleotides #852 through #1511 (SEQ ID NO. 32) encoding the amino acid sequence shown in SEQ ID NO. 10.

The location of the AmEPV tk gene maps in the *EcoRI*-Q fragment near the left end of the physical map of the AmEPV genome (Figure 1) (see also, Hall, R.L., et al. [1990] *Arch. Virol.* 110:77-90, incorporated herein by reference). Because of the orientation of the gene within the AmEPV genome, transcription of the gene is likely to occur toward the terminus. There are believed to be similar tk genes, or variations thereof, in other systems, including mammalian systems. As set out in detail in the examples below, the tk gene of AmEPV was identified through direct microsequencing of the protein, and the results were used for the design of oligonucleotide probes.

The term "polynucleotide sequences" when used with reference to the invention can include the entire EPV spheroidin or tk genes with regulatory sequences flanking the coding sequences. The illustrated AmEPV sequences are also encompassed by that term. Also included in the definition are fragments of the coding sequences with flanking regulatory sequences. The definition also encompasses the regulatory sequences only, e.g., the promoter sequences, transcription sites, termination sequences, and other regulatory sequences.

Sequences of the invention may also include all or portions of the spheroidin or tk genes linked in frame to a heterologous gene sequence. Additionally, polynucleotide sequences of the invention may include sequences of the spheroidin or tk genes into which have been inserted a foreign or heterologous gene sequence, so that the EPV sequences flank the heterologous gene sequence.

Polynucleotide sequences of this invention also include sequences which are capable of hybridizing to the sequences of Figures 2 and 3, under stringent

conditions. Also sequences capable of hybridizing to the sequences of Figures 2 and 3 under non-stringent conditions may fall within this definition providing that the biological or regulatory characteristics of the sequences of Figures 2 and 3, respectively, are retained. Examples of stringent and non-stringent conditions of hybridization are conventional (see, e.g., Sambrook et al. [1989] *Molecular Cloning. A Laboratory Manual*, 2d edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

Similarly, polynucleotide sequences of this invention also include variants, including allelic variations (naturally-occurring base changes in the EPV species population which may or may not result in an amino acid change) of DNA sequences encoding the spheroidin or tk protein sequences or other ORFs or regulatory sequences illustrated in Figures 2 and 3. Similarly, DNA sequences which encode spheroidin or tk proteins of the invention but which differ in codon sequence due to the degeneracies of the genetic code or variations in the DNA sequences which are caused by point mutations or by induced modifications to, for example, enhance a biological property or the usefulness of a desired polynucleotide sequence encoded thereby are also encompassed in the invention.

Utilizing the sequence data in Figures 2 or 3, as well as the denoted characteristics of spheroidin or thymidine kinase, it is within the skill of the art to obtain other DNA sequences encoding these polypeptides. For example, the structural gene may be manipulated by varying individual nucleotides, while retaining the same amino acid(s), or varying the nucleotides, so as to modify the amino acids, without loss of utility. Nucleotides may be substituted, inserted, or deleted by known techniques, including, for example, *in vitro* mutagenesis and primer repair.

The structural gene may be truncated at its 3'-terminus and/or its 5'-terminus. It may also be desirable to ligate a portion of the polypeptide sequence to a heterologous coding sequence, and thus to create a fusion peptide.

The polynucleotide sequences of the present invention may be prepared by a variety of techniques well known to those skilled in the art. The sequences may be prepared synthetically or can be derived from viral RNA or from available

cDNA-containing plasmids by chemical and genetic engineering techniques or combinations thereof which are standard in the art.

The Entomopoxvirus proteins—spheroidin, thymidine kinase and their respective regulatory sequences, as described herein—may be encoded by polynucleotide sequences that differ in sequence from the sequences of Figures 2 and 3 due to, for example, natural allelic or species variations. Thus, the terms spheroidin or tk polypeptides also refer to any of the naturally occurring sequences and various analogs, e.g., processed or truncated sequences or fragments, including the mature spheroidin or tk polypeptides and mutant or modified polypeptides or fragments that retain the same utility and preferably have homology to Figure 2 or 3, respectively, of at least 75%, more preferably 90%, and most preferably 95%.

Another aspect of the present invention is provided by the proteins encoded by the EPV spheroidin and tk polynucleotide sequences. Putative amino acid sequences of the two EPV proteins as well as additional putative proteins encoded by the ORFs of these sequences which are identified in Figures 2 and 3, respectively. EPV spheroidin has no significant amino acid homology to any previously reported protein, including the polyhedrin protein of baculovirus. We have found that both spheroidin and tk are nonessential proteins, which makes them desirable as sites for insertion of exogenous DNA.

Comparison of the AmEPV tk amino acid sequence with other tk genes reveals that the AmEPV tk gene is not highly related to any of the vertebrate poxvirus tk genes (43.4 to 45.7%). The relatedness of the vertebrate tk proteins to AmEPV is still lower (39.3 to 41.0%), while African Swine Fever (ASF) showed the least homology of all the tk proteins tested (31.4%). Although ASF has many similarities to poxviruses, and both ASF and AmEPV infect invertebrate hosts, the tk genes indicate little commonality and/or indication of common origin stemming from invertebrate hosts.

The spheroidin and thymidine kinase polypeptide sequences may include isolated naturally-occurring spheroidin or tk amino acid sequences identified herein or deliberately modified sequences which maintain the biological or regulatory functions, or other utility, of the AmEPV polypeptides, respectively

identified in Figures 2 and 3. Therefore, provided that the utilities of these polypeptides are retained in whole or part despite such modifications, this invention encompasses the use of all amino acid sequences disclosed herein as well as variants thereof retaining spheroidin or tk utility. Similarly, proteins or functions encoded by the other spheroidin or tk ORFs may include sequences containing amino acid modifications but which retain their regulatory or other biological functions, or other utility.

Examples of such modifications include polypeptides with amino acid variations from the natural amino acid sequence of Entomopoxvirus spheroidin or thymidine kinase; in particular, conservative amino acid replacements. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are generally divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) non-polar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid will not have a major effect on biological activity, especially if the replacement does not involve an amino acid at an active site of the polypeptides.

As used herein, the term "polypeptide" refers to a polymer of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. The phrase "polypeptide and variants thereof" includes post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The proteins or polypeptides of the present invention may be expressed in host cells and purified from the cells or media by conventional means (Sambrook *et al.*, *supra*).

5 This invention also relates to novel viral recombinant polynucleotide molecules or vectors, which permit the expression of heterologous genes in a selected host cell. Such a polynucleotide vector of the invention comprises the polynucleotide sequence encoding all or a portion of the spheroidin or tk gene, the RNA polymerase from a selected poxvirus, and the polynucleotide sequence encoding a desired heterologous gene. Preferably, the sequence includes the  
10 regulatory region, and most preferably, the promoter region, of either the EPV spheroidin or tk gene. In addition, the source of the polymerase is not limited to EPV; rather, any poxvirus RNA polymerase may be utilized.

Therefore, the viral vectors may contain other viral elements contributed by another poxvirus, either vertebrate or invertebrate, with the only EPV  
15 sequences being provided by the presence of the EPV spheroidin or tk gene sequences, or fragments thereof. Numerous conventional expression viral vectors and expression systems are known in the art. Particularly desirable vectors systems are those of vertebrate or invertebrate poxviruses. The Entomopoxvirus spheroidin and tk gene regulatory sequences may be used in other virus vector  
20 systems which contain a poxvirus RNA polymerase to enhance the performance of those systems, e.g., in vaccinia vectors. Methods for the construction of expression systems, in general, and the components thereof, including expression vectors and transformed host cells, are within the art. See, generally, methods described in standard texts, such as Sambrook *et al.*, *supra*. The present invention  
25 is therefore not limited to any particular viral expression system or vector into which a polynucleotide sequence of this invention may be inserted, provided that the vector or system contains a poxvirus RNA polymerase.

The vectors of the invention provide a helper independent vector system, that is, the presence or absence of a functional spheroidin or tk gene in a poxvirus  
30 contributing elements to the vector, e.g., contributing the RNA polymerase, does not affect the usefulness of the resulting recombinant viral vector. Because both spheroidin and tk are non-essential genes, the viral vectors of this invention do not

require the presence of any other viral proteins, which in helper-dependent systems are contributed by additional viruses to coinfect the selected host cell.

Selected host cells which, upon infection by the viral vectors will permit expression of the heterologous gene, include insect and mammalian cells. Specifically, if the viral vector comprises the EPV spheroidin or tk gene sequences of the invention inserted into any member of the family Entomopoxvirinae, e.g., EPVs of any species, the host cell will be limited to cells of insects normally infected by EPVs. If the viral vector comprises the EPV spheroidin or tk gene sequences of the invention inserted into a vertebrate poxvirus, such as vaccinia or swinepox, the host cells may be selected from among the mammalian species normally infected by the wild-type vertebrate poxvirus. Most desirably, such mammalian cells may include human cells, rodent cells and primate cells, all known and available to one of skill in the art.

According to one aspect of the subject invention, therefore, vectors of the present invention may utilize a fragment of the polynucleotide sequence of EPV spheroidin, particularly the promoter and ancillary regulatory sequences which are responsible for the naturally high levels of expression of the gene. Desirably, spheroidin sequences may be found within the sequence of Figure 2 (SEQ ID NO. 1), more particularly within the region of nucleotides # 2781 through 3199 (SEQ ID NO. 22). Smaller fragments within that region may also provide useful regulatory sequences. The desired spheroidin promoter sequence can be utilized to produce large quantities of a desired protein by placing it in operative association with a selected heterologous gene in viral expression vectors capable of functioning in either a vertebrate or invertebrate host cell.

As used herein, the term "operative association" defines the relationship between a regulatory sequence and a selected protein gene, such that the regulatory sequence is capable of directing the expression of the protein in the appropriate host cell. One of skill in the art is capable of operatively associating such sequences by resort to conventional techniques.

Where the spheroidin polynucleotide sequence in the vector contains all or a portion of the spheroidin coding sequence in association with, or linked to, the heterologous gene, the resulting protein expressed in the host cell may be a

fusion protein consisting of all or a portion of the spheroidin protein and the heterologous protein. Where the spheroidin polynucleotide sequence in the vector does not contain sufficient coding sequence for the expression of a spheroidin protein or peptide fragment, the heterologous protein may be produced alone.

5 In an analogous manner, the promoter and regulatory sequences of tk (Figure 3 SEQ ID NO. 8) may be employed in the construction of an expression vector to drive expression of a heterologous protein, or a fusion protein, in a selected known expression system. These tk regulatory sequences are desirably obtained from the sequence of Figure 3 (SEQ ID NO. 8), particularly in the  
10 fragment occurring between nucleotide #750 through 890 (SEQ ID NO. 30). Smaller fragments within that region may also provide useful regulatory sequences.

An advantage of the use of the novel EPV spheroidin or tk promoter sequences of this invention is that these regulatory sequences are capable of operating in a vertebrate poxvirus (e.g., vaccinia)-mammalian cell expression  
15 vector system. For example, the strong spheroidin promoter can be incorporated into the vaccinia virus system through homologous recombination. Unlike the promoter for the baculovirus polyhedrin gene, the promoter for the EPV spheroidin gene can be utilized directly in the vaccinia or swinepox virus expression vector.

20 To construct a vector according to the present invention, the spheroidin or tk polynucleotide sequence may be isolated and purified from a selected Entomopoxvirus, e.g., AmEPV, and digested with appropriate restriction endonuclease enzymes to produce a fragment comprising all or part of the spheroidin or tk gene. Alternatively such a fragment may be chemically  
25 synthesized.

Still alternatively, the desired AmEPV sequences may be obtained from bacterial cultures containing the plasmids pRH512, pMEGtk-1 or pRH7. The construction of the plasmid pRH512 is described in the examples below. This plasmid contains the 4.51 kb *Bgl*II fragment AmEPV DNA sequence inserted into  
30 a *Bam*HI site in the conventional vector pUC9. The plasmid pRH7 was constructed by digesting AmEPV genomic DNA, obtained as described in Example 1, with *Bsp*1286I, and the resulting fragments with *Hae*II. T4 DNA

polymerase is employed to blunt end the AmEPV DNA and the fragment containing the spheroidin gene is ligated to the large fragment of a *Sma*I digested pUC9 fragment. This fragment contains the entire spheroidin open reading frame and some flanking sequence, included within the nucleotide sequence spanning #2274-6182 (SEQ ID NO. 33) of Figure 2. The construction of plasmid pMEGtk-1 comprising the regulatory sequences of the tk gene as well as the structural gene is described below in the Example 8. It was constructed by inserting the *Eco*RI-Q fragment of AmEPV into the conventional vector pUC18.

Bacterial cultures containing plasmids pRH512, pMEG tk-1, and pRH7 have been deposited in the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland, USA. The deposited cultures are as follows:

| <u>Culture</u>                                      | <u>Accession No.</u> | <u>Deposit Date</u> |
|-----------------------------------------------------|----------------------|---------------------|
| <i>E. coli</i> SURE strain<br>(Stratagene) pMEG-tk1 | ATCC 68532           | February 26, 1991   |
| <i>E. coli</i> SURE strain<br>(Stratagene) pRH512   | ATCC 68533           | February, 26 1991   |
| <i>E. coli</i> SURE strain<br>(Stratagene) pRH7     | ATCC 68902           | January 28, 1992    |

The plasmids can be obtained from the deposited bacterial cultures by use of standard procedures, for example, using cleared lysate-isopycnic density gradient procedures, and the like.

These ATCC deposits were made under conditions that assure that access to the cultures will be available during the pendency of this patent application to one determined by the Commissioner of Patents and Trademark to be entitled thereto under 37 CFR 1.14 and 35 USC 122. The deposits will be available as required by foreign patent laws in countries wherein counterparts of the subject application, or its progeny, are filed. However, it should be understood that the availability of a deposit does not constitute a license to practice the subject invention in derogation of patent rights granted by government action.

Further, the subject culture deposits will be stored and made available to the public in accord with the provisions of the Budapest Treaty for the Deposit of Microorganisms, i.e., they will be stored with all the care necessary to keep them viable and uncontaminated for a period of at least five years after the most recent request for the furnishing of a sample of a deposit, and in any case, for a period of at least 30 (thirty) years after the date of deposit or for the enforceable life of any patent which may issue disclosing the cultures. The depositor acknowledges the duty to replace the deposit(s) should the depository be unable to furnish a sample when requested, due to the condition of a deposit. All restrictions on the availability to the public of the subject culture deposits will be irrevocably removed upon the granting of a patent disclosing them.

The molecular biology procedures referred to herein in describing construction of the vectors of this invention are standard, well-known procedures. The various methods employed in the preparation of the plasmid vectors and transformation or infection of host organisms are well-known in the art. These procedures are all described in, for example, Sambrook et al., cited above. Thus, it is within the skill of those in the genetic engineering art to extract DNA from microbial cells, perform restriction enzyme digestions, electrophorese DNA fragments, tail and anneal plasmid and insert DNA, ligate DNA, transform cells, prepare plasmid DNA, electrophorese proteins, and sequence DNA.

Because the AmEPV genome has no known unique restriction sites into which selected genes may be effectively introduced in a site-specific manner so as to be under the control of the spheroidin or tk promoter sequences, such restriction sites must be introduced into desired sites in the selected EPV polynucleotide sequence. For example, the unique *Bst*B1 site located at nucleotide #3172 downstream from the start of the spheroidin gene is the closest site to genetically engineer a usable insertion sequence for cloning. Therefore, restriction sites closer to the initiating Met of the spheroidin gene must be deliberately inserted.

Methods for the insertion of restriction sites are known to those of skill in the art and include, the use of an intermediate shuttle vector, e.g., by cloning the EPV sequence into the site of an appropriate cloning vehicle. It will be

recognized by those skilled in the art that any suitable cloning vehicle may be utilized provided that the spheroidin or tk gene and flanking viral DNA may be functionally incorporated.

5 A spheroidin shuttle vector may be constructed to include elements of the spheroidin structural gene, a cloning site located or introduced in the gene to enable the selected heterologous gene to be properly inserted into the viral genome adjacent to, and under the control of, the spheroidin promoter, and flanking viral DNA linked to either side of the spheroidin gene to facilitate insertion of the spheroidin-foreign gene-flanking sequence into another expression  
10 vector. The presence of flanking viral DNA also facilitates recombination with the wild type Entomopoxvirus, allowing the transfer of a selected gene into a replicating viral genome.

The shuttle vectors may thereafter be modified for insertion of a selected gene by deleting some or all of the sequences encoding for spheroidin or tk  
15 synthesis near the respective transcriptional start sites. Examples of such sites in spheroidin are nucleotides #3077 and 3080 and in tk includes nucleotide #809. Conventional procedures are available to delete spheroidin or tk coding sequences.

As an alternative to or in addition to the restriction site, a variety of  
20 synthetic or natural oligonucleotide linker sequences may be inserted at the site of the deletion. A polynucleotide linker sequence, which may be either a natural or synthetic oligonucleotide, may be inserted at the site of the deletion to allow the coupling of DNA segments at that site. One such linker sequence may provide an appropriate space between the two linked sequences, e.g., between the  
25 promoter sequence and the gene to be expressed. Alternatively, this linker sequence may encode, if desired, a polypeptide which is selectively cleavable or digestible by conventional chemical or enzymatic methods. For example, the selected cleavage site may be an enzymatic cleavage site, including sites for cleavage by a proteolytic enzyme, such as enterokinase, factor Xa, trypsin, collagenase and thrombin. Alternatively, the cleavage site in the linker may be  
30 a site capable of being cleaved upon exposure to a selected chemical, e.g. cyanogen bromide or hydroxylamine. The cleavage site, if inserted into a linker

useful in the sequences of this invention, does not limit this invention. Any desired cleavage site, of which many are known in the art, may be used for this purpose. In another alternative, the linker sequence may encode one or a series of restriction sites.

5           It will be recognized by those skilled in the art who have the benefit of this disclosure that linker sequences bearing an appropriate restriction site need not be inserted in place of all or a portion of the spheroidin structural sequence, and that it would be possible to insert a linker in locations in the Entomopoxvirus genome such that both the sequence coding for the selected polypeptide and the  
10           spheroidin structural sequence would be expressed. For instance, the sequence coding for the selected polypeptide could be inserted into the tk gene in place of all or a portion of the tk structural sequence and under the transcriptional control of the tk promoter.

          Polymerase chain reaction (PCR) techniques can also be used to introduce  
15           convenient restriction sites into the EPV DNA, as well as to amplify specific regions of the EPV DNA. These techniques are well known to those skilled in this art. See, for example, *PCR Protocols: A Guide to Methods and Applications*, M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, (1990).

          By use of these techniques, a variety of alternative modified shuttle vectors  
20           into which a selected gene or portion thereof may be incorporated may be suitably utilized in the present invention.

          As one embodiment of the invention, therefore, the polynucleotide sequence, described above, may be used as a shuttle vector to transfer a selected heterologous gene to a selected virus. In this embodiment, the polynucleotide  
25           sequence encoding the EPV spheroidin gene or EPV tk gene, or a fragment thereof, is linked to a heterologous gene. The polynucleotide sequence further contains a flanking region on either side of the spheroidin-heterologous gene or tk-heterologous gene to enable ready transfer into a selected virus. This resulting construct is termed a cassette. Such a flanking region may be derived from EPV,  
30           or alternatively, may be complementary to the target virus. For example, if it is desirable to insert a selected heterologous gene into a vaccinia virus to create a recombinant virus, one would utilize flanking regions complementary to the

targeted vaccinia virus. Similarly if the heterologous gene is inserted within the EPV spheroidin or tk gene, so that the selected EPV regulatory sequence and heterologous gene are flanked by the EPV gene's own sequences, this cassette may be used for transfer into a wild type EPV having homologous sequences to the flanking sequences.

The insertion or linkage of the foreign gene into the tk or spheroidin sequences of the present invention or the linkage of flanking sequences foreign to the spheroidin or tk genes may be accomplished as described above. The vectors of the subject invention may use cDNA clones of foreign genes, because poxvirus genes contain no introns, presumably as a consequence of a totally cytoplasmic site of infection.

In accordance with standard cloning techniques, any selected gene may be inserted into the vector at an available restriction site to produce a recombinant shuttle vector. Virtually any gene of interest could be inserted into the vectors described herein in order to obtain high expression of the desired protein. The spheroidin gene product may be useful as a particulate biological carrier for foreign gene antigens. Thus, a foreign gene fused to the spheroidin gene may be useful as a method to produce a foreign protein attached to an effective vaccine carrier. Restriction sites in the fragment may thereafter be removed so as to produce a preferred spheroidin or tk shuttle vector, having one or more cleavage or cloning sites located in the 3' direction downstream from the spheroidin promoter sequence. Thus, the present invention is not limited by the selection of the heterologous gene.

Alternatively, a vector of this invention may comprise a heterologous gene which is inserted into all or a portion of the EPV spheroidin or tk protein encoding sequence to interrupt the protein's natural processing. However, when the vector is transferred to another virus which contains a wild-type spheroidin or tk gene, expression of the inserted heterologous gene is obtained. Thus, the Entomopoxvirus spheroidin gene (Figure 2; SEQ ID NO. 1) and/or the tk gene (Figure 3; SEQ ID NO. 8) can be used as the location for the insertion of exogenous (heterologous) DNA in any of the above-mentioned expression systems. A shuttle vector so constructed may be useful as a marker for research and

production techniques for identifying the presence of successfully integrated heterologous genes into the selected expression system.

5 The tk gene is a particularly desirable site for insertion of a selected heterologous gene. Unlike spheroidin, tk is produced early in infection and in lesser quantities. Additionally, many poxviruses possess tk genes which may be sufficiently homologous to the EPV tk to provide easy recombination. For example, in vaccinia virus expression systems for mammalian cells, the vaccinia tk gene is a common insertion site. Therefore, the use of this gene is particularly desirable for construction of a shuttle vector to shuttle selected genes directly  
10 between vector systems. More specifically, a foreign gene may be desirably inserted into the EPV tk gene sequence between nucleotide #460 and #560 (See Figure 3).

Insertion of cassettes containing foreign genes into wild-type poxviruses can be accomplished by homologous recombination. The homologous recombination  
15 techniques used to insert the genes of interest into the viruses of the invention are well known to those skilled in the art. The shuttle vectors, when co-infected into host cells with a wild-type virus, transfer the cassette containing the selected gene into the virus by homologous recombination, thereby creating recombinant virus vectors.

20 Expression of a selected gene is accomplished by infecting susceptible host insect cells with the recombinant viral vector of this invention in an appropriate medium for growth. An EPV expression vector is propagated in insect cells or insects through replication and assembly of infectious virus particles. These infectious vectors can be used to produce the selected gene in suitable insect cells,  
25 thus facilitating the efficient expression of the selected DNA sequence in the infected cell. If the EPV spheroidin gene (or tk gene)-heterologous gene fragment is inserted into a vertebrate poxvirus by the same methods as described above, the recombinant virus may be used to infect mammalian cells and produce the heterologous protein in the mammalian cells.

30 For example, a gene inserted into the tk site of a vaccinia virus system could be transferred directly to the tk locus of an Entomopoxvirus vector of the subject invention or vice versa. This shuttling could be accomplished, for example,

using homologous recombination. Similarly insertion of a selected gene into the spheroidin gene or tk gene in a viral vector permits the gene to be shuttled into other viruses having homologous spheroidin or tk sequences, respectively.

The following description illustrates an exemplary vector of this invention, employing the gene coding for human  $\beta$ -interferon (IFN- $\beta$ ) synthesis as the heterologous gene. A DNA fragment containing the IFN- $\beta$  gene is prepared conventionally with restriction enzyme digested or blunt ended termini and cloned into a suitable site in the AmEPV spheroidin gene, into which a restriction site has been engineered by the methods described above.

The insertion of the IFN- $\beta$  gene produces a hybrid or fused spheroidin-IFN- $\beta$  gene capable of producing a fused polypeptide product if only a portion of the spheroidin gene was deleted as described above. If the entire spheroidin structural sequence was deleted, only interferon will be produced. Further, the hybrid gene may comprise the spheroidin promoter, the IFN- $\beta$  protein coding sequences, and sequences encoding a portion of the polypeptide sequence of the spheroidin protein, provided all such coding sequences are not deleted from the particular shuttle vector utilized.

The resulting shuttle vector contains the AmEPV spheroidin gene sequence coupled to the IFN- $\beta$  gene. The hybrid spheroidin-IFN- $\beta$  gene of the recombinant shuttle vector is thereafter transferred into the genome of an appropriate Entomopoxvirus, such as the preferred Entomopoxvirus AmEPV, to produce a recombinant viral expression vector capable of expressing the gene encoding for  $\beta$ -interferon in a host insect cell. Transfer of the hybrid gene to a wild-type virus is accomplished by processes which are well known to those skilled in the art. For example, appropriate insect cells may be infected with the wild type Entomopoxvirus. These infected cells are then transfected with the shuttle vector of the subject invention. These procedures are described, for example, in *DNA Cloning: A Practical Approach*, Vol. II, Edited by D. M. Glover, Chapter 7, 1985. A person skilled in the art could choose appropriate insect cells to be used according to the subject invention. By way of example, salt marsh caterpillars and cultured gypsy moth cells can be used.

During replication of the AmEPV DNA after transfection, the hybrid gene is transferred to the wild-type AmEPV by homologous recombination between the recombinant shuttle vector and AmEPV DNA. Accordingly, a mixture is produced comprising wild-type, nonrecombinant EPVs and recombinant EPVs capable of expressing the IFN- $\beta$  gene.

While transfection is the preferred process for transfer of the hybrid gene into the EPV genome, it will be understood by those skilled in the art that other procedures may be suitably utilized so as to effect the insertion of the gene into the EPV genome and that recombination may be accomplished between the recombinant shuttle vector and other strains of EPV (or other poxviruses) so long as there is sufficient homology between the sequence of the hybrid gene and the corresponding sequence of the other strain to allow recombination to occur.

The preferred recombinant AmEPV expression vector, comprising a hybrid spheroidin-IFN- $\beta$  gene incorporated into the AmEPV genome, can thereafter be selected from the mixture of nonrecombinant and recombinant Entomopoxviruses. The preferred, but by no means only, method of selection is by screening for plaques formed by host insect cells infected with viruses that do not produce viral occlusions. Selection may be performed in this manner because recombinant EPV viruses which contain the spheroidin protein coding sequences interrupted by the heterologous gene are defective in the production of viral occlusions due to the insertional inactivation of the spheroidin gene.

Also, the selection procedure may involve the use of the  $\beta$ -galactosidase gene to facilitate color selection. This procedure involves the incorporation of the *E. coli*  $\beta$ -galactosidase gene into the shuttle vector and is well known to those skilled in the art. This technique may be of particular value if the exogenous DNA is inserted into the tk gene so that the spheroidin gene is still expressed. It will be recognized by those skilled in the art that alternative selection procedures may also be utilized in accordance with the present invention.

Accordingly, the DNA from a recombinant virus is thereafter purified and may be analyzed with appropriate restriction enzymes, or PCR technology, to confirm that the recombinant AmEPV vector has an insertion of the selected gene in the proper location.

The vectors and methods provided by the present invention are characterized by several advantages over known vectors and vector systems. Advantageously, such EPV viral vectors of the present invention are not oncogenic or tumorigenic in mammals. Also, the regulatory signals governing *Amsacta moorei* Entomopoxvirus (AmEPV) gene expressions are similar to those of vaccinia. Therefore, it is possible to transfer the strongly expressed spheroidin gene, or the thymidine kinase gene, as an expression cassette, not only in insect cells, but for use in vertebrate poxviruses such as vaccinia and swinepox virus.

Based on reported data with vaccinia, herpes and baculovirus vector systems, which suggest that up to 30 kb can be transferred without disrupting the vector viability, the normal limitation on the amount of exogenous DNA which can be packaged into a virus is not anticipated to be encountered when using the novel EPV vectors and methods of the subject invention.

Another advantage is that for the novel vectors of the subject invention, the transcription and translation of foreign proteins is totally cytoplasmic. Still another advantage lies in the expression power of the EPV spheroidin regulatory sequences, which when in operative association with a heterologous gene in a vector of this invention, can be used to produce high levels of heterologous protein expression in the selected host cell.

The EPV vectors of this invention and methods for employing them to express selected heterologous proteins in insect or mammalian cells, as described above, are characterized by the advantage of replication in insect cells, which avoids the use of mammalian viruses, thereby decreasing the possibility of contamination of the product with mammalian virus. The expression system of this invention is also a helper independent virus expression vector system. These two characteristics are shared by known baculovirus expression systems. However, as shown in Table 1, the EPV expression vector system (EEVS) using the vectors of this invention has some important distinguishing features compared to the baculovirus expression systems (BEVS).

**Table 1. Differences between EEVS and BEVS**

|    | EEVS                                                                          | BEVS                                                                                                                                 |
|----|-------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
|    | cytoplasm                                                                     | nucleus                                                                                                                              |
| 5  | Poxviridae                                                                    | Baculoviridae                                                                                                                        |
|    | spheroidin & thymidine kinase (tk)                                            | polyhedrin & p10                                                                                                                     |
| 10 | (Orthopoxviruses)<br>(Leporipoxviruses)<br>(Suipoxviruses)<br>(Avipoxviruses) | <u>No mammalian counterparts.</u><br>Baculovirus<br>is not known to contain a tk gene. Polyhedrin is not found in mammalian systems. |

The present invention also provides a method for screening recombinant host cells for insertion of heterologous genes by use of the recombinant viral polynucleotide molecules of this invention. The viral molecules containing the selected heterologous gene sequence linked to the polynucleotide sequence encoding less than all of the Entomopoxvirus spheroidin protein. The heterologous gene may be linked to the spheroidin or tk regulatory sequences in the absence of the complete coding sequences, or it may be inserted into the spheroidin or tk gene coding sequences, thus disrupting the coding sequence. The cell infected with the recombinant vector is cultured under conditions suitable for expression of the heterologous protein, either unfused or as a fusion protein with a portion of the spheroidin sequence. The absence of occlusion bodies which would ordinarily be formed by the expression of the intact spheroidin protein indicates the integration of the heterologous gene.

If the viral vector similarly contained either incomplete or interrupted EPV tk encoding sequence, the absence of thymidine kinase function (e.g., resistance to bromodeoxyuridine) formed by the integration of the inactive thymidine kinase sequence indicates the insertion of the heterologous gene.

Alternatively, if a parent virus is deleted of part of its tk or spheroidin gene, and is thereafter mixed with a viral vector containing intact tk or spheroidin fused to the foreign gene, recombinants would express the methotrexate resistance or produce occlusion bodies, respectively, thus indicating integration of the active tk or spheroidin genes and the foreign gene.

The above-described selection procedures provide effective and convenient means for selection of recombinant Entomopoxvirus expression vectors.

Another embodiment of the present invention involves using novel EPV expression systems of the subject invention for insect control. Control of insect pests can be accomplished by employing the vectors and methods of the invention as described above. For example, a gene coding for an selected insect toxin may be inserted into the viral expression vector under the control of the spheroidin or tk regulatory sequences or within either of the two genes for purposes of recombination into a selected virus having homologous flanking regions.

Genes which code for insect toxins are well known to those skilled in the art. An exemplary toxin gene isolated from *Bacillus thuringiensis* (*B.t.*) can be used according to the subject invention. *B.t.* genes are described, for example, in U. S. Patent Nos. 4,775,131 and 4,865,981. Other known insect toxins may also be employed in this method.

The resulting EPV vector containing the toxin gene is applied to the target pest or its surroundings. Advantageously, the viral vector will infect the target pest, and large quantities of the toxin will be produced, thus resulting in the control of the pest. Particularly large quantities of the toxin protein can be produced if the regulatory sequences of the Entomopoxvirus spheroidin gene are used to express the toxin.

Alternatively, the spheroidin gene can be left intact and the toxin gene inserted into a different Entomopoxvirus gene such as the tk gene. In this construct, the toxin will be produced by the system and then effectively coated or encapsulated by the natural viral production of spheroidin. This system thus produces a toxin which will advantageously persist in the environment to prolong the availability to the target pest.

In addition to the novel Entomopoxvirus expression vectors and methods for their use described herein, the subject invention pertains to the use of novel regulatory elements from Entomopoxvirus to construct novel chimeric vaccinia and swinepox vaccines and expression systems which are functional across genera of mammalian poxviruses. The polynucleotide sequences of the invention can also be used with viral vaccines, e.g., known vaccinia virus vaccines, to enhance the effectiveness of these vaccines. Such vaccines have been described for use in controlling rabies and other infectious diseases in mammals. Specifically, it is anticipated that the introduction of the EPV spheroidin promoter sequences into known viral vectors which are used to express selected proteins in a mammalian host *in vivo* may enable the powerful spheroidin promoter to increase expression of the protein in the viral vaccine. This aspect of the invention provides a significant improvement over other expression systems, including the baculovirus expression system (BEVS).

The following examples illustrate the compositions and procedures, including the best mode, for practicing the invention. These examples, should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted. The restriction enzymes disclosed herein can be purchased from Bethesda Research Laboratories, Gaithersburg, MD, or New England Biolabs, Beverly, MA. The enzymes are used according to the instructions provided by the supplier. Klenow fragment of DNA polymerase, T4 polynucleotide kinase, and T4 DNA ligase were obtained from New England Biolabs and Promega.

Example 1 – Preparation of AmEPV DNA

The replication of AmEPV has been described previously (Goodwin, R.H., et al. [1990] *J. Invertebr. Pathol.* 56:190-205). The gypsy moth (*Lymantria dispar*) cell line IPLB-LD-652 (Insect Pathology Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD) is maintained at 26 to 28°C in EX-CELL 400 (JRH Biosciences, Lenexa, KS) supplemented with 10% fetal bovine serum, 100 U of penicillin, and 100 µg of streptomycin per ml. Other insect cell lines are well known to those skilled in the art and can be used according to the subject invention.

The AmEPV inoculum for cell culturing was from an AmEPV-infected, freeze-dried *E. acraea* larva stored at -70°C (Hall, R.L., et al. [1990] *Arch. Virol.* 110:77-90). The larva was crushed and macerated in 5 ml of EX-CELL 400 (with penicillin and streptomycin but without fetal bovine serum) to which 0.003 g of cysteine-HCl had been added to prevent melanization. The debris was pelleted at 200 x g for 5 minutes, and the supernatant was passed through a 0.45-µm-pore-size filter.

The gypsy moth cells were infected with AmEPV by addition of the inoculum to a confluent monolayer of cells (about 0.1 to 1 PFU per cell), with occasional agitation of the dish during the first day. Infected cells were harvested 5 to 6 days postinfection.

AmEPV DNA was prepared from the infected cells by one of two methods. The first method involved *in situ* digestion of infected cells embedded within agarose plugs, after which the released cellular and viral DNAs were separated by pulsed-field electrophoresis (Bio-Rad CHEF-II-DR system). IPLB-LD-652 cells were infected with first-cell-culture-passage AmEPV. Infected cells were harvested 6 days postinfection by centrifugation at 200 x g for 5 minutes, rinsed, and resuspended in modified Hank's phosphate-buffered saline (PBS), which contained 15 g of glucose per liter, but no Ca<sup>2+</sup> or Mg<sup>2+</sup>.

For embedding of the infected cells in agarose plugs, 1% SeaPlaque GTG agarose (prepared in modified Hank's PBS and equilibrated at 37°C) was mixed 1:1 with infected cells to yield 5 x 10<sup>6</sup> cells per ml in 0.5% agarose. Digestion to release DNA was done by gentle shaking of the inserts in 1% Sarkosyl-0.5 M

EDTA-1 mg of proteinase K per ml at 50°C for 2 days (Smith, C.L., et al. [1987] *Methods Enzymol.* 151:461-489). The CHEF-II-DR parameters for DNA separation were 180 V, a pulse ratio of 1, 50 initial and 90 second final pulse times, and a run time of 20 to 25 hours at 4°C. The separating gel was 1%  
5 SeaKem GTG agarose in 0.5x TBE buffer (Sambrook *et al.*, *supra*). Viral DNA bands were visualized by ethidium bromide staining and electroeluted (Allington, W.B., et al. [1978] *Anal. Biochem.* 85:188-196). The recovered DNA was used for plasmid cloning following ethanol precipitation.

The second method of viral DNA preparation used the extracellular virus  
10 found in the infected-cell-culture supernatant. The supernatant from 10-day-postinfection cell cultures was clarified by centrifugation at 200 x g for 5 minutes. Virus was collected from the supernatant by centrifugation at 12,000 x g. Viral pellets were resuspended in 6 ml of 1x TE. DNase I and RNase A (10 and 20 µg/ml final concentrations, respectively) were added, and the mixture was  
15 incubated at 37°C for 30 minutes. The mixture was heated to 50°C for 15 minutes. SDS and proteinase K (1% and 200 µg/ml, respectively) were then added. The sample was incubated to 50°C overnight and extracted three times with buffer-saturated phenol and once with SEVAG (Sambrook *et al.*, *supra*). The DNA was ethanol precipitated and resuspended in 1x TE (pH 8).

20 For routine virus quantitation, 1 ml of an appropriate virus dilution (prepared in unsupplemented EX-CELL 400) was added to a confluent monolayer of cells in a 60 mm culture dish, with intermittent agitation over a 5 hour adsorption period at 26 to 28°C. The virus inoculum was removed, and 5 ml of a 0.75% SeaPlaque agarose (FMC BioProducts, Rockland, ME) overlay  
25 prepared with 2x EX-CELL 400 and equilibrated at 37°C was added to the monolayer. Plaques were visualized after 5 days of incubation at 26°C by inspection with a stereomicroscope.

The DNA prepared according to either method was then cut with a variety of restriction endonuclease enzymes, e.g., *Bam*HI, *Eco*RI, *Hind*III, *Pst*I and *Xho*I,  
30 generating the various fragments which appear on the physical map of Figure 1. Hereafter, reference to each restriction fragment will refer to the enzyme and the

applicable letter, e.g., *Bam*HI-A through *Bam*HI-E, *Eco*RI-A through *Eco*RI-S, etc.

#### Example 2 – Production of Spheroidin Polypeptide

5           To localize the spheroidin gene, a purified preparation of occlusion bodies (OBs) from infected caterpillars was solubilized and subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, J.K. [1970] *Nature* (London) 227:680-685) with a 4% acrylamide stacking gel and a 7.5% separating gel. The acrylamide used to prepare spheroidin for protein  
10           microsequencing was deionized with AG501X8 resin (Bio-Rad, Richmond, CA). The gels were polymerized overnight at 4°C. For sample preparation, 2x Laemmli sample buffer consisting of 125 mM Tris-HCl (pH 6.8), 4% SDS (w/v), 10%  $\beta$ -mercaptoethanol (v/v), and 20% glycerol (v/v) was used.

          OB suspension samples were diluted 1:1 with 2x Laemmli sample buffer  
15           and boiled for 5 minutes. Several lanes of an OB protein preparation were separated electrophoretically. The spheroidin protein (113 kDa) was the predominant protein of the purified OBs. Spheroidin within SDS-polyacrylamide gels was tested for glycosylation by periodic acid-Schiff staining (Zacharius, R.M., et al. [1969] *Anal. Biochem.* 30:149-152).

20           Following electrophoretic separation, several lanes in the unstained gel were transferred to an Immobilon polyvinylidene difluoride (PVDF) membrane with a Bio-Rad TransBlot apparatus at 90 V for 2 hours in a buffer consisting of 10 mM morpholinepropanesulfonic acid (pH 6.0) and 20% methanol. Spheroidin was visualized on the PVDF membrane by Coomassie blue staining.

25           The region of the PVDF membrane containing spheroidin was excised from the membrane, and direct protein microsequencing was done with an Applied Biosystems gas-phase sequencer. Microsequencing of the intact protein was unsuccessful, presumably because the N terminus of the protein was blocked.

          Cyanogen bromide cleavage was performed on samples of spheroidin  
30           eluted from the PVDF membrane to generate internal peptide fragments for sequencing. Major polypeptides of 15, 9, 8, and 6.2 kDa were produced.

### Example 3 - Sequencing, Hybridizations

All DNA sequencing was done by the dideoxy chain termination method (Sanger, F., et al. [1977] *Proc. Natl. Acad. Sci. USA* 74:5463-5467) with [ $\alpha$ -<sup>35</sup>S]dATP and Sequenase (US Biochemical, Cleveland, OH). Standard sequencing reactions with Sequenase were carried out in accordance with the instructions of the supplier, US Biochemical.

A reliable amino acid sequence was obtained from the 9, 8, and 6.2 kDa polypeptides produced as described in Example 2. The 8 and 9 kDa polypeptides represented overlapping partial CNBr cleavage products which together yielded the longest continuous amino acid sequence: Met-Ala-(Asn or Arg)-Asp-Leu-Val-Ser-Leu-Leu-Phe-Met-(Asn or Arg)-(?) -Tyr-Val-(Asn?) -Ile-Glu-Ile-Asn-Glu-Ala-Val-(?) - (Glu?) (SEQ ID NO. 34). The amino acid sequence obtained from the 6.2 kDa fragment was Met-Lys-Ile-Thr-Ser-Ser-Thr-Glu-Val-Asp-Pro-Glu-Tyr-Val-(Thr or Ile)-Ser-(Asn?) (SEQ ID NO. 35). A partial sequence for the 15 kDa fragment was also obtained: (Asn?) -Ala-Leu-Phe-(Phe?) (Asn?) -Val-Phe (SEQ ID NO. 36). The question marks in the above sequences indicated undetermined or unconfirmed amino acids. All sequences were ultimately located within the spheroidin gene sequence.

### Example 4 - Plasmid pRH512

A *Bgl*III AmEPV DNA library was prepared by digesting the genomic AmEPV DNA with *Bgl*III according to manufacturer's instructions. Plasmid pUC9 (GIBCO; Bethesda Research Labs) was *Bam*HI-digested and phosphatase-treated. The genomic *Bgl*III cut AmEPV was shotgun cloned into the *Bam*HI site of pUC9. *Escherichia coli* SURE (Stratagene, La Jolla, CA) was transformed by electroporation with a Bio-Rad Gene Pulser following the instructions provided by the manufacturer with the shotgun ligation, containing a variety of recombinant plasmids. Mini-preparations of plasmids were made by a conventional alkaline lysis procedure (Sambrook *et al.*, *supra*). These plasmids were cut with *Eco*RI-*Sal*I to release the insert and run on a gel. The resulting plasmid DNA was southern blotted to a nylon membrane, producing a number of clones.

Among the fragments produced from the restriction enzyme digestions of the genomic DNA was a 4.4 *Bgl*II fragment and an *Eco*RI-D fragment. In order to locate a desirable clone from among those produced above, the sequence derived from the 6.2 kDa CNBr fragment was used to design a degenerate oligonucleotide for use as a hybridization probe to locate the spheroidin gene in a clone. The nucleotide sequence of this probe called RM58 (SEQ ID NO. 12) was GAAGT7GATCC7GAATATGT, where 5 represents A or G, 6 represents C or T, and 7 represents A, G, C, or T. The peptide sequence of the probe was: Glu-Val-Asp-Pro-Glu-Tyr (SEQ ID NO. 37).

The DNA probe was radiolabeled either with [ $\alpha$ -<sup>32</sup>P]dCTP by the random oligonucleotide extension method (Feinberg, A.P., et al. [1983] *Anal. Biochem.* 132:6-13) or with [ $\gamma$ -<sup>32</sup>P]ATP and T4 polynucleotide kinase (Sambrook *et al.*, *supra*). These same procedures were used for all other oligonucleotide probes described below. Both types of probes were purified by passage through spun columns of Sephadex G-50.

Southern transfer was done with Hybond-N (Amersham); the transferred DNA was fixed to the membrane by UV cross-linking. Southern hybridization was performed both with transferred DNA including the restriction fragments described above, as well as the *Bgl*II library of AmEPV DNA cloned into *Bam*HI-digested plasmid pUC9 as described above. Hybridization with the oligonucleotide probe was done at 37 or 45°C with BLOTTO (Sambrook *et al.*, *supra*) and was followed by two washes at room temperature with 0.3 M NaCl-0.06 M Tris (pH 8)-2 mM EDTA for 5 minutes.

The RM58 probe (SEQ ID NO. 12) hybridized to the 4.4 kb *Bgl*II fragment and the *Eco*RI-D fragment of AmEPV DNA (See Figure 1). A plasmid produced by the shotgun cloning, recombinant pRH512 (a *Bgl*II 4.56 kb fragment inserted into the *Bam*HI site of pUC9 which contains about 1.5 kb of the 5' end of the spheroidin gene) was also identified by this hybridization with the RM58 oligonucleotide (SEQ ID NO. 12).

The 4.51 kb pRH512 *Bgl*II insert was isolated, radiolabeled as described above, and hybridized back to various AmEPV genomic digests as follows. The DNA-DNA hybridization was done at 65°C with BLOTTO (Sambrook *et al.*,

*supra*) and was followed by two washes at room temperature with 0.3 M NaCl-0.06 M Tris (pH 8)-2 mM EDTA for 5 minutes, two washes for 15 minutes each at 65°C but with 0.4% SDS added, and two washes at room temperature with 0.03 M NaCl-0.06 M Tris (pH 8)-0.2 mM EDTA. Hybridization was observed to the

5 *Bam*HI-A, *Eco*RI-D, *Hind*III-G and -J, *Pst*I-A, and *Xho*I-B fragments of AmEPV DNA. The results of these hybridizations indicated that the 4.51 kb fragment in pRH512 was substantially identical to the 4.4 kb fragment produced by *Bgl*II digestion of genomic DNA.

The 4.51 kb *Bgl*II insert of pRH512 was thereafter sequenced by two

10 procedures. One is the double-stranded plasmid sequencing method (Hattori, M., *et al.* [1986] *Anal. Biochem.* 152:232-238) performed with "miniprep" (Sambrook *et al., supra*) DNA and 1 pmol of universal, reverse, or custom-designed oligonucleotide primer in each sequencing reaction. Nested exonuclease II deletions (Henikoff, S. [1987] *Methods Enzymol.* 155:156-165) were used to

15 sequence plasmid pRH512 according to this method. Deletions were made from the universal primer end. For making these deletions, the DNA was cut with *Eco*RI, filled in with  $\alpha$ -thiophosphate dNTPs (Putney, S.D., *et al.* [1981] *Proc. Natl. Acad. Sci. USA* 78:7350-7354) by use of the Klenow fragment of *E. coli* DNA polymerase, cut with *Sma*I, and treated with exonuclease III. Samples were

20 removed every 30 seconds, re-ligated, and used to transform *E. coli* SURE cells by electroporation. Sequencing reactions were carried out with the universal primer.

When a primer complementary to that sequence was prepared and used to sequence back through the RM58 binding site (bases 3983 to 4002), the

25 generated sequence, when translated, yielded the amino acid sequence generated from microsequencing the 6.2 kDa CNBr polypeptide fragment.

A second sequencing method was performed using a combination of M12 shotgun sequencing with standard and universal and reverse M13 primers into M13 phage to permit single-stranded sequencing as follows. Plasmid pRH512 was

30 sonicated to produce random fragments, repaired with bacteriophage T4 DNA polymerase, and these fragments were shotgun cloned into *Sma*I-cut M13mp19 (GIBCO). Plaque lifts were screened with a radiolabeled probe prepared from

the 4.5 kb insert found in pRH512 to identify appropriate clones for shotgun single stranded sequencing (see, e.g., Sambrook *et al.*, *supra*). Sequencing of the *Bgl*III insert of pRH512 isolated it to nucleotides # 0 to 4505, thus extending the sequence 5' and 3' to the spheroidin gene (Figure 2).

5

#### Example 5 – Obtaining Additional AmEPV Sequence

A *Dra*I AmEPV DNA library was prepared by digesting genomic DNA with *Dra*I. These *Dra*I fragments were shotgun cloned into *Sma*I-digested, phosphatase-treated vector M13mp19. Preparations of M13 virus and DNA were made by standard procedures (Sambrook *et al.*, *supra*). Ligation and heat shock transformation procedures were performed conventionally (Sambrook *et al.*, *supra*), resulting in the shotgun cloned fragments being transformed into the bacterial strain, *E. coli* UT481 (University of Tennessee) or the SURE strain.

Standard PCR (Innis *et al.*, *supra*) with 400 ng of genomic AmEPV DNA as a template was used to prepare a probe to identify a 586 bp *Dra*I clone from nitrocellulose filter replicas (plaque lifts) (Micron Separations, Inc.) of the M13 shotgun library of *Dra*I-cut AmEPV fragments. This was done to isolate a clone spanning a central unsequenced region of the spheroidin gene. The standard PCR primers used for this reaction were RM92 (SEQ ID NO. 15) (GCCTGGTTGGGTAACACCTC) and RM118 (SEQ ID NO. 16) (CTGCTAGATTATCTACTCCG). This sequencing revealed that there was a single *Hind*III site at base 931 and that the 3' end of the spheroidin open reading frame (ORF) was truncated (Figure 2).

The technique of inverse polymerase chain reaction (PCR) (Innis, M.A., et al., [1990] *PCR protocol, a guide to methods and applications*, Academic Press, Inc. San Diego, CA) was used with *Cla*I-digested AmEPV DNA fragments which were ligated into a circle, to prepare a probe to identify clones containing a flanking sequence or to verify the absence of an intervening sequence between adjacent clones. The primers used in inverse PCR were RM82 and RM83, which were taken from the pRH512 sequence. The sequence of RM82 (SEQ ID NO. 13) was TTTCAAATTAACCTGGCAACC and that of RM83 (SEQ ID NO. 14) was GGGATGGATTTTAGATTGCG.

The specific PCR reaction conditions for 34 cycles were as follows: 30 seconds at 94°C for denaturation, 30 seconds at 37°C for annealing, and 1.5 minutes at 72°C for extension. Finally, the samples were incubated at 72°C to 8.5 minutes to complete extensions. The concentration of each primer was 1  $\mu$ M.

5       The resulting 2.2 kb inverse PCR product was digested with *Cla*I, and a 1.7 kb fragment was gel purified. The 1.7 kb PCR fragment was sequenced with RM83 as a primer. Additional PCR primers were made to the new sequence as it was identified. The sequencing process employed Sequenase, 5 pmol of each primer, and 10 to 50 ng of template. Prior to being sequenced, the PCR products  
10       were chloroform extracted and purified on spun columns (Sambrook *et al.*, *supra*) of Sephacryl S-400. The DNA sequence was assembled and aligned, and consensus sequence was produced (Staden, R. [1982] *Nucleic Acids Res.* 10:4731-4751). Both strands were completely sequenced; the PCR product sequence was verified by conventional sequence.

15       The relevant *Cla*I sites of the 1.7 kb PCR fragment are at positions 3485 and 6165. This fragment was radiolabeled and used as a probe to locate additional clones, i.e., pRH827 (307 bp), pRH85 (1.88 kb), and pRH87 (1.88 kb) from the *Bgl*II fragment library. Plasmids pRH85 and pRH87 were sequenced using the same nested exonuclease II deletions and sequencing procedure, as  
20       described above for pRH512. Sequencing of the inverse PCR products with custom-designed primers confirmed that plasmids pRH85 and pRH87 represented the same 1.88 kb *Bgl*II DNA insert in opposite orientations, but also revealed a missing 80 bp between pRH827 and pRH85. This 80 bp DNA fragment was identified in the *Dra*I fragment, as extending from bases 4543 to 5128 cloned into  
25       M13.

      The orientation of the spheroidin ORF on the physical map is shown in Figure 1. It is interesting to note that the 1.7 kb inverse PCR fragment only hybridized to the AmEPV *Hind*III-G fragment. The amino acid sequence derived from the 8 and 9 kDa overlapping CnBr-generated polypeptides is found from  
30       nucleotide positions 4883 to 4957 (SEQ ID NO. 38). That derived from the 6.2 kDa polypeptide is found from nucleotides 3962 to 4012 (SEQ ID NO. 39), and that derived from the 15 kDa polypeptide is found from nucleotides 4628 to 4651

(SEQ ID NO. 40). Therefore, all sequences obtained from protein microsequencing were ultimately found to lie within the spheroidin ORF.

#### Example 6 – Spheroidin Gene Transcription

5           The start site for spheroidin gene transcription was determined. A primer complementary to the spheroidin gene sequence beginning 65 bp downstream of the predicted initiating methionine was prepared and used for a series of primer extensions.

##### A. Preparation of RNA and primer extension reactions.

10           Six 150 mm dishes of subconfluent cells were prepared. The culture media were aspirated, and 2 ml of viral inoculum was added to each dish. The virus concentration was about 0.1 to 1 PFU per cell. The dishes were occasionally agitated during a 3 hour adsorption period. At the end of this period, the cells were rinsed with 5 ml of modified PBS. The media were replaced, and the  
15           infected cells were incubated for 72 hours at 27°C. Total RNA from the infected cells was isolated by the guanidinium thiocyanate-caesium chloride procedure (Chirgwin, J.M., et al. [1979] *Biochemistry* 18:5294-5299).

            Primer extension reactions were carried out with primer RM165 (SEQ ID NO. 17), a 35-base oligonucleotide (GTTCGAAACAAGTATTTTCATCTTTT  
20           AAATAAATC) beginning and ending 100 and 65 bp downstream, respectively, of the initiating methionine codon found in the TAAATG motif. The primer was end labeled with [ $\gamma$ -<sup>32</sup>P]ATP and T4 polynucleotide kinase and purified on a “spun column” (Sambrook *et al.*, *supra*). For annealing, 40  $\mu$ g of total infected-cell RNA and 10<sup>6</sup> cpm of radiolabeled primer were coprecipitated with ethanol.  
25           The pellet was resuspended in 25  $\mu$ l of hybridization buffer (80% formamide, 40 mM piperazine-*N,N'*-bis(2-ethanesulfonic acid) (pH 6.4), 400 mM NaCl, 1 mM EDTA (pH 8.0)], denatured at 72°C for 15 minutes, and incubated at 30°C for 18 hours.

            For primer extension, the RNA-primer hybrids were ethanol precipitated,  
30           resuspended, and used for five individual reactions. Each reaction contained 8  $\mu$ g of total infected-cell RNA, 50 mM Tris-HCl, (pH 8.3), 50 mM KCl, 10 mM dithiothreitol, 10 mM MgCl<sub>2</sub>, 4 U of avian myeloblastosis virus reverse

transcriptase (Life Sciences), 8 U of RNasin (Promega), 0.25 mM each deoxynucleoside triphosphate (dNTP), and the appropriate dideoxynucleoside triphosphate (ddNTP), except for a control reaction, which contained no ddNTP. The dNTP/ddNTP ratios were 4:1, 5:1, 5:1, and 2:1, for the C, T, A, and G reactions, respectively. The reactions were carried out at 42°C for 30 minutes.

One microliter of chase buffer (4  $\mu$ l of 5 mM dNTP mixture and 1  $\mu$ l of 20-U/ $\mu$ l reverse transcriptase) was added to each reaction mixture, which was then incubated for an additional 30 minutes at 42°C. Reaction products were separated on a sequencing gel (8% acrylamide containing 7 M urea) and visualized by autoradiography. Complementarity was observed until the AAA of the upstream TAAATG motif, indicating that transcription of the gene initiates within the TAAATG element of the proposed late promoter element. Immediately upstream is a 5' tract of noncoded poly(A) on the transcripts. The average length of the poly(A) is greater than 6bp.

#### Example 7 – Analysis of Spheroidin Sequence

The spheroidin ORF (G5R) was initially identified by sequencing back through the RM58 oligonucleotide primer binding region as described above. Examination of the AmEPV spheroidin gene sequence (ORF G5R) revealed a potential ORF of 3.0 kb capable of encoding 1,003 amino acids or a protein of about 115 kDa. The ORF consists of 29% G+C, in contrast to the 18.5% reported for the entire AmEPV genome (Langridge, W.H.R., R.F. Bozarth, D.W. Roberts [1977] Virology 76:616-620). Inspection of the 92 bases upstream of the initiating ATG revealed only 7 G or C residues. Also detected was the presence of known vertebrate poxvirus regulatory sequences within the 92 bp 5' of the spheroidin ORF. Included are three TTTT TNT early gene termination signals and TAAATG, which presumably represents a late transcription start signal used to initiate transcription and translation of the spheroidin gene. Several adjacent translation termination codons are also present within the 92 bp upstream of the spheroidin ORF.

Analysis of the sequence upstream of the spheroidin gene revealed four additional potential ORFs, G1L (SEQ ID NO. 25), G2R (SEQ ID NO. 23), G3L (SEQ ID NO. 26), and G4R (SEQ ID NO. 24), discussed above. The putative amino acid sequences of these ORFs are reported in Figure 2 (SEQ ID NO. 2, 3, 4 and 5, respectively). No significant homologies were found for the small potential polypeptides encoded by ORF G2R (SEQ ID NO. 23) or G3L (SEQ ID NO. 26). ORF G1L (SEQ ID NO. 25), however, exhibited a significant degree of homology to ORF 17 found within the *HindIII*-I fragment of vaccinia virus, whose function is unknown. ORF G4R (SEQ ID NO. 24) showed homology to ORF HM3 of capripoxvirus. In vaccinia virus, the ORF HM3 homolog was found very near the site of an incomplete ATI gene. The partial G6L ORF (SEQ ID NO. 27) to the right of the spheroidin gene exhibited good homology to vaccinia virus NTPase I. Much better homology (78.4% identity over 162 amino acids) was found between the partial G6L ORF (SEQ ID NO. 27) and NPH I of CbEPV (Yuen, L., et al. [1991] *Virology* 182:403-406).

#### Example 8 – Isolation and Sequencing of the AmEPV EcoRI-Q Fragment Containing the tk Gene

Sequencing of the *EcoRI*-Q fragment of genomic AmEPV of Example 1 was performed using techniques described above for spheroidin. The sequencing showed 1511 bp containing two complete and one partial ORF. Analysis of the DNA sequence of ORF Q2 (SEQ ID NO. 28) indicates the sites where the identifying degenerate oligonucleotides (RM03 SEQ ID NO. 18 and RM04 SEQ ID NO. 19) might hybridize. Two oligonucleotides, RM03 and RM04, based on different but strongly conserved regions of the tk genes of several poxviruses and vertebrates (Upton, C., et al. [1986] *J. Virology* 60:920-927; Boyle, D.B., et al. [1987] *Virology* 156:355-365) were prepared by the methods referred to above. RM03 was the 32-fold degenerate oligonucleotide (SEQ ID NO. 18) GA(T/C)GA(G/A)GG(G/A)GG(G/A)CA(G/A)TT(C/T)TT corresponding to the amino acid residues in the vaccinia tk protein from the aspartic acid at position 82 to the phenylalanine at position 87. RM04 (SEQ ID NO. 19) was (GGNCCCATGTT(C/T)TCNGG with 32-fold degeneracy and corresponded to

the region from the glycine at position 11 to the glycine at position 16 in vaccinia. These probes were radiolabeled as described above for the RH58 probe.

5 The AmEPV thymidine kinase (tk) gene was identified by hybridization with the degenerate oligonucleotide probes RM03 and RM04 to a Southern blot of the *EcoRI*-digested EPV DNA. The *EcoRI* band of interest (*EcoRI*-Q) was isolated, purified, and ligated into a pUC18 vector (GIBCO), previously digested with *EcoRI* and treated with calf intestinal alkaline phosphatase. Recombinant clones were identified by the size of the insert and by hybridization to the radioactive labeled oligonucleotide probes.

10 One such clone was called pMEGtk-1. The recombinant clones containing the *EcoRI*-Q fragment oriented in both directions relative to the pUC18 vector sequences were used for sequencing. Sequential nested deletions were generated by the method of Henikoff, cited above, as described for pRH512. These clones were used for sequencing the entire *EcoRI*-Q fragment.

15 Subsequently, these oligonucleotides and another, RM129 is a non-degenerate oligonucleotide GGTGCAAAATCTGATATTTC (SEQ ID NO. 20) prepared from the ORF Q1, were employed as sequencing primers to confirm their positioning as indicated in ORF Q2 (SEQ ID NO. 28). ORF Q2 potentially encodes for a protein of 182 amino acids (21.2 kDa) (SEQ ID NO. 10). ORF Q3  
20 potentially encodes a polypeptide of at least 68 amino acids but is incomplete and is transcribed in the opposite direction from ORF Q2. ORF Q1 (SEQ ID NO. 31) potentially encodes a small peptide of 66 amino acids (7.75 kDa) (SEQ ID NO. 9).

Further analysis of the *EcoRI*-Q fragment reveals several other points.  
25 First, the A+T content is very high (80%). For ORF Q2, the 100 nucleotides upstream of the start codon for translation are 90% A+T. Some potential poxvirus transcription signals were found between ORFs Q1 and Q2. The five bases immediately preceding the start codon for ORF Q1 are TAAATG which comprise a consensus late poxvirus promoter. A potential poxvirus early  
30 transcription termination signal sequence (TTTTTAT) is located 2 nt past the translation stop codon of Q2.

The deduced amino acid sequence for the tk encoded by the ORF Q2 of the *EcoRI*-Q fragment can be compared to the tk genes for the poxviruses swine pox (Schnitzlein, W.M., et al. [1991] *Viol.* 181:727-732; Feller, J.A., et al. [1991] *Viol.* 183:578-585); fowlpox (Boyle *et al.*, *supra*; Binns, M.M., *et al.* [1988] *J. Gen. Virol.* 69:1275-1283); vaccinia (Weir, J.P., *et al.* [1983] *J. Virol.* 46:530-537; Hruby, D.E., *et al.* [1983] *Proc. Natl. Acad. Sci. USA* 80:3411-3415); variola and monkeypox (Esposito, J.J., et al. [1984] *Viol.* 135:561-567); capripoxvirus (Gershon, P.D., et al. [1989] *J. Gen. Virol.* 70:525-533); Shope fibroma virus (Upton *et al.*, *supra*); the cellular thymidine kinases of humans (Bradshaw, H.D., et al. [1984] *Mol. Cell. Biol.* 4:2316-2320; Flemington, E., *et al.* [1987] *Gene* 52:267-277); the tk of mouse (Lin, P.F., et al. [1985] *Mol Cell. Biol.* 5:3149-3156); the tk of chicken (Kwoh, T.J., *et al.* [1984] *Nucl. Acids Res.* 12:3959-3971); ASF (Blasco, R., *et al.* [1990] *Viol.* 178:301-304; Martin Hernandez, A.M., *et al.* [1991] *J. Virol.* 65:1046-1052).

15

#### Example 9 – Expression of the AmEPV tk Gene in a Vaccinia Virus

The AmEPV tk gene was tested functionally by cloning the gene into a vaccinia virus strain tk<sup>-</sup> mutant, as follows.

The *EcoRI*-Q fragment of AmEPV, described above, was inserted in both possible orientations into shuttle plasmid pHGN3.1 (Bloom, D.D., et al. [1991] *J. Virol.* 65:1530-1542) which had been isolated from bacterial cells by the alkaline lysis method. This *EcoRI*-Q DNA fragment contains the AmEPV tk open reading frame (ORF). The cloning was performed conventionally. The resulting plasmid was designated pHGN3.1/*EcoRI*-Q.

The plasmid was transfected by Lipofectin (GIBCO) as described specifically below into mammalian cells infected with vaccinia virus. The cells were either rat tk<sup>-</sup>, human 143 tk<sup>-</sup>, or CV-1 cell lines onto which the vaccinia virus VSC8 was propagated. These cells were maintained in Eagle's Minimal Essential Medium with Earle's salts (Massung *et al.* [1991] *Viol.* 180:347-354, incorporated by reference herein).

30

The VSC8 vaccinia strain (Dr. Bernard Moss) contains the  $\beta$ -galactosidase gene driven by the vaccinia P<sub>11</sub> promoter (P<sub>11</sub>-Lac Z cassette) inserted into the

viral tk gene. While VSC8 contains an inactive tk gene due to the insertion of the  $\beta$ -galactosidase, portions of the vaccinia tk sequence remain. VSC8 is thus tk<sup>-</sup> and, upon staining with X-Gal (5-bromo-4-chloro-3-indoyl- $\beta$ -D-galactopyranoside), will form blue plaques ( $\beta$ -galactosidase positive).

5           Cells were grown to 80% confluence ( $4 \times 10^6$  per 60 mm dish). Lipofectin solution (20  $\mu$ g of Lipofectin in 50  $\mu$ l of dH<sub>2</sub>O) was added to 10  $\mu$ g plasmid DNA (pHGN3.1/AmEPV *Eco*RI-Q) in 50  $\mu$ l of dH<sub>2</sub>O and incubated for 15 minutes at room temperature. After a 2 hour period of viral adsorption (m.o.i. of 2, 37°C), the monolayers were washed three times with serum-free OptiMEM. Three  
10          milliliters of serum-free OptiMEM was then added to each 60 mm dish. The Lipofectin/DNA mixture was slowly added dropwise with gentle swirling and incubated an additional 12 to 18 hours at 37°C. Fetal bovine serum was then added (10% final) and the infected cells were harvested at 48 hours postinfection.

            Recombinant viruses, containing the *Eco*RI-Q fragment inserted into the  
15          hemagglutinin (HA) gene of vaccinia, were identified by hybridization of AmEPV *Eco*RI-Q fragments, radioactively labeled by procedures described above, to replicas of nitrocellulose "lifts" of virus plaques from the infected monolayer. Potential recombinants were isolated from replica filters and plaque-purified several times before testing.

20          The tk of AmEPV exhibits some degree of homology with the tk of vaccinia. To confirm that insertion of the AmEPV tk gene was within the HA gene of vaccinia rather than within residual tk sequences remaining in VSC8, the recombinants were examined by a series of Southern hybridizations to *Hind*III digests of the various viruses. When DNA from wild-type virus was hybridized to  
25          a vaccinia virus tk probe, hybridization was observed exclusively within the  $\approx 5$  kb *Hind*III-J fragment of vaccinia.

            When either VSC8 or either of the AmEPV tk containing recombinants was examined using the vaccinia tk probe, hybridization occurred instead to an  $\approx 8$  kb fragment consistent with insertion of the  $\approx 3.1$  kb  $\beta$ -galactosidase cassette into  
30          the TK gene of the *Hind*III-J fragment. Hybridization of the *Hind*III digests of the same three viruses to radiolabeled AmEPV *Eco*RI-Q DNA resulted in hybridization to a large-molecular-weight DNA fragment in the recombinant

containing the AmEPV TK gene, which we have shown corresponds to the *HindIII*-A fragment. These results suggest that insertion of the AmEPV TK gene into vaccinia occurred within the HA gene contained within the large (>23 kb) *HindIII*-A fragment of the virus as expected. Identical results were observed for VSC8::TKII. It is also interesting to note the lack of hybridization of the vaccinia virus (VV) and AmEPV TK probes to the heterologous poxvirus TK genes under these conditions, even though there is significant sequence homology between the vaccinia and the AmEPV TK genes.

Plaque-purified vaccinia recombinants were tested for growth on human 143 TK<sup>-</sup> cells in the presence of methotrexate. Under these conditions, only TK<sup>+</sup> virus will grow and produce plaques. Both recombinants (VSC8::AmEPV TKI and VSC8::AmEPV TKII) representing the two orientations of the AmEPV *EcoRI*-Q fragment gave plaques in the presence of methotrexate. All plaques from both recombinant viruses which grew in the presence of methotrexate also stained blue upon staining with X-Gal, suggesting that the TK<sup>+</sup> phenotype of these recombinants is most likely due to a functional TK gene contained within the AmEPV *EcoRI*-Q fragment. Since functionality is independent of the orientation of the AmEPV *EcoRI*-Q fragment, it is plausible that an AmEPV TK gene promoter contained within the *EcoRI*-Q fragment may function in vaccinia.

#### Example 10 – Orientation of the TK Gene Within the AmEPV Genome

In order to determine the orientation of the Q1, Q2, Q3 ORFs within the *EcoRI*-Q fragment relative to the genome, a DNA primer-mediated extension reaction was performed. The *EcoRI*-Q fragment is contained within the *PstI*-F fragment. An oligonucleotide identical in sequence to bases 182-163 (RM129, SEQ ID NO. 12) was prepared and hybridized to a heat-denatured *PstI* digest of genomic AmEPV DNA. The partial hybrid was then extended with the Klenow fragment of *E. coli* DNA polymerase in the presence of radiolabeled substrates. Extension will terminate at the end of the *PstI*-F fragment.

The radiolabeled product was then hybridized to an *EcoRI* digest of AmEPV DNA. If orientation of the gene is such that the tk ORF reads toward the end of the genome, hybridization would be expected to the *EcoRI*-E fragment;

whereas if the gene is read toward the center of the genome, hybridization would be expected to the *EcoRI*-I fragment.

The results indicate hybridization not only to the *EcoRI*-E fragment, but also to the *EcoRI*-A fragment. These results infer that the orientation of the tk gene is with reading toward the left end of the genome. Hybridization of the run-off extension product also to the *EcoRI*-A fragment is consistent with the presence of an inverted terminal repetition, common in poxviruses, with identical sequences residing in both the *EcoRI*-A and the *EcoRI*-E fragments.

The optimal growth temperature for AmEPV in the laboratory is 28°C, whereas that of the vertebrate poxviruses is 37°C. As described herein, when the AmEPV DNA fragment containing the entire tk gene was cloned into the tk<sup>-</sup> strain of vaccinia virus, the recombinant virus was capable of growing at 37°C in the presence of methotrexate (Sigma), indicative of a tk<sup>+</sup> phenotype. This example demonstrates that the Entomopoxvirus tk gene can be successfully transferred into mammalian expression systems, and that AmEPV tk is functionally active over a considerable temperature range.

#### Example 11 – Isolation and Sequencing of CbEPV and CfEPV Spheroidin Genes

Viruses and cell culture. *Choristoneura* EPVs, CbEPV and CfEPV, occlusion bodies (OBs) were obtained from Dr. Basil M. Arif, Forest Pest Management Institute, Sault Ste. Marie, Ontario, Canada, and can be obtained from other sources as well. The AmEPV OBs were the same as those used previously (Hall and Moyer, 1991). The gypsy moth cell line, IPLB-LD-652, used for AmEPV replication was maintained in EX-CELL 401 (JRH Biosciences, Lenexa, KS) supplemented with 10% fetal bovine serum. Wild type cowpox virus, Brighton red strain, was produced in CV-1 cells cultured in Eagle's MEM with Earle's salts.

SDS-PAGE, occlusion body solubilization, and total cell protein preparation. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of proteins was with a 4% acrylamide stacking gel and a 7.5% separating gel. For sample preparation, 2X Laemmli sample buffer consisting of 125 mM Tris-HCl (pH 6.8), 4% SDS (wt/vol), 10%  $\beta$ -mercaptoethanol (vol/vol), 20%

glycerol (vol/vol), and 0.05% bromphenol blue (wt/vol) was used. Quantitation of OBs was achieved by counting a diluted suspension in a hemocytometer. The OB suspension (varying from  $8 \times 10^6$  to  $1 \times 10^8$  OBs/ml) usually prepared in 1X TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), was diluted 1:1 with the 2X Laemmli sample buffer and immediately boiled for 5 or 10 min. For studies designed to assay for spheroidin degradation, the OBs plus sample buffer were left at room temperature for various time periods before boiling. After electrophoretic separation, the proteins were visualized by Coomassie blue staining.

The preparation of total uninfected or infected cell protein for SDS-PAGE utilized a 2X lysis buffer consisting of 160 mM Tris-HCl (pH 9), 4% SDS (wt/vol), 4%  $\beta$ -mercaptoethanol (vol/vol), 10% glycerol (vol/vol), 5 M urea, 20 mM EDTA (pH 8), and 0.01% bromphenol blue (wt/vol). Cells were scraped from a 150 mm diameter culture dish using a rubber policeman and centrifuged at  $1000 \times g$  for 5 min at 4°C. Cells were washed in appropriate cold phosphate buffered saline (PBS) and recentrifuged. Insect cells were harvested in Hank's PBS which contained 15 g glucose per liter. The cell pellets were resuspended in a final volume of 500  $\mu$ l of PBS which contained 5  $\mu$ l each per ml of PMSF (phenylmethyl sulfonyl fluoride, 0.1 M stock in 95% ethanol) and aprotinin (0.3 mg/ml stock). The cells were probe sonicated on ice for 15 sec., five hundred microliters of 2X lysis buffer was added, and the preparation sonicated as before. Samples were stored at -70°C and boiled before loading on an SDS-PAGE gel.

Preparation of EPV DNAs. CfEPV DNA was obtained from Dr. Basil M. Arif, Forest Pest Management Institute, Sault Ste. Marie, Ontario, Canada. CbEPV DNA used in the polymerase chain reaction (PCR) studies was prepared directly from occlusion bodies by the agarose *in situ* method (Hall and Hink, 1990) and extracted by "freeze-squeeze" (Sambrook *et al.*, 1989). AmEPV DNA was prepared from infected LD-652 cell culture by concentrating extracellular virus which was then treated with DNase I and RNase A prior to dissolution of the occlusion bodies by SDS and proteinase K digestion (Gruidl *et al.*, 1992).

PCR Primers and Reactions. PCR, product purification, and dideoxy sequencing were done as described previously (Hall and Moyer, 1991) except that the specific reaction conditions for 34 cycles were 1 min at 94°C for denaturation,

1 min at 37°C for annealing, and 2 min at 72°C for extension. Finally, samples were incubated at 72°C for 9 min to complete extensions. Typically, 50 to 100 ng of template DNA was used in the PCRs. Custom oligonucleotide primers RM82, RM83, RM92 and RM118 were described previously (Hall and Moyer, 1991).  
5 Sequences of these and other AmEPV spheroidin specific PCR primers (5' to 3', base pair numbers from Hall and Moyer, 1991) are summarized in Table 2.

Primer RM206 (AGATGATGATTAAAGTGTGG) (SEQ ID NO. 50) was from bases 2379 through 2398 and RM212 (GATAATGATACTCCGGTTGC) (SEQ ID NO. 51) from bases 2077 through 2096 of the CbEPV NPH I sequence (Yuen *et al.*, 1991).

5        Cloning and double strand plasmid sequencing. BglII clones in both orientations (1.06 kb and 1.78 kb) covering the unsequenced 5' end of the AmEPV NPH I gene were selected from an AmEPV BglII fragment library (Hall and Moyer, 1991) by hybridization with an AmEPV 13 kb HindIII fragment probe. Plasmids were cloned in the *Escherichia coli* SURE strain (Stratagene, La Jolla, CA). Plasmids were sequenced by use of exonuclease III deletions and dideoxy sequencing as described (Hall and Moyer, 1991). Both strands were completely sequenced.

10        Radiolabeled probes, Southern blotting, and hybridization. Random oligolabeling of DNA for probes, Southern transfer and hybridization (at 65°C with BLOTTO) were as described previously (Hall and Moyer, 1991).

15        Protein microsequencing. Lys-c endoprotease digestion of the 115 kDa protein of CbEPV OBs purified and recovered from SDS-PAGE gels was used to generate internal peptide fragments for sequencing on a gas phase sequencer.

20        Antibody preparation and immunoblotting. A preparation of total occlusion body antigens was prepared by solubilizing purified AmEPV occlusion bodies purified from infected cell cultures. Rabbits were then intradermally injected with 100-200 µg of antigen per rabbit in Freund's complete adjuvant. One month later, the rabbits were boosted with about 500 µg each of the same antigen in incomplete adjuvant. After an additional 3 weeks, the rabbits were  
25        boosted with 200 µg of antigen in incomplete adjuvant.

Eleven days later, the immune serum was collected. This serum is referred to as occlusion body antisera. Monospecific spheroidin antibodies were prepared from this serum based on immunoaffinity of individual antibodies for purified spheroidin (Harlow and Lane, 1988). Samples of the sera were adsorbed to a  
30        nitrocellulose blot prepared from a preparative SDS-PAGE gel of solubilized AmEPV occlusion bodies. The section of the blot containing the 115 kDa AmEPV spheroidin and bound antibodies was then excised. The monospecific

AmEPV spheroidin antibodies were eluted using 100 mM glycine, pH 2.5. After neutralization and dilution, the monospecific spheroidin antibodies were used to probe Western blots.

For immunoblotting, duplicate samples run on SDS-PAGE gels were prepared. One-half of the gel was stained with Coomassie blue, and the other half (containing prestained molecular mass markers) was transferred to nitrocellulose membrane in Tris-glycine buffer (Harlow and Lane, 1988) using the BioRad Trans Blot at 250 mA for 3 hours. The blot was blocked for 2 hr using BLOTTO (0.5% nonfat dry milk in TBS: 0.01 M Tris, pH8, 0.15 M NaCl). Dilutions of the antibody were prepared in BLOTTO and adsorbed to the blot overnight. The blot was washed 3 times in TBS at room temperature, and the secondary antibody (goat anti-rabbit conjugated to alkaline phosphatase) at a 1:1000 dilution was adsorbed to the blot for 1½ hr. The blot was washed with TBS as previously. Secondary antibody reactions and color development was as described (Harlow and Lane, 1988).

SDS-PAGE of solubilized occlusion bodies. When purified occlusion bodies of CbEPV, CfEPV and AmEPV are solubilized in Laemmli sample buffer, boiled immediately, and analyzed on SDS-PAGE gels, Coomassie blue staining shows the major protein to be about 115 kDa in each case. The *Choristoneura* EPVs show a second 47 kDa protein in lesser amounts. Other minor proteins were also observed.

The purported CbEPV spheroidin gene has a coding capacity of 47 kDa (Yuen *et al.*, 1990) despite the fact that the observed size of the corresponding spheroidin appears to be 115 kDa, which is similar to that observed for the AmEPV spheroidin. This discrepancy has been explained by suggesting that the CbEPV spheroidin exists as relatively unstable dimers, which dissociate under a variety of conditions to monomers of 47 kDa (Yuen *et al.*, 1990). The AmEPV spheroidin of 115 kDa, however, shows no such propensity for dissociation.

Prior to dissolution, occlusion bodies are stable and routinely stored at 4°C in buffer. The only discernible difference in methods of occlusion body solubilization and preparation is the incubation time in SDS buffer before boiling. The relative instability of the various spheroidins was evaluated by incubating the

occlusion bodies of all three EPVs at room temperature for up to two hours in SDS solubilizing buffer before boiling the samples. While some degradation of the 115 kDa protein was observed for the CfEPV OB preparation, little if any degradation of the CbEPV or AmEPV preparations was observed. The CfEPV 115 kDa protein was degraded to a variety of smaller proteins but not in a fashion suggesting a relationship to the 47 kDa protein. Whether the OB suspension was in 1X TE (10 mM Tris, 1 mM EDTA, pH8) or in deionized water prior to dissolution has no effect on the subsequent degradation CfEPV occlusion bodies.

Discovery of an AmEPV spheroidin gene homolog in CbEPV and CfEPV.

Selected oligonucleotide primers derived from within various regions of the AmEPV spheroidin gene were selected as appropriate primer pairs for PCR to look for the spheroidin gene in CbEPV and CfEPV. The relative positions of these primers within the spheroidin gene are shown in Figure 5. Table 2 lists the specific primer pairs and sequences, the expected size of the PCR products based on the AmEPV spheroidin sequences, and the results when these primers were used in conjunction with CbEPV or CfEPV templates. From Table 2, for CbEPV template 10 out 19 primer pairs resulted in an appropriate size product expected if the two genes were similar. For CfEPV this was 13 out of 19 primer pairs. We chose one of the products (1 kb) generated from CfEPV DNA by primer pair 1-13 (in Table 2) for further analyses.

This PCR product was radiolabeled and used as a probe for a blot containing restriction fragments of both CfEPV and AmEPV DNAs. All hybridizations to CfEPV showed predominant, specific hybridization signals. The CfEPV derived probe also shows appropriate, discrete hybridizations to AmEPV DNA; i.e., a 13 kb *Hind*III fragment, a 20 kb *Eco*RI fragment, a 4.5 kb *Bgl*II fragment, and a 4.5 kb *Bst*BI fragment. This pattern of hybridization to AmEPV is that expected for hybridization to the AmEPV spheroidin gene.

Partial sequence of a spheroidin-like gene in CbEPV or CfEPV. Further indications of the existence for an AmEPV spheroidin gene homolog in the genome *Choristoneura* EPVs come from PCR product sequencing of the 1 kb *Choristoneura* PCR products (primer pair 1-13; Table 2). The resulting sequences derived from the CbEPV or CfEPV DNA with a comparison to the AmEPV

sequence is shown in Figure 6A. When the deduced amino acid sequence of this region is compared (Figure 6B), a very high degree of homology is found between all three viruses.

5        The spheroidin-like gene in CbEPV and CfEPV is expressed. Samples of  
the CbEPV  $\approx$ 115 kDa protein were isolated from SDS-polyacrylamide gels, and  
treated with lys-c endoprotease to generate peptides for microsequencing. Three  
of the resulting peptides were analyzed, and the amino acid sequence was  
compared to the spheroidin of AmEPV. The CbEPV sequences obtained were  
homologous to three corresponding regions of the AmEPV spheroidin (Figure  
10    6C). These results demonstrate that the *Choristoneura* viruses not only contain  
a spheroidin-like gene, but that gene is expressed to yield a polypeptide within  
occlusion bodies of similar size and sequence to the previously-identified  
spheroidin protein of AmEPV.

We have also addressed the question of whether the AmEPV spheroidin  
15    homolog in CbEPV and CfEPV is expressed by a Western blot analysis of the  
proteins of CbEPV and CfEPV using antibody derived against either AmEPV  
occlusion bodies or monospecific sera against AmEPV spheroidin. Sera directed  
against purified AmEPV occlusion bodies recognize proteins of  $\approx$ 115 kDa in both  
CbEPV and AmEPV. Stronger signals are observed in the AmEPV samples as  
20    expected. An AmEPV protein of 38 kDa is also recognized in the samples.  
Weak binding is also observed to the abundantly expressed protein of the  
vertebrate cowpox virus which was used as a control. When immunoaffinity  
purified, monospecific AmEPV spheroidin sera is used, the 115 kDa protein of  
CbEPV also cross-reacts. Similar results were obtained with CfEPV. These  
25    results also support the conclusion that the two *Choristoneura* viruses and AmEPV  
encode a very similar major occlusion body protein of 115 kDa which in AmEPV  
corresponds to the spheroidin gene.

Co-linearity of AmEPV, CbEPV, and CfEPV maps in the spheroidin  
region. We have shown that the gene adjacent to the 3' terminus of the AmEPV  
30    spheroidin gene is NPH I (NTPase I) in polarity opposite to that of the spheroidin  
gene. A NPH I gene from CbEPV has been sequenced (Yuen *et al.*, 1991).

The sequence of the NPH I (or NTPase I) gene of AmEPV is presented in Figure 7. The 3' end sequence of this gene is provided in Figure 1 and the nucleotide numbering system depicted in Figure 7 results from appending the 5' end of the sequence to the sequence provided in Figure 1.

5           When the complete coding sequence and deduced amino acid sequence of the AmEPV NPH I gene was compared to the already published CbEPV NPH I gene (Yuen *et al.*, 1991), the two genes show 89% amino acid and 86% nucleotide identity. Both proteins have a deduced 648 amino acids, with the major difference being that AmEPV has deleted amino acid number 127 and has  
10           one extra amino acid at the very end of the sequence. Both genes show the typical poxvirus late gene promoter sequence motif, TAAATG, at the beginning of the open reading frame as well as the A+T rich sequence upstream of the gene. Of the 30 bases preceding the starting ATG, only a single G differentiates the AmEPV and CbEPV. The intergenic region between the spheroidin and NPH  
15           I genes begins to diverge immediately following the NPH I open reading frame (ORF) at the downstream 3' end.

          The proximity of the spheroidin and NPH I genes in both *Choristoneura* viruses was tested by performing a series of PCR reactions using the PCR primers shown in Figure 8 which were designed based on the AmEPV gene arrangement  
20           and were used with either CfEPV or CbEPV. Primers RM206 and RM212 are described above in this example, RM83 and RM161 in Table 2. Primer pair RM161-RM212 failed to give a PCR product with CbEPV DNA indicating sequence differences between CbEPV and CfEPV in the RM161 primer binding region. However, the other three reactions generated PCR products ranging in  
25           size from  $\approx$ 1-2.5 kb, indicating that the NPH I and spheroidin genes are adjacent and arranged similarly to AmEPV. Based on the PCR products, the intergenic distance between the *Choristoneura* spheroidin and NPH I genes appears to be at least about 450 bp longer than the same region in AmEPV. The increased size of this intergenic region is large enough to perhaps allow for a small ORF to be  
30           present between the two genes in the *Choristoneura* viruses. However, PCR product sequencing of the RM83 - RM212 and the RM161 - RM206 PCR products using RM206 as the primer failed to show any ORF of significant size.

Our results suggest that the published *Choristoneura* EPV gene identified by Yuen *et al.* (1990) as the spheroidin gene, is incorrect. Our evidence shows instead that the spheroidin gene of *Choristoneura* EPVs is virtually identical to that of AmEPV (Figure 6), i.e., encodes a 115 kDa protein and is expressed.

5 Hence this protein is highly conserved amongst all three viruses.

The AmEPV spheroidin gene found in the *Choristoneura* EPV genomes and the NPH I genes were found to be immediately adjacent to each other and in opposite polarity in all three viruses. Although the intergenic distance between the two genes is somewhat different between the two *Choristoneura* viruses and  
10 AmEPV, it would appear that the genes in this region of the viral DNA are co-linear.

We have shown that the linear conserved core of genes found in vertebrate poxviruses is not maintained in AmEPV. Based on the data presented here, it appears that the entomopoxviruses have evolved and maintained a common core  
15 of co-linear genes, different from their vertebrate counterparts.

It should be understood that the examples and embodiments described herein are for illustrative purposes only. Various modifications or changes in light thereof will be suggested to persons skilled in the art by this specification. The  
20 subject invention encompasses recombinant polynucleotide sequences, plasmids, vectors, and transformed hosts which are equivalent to those which are specifically exemplified herein in that the characteristic expression features are retained in said equivalent constructs even if inconsequential modifications to the DNA sequence have been made. For example, it is within the skill of a person trained  
25 in the art to use a fragment of the spheroidin gene's non-coding region which is upstream of the structural gene in order to achieve the desired level of expression. Such fragments of the regulatory sequences fall within the scope of the current invention, so long as the desired level of expression which is characteristic of this system is retained. Furthermore, inconsequential changes to the nucleotide  
30 sequences can be made without affecting the disclosed functions of these sequences. Such modifications also fall within the scope of the current invention

and are to be included within the spirit and purview of this application and the scope of the appended claims.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Moyer, Richard W.  
Hall, Richard L.  
Gruidl, Michael E.
- (ii) TITLE OF INVENTION: Novel Entomopoxvirus Expression System
- (iii) NUMBER OF SEQUENCES: 66
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: David R. Saliwanchik
  - (B) STREET: 2421 N.W. 41st Street, Suite A-1
  - (C) CITY: Gainesville
  - (D) STATE: FL
  - (E) COUNTRY: USA
  - (F) ZIP: 32606
- (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: US 07/991,867
  - (B) FILING DATE: 12-DEC-1992
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: WO 92/14818
  - (B) FILING DATE: 12-FEB-1992
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 07/827,685
  - (B) FILING DATE: 30-JAN-1992
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 07/657,584
  - (B) FILING DATE: 19-FEB-1991
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Saliwanchik, David R.
  - (B) REGISTRATION NUMBER: 31,794
  - (C) REFERENCE/DOCKET NUMBER: UF114.C3
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 904-375-8100
  - (B) TELEFAX: 904-372-5800

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8457 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: DNA (genomic)
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Amsacta moorei entomopoxvirus
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: complement (65..1459)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1474..2151
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: complement (2239..2475)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 2502..2987
- (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 3080..6091

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: complement (6277..6768)

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

|                                                                   |      |
|-------------------------------------------------------------------|------|
| AGATCTGATG TTCTATATAT AGTACAAATT TGTATGATTA ATTGATATTT TAAAATTCAA | 60   |
| GATATTAAAT ATTAGATTCT AAACATTCT TCTCATTATC AATATAACTA TCATAATCAT  | 120  |
| TTTTTATTTT ACTACATACA TTCATAATTC TATTACTATT TTTTTTATAC ATATCTATTA | 180  |
| ATTCCATAAA CTTTTATTT TTTATATTAA ATATTTCTAA TGTATTTTAA AATTCGTCAA  | 240  |
| TACTATTAAT ATCATATCTA GAAATAAATA ATGCACCTCT ATAACACTA GCCAATAAAT  | 300  |
| CACCAATAAA ACTCATAGAA TAATATAATT TTTTAAATTC AAATTTAGAT TTTATGTTGA | 360  |
| AATAAACTAT ATAATATAAA AATATTATAT TAAACATACC ACAATCGGGA CTATCATATT | 420  |
| GTAATTCAAA AGTATTAAAA AAGTAATAAT TTACATTTTT AAATATATCA TTTAAATATT | 480  |
| CTGATAGTAC ATCAATGTAT AAATAAGCAT AATTAGTATT AGGAGTACTA TTGTAGTGTT | 540  |
| TATGGCTTTT TATAGTCATA TCAGATTCAA TAAACATATA TTTTTTATTT TGTTTTATAA | 600  |
| GTTCTGGTAT ATAACCACTA CTATTAATAA AGTATGCAGC TTTTTTATCT TTATCAAAGT | 660  |
| GTTTATCTAT TACGCAACAA GTAAAATGAT CATTATAAAT TATAGGAAAC ATAAAAATC  | 720  |
| TTTTTTTATC ATTCATTAAA AAAAATTTTA CTCTATCTTC AAGTTTATAG CATCTCATAG | 780  |
| ATGAAGCTAC TGTAGCAATA TTTTATCAG TTTTTCAAA TAAAATCAA TGAAAATAAT    | 840  |
| CATAATCTGT ATTAATCATA GTTAATGGAT ATATACAATT ATATATATCT CCCGAACCTA | 900  |
| ACCATGTAGA TTTATCATGT TTTCTTGGGT AAGCTTTAGG TTTAGGATTA AATCCCAAAG | 960  |
| GCGGTATTCC TATTTGAGCA TCCAAATCAT CATAAATTGT GGCAAATGTA GAAAAATCTC | 1020 |
| TTGTTTTGGA TAATCTGAT TTTAGAAAAG ACTTCTCAT ATATACTAAT GGAATGCCTT   | 1080 |
| TATATTTTTT AGATGTAATA AAAGTATTAA TATTATATT TTTATCTTGT AAATATTTTT  | 1140 |
| TTATAGTCCA AAATAGAAAA AATTTTCTTT TAATATTATT TTCAAATTA ATATTATTAA  | 1200 |
| TATGATTTGG ATCTAAAACT AATTCATTAT ATAATATTTT CAAGTATTTT ATAGGTATAA | 1260 |
| ATGTTACTTT ACCTCTTGTT TCATCATCAT CATCTATTTT TTCTAATATA GCTATATTTG | 1320 |
| CATTAGTATT ATATTTAATA GGATTTATAA AATATACCAT ATTATCTATT TTAATAAAAA | 1380 |
| ATAACATAGA CATAAAATTA ATACCAGATT CTGGCATTTT TAAATTTTAA TTTGGAAATC | 1440 |
| TTCTAATTTT ATTATTCATT ATTTATTTAA TAA ATG TTT CTA GTT TAT TTC AAT  | 1494 |
| Met Phe Leu Val Tyr Phe Asn                                       |      |
| 1 5                                                               |      |
| ACA TTT TTA ATA ATA ATT TTA TTA TTT GGT ATT ATA GGT ATT TAT ATA   | 1542 |
| Thr Phe Leu Ile Ile Ile Leu Leu Phe Gly Ile Ile Gly Ile Tyr Ile   |      |
| 10 15 20                                                          |      |
| TTA ACA TTT GTG TTT AAT ATA GAT TTT TTA ATA AAT AAT AAT AAA ATA   | 1590 |
| Leu Thr Phe Val Phe Asn Ile Asp Phe Leu Ile Asn Asn Asn Lys Ile   |      |
| 25 30 35                                                          |      |
| TAT ATA TTA TCA TAT AAC GCA ACT AAT ATA AAC AAT ATA AAT AAT TTA   | 1638 |
| Tyr Ile Leu Ser Tyr Asn Ala Thr Asn Ile Asn Asn Ile Asn Asn Leu   |      |
| 40 45 50 55                                                       |      |
| AAT TTA TAC GAT TAT TCA GAT ATT ATA TTT TTG ACA AAT TTT AAC ATA   | 1686 |
| Asn Leu Tyr Asp Tyr Ser Asp Ile Ile Phe Leu Thr Asn Phe Asn Ile   |      |
| 60 65 70                                                          |      |
| AAT AAT AAT CTT TTA GTA ACA CAA GCT AAT AAT TTA CAA GAT ATA CCA   | 1734 |
| Asn Asn Asn Leu Leu Val Thr Gln Ala Asn Asn Leu Gln Asp Ile Pro   |      |
| 75 80 85                                                          |      |

61

|                   |                   |                   |                   |                  |                   |                   |                   |                  |                   |                       |                   |                   |                  |                   |                   |      |  |            |            |            |            |            |      |
|-------------------|-------------------|-------------------|-------------------|------------------|-------------------|-------------------|-------------------|------------------|-------------------|-----------------------|-------------------|-------------------|------------------|-------------------|-------------------|------|--|------------|------------|------------|------------|------------|------|
| ATA<br>Ile        | TTT<br>Phe        | AAT<br>Asn<br>90  | GTA<br>Val        | AAT<br>Asn       | AAT<br>Asn        | ATT<br>Ile        | ATA<br>Ile<br>95  | TCT<br>Ser       | AAT<br>Asn        | CAA<br>Gln            | TAT<br>Tyr        | AAT<br>Asn<br>100 | TTT<br>Phe       | TAT<br>Tyr        | TCA<br>Ser        | 1782 |  |            |            |            |            |            |      |
| GCG<br>Ala        | TCT<br>Ser<br>105 | AGT<br>Ser        | AAT<br>Asn        | AAT<br>Asn       | GTA<br>Val        | AAT<br>Asn<br>110 | ATA<br>Ile        | TTA<br>Leu       | TTA<br>Leu        | GGA<br>Gly<br>115     | TTA<br>Leu        | AGA<br>Arg        | AAA<br>Lys       | ACA<br>Thr        | TTA<br>Leu        | 1830 |  |            |            |            |            |            |      |
| AAT<br>Asn<br>120 | ATA<br>Ile        | AAT<br>Asn        | AGA<br>Arg        | AAT<br>Asn       | CCA<br>Pro<br>125 | TTT<br>Phe        | TTA<br>Leu        | TTA<br>Leu       | TTT<br>Phe        | AGA<br>Arg<br>130     | AAT<br>Asn        | ACA<br>Thr        | TCT<br>Ser       | CTA<br>Leu        | GCT<br>Ala<br>135 | 1878 |  |            |            |            |            |            |      |
| ATA<br>Ile        | GTT<br>Val        | TTC<br>Phe        | AAT<br>Asn<br>140 | AAT<br>Asn       | AAT<br>Asn        | GAA<br>Glu        | ACT<br>Thr        | TTT<br>Phe       | CAC<br>His<br>145 | TGT<br>Cys            | TAT<br>Tyr        | ATA<br>Ile        | AGT<br>Ser       | TCA<br>Ser<br>150 | AAT<br>Asn        | 1926 |  |            |            |            |            |            |      |
| CAA<br>Gln        | AAT<br>Asn        | AGT<br>Ser        | GAT<br>Asp<br>155 | GTA<br>Val       | TTA<br>Leu        | GAT<br>Asp        | ATA<br>Ile<br>160 | GTA<br>Val       | TCA<br>Ser        | CAT<br>His            | ATA<br>Ile        | GAA<br>Glu<br>165 | TTT<br>Phe       | ATG<br>Met        | AAA<br>Lys        | 1974 |  |            |            |            |            |            |      |
| TCT<br>Ser        | AGA<br>Arg        | TAT<br>Tyr<br>170 | AAT<br>Asn        | AAA<br>Lys       | TAT<br>Tyr        | GTA<br>Val        | ATT<br>Ile<br>175 | ATA<br>Ile       | GGA<br>Gly        | GAA<br>Glu            | ATA<br>Ile        | CCC<br>Pro<br>180 | GTA<br>Val       | AAT<br>Asn        | AAT<br>Asn        | 2022 |  |            |            |            |            |            |      |
| AAT<br>Asn<br>185 | ATA<br>Ile        | TCT<br>Ser        | ATT<br>Ile        | AAT<br>Asn       | AAT<br>Asn        | ATA<br>Ile<br>190 | TTA<br>Leu        | AAT<br>Asn       | AAT<br>Asn        | TTT<br>Phe            | GCT<br>Ala<br>195 | ATT<br>Ile        | ATA<br>Ile       | ACT<br>Thr        | AAT<br>Asn        | 2070 |  |            |            |            |            |            |      |
| GTG<br>Val<br>200 | AGA<br>Arg        | TTA<br>Leu        | ATA<br>Ile        | GAT<br>Asp       | AAA<br>Lys<br>205 | TAT<br>Tyr        | AAC<br>Asn        | TCT<br>Ser       | ATA<br>Ile        | ATA<br>Ile<br>210     | TCA<br>Ser        | TTT<br>Phe        | TTA<br>Leu       | AAT<br>Asn        | ATC<br>Ile<br>215 | 2118 |  |            |            |            |            |            |      |
| AAC<br>Asn        | GTA<br>Val        | GGA<br>Gly        | ACA<br>Thr<br>220 | CTT<br>Leu       | TTT<br>Phe        | GTC<br>Val        | ATA<br>Ile        | AAT<br>Asn       | CCA<br>Pro<br>225 | TAATATTTAG TAATAATCAC |                   |                   |                  |                   |                   | 2168 |  |            |            |            |            |            |      |
| TAACATATTT        |                   |                   | TTTATTAAAA        |                  |                   | TGAATAAAAT        |                   |                  | ATATATTGTT        |                       |                   | ATTGTCAATA        |                  |                   | TTTTATATCA        |      |  | 2228       |            |            |            |            |      |
| TTTTACAGTC        |                   |                   | TTATTTTTTT        |                  |                   | TTTTTGCTTT        |                   |                  | TAGGTATAAT        |                       |                   | TTTACCTTCT        |                  |                   | AAACGTTTAT        |      |  | 2288       |            |            |            |            |      |
| CTCCCCAAAC        |                   |                   | ATCTACAGTA        |                  |                   | GATGGTTTAT        |                   |                  | TAGATTCTGT        |                       |                   | GTTATACACA        |                  |                   | TCTGCTGGAT        |      |  | 2348       |            |            |            |            |      |
| TTGCGGCATT        |                   |                   | TGTATCCAAA        |                  |                   | CCATAATATC        |                   |                  | CAGGTCTATA        |                       |                   | ATTATCTTTA        |                  |                   | AAAACCTGGG        |      |  | 2408       |            |            |            |            |      |
| ATTGAGATAC        |                   |                   | TTCTTCAGTT        |                  |                   | TTTAAATTAT        |                   |                  | TAAAATATCC        |                       |                   | AAGATTATTT        |                  |                   | TTTTTTGATG        |      |  | 2468       |            |            |            |            |      |
| AAGACATAAT        |                   |                   | TGATATTATA        |                  |                   | ATACTTTATA        |                   |                  | GAT<br>Met<br>1   |                       |                   | ATG<br>Ser        |                  |                   | TCA<br>Ile        |      |  | ATA<br>Phe | TTT<br>Ile | ATC<br>Ile | TAC<br>Tyr | TAT<br>Tyr | 2522 |
| ATT<br>Ile        | TTC<br>Phe<br>10  | AAC<br>Asn        | AAT<br>Asn        | AGA<br>Arg       | TTT<br>Phe        | TAT<br>Tyr<br>15  | ATA<br>Ile<br>15  | TAT<br>Tyr       | AAA<br>Lys        | AGA<br>Arg            | ATG<br>Met<br>20  | AAT<br>Asn<br>20  | ACT<br>Thr       | GTA<br>Val        | CAA<br>Gln        | 2570 |  |            |            |            |            |            |      |
| ATT<br>Ile        | TTA<br>Leu<br>25  | GTT<br>Val        | GTC<br>Val        | ATA<br>Ile       | TTA<br>Leu        | ATA<br>Ile<br>30  | ACA<br>Thr        | ACA<br>Thr       | GCA<br>Ala        | TTA<br>Leu            | TCT<br>Ser<br>35  | TTT<br>Phe        | CTA<br>Leu       | GTT<br>Val        | TTT<br>Phe        | 2618 |  |            |            |            |            |            |      |
| CAA<br>Gln<br>40  | TTA<br>Leu        | TGG<br>Trp        | TAT<br>Tyr        | TAT<br>Tyr       | GCC<br>Ala<br>45  | GAA<br>Glu        | AAT<br>Asn        | TAC<br>Tyr       | GAA<br>Glu        | TAT<br>Tyr<br>50      | ATA<br>Ile        | TTA<br>Leu        | AGA<br>Arg       | TAT<br>Tyr        | AAT<br>Asn<br>55  | 2666 |  |            |            |            |            |            |      |
| GAT<br>Asp        | ACA<br>Thr        | TAT<br>Tyr        | TCA<br>Ser        | AAT<br>Asn<br>60 | TTA<br>Leu        | CAA<br>Gln        | TTT<br>Phe        | GCG<br>Ala       | AGA<br>Arg<br>65  | AGC<br>Ser            | GCA<br>Ala        | AAT<br>Asn        | ATA<br>Ile       | AAT<br>Asn<br>70  | TTT<br>Phe        | 2714 |  |            |            |            |            |            |      |
| GAT<br>Asp        | GAT<br>Asp        | TTA<br>Leu        | ACT<br>Thr<br>75  | GTT<br>Val       | TTT<br>Phe        | GAT<br>Asp        | CCC<br>Pro        | AAC<br>Asn<br>80 | GAT<br>Asp        | AAT<br>Asn            | GTT<br>Val        | TTT<br>Phe        | AAT<br>Asn<br>85 | GTT<br>Val        | GAA<br>Glu        | 2762 |  |            |            |            |            |            |      |
| GAA<br>Glu        | AAA<br>Lys        | TGG<br>Trp<br>90  | CGC<br>Arg        | TGT<br>Cys       | GCT<br>Ala        | TCA<br>Ser        | ACT<br>Thr<br>95  | AAT<br>Asn       | AAT<br>Asn        | AAT<br>Asn            | ATA<br>Ile        | TTT<br>Phe<br>100 | TAT<br>Tyr       | GCA<br>Ala        | GTT<br>Val        | 2810 |  |            |            |            |            |            |      |
| TCA<br>Ser        | ACT<br>Thr<br>105 | TTT<br>Phe        | GGA<br>Gly        | TTT<br>Phe       | TTA<br>Leu        | AGT<br>Ser<br>110 | ACA<br>Thr        | GAA<br>Glu       | AGT<br>Ser        | ACT<br>Thr            | GGT<br>Gly<br>115 | ATT<br>Ile        | AAT<br>Asn       | TTA<br>Leu        | ACA<br>Thr        | 2858 |  |            |            |            |            |            |      |
| TAT<br>Tyr<br>120 | ACA<br>Thr        | AAT<br>Asn        | TCT<br>Ser        | AGA<br>Arg       | GAT<br>Asp<br>125 | TGT<br>Cys        | ATT<br>Ile        | ATA<br>Ile       | GAT<br>Asp        | TTA<br>Leu<br>130     | TTT<br>Phe        | TCT<br>Ser        | AGA<br>Arg       | ATT<br>Ile        | ATA<br>Ile<br>135 | 2906 |  |            |            |            |            |            |      |

|                                                                                                                                                       |      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| AAA ATA GTA TAT GAT CCT TGT ACT GTC GAA ACA TCT AAC GAT TGT AGA<br>Lys Ile Val Tyr Asp Pro Cys Thr Val Glu Thr Ser Asn Asp Cys Arg<br>140 150         | 2954 |
| TTA TTA AGA TTA TTG ATG GCC AAT ACA TCA TAAATACATT ATAATATTAT<br>Leu Leu Arg Leu Leu Met Ala Asn Thr Ser<br>155 160                                   | 3004 |
| TATAATATCA ATCATAATTT TTATATATAT TTTATCTAAA AGGACTTTTT ATTTTTTATA                                                                                     | 3064 |
| TATTAATAAT AATAA ATG AGT AAC GTA CCT TTA GCA ACC AAA ACA ATA AGA<br>Met Ser Asn Val Pro Leu Ala Thr Lys Thr Ile Arg<br>1 5 10                         | 3115 |
| AAA TTA TCA AAT CGA AAA TAT GAA ATA AAG ATT TAT TTA AAA GAT GAA<br>Lys Leu Ser Asn Arg Lys Tyr Glu Ile Lys Ile Tyr Leu Lys Asp Glu<br>15 20 25        | 3163 |
| AAT ACT TGT TTC GAA CGT GTA GTA GAT ATG GTA GTT CCA TTA TAT GAT<br>Asn Thr Cys Phe Glu Arg Val Val Asp Met Val Val Pro Leu Tyr Asp<br>30 35 40        | 3211 |
| GTG TGT AAT GAA ACT TCT GGT GTT ACT TTA GAA TCA TGT AGT CCA AAT<br>Val Cys Asn Glu Thr Ser Gly Val Thr Leu Glu Ser Cys Ser Pro Asn<br>45 50 55 60     | 3259 |
| ATA GAA GTA ATT GAA TTA GAC AAT ACT CAT GTT AGA ATC AAA GTT CAC<br>Ile Glu Val Ile Glu Leu Asp Asn Thr His Val Arg Ile Lys Val His<br>65 70 75        | 3307 |
| GGC GAT ACA TTA AAA GAA ATG TGT TTT GAA TTA TTG TTC CCG TGT AAT<br>Gly Asp Thr Leu Lys Glu Met Cys Phe Glu Leu Leu Phe Pro Cys Asn<br>80 85 90        | 3355 |
| GTA AAC GAA GCC CAA GTA TGG AAA TAT GTA AGT CGA TTA TTG CTA GAT<br>Val Asn Glu Ala Gln Val Trp Lys Tyr Val Ser Arg Leu Leu Asp<br>95 100 105          | 3403 |
| AAT GTA TCA CAT AAT GAC GTA AAA TAT AAA TTA GCT AAT TTT AGA CTG<br>Asn Val Ser His Asn Asp Val Lys Tyr Lys Leu Ala Asn Phe Arg Leu<br>110 115 120     | 3451 |
| ACT CTT AAT GGA AAA CAT TTA AAA TTA AAA GAA ATC GAT CAA CCG CTA<br>Thr Leu Asn Gly Lys His Leu Lys Leu Lys Glu Ile Asp Gln Pro Leu<br>125 130 135 140 | 3499 |
| TTT ATT TAT TTT GTC GAT GAT TTG GGA AAT TAT GGA TTA ATT ACT AAG<br>Phe Ile Tyr Phe Val Asp Asp Leu Gly Asn Tyr Gly Leu Ile Thr Lys<br>145 150 155     | 3547 |
| GAA AAT ATT CAA AAT AAT AAT TTA CAA GTT AAC AAA GAT GCA TCA TTT<br>Glu Asn Ile Gln Asn Asn Asn Leu Gln Val Asn Lys Asp Ala Ser Phe<br>160 165 170     | 3595 |
| ATT ACT ATA TTT CCA CAA TAT GCG TAT ATT TGT TTA GGT AGA AAA GTA<br>Ile Thr Ile Phe Pro Gln Tyr Ala Tyr Ile Cys Leu Gly Arg Lys Val<br>175 180 185     | 3643 |
| TAT TTA AAT GAA AAA GTA ACT TTT GAT GTA ACT ACA GAT GCA ACT AAT<br>Tyr Leu Asn Glu Lys Val Thr Phe Asp Val Thr Thr Asp Ala Thr Asn<br>190 195 200     | 3691 |
| ATT ACT TTA GAT TTT AAT AAA TCT GTT AAT ATC GCA GTA TCA TTC CTT<br>Ile Thr Leu Asp Phe Asn Lys Ser Val Asn Ile Ala Val Ser Phe Leu<br>205 210 215 220 | 3739 |
| GAT ATA TAT TAC GAA GTT AAT AAT AAT GAA CAA AAA GAT TTA TTA AAA<br>Asp Ile Tyr Tyr Glu Val Asn Asn Asn Glu Gln Lys Asp Leu Leu Lys<br>225 230 235     | 3787 |
| GAT TTA CTT AAG AGA TAC GGT GAA TTT GAA GTC TAT AAC GCA GAT ACT<br>Asp Leu Leu Lys Arg Tyr Gly Glu Phe Glu Val Tyr Asn Ala Asp Thr<br>240 245 250     | 3835 |
| GGA TTA ATT TAT GCT AAA AAT CTA AGT ATT AAA AAT TAT GAT ACT GTG<br>Gly Leu Ile Tyr Ala Lys Asn Leu Ser Ile Lys Asn Tyr Asp Thr Val<br>255 260 265     | 3883 |
| ATT CAA GTA GAA AGG TTG CCA GTT AAT TTG AAA GTT AGA GCA TAT ACT<br>Ile Gln Val Glu Arg Leu Pro Val Asn Leu Lys Val Arg Ala Tyr Thr<br>270 275 280     | 3931 |

63

|                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |      |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------|
| AAG<br>Lys<br>285 | GAT<br>Asp        | GAA<br>Glu        | AAT<br>Asn        | GGT<br>Gly        | CGC<br>Arg<br>290 | AAT<br>Asn        | CTA<br>Leu        | TGT<br>Cys        | TTG<br>Leu        | ATG<br>Met<br>295 | AAA<br>Lys        | ATA<br>Ile        | ACA<br>Thr        | TCT<br>Ser        | AGT<br>Ser<br>300 | 3979 |
| ACA<br>Thr        | GAA<br>Glu        | GTA<br>Val        | GAC<br>Asp        | CCC<br>Pro<br>305 | GAG<br>Glu        | TAT<br>Tyr        | GTA<br>Val        | ACT<br>Thr        | AGT<br>Ser<br>310 | AAT<br>Asn        | AAT<br>Asn        | GCT<br>Ala        | TTA<br>Leu        | TTG<br>Leu<br>315 | GGT<br>Gly        | 4027 |
| ACG<br>Thr        | CTC<br>Leu        | AGA<br>Arg        | GTA<br>Val<br>320 | TAT<br>Tyr        | AAA<br>Lys        | AAG<br>Lys        | TTT<br>Phe        | GAT<br>Asp<br>325 | AAA<br>Lys        | TCT<br>Ser        | CAT<br>His        | TTA<br>Leu        | AAA<br>Lys<br>330 | ATT<br>Ile        | GTA<br>Val        | 4075 |
| ATG<br>Met        | CAT<br>His        | AAC<br>Asn<br>335 | AGA<br>Arg        | GGA<br>Gly        | AGT<br>Ser        | GGT<br>Gly        | AAT<br>Asn<br>340 | GTA<br>Val        | TTT<br>Phe        | CCA<br>Pro        | TTA<br>Leu        | AGA<br>Arg<br>345 | TCA<br>Ser        | TTA<br>Leu        | TAT<br>Tyr        | 4123 |
| CTG<br>Leu        | GAA<br>Glu<br>350 | TTG<br>Leu        | TCT<br>Ser        | AAT<br>Asn        | GTA<br>Val<br>355 | AAA<br>Lys<br>355 | GGA<br>Gly        | TAT<br>Tyr        | CCA<br>Pro        | GTT<br>Val        | AAA<br>Lys<br>360 | GCA<br>Ala        | TCT<br>Ser        | GAT<br>Asp        | ACT<br>Thr        | 4171 |
| TCG<br>Ser<br>365 | AGA<br>Arg        | TTA<br>Leu        | GAT<br>Asp        | GTT<br>Val        | GGT<br>Gly<br>370 | ATT<br>Ile        | TAC<br>Tyr        | AAA<br>Lys        | TTA<br>Leu        | AAT<br>Asn<br>375 | AAA<br>Lys        | ATT<br>Ile        | TAT<br>Tyr        | GTA<br>Val        | GAT<br>Asp<br>380 | 4219 |
| AAC<br>Asn        | GAC<br>Asp        | GAA<br>Glu        | AAT<br>Asn        | AAA<br>Lys<br>385 | ATT<br>Ile        | ATA<br>Ile        | TTG<br>Leu        | GAA<br>Glu<br>390 | ATT<br>Ile        | GAA<br>Glu        | GCA<br>Ala        | GAA<br>Glu        | TAT<br>Tyr<br>395 | AGA<br>Arg        |                   | 4267 |
| TGC<br>Cys        | GGA<br>Gly        | AGA<br>Arg        | CAA<br>Gln<br>400 | GTA<br>Val        | TTC<br>Phe        | CAC<br>His        | GAA<br>Glu        | CGT<br>Arg<br>405 | GTA<br>Val        | AAA<br>Lys        | CTT<br>Leu        | AAT<br>Asn        | AAA<br>Lys<br>410 | CAC<br>His        | CAA<br>Gln        | 4315 |
| TGT<br>Cys        | AAA<br>Lys<br>415 | TAT<br>Tyr        | ACT<br>Thr        | CCC<br>Pro        | AAA<br>Lys        | TGT<br>Cys        | CCA<br>Pro<br>420 | TTC<br>Phe        | CAA<br>Gln        | TTT<br>Phe        | GTT<br>Val        | GTA<br>Val<br>425 | AAC<br>Asn        | AGC<br>Ser        | CCA<br>Pro        | 4363 |
| GAT<br>Asp<br>430 | ACT<br>Thr        | ACG<br>Thr        | ATT<br>Ile        | CAC<br>His        | TTA<br>Leu        | TAT<br>Tyr<br>435 | GGT<br>Gly        | ATT<br>Ile        | TCT<br>Ser        | AAT<br>Asn        | GTT<br>Val<br>440 | TGT<br>Cys        | TTA<br>Leu        | AAA<br>Lys        | CCT<br>Pro        | 4411 |
| AAA<br>Lys<br>445 | GTA<br>Val        | CCC<br>Pro        | AAA<br>Lys        | AAT<br>Asn        | TTA<br>Leu<br>450 | AGA<br>Arg        | CTT<br>Leu        | TGG<br>Trp        | GGA<br>Gly        | TGG<br>Trp<br>455 | ATT<br>Ile        | TTA<br>Leu        | GAT<br>Asp        | TGC<br>Cys        | GAT<br>Asp<br>460 | 4459 |
| ACT<br>Thr        | TCT<br>Ser        | AGA<br>Arg        | TTT<br>Phe        | ATT<br>Ile<br>465 | AAA<br>Lys        | CAT<br>His        | ATG<br>Met        | GCT<br>Ala        | GAT<br>Asp<br>470 | GGA<br>Gly        | TCT<br>Ser        | GAT<br>Asp        | GAT<br>Asp        | TTA<br>Leu<br>475 | GAT<br>Asp        | 4507 |
| CTT<br>Leu        | GAC<br>Asp        | GTT<br>Val        | AGG<br>Arg<br>480 | CTT<br>Leu        | AAT<br>Asn        | AGA<br>Arg        | AAT<br>Asn        | GAT<br>Asp<br>485 | ATA<br>Ile        | TGT<br>Cys        | TTA<br>Leu        | AAA<br>Lys        | CAA<br>Gln<br>490 | GCC<br>Ala        | ATA<br>Ile        | 4555 |
| AAA<br>Lys        | CAA<br>Gln<br>495 | CAT<br>His        | TAT<br>Tyr        | ACT<br>Thr        | AAT<br>Asn        | GTA<br>Val        | ATT<br>Ile<br>500 | ATA<br>Ile        | TTA<br>Leu        | GAG<br>Glu        | TAC<br>Tyr        | GCA<br>Ala<br>505 | AAT<br>Asn        | ACA<br>Thr        | TAT<br>Tyr        | 4603 |
| CCA<br>Pro<br>510 | AAT<br>Asn        | TGC<br>Cys        | ACA<br>Thr        | TTA<br>Leu        | TCA<br>Ser        | TTG<br>Leu<br>515 | GGT<br>Gly        | AAT<br>Asn        | AAT<br>Asn        | AGA<br>Arg        | TTT<br>Phe<br>520 | AAT<br>Asn        | AAT<br>Asn        | GTA<br>Val        | TTT<br>Phe        | 4651 |
| GAT<br>Asp<br>525 | ATG<br>Met        | AAT<br>Asn        | GAT<br>Asp        | AAC<br>Asn        | AAA<br>Lys<br>530 | ACT<br>Thr        | ATA<br>Ile        | TCT<br>Ser        | GAG<br>Glu        | TAT<br>Tyr<br>535 | ACT<br>Thr        | AAC<br>Asn        | TTT<br>Phe        | ACA<br>Thr        | AAA<br>Lys<br>540 | 4699 |
| AGT<br>Ser        | AGA<br>Arg        | CAA<br>Gln        | GAC<br>Asp        | CTT<br>Leu<br>545 | AAT<br>Asn        | AAC<br>Asn        | ATG<br>Met        | TCA<br>Ser        | TGT<br>Cys<br>550 | ATA<br>Ile        | TTA<br>Leu        | GGA<br>Gly        | ATA<br>Ile<br>555 | AAC<br>Asn        | ATA<br>Ile        | 4747 |
| GGT<br>Gly        | AAT<br>Asn        | TCC<br>Ser        | GTA<br>Val<br>560 | AAT<br>Asn        | ATT<br>Ile        | AGT<br>Ser        | AGT<br>Ser        | TTG<br>Leu<br>565 | CCT<br>Pro        | GGT<br>Gly        | TGG<br>Trp        | GTA<br>Val<br>570 | ACA<br>Thr        | CCT<br>Pro        | CAC<br>His        | 4795 |
| GAA<br>Glu        | GCT<br>Ala<br>575 | AAA<br>Lys        | ATT<br>Ile        | CTA<br>Leu        | AGA<br>Arg        | TCT<br>Ser        | GGT<br>Gly<br>580 | TGT<br>Cys        | GCT<br>Ala        | AGA<br>Arg        | GTT<br>Val        | AGA<br>Arg<br>585 | GAA<br>Glu        | TTT<br>Phe        | TGT<br>Cys        | 4843 |
| AAA<br>Lys<br>590 | TCA<br>Ser        | TTC<br>Phe        | TGT<br>Cys        | GAT<br>Asp        | CTT<br>Leu        | TCT<br>Ser<br>595 | AAT<br>Asn        | AAG<br>Lys        | AGA<br>Arg        | TTC<br>Phe        | TAT<br>Tyr<br>600 | GCT<br>Ala        | ATG<br>Met        | GCT<br>Ala        | AGA<br>Arg        | 4891 |
| GAT<br>Asp<br>605 | CTC<br>Leu        | GTA<br>Val        | AGT<br>Ser        | TTA<br>Leu        | CTA<br>Leu<br>610 | TTT<br>Phe        | ATG<br>Met        | TGT<br>Cys        | AAC<br>Asn        | TAT<br>Tyr<br>615 | GTT<br>Val        | AAT<br>Asn        | ATT<br>Ile        | GAA<br>Glu        | ATT<br>Ile<br>620 | 4939 |

64

|                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |                   |      |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------|
| AAC<br>Asn        | GAA<br>Glu        | GCA<br>Ala        | GTA<br>Val        | TGC<br>Cys<br>625 | GAA<br>Glu        | TAT<br>Tyr        | CCT<br>Pro        | GGA<br>Gly        | TAT<br>Tyr<br>630 | GTC<br>Val        | ATA<br>Ile        | TTA<br>Leu        | TTC<br>Phe<br>635 | GCA<br>Ala        | AGA<br>Arg        | 4987 |
| GCT<br>Ala        | ATT<br>Ile        | AAA<br>Lys        | GTA<br>Val<br>640 | ATT<br>Ile        | AAT<br>Asn        | GAT<br>Asp        | TTA<br>Leu        | TTA<br>Leu<br>645 | TTA<br>Leu        | ATT<br>Ile        | AAC<br>Asn        | GGA<br>Gly        | GTA<br>Val<br>650 | GAT<br>Asp        | AAT<br>Asn        | 5035 |
| CTA<br>Leu        | GCA<br>Ala        | GGA<br>Gly<br>655 | TAT<br>Tyr        | TCA<br>Ser        | ATT<br>Ile        | TCC<br>Ser        | TTA<br>Leu<br>660 | CCT<br>Pro        | ATA<br>Ile        | CAT<br>His        | TAT<br>Tyr        | GGA<br>Gly<br>665 | TCT<br>Ser        | ACT<br>Thr        | GAA<br>Glu        | 5083 |
| AAG<br>Lys        | ACT<br>Thr<br>670 | CTA<br>Leu        | CCA<br>Pro        | AAT<br>Asn        | GAA<br>Glu        | AAG<br>Lys<br>675 | TAT<br>Tyr        | GGT<br>Gly        | GGT<br>Gly        | GTT<br>Val        | GAT<br>Asp<br>680 | AAG<br>Lys        | AAA<br>Lys        | TTT<br>Phe        | AAA<br>Lys        | 5131 |
| TAT<br>Tyr<br>685 | CTA<br>Leu        | TTC<br>Phe        | TTA<br>Leu        | AAG<br>Lys        | AAT<br>Asn        | AAA<br>Lys<br>690 | CTA<br>Leu        | AAA<br>Lys        | GAT<br>Asp        | TTA<br>Leu<br>695 | ATG<br>Met        | CGT<br>Arg        | GAT<br>Asp        | GCT<br>Ala        | GAT<br>Asp<br>700 | 5179 |
| TTT<br>Phe        | GTC<br>Val        | CAA<br>Gln        | CCT<br>Pro        | CCA<br>Pro<br>705 | TTA<br>Leu        | TAT<br>Tyr        | ATT<br>Ile        | TCT<br>Ser        | ACT<br>Thr<br>710 | TAC<br>Tyr        | TTT<br>Phe        | AGA<br>Arg        | ACT<br>Thr        | TTA<br>Leu<br>715 | TTG<br>Leu        | 5227 |
| GAT<br>Asp        | GCT<br>Ala        | CCA<br>Pro        | CCA<br>Pro<br>720 | ACT<br>Thr        | GAT<br>Asp        | AAT<br>Asn        | TAT<br>Tyr        | GAA<br>Glu<br>725 | AAA<br>Lys        | TAT<br>Tyr        | TTG<br>Leu        | GTT<br>Val        | GAT<br>Asp<br>730 | TCG<br>Ser        | TCC<br>Ser        | 5275 |
| GTA<br>Val        | CAA<br>Gln        | TCA<br>Ser<br>735 | CAA<br>Gln        | GAT<br>Asp        | GTT<br>Val        | CTA<br>Leu        | CAG<br>Gln<br>740 | GGT<br>Gly        | CTG<br>Leu        | TTG<br>Leu        | AAT<br>Asn        | ACA<br>Thr<br>745 | TGT<br>Cys        | AAT<br>Asn        | ACT<br>Thr        | 5323 |
| ATT<br>Ile        | GAT<br>Asp<br>750 | ACT<br>Thr        | AAT<br>Asn        | GCT<br>Ala        | AGA<br>Arg        | GTT<br>Val<br>755 | GCA<br>Ala        | TCA<br>Ser        | AGT<br>Ser        | GTT<br>Val        | ATT<br>Ile<br>760 | GGA<br>Gly        | TAT<br>Tyr        | GTT<br>Val        | TAT<br>Tyr        | 5371 |
| GAA<br>Glu<br>765 | CCA<br>Pro        | TGC<br>Cys        | GGA<br>Gly        | ACA<br>Thr<br>770 | TCA<br>Ser        | GAA<br>Glu        | CAT<br>His        | AAA<br>Lys        | ATT<br>Ile        | GGT<br>Gly<br>775 | TCA<br>Ser        | GAA<br>Glu        | GCA<br>Ala        | TTG<br>Leu        | TGT<br>Cys<br>780 | 5419 |
| AAA<br>Lys        | ATG<br>Met        | GCT<br>Ala        | AAA<br>Lys        | GAA<br>Glu<br>785 | GCA<br>Ala        | TCT<br>Ser        | AGA<br>Arg        | TTA<br>Leu        | GGA<br>Gly<br>790 | AAT<br>Asn        | CTA<br>Leu        | GGT<br>Gly        | TTA<br>Leu        | GTA<br>Val<br>795 | AAT<br>Asn        | 5467 |
| CGT<br>Arg        | ATT<br>Ile        | AAT<br>Asn        | GAA<br>Glu<br>800 | AGT<br>Ser        | AAT<br>Asn        | TAC<br>Tyr        | AAC<br>Asn        | AAA<br>Lys<br>805 | TGT<br>Cys        | AAT<br>Asn        | AAA<br>Lys        | TAT<br>Tyr        | GGT<br>Gly<br>810 | TAT<br>Tyr        | AGA<br>Arg        | 5515 |
| GGA<br>Gly        | GTA<br>Val        | TAC<br>Tyr<br>815 | GAA<br>Glu        | AAT<br>Asn        | AAC<br>Asn        | AAA<br>Lys        | CTA<br>Leu<br>820 | AAA<br>Lys        | ACA<br>Thr        | AAA<br>Lys        | TAT<br>Tyr        | TAT<br>Tyr<br>825 | AGA<br>Arg        | GAA<br>Glu        | ATA<br>Ile        | 5563 |
| TTT<br>Phe        | GAT<br>Asp<br>830 | TGT<br>Cys        | AAT<br>Asn        | CCT<br>Pro        | AAT<br>Asn        | AAT<br>Asn<br>835 | AAT<br>Asn        | AAT<br>Asn        | GAA<br>Glu        | TTA<br>Leu        | ATA<br>Ile<br>840 | TCC<br>Ser        | AGA<br>Arg        | TAT<br>Tyr        | GGA<br>Gly        | 5611 |
| TAT<br>Tyr<br>845 | AGA<br>Arg        | ATA<br>Ile        | ATG<br>Met        | GAT<br>Asp        | TTA<br>Leu<br>850 | CAT<br>His        | AAA<br>Lys        | ATT<br>Ile        | GGA<br>Gly        | GAA<br>Glu<br>855 | ATT<br>Ile        | TTT<br>Phe        | GCA<br>Ala        | AAT<br>Asn        | TAC<br>Tyr<br>860 | 5659 |
| GAT<br>Asp        | GAA<br>Glu        | AGT<br>Ser        | GAA<br>Glu        | TCT<br>Ser<br>865 | CCT<br>Pro        | TGC<br>Cys        | GAA<br>Glu        | CGA<br>Arg        | AGA<br>Arg<br>870 | TGT<br>Cys        | CAT<br>His        | TAC<br>Tyr        | TTG<br>Leu        | GAA<br>Glu<br>875 | GAT<br>Asp        | 5707 |
| AGA<br>Arg        | GGT<br>Gly        | CTT<br>Leu        | TTA<br>Leu<br>880 | TAT<br>Tyr        | GGT<br>Gly        | CCT<br>Pro        | GAA<br>Glu        | TAT<br>Tyr<br>885 | GTA<br>Val        | CAT<br>His        | CAC<br>His        | AGA<br>Arg        | TAT<br>Tyr<br>890 | CAA<br>Gln        | GAA<br>Glu        | 5755 |
| TCA<br>Ser        | TGT<br>Cys        | ACG<br>Thr<br>895 | CCT<br>Pro        | AAT<br>Asn        | ACG<br>Thr        | TTT<br>Phe        | GGA<br>Gly<br>900 | AAT<br>Asn        | AAC<br>Asn        | ACA<br>Thr        | AAT<br>Asn        | TGT<br>Cys<br>905 | GTA<br>Val        | ACA<br>Thr        | AGA<br>Arg        | 5803 |
| AAT<br>Asn        | GGT<br>Gly<br>910 | GAA<br>Glu        | CAA<br>Gln        | CAC<br>His        | GTA<br>Val        | TAC<br>Tyr<br>915 | GAA<br>Glu        | AAT<br>Asn        | AGT<br>Ser        | TGT<br>Cys        | GGA<br>Gly<br>920 | GAT<br>Asp        | AAT<br>Asn        | GCA<br>Ala        | ACA<br>Thr        | 5851 |
| TGT<br>Cys<br>925 | GGA<br>Gly        | AGA<br>Arg        | AGA<br>Arg        | ACA<br>Thr        | GGA<br>Gly<br>930 | TAT<br>Tyr        | GGA<br>Gly        | AGA<br>Arg        | AGA<br>Arg        | AGT<br>Ser<br>935 | AGG<br>Arg        | GAT<br>Asp        | GAA<br>Glu        | TGG<br>Trp        | AAT<br>Asn<br>940 | 5899 |
| GAC<br>Asp        | TAT<br>Tyr        | AGA<br>Arg        | AAA<br>Lys        | CCC<br>Pro<br>945 | CAC<br>His        | GTT<br>Val        | TAT<br>Tyr        | GAC<br>Asp        | AAT<br>Asn<br>950 | TGT<br>Cys        | GCC<br>Ala        | GAT<br>Asp        | GCA<br>Ala        | AAT<br>Asn<br>955 | AGT<br>Ser        | 5947 |

65

|                                                                        |      |
|------------------------------------------------------------------------|------|
| TCA TCT TCA GAT AGC TGT TCA GAC AGT AGT AGT AGT AGT GAA TCT GAA        | 5995 |
| Ser Ser Ser Asp Ser Cys Ser Asp Ser Ser Ser Ser Ser Glu Ser Glu        |      |
| 960 965 970                                                            |      |
| TCT GAT TCA GAT GGA TGT TGC GAC ACA GAT GCT AGT TTA GAT TCT GAT        | 6043 |
| Ser Asp Ser Asp Gly Cys Cys Asp Thr Asp Ala Ser Leu Asp Ser Asp        |      |
| 975 980 985                                                            |      |
| ATT GAA AAT TGT TAT CAA AAT CCA TCA AAA TGT GAT GCA GGA TGC TAAATGAAAT | 6098 |
| Ile Glu Asn Cys Tyr Gln Asn Pro Ser Lys Cys Asp Ala Gly Cys            |      |
| 990 995 1000                                                           |      |
| TTAATATTAT ATAATATTAA CTTACAAGTT ATAAAAATCA TTAAAATGAT TTTTAAAAAT      | 6158 |
| GATATTATCG ATAGTTGTGA TAATGTGCTC TTTTATTTTA TTAATTGCGA TGATTATAAT      | 6218 |
| ATTATCTTTT AGATATATTT AATATTAATT ATAAATCGAC TGACAATAAT ATTTATTCCT      | 6278 |
| ATTCATAATA ATCATCTGCT ATATATATTA ATGTATCATT CTCTATTATA AATATAGGTA      | 6338 |
| TATTGTCTTT ATCAATCATT AATTTTGCTA CAGCTGTATT ATCTTTATAT ACTATATTTG      | 6398 |
| TGTCTTTGTT TAATAAACCT TTTAATATAG TGGCTCTATC ATAATCTTTA CAATATGATA      | 6458 |
| TGGGATATAA TTTTATATTA ATAATAACAT TAGATACGTT CATTTCTTTC ATTCTAGTTT      | 6518 |
| TACGTATTGT GTCAAAAATT ATTTTCATTTT CTGCTGGTTC TATATATTTA TATGTGTTAT     | 6578 |
| GAATAGATTC GATAGATGAT GATTTTAATA AATCAAATAT AACATTTATT TTACCTTGTT      | 6638 |
| TATCTTTTAT AATATCTAAT ATTTCTTTAT CTACAGATTT TCTGTTGTTG GTATATGATA      | 6698 |
| TTAAAAAATG AACGTTAACA TATCTATATT CTTGTGGTAA ATCTTTATGA GAATTTAATC      | 6758 |
| TTATAGATCT TCCTATTATT TGTTTTAATT CTGATTCATT CCACGGCATA TCTAATATAA      | 6818 |
| TTATATCATT AATACATTTG AATGATATGC CTTCAGATCC AGCGTAAGAA AATATGCAAA      | 6878 |
| CTTTTACTTT TTTACCATTA TTATTTTCAT AATTATTATA TTCGTTTAAT TCATTATCTC      | 6938 |
| TAGTTTTTAA AGTTTGTCTA GAATATTCAA TATAAGAAAT ATTTAAACAA TTTAAATAAC      | 6998 |
| ATTTTAAACT TGATATTCCT TCAAAATTAA CTAAAGGTTT AAATATTAAT ACTTTTCCTC      | 7058 |
| TCGAATTTAA AATTATTTTA CAAGTTTCTA TATATTTACA CGAATATTGA TATAATATAT      | 7118 |
| TATAATTATT TATATCAGTG ATTGGTAAAT TAGTTTTTAT TTTTATATTA TCATTTTTAA      | 7178 |
| AACTTTCAAT AAAAGATTCA GAGAAATTAA TATTTTTTGT AAACTCGGAA AATTCAGCAA      | 7238 |
| GTTTTCTTTT AATCATATCA TTATATTCTA TATTATCTAA ATCTCCTTTT ATTTAAGAT       | 7298 |
| CATAAAAAGC AAATGAAGAT ATTAATCTTC TCATAGTTTT TAAACCACCT AATTCAGTTT      | 7358 |
| TATAATCATA TTTTCTGCC ATATTATATA ATTTAGATTG CTCATCTGAC ATAATTATAT       | 7418 |
| TATGATAAAA TATATTTTTT TTTGCATATC CATCTATATA ATTTGTTTCT GTTAAACTAT      | 7478 |
| CTGCTTCTAT TAATCTTTTA TAAGAACATA TAGCTAATAA TGTTTCTCTT AATTCCTTAA      | 7538 |
| AATTAATTAA CTTTCCATTA TTTATATATT CTTCTTTTAT ATTCATAACA TTTGGTCTAA      | 7598 |
| GTAAACCTAT TAAATTATTA AATTCAGAAA TATTATTAGT TACTGGAGTA GCGGACATAC      | 7658 |
| ATAATATTTT ATTATTTTCG AAATTGCTA ATTTTATTA TTTTATATA ATAGGAGTAA         | 7718 |
| AATTTCTTTC GTTATTATCT TTTTAAACAG TTCTTGATAT TAATTTATGA ACTTCGTCTA      | 7778 |
| TTATTATTAG TAATCTACTT TTTTATTATA GAGAACTTTC TATAGATCTA TATATATTAT      | 7838 |
| TAAATTTATC TAAACTAGAT GACGAATCAT AATATATAAA TTTTATATTA CTGGTATCTG      | 7898 |
| ATATATATGA TCTTATAGTA TTTAACCAAG GATCTATGTA TAATGATTTT TTAATAAATA      | 7958 |
| TTAAAATTAT CCATCTTGGA AATAATTCTT TTATATATTT TATAATATAC ACAGCAGTTA      | 8018 |
| ATGTTTTTCC CATACCAGTA TCCCAAAATA ATAACATACT ATTCAAATTT TTTAATCCTA      | 8078 |
| TGAATATTCT ACTTACAAAA TATTGATAAT CTTGTAATGT AATTTTCAGTA TTTGTAATAT     | 8138 |
| TATTCATAAT TTTATTAGGC AAATGTTGTG TTTTATCAAG TGCATAATTT ATATGTTTAC      | 8198 |
| CAACAATAGA ATCTAATGCA AACATTTAGT TATATAAAAA ATAATATTTA TATTAACCTA      | 8258 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| AGATGTTTCA TTAATTTTAT GTCTGTGATG TGGAGTTAAA ACCCAAGATA TTGATATATC | 8318 |
| TATATCATTATTCTTCTTT TGAATCTATG TCTATCAATC GCAAATTTAT CCCAGTATAA   | 8378 |
| TTTTCGAGTT TGTTTTGCAG CATATAACCA AACATACATA ATGTGGAGTT TTGGTGGTTC | 8438 |
| GGATGAAAAG CGTACTTTT                                              | 8457 |

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 464 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Asn | Asn | Lys | Ile | Arg | Arg | Phe | Pro | Asn | Lys | Asn | Leu | Lys | Met | Pro |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Glu | Ser | Gly | Ile | Asn | Phe | Met | Ser | Met | Leu | Phe | Phe | Ser | Lys | Ile | Asp |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Asn | Met | Val | Tyr | Phe | Ile | Asn | Pro | Ile | Lys | Tyr | Asn | Thr | Asn | Ala | Asn |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Ile | Ala | Ile | Leu | Glu | Lys | Ile | Asp | Asp | Asp | Asp | Glu | Thr | Arg | Gly | Lys |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Val | Thr | Phe | Ile | Pro | Ile | Lys | Tyr | Leu | Glu | Ile | Leu | Tyr | Asn | Glu | Leu |
|     | 65  |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| Val | Leu | Asp | Pro | Asn | His | Ile | Asn | Asn | Ile | Asn | Phe | Glu | Asn | Asn | Ile |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Lys | Arg | Lys | Phe | Phe | Leu | Phe | Trp | Thr | Ile | Lys | Lys | Tyr | Leu | Gln | Asp |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Lys | Asn | Ile | Asn | Ile | Asn | Thr | Phe | Ile | Thr | Ser | Lys | Lys | Tyr | Lys | Gly |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Ile | Pro | Leu | Val | Tyr | Met | Arg | Lys | Ser | Phe | Leu | Lys | Ser | Glu | Leu | Ser |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Lys | Thr | Arg | Asp | Phe | Ser | Thr | Phe | Ala | Thr | Ile | Tyr | Asp | Asp | Leu | Asp |
|     | 145 |     |     |     | 150 |     |     |     |     | 155 |     |     |     | 160 |     |
| Ala | Gln | Ile | Gly | Ile | Pro | Pro | Leu | Gly | Phe | Asn | Pro | Lys | Pro | Lys | Ala |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Tyr | Pro | Arg | Lys | His | Asp | Lys | Ser | Thr | Trp | Leu | Ser | Ser | Gly | Asp | Ile |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Tyr | Asn | Cys | Ile | Tyr | Pro | Leu | Thr | Met | Ile | Asn | Thr | Asp | Tyr | Asp | Tyr |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Phe | His | Leu | Ile | Leu | Phe | Glu | Lys | Thr | Asp | Lys | Asn | Ile | Ala | Thr | Val |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Ala | Ser | Ser | Met | Arg | Cys | Tyr | Lys | Leu | Glu | Asp | Arg | Val | Lys | Phe | Phe |
|     | 225 |     |     |     | 230 |     |     |     |     | 235 |     |     |     | 240 |     |
| Leu | Met | Asn | Asp | Lys | Lys | Arg | Phe | Phe | Met | Phe | Pro | Ile | Ile | Tyr | Asn |
|     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Asp | His | Phe | Thr | Cys | Cys | Val | Ile | Asp | Lys | His | Phe | Asp | Lys | Asp | Lys |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Lys | Ala | Ala | Tyr | Phe | Phe | Asn | Ser | Ser | Gly | Tyr | Ile | Pro | Glu | Leu | Ile |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Lys | Gln | Asn | Lys | Lys | Tyr | Met | Phe | Ile | Glu | Ser | Asp | Met | Thr | Ile | Lys |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Ser | His | Lys | His | Tyr | Asn | Ser | Thr | Pro | Asn | Thr | Asn | Tyr | Ala | Tyr | Leu |
|     | 305 |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Tyr | Ile | Asp | Val | Leu | Ser | Glu | Tyr | Leu | Asn | Asp | Ile | Phe | Lys | Asn | Val |
|     |     |     | 325 |     |     |     |     |     | 330 |     |     |     |     | 335 |     |

67

Asn Tyr Tyr Phe Phe Asn Thr Phe Glu Leu Gln Tyr Asp Ser Pro Asp  
                   340                  345                  350  
 Cys Gly Met Phe Asn Ile Ile Phe Leu Tyr Tyr Ile Val Tyr Phe Asn  
                   355                  360                  365  
 Ile Lys Ser Lys Phe Glu Phe Lys Lys Leu Tyr Tyr Ser Met Ser Phe  
                   370                  375                  380  
 Ile Gly Asp Leu Leu Ala Ser Ser Tyr Arg Gly Ala Leu Phe Ile Ser  
                   385                  390                  395                  400  
 Arg Tyr Asp Ile Asn Ser Ile Asp Glu Phe Lys Asn Thr Leu Glu Ile  
                   405                  410                  415  
 Phe Asn Ile Lys Asn Lys Lys Phe Met Glu Leu Ile Asp Met Tyr Lys  
                   420                  425                  430  
 Lys Asn Ser Asn Arg Ile Met Asn Val Cys Ser Lys Ile Lys Asn Asp  
                   435                  440                  445  
 Tyr Asp Ser Tyr Ile Asp Asn Glu Lys Asn Ser Leu Glu Ser Asn Ile  
                   450                  455                  460

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 225 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

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

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

```

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

```

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 161 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

```

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

```

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1003 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

```

Met  Ser  Asn  Val  Pro  Leu  Ala  Thr  Lys  Thr  Ile  Arg  Lys  Leu  Ser  Asn
 1      5      10

```

Arg Lys Tyr Glu<sub>20</sub> Ile Lys Ile Tyr Leu<sub>25</sub> Lys Asp Glu Asn Thr<sub>30</sub> Cys Phe  
 Glu Arg Val<sub>35</sub> Val Asp Met Val Val<sub>40</sub> Pro Leu Tyr Asp Val<sub>45</sub> Cys Asn Glu  
 Thr Ser<sub>50</sub> Gly Val Thr Leu Glu<sub>55</sub> Ser Cys Ser Pro Asn<sub>60</sub> Ile Glu Val Ile  
 Glu<sub>65</sub> Leu Asp Asn Thr His<sub>70</sub> Val Arg Ile Lys Val<sub>75</sub> His Gly Asp Thr Leu<sub>80</sub>  
 Lys Glu Met Cys Phe<sub>85</sub> Glu Leu Leu Phe Pro<sub>90</sub> Cys Asn Val Asn Glu<sub>95</sub> Ala  
 Gln Val Trp Lys<sub>100</sub> Tyr Val Ser Arg Leu<sub>105</sub> Leu Leu Asp Asn Val<sub>110</sub> Ser His  
 Asn Asp Val<sub>115</sub> Lys Tyr Lys Leu Ala<sub>120</sub> Asn Phe Arg Leu Thr<sub>125</sub> Leu Asn Gly  
 Lys His<sub>130</sub> Leu Lys Leu Lys Glu<sub>135</sub> Ile Asp Gln Pro Leu<sub>140</sub> Phe Ile Tyr Phe  
 Val<sub>145</sub> Asp Asp Leu Gly Asn<sub>150</sub> Tyr Gly Leu Ile Thr<sub>155</sub> Lys Glu Asn Ile Gln<sub>160</sub>  
 Asn Asn Asn Leu<sub>165</sub> Gln Val Asn Lys Asp Ala<sub>170</sub> Ser Phe Ile Thr Ile<sub>175</sub> Phe  
 Pro Gln Tyr Ala<sub>180</sub> Tyr Ile Cys Leu Gly<sub>185</sub> Arg Lys Val Tyr Leu<sub>190</sub> Asn Glu  
 Lys Val Thr<sub>195</sub> Phe Asp Val Thr Thr<sub>200</sub> Asp Ala Thr Asn Ile<sub>205</sub> Thr Leu Asp  
 Phe Asn<sub>210</sub> Lys Ser Val Asn Ile<sub>215</sub> Ala Val Ser Phe Leu<sub>220</sub> Asp Ile Tyr Tyr  
 Glu<sub>225</sub> Val Asn Asn Asn Glu<sub>230</sub> Gln Lys Asp Leu Leu<sub>235</sub> Lys Asp Leu Leu Lys<sub>240</sub>  
 Arg Tyr Gly Glu Phe<sub>245</sub> Glu Val Tyr Asn Ala<sub>250</sub> Asp Thr Gly Leu Ile<sub>255</sub> Tyr  
 Ala Lys Asn Leu<sub>260</sub> Ser Ile Lys Asn Tyr<sub>265</sub> Asp Thr Val Ile Gln Val Glu  
 Arg Leu Pro<sub>275</sub> Val Asn Leu Lys Val<sub>280</sub> Arg Ala Tyr Thr Lys<sub>285</sub> Asp Glu Asn  
 Gly Arg<sub>290</sub> Asn Leu Cys Leu Met<sub>295</sub> Lys Ile Thr Ser Ser<sub>300</sub> Thr Glu Val Asp  
 Pro Glu Tyr Val Thr Ser<sub>310</sub> Asn Asn Ala Leu Leu<sub>315</sub> Gly Thr Leu Arg Val<sub>320</sub>  
 Tyr Lys Lys Phe Asp<sub>325</sub> Lys Ser His Leu Lys<sub>330</sub> Ile Val Met His Asn<sub>335</sub> Arg  
 Gly Ser Gly Asn<sub>340</sub> Val Phe Pro Leu Arg<sub>345</sub> Ser Leu Tyr Leu Glu<sub>350</sub> Leu Ser  
 Asn Val Lys<sub>355</sub> Gly Tyr Pro Val Lys<sub>360</sub> Ala Ser Asp Thr Ser<sub>365</sub> Arg Leu Asp  
 Val Gly<sub>370</sub> Ile Tyr Lys Leu Asn<sub>375</sub> Lys Ile Tyr Val Asp<sub>380</sub> Asn Asp Glu Asn  
 Lys<sub>385</sub> Ile Ile Leu Glu Glu<sub>390</sub> Ile Glu Ala Glu Tyr<sub>395</sub> Arg Cys Gly Arg Gln<sub>400</sub>  
 Val Phe His Glu Arg<sub>405</sub> Val Lys Leu Asn Lys<sub>410</sub> His Gln Cys Lys Tyr<sub>415</sub> Thr  
 Pro Lys Cys Pro<sub>420</sub> Phe Gln Phe Val Val<sub>425</sub> Asn Ser Pro Asp Thr<sub>430</sub> Thr Ile  
 His Leu Tyr<sub>435</sub> Gly Ile Ser Asn Val<sub>440</sub> Cys Leu Lys Pro Lys<sub>445</sub> Val Pro Lys  
 Asn Leu Arg Leu Trp Gly Trp<sub>455</sub> Ile Leu Asp Cys Asp<sub>460</sub> Thr Ser Arg Phe

Ile Lys His Met Ala Asp Gly Ser Asp Asp Leu Asp Leu Asp Val Arg  
 465 470 475  
 Leu Asn Arg Asn Asp Ile Cys Leu Lys Gln Ala Ile Lys Gln His Tyr  
 485 490 495  
 Thr Asn Val Ile Ile Leu Glu Tyr Ala Asn Thr Tyr Pro Asn Cys Thr  
 500 505 510  
 Leu Ser Leu Gly Asn Asn Arg Phe Asn Asn Val Phe Asp Met Asn Asp  
 515 520 525  
 Asn Lys Thr Ile Ser Glu Tyr Thr Asn Phe Thr Lys Ser Arg Gln Asp  
 530 535 540  
 Leu Asn Asn Met Ser Cys Ile Leu Gly Ile Asn Ile Gly Asn Ser Val  
 545 550 555  
 Asn Ile Ser Ser Leu Pro Gly Trp Val Thr Pro His Glu Ala Lys Ile  
 565 570 575  
 Leu Arg Ser Gly Cys Ala Arg Val Arg Glu Phe Cys Lys Ser Phe Cys  
 580 585 590  
 Asp Leu Ser Asn Lys Arg Phe Tyr Ala Met Ala Arg Asp Leu Val Ser  
 595 600 605  
 Leu Leu Phe Met Cys Asn Tyr Val Asn Ile Glu Ile Asn Glu Ala Val  
 610 615 620  
 Cys Glu Tyr Pro Gly Tyr Val Ile Leu Phe Ala Arg Ala Ile Lys Val  
 625 630 635  
 Ile Asn Asp Leu Leu Leu Ile Asn Gly Val Asp Asn Leu Ala Gly Tyr  
 645 650 655  
 Ser Ile Ser Leu Pro Ile His Tyr Gly Ser Thr Glu Lys Thr Leu Pro  
 660 665 670  
 Asn Glu Lys Tyr Gly Gly Val Asp Lys Lys Phe Lys Tyr Leu Phe Leu  
 675 680 685  
 Lys Asn Lys Leu Lys Asp Leu Met Arg Asp Ala Asp Phe Val Gln Pro  
 690 695 700  
 Pro Leu Tyr Ile Ser Thr Tyr Phe Arg Thr Leu Leu Asp Ala Pro Pro  
 705 710 715 720  
 Thr Asp Asn Tyr Glu Lys Tyr Leu Val Asp Ser Ser Val Gln Ser Gln  
 725 730 735  
 Asp Val Leu Gln Gly Leu Leu Asn Thr Cys Asn Thr Ile Asp Thr Asn  
 740 745 750  
 Ala Arg Val Ala Ser Ser Val Ile Gly Tyr Val Tyr Glu Pro Cys Gly  
 755 760 765  
 Thr Ser Glu His Lys Ile Gly Ser Glu Ala Leu Cys Lys Met Ala Lys  
 770 775 780  
 Glu Ala Ser Arg Leu Gly Asn Leu Gly Leu Val Asn Arg Ile Asn Glu  
 785 790 795 800  
 Ser Asn Tyr Asn Lys Cys Asn Lys Tyr Gly Tyr Arg Gly Val Tyr Glu  
 805 810 815  
 Asn Asn Lys Leu Lys Thr Lys Tyr Tyr Arg Glu Ile Phe Asp Cys Asn  
 820 825 830  
 Pro Asn Asn Asn Asn Glu Leu Ile Ser Arg Tyr Gly Tyr Arg Ile Met  
 835 840 845  
 Asp Leu His Lys Ile Gly Glu Ile Phe Ala Asn Tyr Asp Glu Ser Glu  
 850 855 860  
 Ser Pro Cys Glu Arg Arg Cys His Tyr Leu Glu Asp Arg Gly Leu Leu  
 865 870 875 880  
 Tyr Gly Pro Glu Tyr Val His His Arg Tyr Gln Glu Ser Cys Thr Pro  
 885 890 895  
 Asn Thr Phe Gly Asn Asn Thr Asn Cys Val Thr Arg Asn Gly Glu Gln  
 900 905 910

71

His Val Tyr Glu Asn Ser Cys Gly Asp Asn Ala Thr Cys Gly Arg Arg  
 915 920 925  
 Thr Gly Tyr Gly Arg Arg Ser Arg Asp Glu Trp Asn Asp Tyr Arg Lys  
 930 935 940  
 Pro His Val Tyr Asp Asn Cys Ala Asp Ala Asn Ser Ser Ser Ser Asp  
 945 950 955  
 Ser Cys Ser Asp Ser Ser Ser Ser Glu Ser Glu Ser Asp Ser Asp  
 965 970 975  
 Gly Cys Cys Asp Thr Asp Ala Ser Leu Asp Ser Asp Ile Glu Asn Cys  
 980 985 990  
 Tyr Gln Asn Pro Ser Lys Cys Asp Ala Gly Cys  
 995 1000

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 163 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

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

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1511 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Amsacta moorei entemopoxvirus

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: complement (18..218)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: complement (234..782)

## (ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 852..1511

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

|                                                                   |      |
|-------------------------------------------------------------------|------|
| GAATTC AAGT TAAATATTTA TAAACAACAA TCATATTTT TTAAGAATC TAATAAATTT  | 60   |
| TTTAACATTT TATTATTATT TGATAATTGT TTATTTAATT CGTTATTGAT ATTAACAATA | 120  |
| TTATTTATCA TTTTACCTAT TTTTTTTTTT CTATCTACTA ACGAAATATC AGATTTTGCA | 180  |
| CCTTCAATAT CAGAATAATA ATTATCATTA TTTTGCATTT ATGAATAAAA ATATTAATAT | 240  |
| GAATTATTAT AACATAATCT ACACACAGGA ACATATAAAT CTTGTCCACC TATTTCAATT | 300  |
| ATTTGATTTT TATTATGTTT TTTAATTGTA AAAGAAGCAT CTTTATAACA AAATTGACAT | 360  |
| ATAGCTTGTA ATTTTTTTTAT TTTTCTACT TTAGGAATTA ATTTTGATAT AGAATTAAAT | 420  |
| ATATTTCTGT TAAAGTCACA ATTTAATCCA GCAACAATAA CTTTTTTTTT ATTATTAGCC | 480  |
| ATTTTATCAC AAAATTGTTT TAAATCATTT TCTTCAAAAA ATTGACACTC ATCTATGCCA | 540  |
| ATAATATCAT AATTATCTAC GATATTGATT TCATTAATTA AATTATTTGT TTTAATGTAT | 600  |
| AAATATTCTT TATTTAATAT ATTTCCGTCA TGATTTATTA TATTTTTATT TATAAATCTA | 660  |
| TTATCTATAT TATGAGTTAT AATTACACAT TTTTGATTAG ATAAAATATA TCTATTAATT | 720  |
| TTTCGCATCA ATTCTGTTGT TTTGCCAGAA AACATAGGAC CAATTATTAA TTCTATCGAC | 780  |
| ATTTTTTTTT ATTATTTGAT ATATTTTTTC AAAAAAAAT TAATCAATGA AAAAAAATA   | 840  |
| AAATTATCAA A ATG GAT TTA CTA AAT TCT GAT ATA ATT TTA ATA AAT ATT  | 890  |
| Met Asp Leu Leu Asn Ser Asp Ile Ile Leu Ile Asn Ile               |      |
| 1 5 10                                                            |      |
| TTA AAA TAT TAT AAT TTA AAA AAA ATA ATA ATA AAC AGA GAT AAT GTT   | 938  |
| Leu Lys Tyr Tyr Asn Leu Lys Lys Ile Ile Ile Asn Arg Asp Asn Val   |      |
| 15 20 25                                                          |      |
| ATT AAT ATT AAT ATA TTA AAA AAA TTA GTT AAT TTA GAA GAA TTG CAT   | 986  |
| Ile Asn Ile Asn Ile Leu Lys Lys Leu Val Asn Leu Glu Glu Leu His   |      |
| 30 35 40 45                                                       |      |
| ATA ATA TAT TAT GAT AAT AAT ATT TTA AAT AAT ATT CCA GAA AAT ATT   | 1034 |
| Ile Ile Tyr Tyr Asp Asn Asn Ile Leu Asn Asn Ile Pro Glu Asn Ile   |      |
| 50 55 60                                                          |      |
| AAA AGT TTA TAT ATT TCA AAT TTA AAT ATT ATT AAT TTA AAT TTT ATA   | 1082 |
| Lys Ser Leu Tyr Ile Ser Asn Leu Asn Ile Ile Asn Leu Asn Phe Ile   |      |
| 65 70 75                                                          |      |
| ACA AAA TTA AAA AAT ATA ACA TAT TTA GAT ATA TCT TAT AAC AAA AAT   | 1130 |
| Thr Lys Leu Lys Asn Ile Thr Tyr Leu Asp Ile Ser Tyr Asn Lys Asn   |      |
| 80 85 90                                                          |      |
| AGC AAT ATA AGT AAT ATT ATA CTA CCA CAT TCT ATA GAA TTT TTA AAT   | 1178 |
| Ser Asn Ile Ser Asn Ile Ile Leu Pro His Ser Ile Glu Phe Leu Asn   |      |
| 95 100 105                                                        |      |
| TGT GAA TCA TGT AAT ATA AAT GAC TAT AAT TTT ATT AAT AAT TTA GTA   | 1226 |
| Cys Glu Ser Cys Asn Ile Asn Asp Tyr Asn Phe Ile Asn Asn Leu Val   |      |
| 110 115 120 125                                                   |      |
| AAT TTA AAA AAA TTA ATA ATA TCT AAA AAT AAA TTT GGT AAC TTT AAT   | 1274 |
| Asn Leu Lys Lys Leu Ile Ile Ser Lys Asn Lys Phe Gly Asn Phe Asn   |      |
| 130 135 140                                                       |      |
| AAT GTT TTT CCT ATT AGT ATA GTT GAG TTA AAT ATG GAA TCA ATA CAA   | 1322 |
| Asn Val Phe Pro Ile Ser Ile Val Glu Leu Asn Met Glu Ser Ile Gln   |      |
| 145 150 155                                                       |      |
| ATA AAA GAT TAT AAA TTT ATA GAA AAA TTA ATT AAT TTA AAA AAA TTA   | 1370 |
| Ile Lys Asp Tyr Lys Phe Ile Glu Lys Leu Ile Asn Leu Lys Lys Leu   |      |
| 160 165 170                                                       |      |
| GAT ATA TCT TTC AAT GTT AAA AAA AAT AAT ATA CAT TTG ATA AAA TTT   | 1418 |
| Asp Ile Ser Phe Asn Val Lys Lys Asn Asn Ile His Leu Ile Lys Phe   |      |
| 175 180 185                                                       |      |

|                                                                 |      |
|-----------------------------------------------------------------|------|
| CCA AAA AGT ATA ACT CAT TTA TGT GAT TAT CAA TCA TAT AAA GAA AAT | 1466 |
| Pro Lys Ser Ile Thr His Leu Cys Asp Tyr Gln Ser Tyr Lys Glu Asn |      |
| 190 195 200 205                                                 |      |
| TAT AAT TAT TTA AAA AAT TTA TCA AAT ATA ATT GAA TAT GAA TTC     | 1511 |
| Tyr Asn Tyr Leu Lys Asn Leu Ser Asn Ile Ile Glu Tyr Glu Phe     |      |
| 210 215 220                                                     |      |

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 66 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

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

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 182 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

Met Ser Ile Glu Leu Ile Ile Gly Pro Met Phe Ser Gly Lys Thr Thr  
 1 5 10 15  
 Glu Leu Met Arg Lys Ile Asn Arg Tyr Ile Leu Ser Asn Gln Lys Cys  
 20 25 30  
 Val Ile Ile Thr His Asn Ile Asp Asn Arg Phe Ile Asn Lys Asn Ile  
 35 40 45  
 Ile Asn His Asp Gly Asn Ile Leu Asn Lys Glu Tyr Leu Tyr Ile Lys  
 50 55 60  
 Thr Asn Asn Leu Ile Asn Glu Ile Asn Ile Val Asp Asn Tyr Asp Ile  
 65 70 75 80  
 Ile Gly Ile Asp Glu Cys Gln Phe Phe Glu Glu Asn Asp Leu Glu Gln  
 85 90 95  
 Phe Cys Asp Lys Met Ala Asn Asn Lys Lys Lys Val Ile Val Ala Gly  
 100 105 110  
 Leu Asn Cys Asp Phe Asn Arg Asn Ile Phe Asn Ser Ile Ser Lys Leu  
 115 120 125  
 Ile Pro Lys Val Glu Lys Ile Lys Lys Leu Gln Ala Ile Cys Gln Phe  
 130 135 140  
 Cys Tyr Lys Asp Ala Ser Phe Thr Ile Lys Lys His Asn Lys Asn Gln  
 145 150 155 160  
 Ile Ile Glu Ile Gly Gly Gln Asp Leu Tyr Val Pro Val Cys Arg Leu  
 165 170 175  
 Cys Tyr Asn Asn Ser Tyr  
 180

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 220 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

```

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

```

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: DNA (genomic)

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

GAAGTNGATC CNGAATATGT

20

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: DNA (genomic)

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

TTTCAAATTA ACTGGCAACC

20

## (2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: DNA (genomic)

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

GGGATGGATT TTAGATTGCG

20

## (2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: DNA (genomic)

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

GCCTGGTTGG GTAACACCTC

20

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: DNA (genomic)

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

CTGCTAGATT ATCTACTCCG

20

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 35 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: DNA (genomic)

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

GTTTCGAAACA AGTATTTTCA TCTTTTAAAT AAATC

35

## (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: DNA (genomic)

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

GAYGARGGRG GRCARTTYTT

20

## (2) INFORMATION FOR SEQ ID NO:19:

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

- (ii) MOLECULE TYPE: DNA (genomic)

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

GGNCCCATGT TYTCNGG

17

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

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

GGTGCAAAAT CTGATATTTTC

20

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3012 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

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

|                                                                    |      |
|--------------------------------------------------------------------|------|
| ATGAGTAACG TACCTTTAGC AACCAAAACA ATAAGAAAAT TATCAAATCG AAAATATGAA  | 60   |
| ATAAGATTT ATTTAAAAGA TGAAATACT TGTTCGAAC GTGTAGTAGA TATGGTAGTT     | 120  |
| CCATTATATG ATGTGTGTAA TGAACTTCT GGTGTTACTT TAGAATCATG TAGTCCAAAT   | 180  |
| ATAGAAGTAA TTGAATTAGA CAATACTCAT GTTAGAATCA AAGTTCACGG CGATACATTA  | 240  |
| AAAGAAATGT GTTTTGAATT ATTGTTCCCG TGTAATGTAA ACGAAGCCCA AGTATGGAAA  | 300  |
| TATGTAAGTC GATTATTGCT AGATAATGTA TCACATAATG ACGTAAAATA TAAATTAGCT  | 360  |
| AATTTTAGAC TGA CTCTTAA TGGAAACAT TTAAATTAA AAGAAATCGA TCAACCGCTA   | 420  |
| TTTATTTATT TTGTCGATGA TTTGGGAAAT TATGGATTAA TTACTAAGGA AAATATTCAA  | 480  |
| AATAATAATT TACAAGTTAA CAAAGATGCA TCATTTATTA CTATATTTCC ACAATATGCG  | 540  |
| TATATTTGTT TAGGTAGAAA AGTATATTTA AATGAAAAG TAACTTTTGA TGTA ACTACA  | 600  |
| GATGCAACTA ATATTACTTT AGATTTTAAT AAATCTGTTA ATATCGCAGT ATCATTCCTT  | 660  |
| GATATATATT ACGAAGTTAA TAATAATGAA CAAAAGATT TATTAAAAGA TTTACTTAAG   | 720  |
| AGATACGGTG AATTTGAAGT CTATAACGCA GATACTGGAT TAATTTATGC TAAAAATCTA  | 780  |
| AGTATTAAAA ATTATGATAC TGTGATTCAA GTAGAAAGGT TGCCAGTTAA TTTGAAAGTT  | 840  |
| AGAGCATATA CTAAGGATGA AAATGGTCGC AATCTATGTT TGATGAAAAT AACATCTAGT  | 900  |
| ACAGAAGTAG ACCCCGAGTA TGTA ACTAGT AATAATGCTT TATTGGGTAC GCTCAGAGTA | 960  |
| TATAAAAAGT TTGATAAATC TCATTTAAAA ATTGTAATGC ATAACAGAGG AAGTGGTAAT  | 1020 |
| GTATTTCCAT TAAGATCATT ATATCTGGAA TTGTCTAATG TAAAAGGATA TCCAGTTAAA  | 1080 |
| GCATCTGATA CTTGAGATT AGATGTTGGT ATTTACAAAT TAAATAAAAT TTATGTAGAT   | 1140 |
| AACGACGAAA ATAAAATTAT ATTGGAAGAA ATTGAAGCAG AATATAGATG CGGAAGACAA  | 1200 |
| GTATTCACG AACGTGTAAA ACTTAATAAA CACCAATGTA AATATACTCC CAAATGTCCA   | 1260 |
| TTCCAATTTG TTGTAAACAG CCCAGATACT ACGATTCACT TATATGGTAT TTCTAATGTT  | 1320 |
| TGTTTTAAAC CTAAAGTACC CAAAAATTTA AGACTTTGGG GATGGATTTT AGATTGCGAT  | 1380 |
| ACTTCTAGAT TTATTAAACA TATGGCTGAT GGATCTGATG ATTTAGATCT TGACGTTAGG  | 1440 |
| CTTAATAGAA ATGATATATG TTTAAAACAA GCCATAAAAC AACATTATAC TAATGTAATT  | 1500 |
| ATATTAGAGT ACGCAAATAC ATATCCAAAT TGCACATTAT CATTGGGTAA TAATAGATTT  | 1560 |

|            |            |            |            |            |             |      |
|------------|------------|------------|------------|------------|-------------|------|
| AATAATGTAT | TTGATATGAA | TGATAACAAA | ACTATATCTG | AGTATACTAA | CTTTACAAAA  | 1620 |
| AGTAGACAAG | ACCTTAATAA | CATGTCATGT | ATATTAGGAA | TAAACATAGG | TAATTCCGTA  | 1680 |
| AATATTAGTA | GTTTGCCTGG | TTGGGTAACA | CCTCACGAAG | CTAAAATTCT | AAGATCTGGT  | 1740 |
| TGTGCTAGAG | TTAGAGAATT | TTGTAAATCA | TTCTGTGATC | TTTCTAATAA | GAGATTCTAT  | 1800 |
| GCTATGGCTA | GAGATCTCGT | AAGTTTACTA | TTTATGTGTA | ACTATGTTAA | TATTGAAATT  | 1860 |
| AACGAAGCAG | TATGCGAATA | TCCTGGATAT | GTCATATTAT | TCGCAAGAGC | TATTAAAGTA  | 1920 |
| ATTAATGATT | TATTATTAAT | TAACGGAGTA | GATAATCTAG | CAGGATATTC | AATTTCTCTA  | 1980 |
| CCTATACATT | ATGGATCTAC | TGAAAAGACT | CTACCAAATG | AAAAGTATGG | TGGTGTGAT   | 2040 |
| AAGAAATTTA | AATATCTATT | CTTAAAGAAT | AAACTAAAAG | ATTTAATGCG | TGATGCTGAT  | 2100 |
| TTTGTCCAAC | CTCCATTATA | TATTTCTACT | TACTTTAGAA | CTTTATTGGA | TGCTCCACCA  | 2160 |
| ACTGATAATT | ATGAAAAATA | TTTGGTTGAT | TCGTCCGTAC | AATCACAAGA | TGTTCTACAG  | 2220 |
| GGTCTGTTGA | ATACATGTAA | TACTATTGAT | ACTAATGCTA | GAGTTGCATC | AAGTGTATT   | 2280 |
| GGATATGTTT | ATGAACCATG | CGGAACATCA | GAACATAAAA | TTGGTTCAGA | AGCATTGTGT  | 2340 |
| AAAATGGCTA | AAGAAGCATC | TAGATTAGGA | AATCTAGGTT | TAGTAAATCG | TATTAATGAA  | 2400 |
| AGTAATTACA | ACAAATGTAA | TAAATATGGT | TATAGAGGAG | TATACGAAAA | TAACAACTA   | 2460 |
| AAAACAAAAT | ATTATAGAGA | AATATTTGAT | TGTAATCCTA | ATAATAATAA | TGAATTAATA  | 2520 |
| TCCAGATATG | GATATAGAAT | AATGGATTTA | CATAAAATTG | GAGAAATTTT | TGCAAAATTAC | 2580 |
| GATGAAAGTG | AATCTCCTTG | CGAACGAAGA | TGTCATTACT | TGGAAGATAG | AGGTCTTTTA  | 2640 |
| TATGGTCCTG | AATATGTACA | TCACAGATAT | CAAGAATCAT | GTACGCCTAA | TACGTTTGGA  | 2700 |
| AATAACACAA | ATTGTGTAA  | AAGAAATGGT | GAACAACACG | TATACGAAAA | TAGTTGTGGA  | 2760 |
| GATAATGCAA | CATGTGGAAG | AAGAACAGGA | TATGGAAGAA | GAAGTAGGGA | TGAATGGAAT  | 2820 |
| GACTATAGAA | AACCCACG   | TTATGACAAT | TGTGCCGATG | CAAATAGTTC | ATCTTCAGAT  | 2880 |
| AGCTGTTT   | CAGTAGTAG  | TAGTAGTGAA | TCTGAATCTG | ATTCAGATGG | ATGTTGCGAC  | 2940 |
| ACAGATGCTA | GTTTAGATT  | TGATATTGAA | AATTGTTATC | AAAATCCATC | AAAATGTGAT  | 3000 |
| GCAGGATGCT | AA         |            |            |            |             | 3012 |

## (2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 419 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: DNA (genomic)

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

|            |            |            |            |             |            |     |
|------------|------------|------------|------------|-------------|------------|-----|
| TCAACTAATA | ATAATATATT | TTATGCAGTT | TCAACTTTTG | GATTTTTTAAG | TACAGAAAGT | 60  |
| ACTGGTATTA | ATTTAACATA | TACAAATCT  | AGAGATTGTA | TTATAGATT   | ATTTTCTAGA | 120 |
| ATTATAAAAA | TAGTATATGA | TCCTTGTA   | CTCGAAACAT | CTAACGATTG  | TAGATTATTA | 180 |
| AGATTATTGA | TGGCCAATAC | ATCATAAATA | CATTATAATA | TTATTATAAT  | ATCAATCATA | 240 |
| ATTTTTATAT | ATATTTTATC | TAAAAGGACT | TTTTATTTTT | TATATATTAA  | TAATAATAAA | 300 |
| TGAGTAACGT | ACCTTTAGCA | ACCAAAACAA | TAAGAAAATT | ATCAAATCGA  | AAATATGAAA | 360 |
| TAAAGATTTA | TTTAAAAGAT | GAAAATACTT | GTTTCGAACG | TGTAGTAGAT  | ATGGTAGTT  | 419 |

## (2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 678 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| ATGTTCTAG TTTATTTCAA TACATTTTAA ATAATAATTT TATTATTTGG TATTATAGGT   | 60  |
| ATTTATATAT TAACATTTGT GTTAAATATA GATTTTTTTAA TAAATAATAA TAAAATATAT | 120 |
| ATATTATCAT ATAACGCAAC TAATATAAAC AATATAAATA ATTTAAATTT ATACGATTAT  | 180 |
| TCAGATATTA TATTTTTGAC AAATTTTAAAC ATAAATAATA ATCTTTTAGT AACACAAGCT | 240 |
| AATAATTTAC AAGATATACC AATATTTAAT GTAAATAATA TTATATCTAA TCAATATAAT  | 300 |
| TTTTATTCAG CGTCTAGTAA TAATGTAAAT ATATTATTAG GATTAAGAAA AACATTAAAT  | 360 |
| ATAAATAGAA ATCCATTTTT ATTATTTAGA AATACATCTC TAGCTATAGT TTTCAATAAT  | 420 |
| AATGAACTT TTCACTGTTA TATAAGTTCA AATCAAAATA GTGATGTATT AGATATAGTA   | 480 |
| TCACATATAG AATTTATGAA ATCTAGATAT AATAAATATG TAATTATAGG AGAAATACCC  | 540 |
| GTAAATAATA ATATATCTAT TAATAATATA TTAAATAATT TTGCTATTAT AACTAATGTG  | 600 |
| AGATTAATAG ATAAATATAA CTCTATAATA TCATTTTTTAA ATATCAACGT AGGAACACTT | 660 |
| TTTGTCAATA ATCCATAA                                                | 678 |

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 486 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| ATGTCAATAT TTATCTACTA TATTTTCAAC AATAGATTTT ATATATATAA AAGAATGAAT | 60  |
| ACTGTACAAA TTTTAGTTGT CATATTAATA ACAACAGCAT TATCTTTTCT AGTTTTTCAA | 120 |
| TTATGGTATT ATGCCGAAA TTACGAATAT ATATTAAGAT ATAATGATAC ATATTCAAAT  | 180 |
| TTACAATTTG CGAGAAGCGC AAATATAAAT TTTGATGATT TAACTGTTTT TGATCCCAAC | 240 |
| GATAATGTTT TTAATGTTGA AGAAAAATGG CGCTGTGCTT CAACTAATAA TAATATATTT | 300 |
| TATGCAGTTT CAACTTTTGG ATTTTAAAGT ACAGAAAGTA CTGGTATTAA TTTAACATAT | 360 |
| ACAAATTCTA GAGATTGTAT TATAGATTTA TTTTCTAGAA TTATAAAAAT AGTATATGAT | 420 |
| CCTTGTAATG TCGAAACATC TAACGATTGT AGATTATTAA GATTATTGAT GGCCAATACA | 480 |
| TCATAA                                                            | 486 |

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1395 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| TTAAATATTA GATTCTAAAC TATTCTTCTC ATTATCAATA TAACTATCAT AATCATTTTT | 60  |
| TATTTTACTA CATAATTCA TAATTCTATT ACTATTTTTT TTATACATAT CTATTAATTC  | 120 |
| CATAAACTTT TTATTTTTTA TATTAAATAT TTCTAATGTA TTTTAAATT CGTCAATACT  | 180 |
| ATTAATATCA TATCTAGAAA TAAATAATGC ACCTCTATAA CTACTAGCCA ATAAATCACC | 240 |
| AATAAACTC ATAGAATAAT ATAATTTTTT AAATTCAAAT TTAGATTTTA TGTTGAAATA  | 300 |

|             |            |            |            |            |            |      |
|-------------|------------|------------|------------|------------|------------|------|
| AACTATATAA  | TATAAAAATA | TTATATTAAA | CATACCACAA | TCGGGACTAT | CATATTGTAA | 360  |
| TTCAAAAAGTA | TTAAAAAAGT | AATAATTTAC | ATTTTAAAT  | ATATCATTTA | AATATTCTGA | 420  |
| TAGTACATCA  | ATGTATAAAT | AAGCATAATT | AGTATTAGGA | GTACTATTGT | AGTGTTTATG | 480  |
| GCTTTTATA   | GTCATATCAG | ATTCAATAAA | CATATATTTT | TTATTTTGTT | TTATAAGTTC | 540  |
| TGGTATATAA  | CCACTACTAT | TAAAAAAGTA | TGCAGCTTTT | TTATCTTTAT | CAAAGTGTTT | 600  |
| ATCTATTACG  | CAACAAGTAA | AATGATCATT | ATAAATTATA | GGAAACATAA | AAAATCTTTT | 660  |
| TTTATCATTC  | ATTAAAAAAA | ATTTTACTCT | ATCTTCAAGT | TTATAGCATC | TCATAGATGA | 720  |
| AGCTACTGTA  | GCAATATTTT | TATCAGTTTT | TTCAAATAAA | ATCAAATGAA | AATAATCATA | 780  |
| ATCTGTATTA  | ATCATAGTTA | ATGGATATAT | ACAATTATAT | ATATCTCCCG | AACTTAACCA | 840  |
| TGTAGATTTA  | TCATGTTTTT | TTGGGTAAGC | TTTAGGTTTA | GGATTAAATC | CCAAAGGCCG | 900  |
| TATTCCTATT  | TGAGCATCCA | AATCATCATA | AATTGTGGCA | AATGTAGAAA | AATCTCTTGT | 960  |
| TTTGGATAAT  | TCTGATTTTA | GAAAAGACTT | TCTCATATAT | ACTAATGGAA | TGCCTTTATA | 1020 |
| TTTTTTAGAT  | GTAATAAAAG | TATTAATATT | TATATTTTTA | TCTTGTAAT  | ATTTTTTTAT | 1080 |
| AGTCCAAAAT  | AGAAAAAATT | TTCTTTTAAT | ATTATTTTCA | AAATTAATAT | TATTAATATG | 1140 |
| ATTTGGATCT  | AAAATAATT  | CATTATATAA | TATTTCCAAG | TATTTTATAG | GTATAAATGT | 1200 |
| TACTTTACCT  | CTTGTTTCAT | CATCATCATC | TATTTTTTCT | AATATAGCTA | TATTTGCATT | 1260 |
| AGTATTATAT  | TTAATAGGAT | TTATAAAATA | TACCATATTA | TCTATTTTAC | TAAAAAATAA | 1320 |
| CATAGACATA  | AAATTAATAC | CAGATTCTGG | CATTTTTTAA | TTTTTATTG  | GAAATCTTCT | 1380 |
| AATTTTATTA  | TTCAT      |            |            |            |            | 1395 |

## (2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 237 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| TTATTTTTTT | TTTTTGCTTT | TAGGTATAAT | TTTACCTTCT | AAACGTTTAT | CTCCCCAAAC | 60  |
| ATCTACAGTA | GATGGTTTAT | TAGATTCTGT | GTTATACACA | TCTGCTGGAT | TTGCGGCATT | 120 |
| TGTATCCAAA | CCATAATATC | CAGGTCTATA | ATTATCTTTA | AAACTTGGG  | ATTGAGATAC | 180 |
| TTCTTCAGTT | TTTAAATTAT | TAAAATATCC | AAGATTATTT | TTTTTTGATG | AAGACAT    | 237 |

## (2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 492 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

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

|            |            |             |            |            |            |     |
|------------|------------|-------------|------------|------------|------------|-----|
| CTATTCATAA | TAATCATCTG | CTATATATAT  | TAATGTATCA | TTCTCTATTA | TAAATATAGG | 60  |
| TATATTGTCT | TTATCAATCA | TTAATTTTGC  | TACAGCTGTA | TTATCTTTAT | ATACTATATT | 120 |
| TGTGTCTTTG | TTTAATAAAC | CTTTTAATAT  | AGTGGCTCTA | TCATAATCTT | TACAATATGA | 180 |
| TATGGGATAT | AATTTTATAT | TAATAATAAC  | ATTAGATACG | TTCATTTCTT | TCATTCTAGT | 240 |
| TTTACGTATT | GTGTCAAAAA | TTATTTTCATT | TTCTGCTGGT | TCTATATATT | TATATGTGTT | 300 |
| ATGAATAGAT | TCGATAGATG | ATGATTTTAA  | TAAATCAAAT | ATAACATTTA | TTTTACCTTG | 360 |

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| TTTATCTTTT ATAATATCTA ATATTTCTTT ATCTACAGAT TTTCTGTTGT TGGTATATGA  | 420 |
| TATTAACAAAA TGAACGTTAA CATATCTATA TTCTTGTGGT AAATCTTTAT GAGAATTTAA | 480 |
| TCTTATAGAT CT                                                      | 492 |

## (2) INFORMATION FOR SEQ ID NO:28:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 549 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: DNA (genomic)

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

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| TTAATATGAA TTATTATAAC ATAATCTACA CACAGGAACA TATAAATCTT GTCCACCTAT | 60  |
| TTCAATTATT TGATTTTTAT TATGTTTTTT AATTGTAAAA GAAGCATCTT TATAACAAAA | 120 |
| TTGACATATA GCTTGTAATT TTTTATTTTT TTCTACTTTA GGAATTAATT TTGATATAGA | 180 |
| ATTAAATATA TTTCTGTAA AGTCACAATT TAATCCAGCA ACAATAACTT TTTTTTTATT  | 240 |
| ATTAGCCATT TTATCACAAA ATTGTTCTAA ATCATTCTCT TCAAAAAATT GACACTCATC | 300 |
| TATGCCAATA ATATCATAAT TATCTACGAT ATTGATTTC AATTAATAAT TATTTGTTTT  | 360 |
| AATGTATAAA TATTCTTTAT TTAATATATT TCCGTCATGA TTTATTATAT TTTTATTTAT | 420 |
| AAATCTATTA TCTATATTAT GAGTTATAAT TACACATTTT TGATTAGATA AAATATATCT | 480 |
| ATTAATTTTT CGCATCAATT CTGTTGTTTT GCCAGAAAAC ATAGGACCAA TTATTAATTC | 540 |
| TATCGACAT                                                         | 549 |

## (2) INFORMATION FOR SEQ ID NO:29:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: DNA (genomic)

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

|                                                                  |    |
|------------------------------------------------------------------|----|
| TTTTTTTTAT TATTGATAT ATTTTTTCAA AAAAAAATTA ATCAATGAAA AAAAAATAAA | 60 |
| ATTATCAAA                                                        | 69 |

## (2) INFORMATION FOR SEQ ID NO:30:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 141 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| AAACATAGGA CCAATTATTA ATTCTATCGA CATTTTTTTT TATTATTGA TATATTTTTT  | 60  |
| CAAAAAAAAA TTAATCAATG AAAAAAAAAAT AAAATTATCA AAATGGATT ACTAAATTCT | 120 |
| GATATAATTT TAATAAATAT T                                           | 141 |

## (2) INFORMATION FOR SEQ ID NO:31:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 201 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| TTATAAACAA CAATCATATT TTTTAAAGA ATCTAATAAA TTTTAAACA TTTTATTATT   | 60  |
| ATTTGATAAT TGTTTATTTA ATTCGTTATT GATATTAACA ATATTATTTA TCATTTTACC | 120 |
| TATTTTTTTT TTTCTATCTA CTAACGAAAT ATCAGATTTT GCACCTTCAA TATCAGAATA | 180 |
| ATAATTATCA TTATTTTGCA T                                           | 201 |

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 660 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| ATGGATTAC TAAATCTGA TATAATTTTA ATAAATATTT TAAATATTA TAATTTAAAA     | 60  |
| AAAATAATA TAAACAGAGA TAATGTTATT AATATTAATA TATTAAAAA ATTAGTTAAT    | 120 |
| TTAGAAGAAT TGCATATAAT ATATTATGAT AATAATATTT TAAATAATAT TCCAGAAAAAT | 180 |
| ATTAAAAGTT TATATATTTT AAATTTAAAT ATTATTAATT TAAATTTTAT AACAAAATTA  | 240 |
| AAAAATATA CATATTTAGA TATATCTTAT AACAAAAATA GCAATATAAG TAATATTATA   | 300 |
| CTACCACATT CTATAGAATT TTTAAATTGT GAATCATGTA ATATAAATGA CTATAATTTT  | 360 |
| ATTAATAATT TAGTAAATTT AAAAAAATTA ATAATATCTA AAAATAAATT TGGTAACTTT  | 420 |
| AATAATGTTT TTCCTATTAG TATAGTTGAG TTAAATATGG AATCAATACA AATAAAAGAT  | 480 |
| TATAAATTTA TAGAAAAATT AATTAATTTA AAAAAATTAG ATATATCTTT CAATGTTAAA  | 540 |
| AAAAATAATA TACATTTGAT AAAATTTCCA AAAAGTATAA CTCATTTATG TGATTATCAA  | 600 |
| TCATATAAAG AAAATTATAA TTATTTAAAA AATTTATCAA ATATAATTGA ATATGAATTC  | 660 |

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3907 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| TTCTAAACGT TTATCTCCCC AAACATCTAC AGTAGATGGT TTATTAGATT CTGTGTTATA  | 60  |
| CACATCTGCT GGATTTGCGG CATTGTATC CAAACCATAA TATCCAGGTC TATAATTATC   | 120 |
| TTTAAAACT TGGGATTGAG ATACTTCTTC AGTTTTTAAA TTATTAAAAT ATCCAAGATT   | 180 |
| ATTTTTTTTT GATGAAGACA TAATTGATAT TATAATACTT TATAGATATG TCAATATTTA  | 240 |
| TCTACTATAT TTTCAACAAT AGATTTTATA TATATAAAG AATGAATACT GTACAAATTT   | 300 |
| TAGTTGTCAT ATTAATAACA ACAGCATTAT CTTTCTAGT TTTTCAATTA TGGTATTATG   | 360 |
| CCGAAAATTA CGAATATATA TTAAGATATA ATGATACATA TTCAAATTTA CAATTTGCCA  | 420 |
| GAAGCGCAAA TATAAATTTT GATGATTTAA CTGTTTTTGA TCCCAACGAT AATGTTTTTA  | 480 |
| ATGTTGAAGA AAAATGGCGC TGTGCTTCAA CTAATAATAA TATATTTTAT GCAGTTTCAA  | 540 |
| CTTTTGGATT TTTAAGTACA GAAAGTACTG GTATTAATTT AACATATACA AATTCATAGAG | 600 |
| ATTGTATTAT AGATTTATTT TCTAGAATTA TAAAAATAGT ATATGATCCT TGTACTGTCC  | 660 |
| AAACATCTAA CGATTGTAGA TTATTAAGAT TATTGATGGC CAATACATCA TAAATACATT  | 720 |

|            |            |            |            |             |            |      |
|------------|------------|------------|------------|-------------|------------|------|
| ATAATATTAT | TATAATATCA | ATCATAATTT | TTATATATAT | TTTATCTAAA  | AGGACTTTTT | 780  |
| ATTTTTTATA | TATTAATAAT | AATAAATGAG | TAACGTACCT | TTAGCAACCA  | AAACAATAAG | 840  |
| AAAATTATCA | AATCGAAAAT | ATGAAATAAA | GATTTATTTA | AAAGATGAAA  | ATACTTGTTT | 900  |
| CGAACGTGTA | GTAGATATGG | TAGTTCCATT | ATATGATGTG | TGTAATGAAA  | CTTCTGGTGT | 960  |
| TACTTTAGAA | TCATGTAGTC | CAAATATAGA | AGTAATTGAA | TTAGACAATA  | CTCATGTTAG | 1020 |
| AATCAAAGTT | CACGGCGATA | CATTAAAAGA | AATGTGTTTT | GAATTATTGT  | TCCCGTGTA  | 1080 |
| TGTAAACGAA | GCCCAAGTAT | GGAAATATGT | AAGTCGATTA | TTGCTAGATA  | ATGTATCACA | 1140 |
| TAATGACGTA | AAATATAAAT | TAGCTAATTT | TAGACTGACT | CTTAATGGAA  | AACATTTAAA | 1200 |
| ATTAAAAGAA | ATCGATCAAC | CGCTATTTAT | TTATTTTGTC | GATGATTGG   | GAAATTATGG | 1260 |
| ATTAATTACT | AAGGAAAATA | TTCAAAATAA | TAATTTACAA | GTTAACAAAG  | ATGCATCATT | 1320 |
| TATTACTATA | TTTCCACAAT | ATGCGTATAT | TTGTTTAGGT | AGAAAAGTAT  | ATTTAAATGA | 1380 |
| AAAAGTAACT | TTTGATGTAA | CTACAGATGC | AACTAATATT | ACTTTAGATT  | TTAATAAATC | 1440 |
| TGTTAATATC | GCAGTATCAT | TCCTTGATAT | ATATTACGAA | GTTAATAATA  | ATGAACAAAA | 1500 |
| AGATTTATTA | AAAGATTTAC | TTAAGAGATA | CGGTGAATTT | GAAGTCTATA  | ACGCAGATAC | 1560 |
| TGGATTAATT | TATGCTAAAA | ATCTAAGTAT | TAAAAATTAT | GATACTGTGA  | TTCAAGTAGA | 1620 |
| AAGGTTGCCA | GTTAATTTGA | AAGTTAGAGC | ATATACTAAG | GATGAAAATG  | GTCGCAATCT | 1680 |
| ATGTTTGATG | AAAATAACAT | CTAGTACAGA | AGTAGACCCC | GAGTATGTAA  | CTAGTAATAA | 1740 |
| TGCTTTATTG | GGTACGCTCA | GAGTATATAA | AAAGTTTGAT | AAATCTCATT  | TAAAAATTGT | 1800 |
| AATGCATAAC | AGAGGAAGTG | GTAATGTATT | TCCATTAAGA | TCATTATATC  | TGGAATTGTC | 1860 |
| TAATGTAAAA | GGATATCCAG | TTAAAGCATC | TGATACTTCG | AGATTAGATG  | TTGGTATTTA | 1920 |
| CAAATTAAT  | AAAATTTATG | TAGATAACGA | CGAAAATAAA | ATTATATTGG  | AAGAAATTGA | 1980 |
| AGCAGAATAT | AGATGCGGAA | GACAAGTATT | CCACGAACGT | GTAAGAACTTA | ATAAACACCA | 2040 |
| ATGTAAATAT | ACTCCCAAAT | GTCCATTCCA | ATTTGTTGTA | AACAGCCCAG  | ATACTACGAT | 2100 |
| TCACTTATAT | GGTATTTCTA | ATGTTTGTTT | AAAACCTAAA | GTACCCAAAA  | ATTTAAGACT | 2160 |
| TTGGGGATGG | ATTTTAGATT | GCGATACTTC | TAGATTTATT | AAACATATGG  | CTGATGGATC | 2220 |
| TGATGATTTA | GATCTTGACG | TTAGGCTTAA | TAGAAATGAT | ATATGTTTAA  | AACAAGCCAT | 2280 |
| AAAACAACAT | TATACTAATG | TAATTATATT | AGAGTACGCA | AATACATATC  | CAAATTGCAC | 2340 |
| ATTATCATTG | GGTAATAATA | GATTTAATAA | TGTATTTGAT | ATGAATGATA  | ACAAAACAT  | 2400 |
| ATCTGAGTAT | ACTAACTTTA | CAAAAAGTAG | ACAAGACCTT | AATAACATGT  | CATGTATATT | 2460 |
| AGGAATAAAC | ATAGGTAATT | CCGTAAATAT | TAGTAGTTTG | CCTGGTTGGG  | TAACACCTCA | 2520 |
| CGAAGCTAAA | ATTCTAAGAT | CTGGTTGTGC | TAGAGTTAGA | GAATTTTGTA  | AATCATTCTG | 2580 |
| TGATCTTTCT | AATAAGAGAT | TCTATGCTAT | GGCTAGAGAT | CTCGTAAGTT  | TACTATTTAT | 2640 |
| GTGTAACAT  | GTTAATATTG | AAATTAACGA | AGCAGTATGC | GAATATCCTG  | GATATGTCAT | 2700 |
| ATTATTCGCA | AGAGCTATTA | AAGTAATTAA | TGATTTATTA | TTAATTAACG  | GAGTAGATAA | 2760 |
| TCTAGCAGGA | TATTCAATTT | CCTTACCTAT | ACATTATGGA | TCTACTGAAA  | AGACTCTACC | 2820 |
| AAATGAAAAG | TATGGTGGTG | TTGATAAGAA | ATTTAAATAT | CTATTCTTAA  | AGAATAAACT | 2880 |
| AAAAGATTTA | ATGCGTGATG | CTGATTTTGT | CCAACCTCCA | TTATATATTT  | CTACTTACTT | 2940 |
| TAGAACTTTA | TTGGATGCTC | CACCAACTGA | TAATTATGAA | AAATATTTGG  | TTGATTCGTC | 3000 |
| CGTACAATCA | CAAGATGTTT | TACAGGTCT  | GTTGAATACA | TGTAATACTA  | TTGATACTAA | 3060 |
| TGCTAGAGTT | GCATCAAGTG | TTATTGGATA | TGTTTATGAA | CCATGCGGAA  | CATCAGAACA | 3120 |
| TAAAATTGGT | TCAGAAGCAT | TGTGTAAAT  | GGCTAAAGAA | GCATCTAGAT  | TAGGAAATCT | 3180 |
| AGGTTTAGTA | AATCGTATTA | ATGAAAGTAA | TTACAACAAA | TGTAATAAAT  | ATGGTTATAG | 3240 |

```

AGGAGTATAC GAAAATAACA AACTAAAAAC AAAATATTAT AGAGAAATAT TTGATTGTAA      3300
TCCTAATAAT AATAATGAAT TAATATCCAG ATATGGATAT AGAATAATGG ATTTACATAA      3360
AATTGGAGAA ATTTTTCGAA ATTACGATGA AAGTGAATCT CCTTGCGAAC GAAGATGTCA      3420
TTACTTGGAA GATAGAGGTC TTTTATATGG TCCTGAATAT GTACATCACA GATATCAAGA      3480
ATCATGTACG CCTAATACGT TTGGAAATAA CACAAATTGT GTAACAAGAA ATGGTGAACA      3540
ACACGTATAC GAAAATAGTT GTGGAGATAA TGCAACATGT GGAAGAAGAA CAGGATATGG      3600
AAGAAGAAGT AGGGATGAAT GGAATGACTA TAGAAAACCC CACGTTTATG ACAATTGTGC      3660
CGATGCAAT AGTTCATCTT CAGATAGCTG TTCAGACAGT AGTAGTAGTA GTGAATCTGA      3720
ATCTGATTCA GATGGATGTT GCGACACAGA TGCTAGTTTA GATTCTGATA TTGAAAATTG      3780
TTATCAAAAT CCATCAAAAT GTGATGCAGG ATGCTAAATG AAATTTAATA TTATATAATA      3840
TTAACTTACA AGTTATAAAA ATCATTAAAA TGATTTTTTA AAATGATATT ATCGATAGTT      3900
GTGATAA                                     3907

```

## (2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: protein
- (ix) FEATURE:
  - (A) NAME/KEY: Region
  - (B) LOCATION: 3
  - (D) OTHER INFORMATION: /note= "This amino acid may be either Asn or Arg."
- (ix) FEATURE:
  - (A) NAME/KEY: Region
  - (B) LOCATION: 12
  - (D) OTHER INFORMATION: /note= "This amino acid may be either Asn or Arg."

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

```

Met Ala Xaa Asp Leu Val Ser Leu Leu Phe Met Xaa Xaa Tyr Val Asn
 1           5           10           15
Ile Glu Ile Asn Glu Ala Val Xaa Glu
          20           25

```

## (2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
  - (A) NAME/KEY: Region
  - (B) LOCATION: 15
  - (D) OTHER INFORMATION: /note= "This amino acid may be either Thr or Ile."

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

```

Met Lys Ile Thr Ser Ser Thr Glu Val Asp Pro Glu Tyr Val Xaa Ser
 1           5           10           15
Asn

```

## (2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8 amino acids
  - (B) TYPE: amino acid

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

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

Asn Ala Leu Phe Phe Asn Val Phe  
1 5

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

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

Glu Val Asp Pro Glu Tyr Val  
1 5

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 66 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

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

ATGGCTAGAG ATCTCGTAAG TTTACTATTT ATGTGTAAC ATGTTAATAT TGAAATTAAC 60  
GAAGCA 66

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 51 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

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

ATGAAAATAA CATCTAGTAC AGAAGTAGAC CCCGAGTATG TAACTAGTAA T 51

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

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

AATAATAGAT TTAATAATGT ATTT 24

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1689 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

|            |            |            |             |             |            |      |
|------------|------------|------------|-------------|-------------|------------|------|
| TCCTATTATT | TGTTTTAATT | CTGATTCATT | CCACGGCATA  | TCTAATATAA  | TTATATCATT | 60   |
| AATACATTTG | AATGATATGC | CTTCAGATCC | AGCGTAAGAA  | AATATGCAA   | CTTTTACTTT | 120  |
| TTTACCATTA | TTATTTTCAT | AATTATTATA | TCGTTTAAAT  | TCATTATCTC  | TAGTTTTTAA | 180  |
| AGTTTTGCTA | GAATATTCAA | TATAAGAAAT | ATTAAAACAA  | TTAAAATAAC  | ATTTTAAACT | 240  |
| TGATATTCCT | TCAAAATTAA | CTAAAGGTTT | AAATATTAAT  | ACTTTTCCTC  | TCGAATTTAA | 300  |
| AATTATTTTA | CAAGTTTCTA | TATATTTACA | CGAATATTGA  | TATAATATAT  | TATAATTATT | 360  |
| TATATCAGTG | ATTGGTAAAT | TAGTTTTTAT | TTTTTATATTA | TCATTTTTTAA | AACTTTCAAT | 420  |
| AAAAGATTCA | GAGAAATTAA | TATTTTTTGT | AAACTCGGAA  | AATTCAGCAA  | GTTTTCTTTT | 480  |
| AATCATATCA | TTATATTCTA | TATTATCTAA | ATCTCCTTTT  | ATTTTAAGAT  | CATAAAAAGC | 540  |
| AAATGAAGAT | ATTAATCTTC | TCATAGTTTT | TAAACCACCT  | AATTCAGTTT  | TATAATCATA | 600  |
| TTTTTCTGCC | ATATTATATA | ATTTAGATTG | CTCATCTGAC  | ATAATTATAT  | TATGATAAAA | 660  |
| TATATTTTTT | TTTGCATATC | CATCTATATA | ATTTGTTTCT  | GTTAAACTAT  | CTGCTTCTAT | 720  |
| TAATCTTTTA | TAAGAACATA | TAGCTAATAA | TGTTTCTCTT  | AATTCCTTAA  | AATTAATTAA | 780  |
| CTTTCCATTA | TTTATATATT | CTTCTTTTAT | ATTCATAACA  | TTTGGTCTAA  | GTAAACCTAT | 840  |
| TAAATTATTA | AATTCAGAAA | TATTATTAGT | TACTGGAGTA  | GCGGACATAC  | ATAATATTTT | 900  |
| ATTATTTTCG | AAATTGCTA  | ATTTTATTAA | TTTTTTATAA  | ATAGGAGTAA  | AATTTCTTTC | 960  |
| GTTATTATCT | TTTTTAACAG | TTCTTGATAT | TAATTTATGA  | ACTTCGTCTA  | TTATTATTAG | 1020 |
| TAATCTACTT | TTTTTATTAA | GAGAACTTTC | TATAGATCTA  | TATATATTAT  | TAAATTTATC | 1080 |
| TAAACTAGAT | GACGAATCAT | AATATATAAA | TTTTATATTA  | CTGGTATCTG  | ATATATATGA | 1140 |
| TCTTATAGTA | TTTAACCAAG | GATCTATGTA | TAATGATTTT  | TTAATAAATA  | TTAAAATTAT | 1200 |
| CCATCTTGGA | AATAATTCTT | TTATATATTT | TATAATATAC  | ACAGCAGTTA  | ATGTTTTTCC | 1260 |
| CATACCAGTA | TCCCAAAATA | ATAACATACT | ATTCAAATTT  | TTAATCCTA   | TGAATATTCT | 1320 |
| ACTTACAAAA | TATTGATAAT | CTTGTAATGT | AATTTTCAGTA | TTTGTAATAT  | TATTCATAAT | 1380 |
| TTTATTAGGC | AAATGTTGTG | TTTTATCAAG | TGCATAATTT  | ATATGTTTAC  | CAACAATAGA | 1440 |
| ATCTAATGCA | AACATTTAGT | TATATAAAAA | ATAATATTTA  | TATTAACCTA  | AGATGTTTCA | 1500 |
| TTAATTTTAT | GTCTGTGATG | TGGAGTTAAA | ACCCAAGATA  | TTGATATATC  | TATATCATT  | 1560 |
| ATTCTTCTTT | TGAATCTATG | TCTATCAATC | GCAAATTTAT  | CCCAGTATAA  | TTTTCGAGTT | 1620 |
| TGTTTTGCAG | CATATAACCA | AACATACATA | ATGTGGAGTT  | TTGGTGGTTC  | GGATGAAAAG | 1680 |
| CGTACTTTT  |            |            |             |             |            | 1689 |

## (2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 485 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

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

86

Met Gly Lys Thr Leu Thr Ala Val Tyr Ile Ile Lys Tyr Ile Lys Glu  
 65 70 75 80  
 Leu Phe Pro Arg Trp Ile Ile Leu Ile Phe Ile Lys Lys Ser Leu Tyr  
 85 90 95  
 Ile Asp Pro Trp Leu Asn Thr Ile Arg Ser Tyr Ile Ser Asp Thr Ser  
 100 105 110  
 Asn Ile Lys Phe Ile Tyr Tyr Asp Ser Ser Ser Ser Leu Asp Lys Phe  
 115 120 125  
 Asn Asn Ile Tyr Arg Ser Ile Glu Ser Ser Leu Asn Lys Lys Ser Arg  
 130 135 140  
 Leu Leu Ile Ile Ile Asp Glu Val His Lys Leu Ile Ser Arg Thr Val  
 145 150 155 160  
 Lys Lys Asp Asn Asn Glu Arg Asn Phe Thr Pro Ile Tyr Lys Lys Leu  
 165 170 175  
 Ile Lys Leu Ala Asn Phe Glu Asn Asn Lys Ile Leu Cys Met Ser Ala  
 180 185 190  
 Thr Pro Val Thr Asn Asn Ile Ser Glu Phe Asn Asn Leu Ile Gly Leu  
 195 200 205  
 Leu Arg Pro Asn Val Met Asn Ile Lys Glu Glu Tyr Ile Asn Asn Gly  
 210 215 220  
 Lys Leu Ile Asn Phe Lys Glu Leu Arg Glu Thr Leu Leu Ala Ile Cys  
 225 230 235 240  
 Ser Tyr Lys Arg Leu Ile Glu Ala Asp Ser Leu Thr Glu Thr Asn Tyr  
 245 250 255  
 Ile Asp Gly Tyr Ala Lys Lys Asn Ile Phe Tyr His Asn Ile Ile Met  
 260 265 270  
 Ser Asp Glu Gln Ser Lys Leu Tyr Asn Met Ala Glu Lys Tyr Asp Tyr  
 275 280 285  
 Lys Thr Glu Leu Gly Gly Leu Lys Thr Met Arg Arg Leu Ile Ser Ser  
 290 295 300  
 Phe Ala Phe Tyr Asp Leu Lys Ile Lys Gly Asp Leu Asp Asn Ile Glu  
 305 310 315 320  
 Tyr Asn Asp Met Ile Lys Arg Lys Leu Ala Glu Phe Ser Glu Phe Thr  
 325 330 335  
 Lys Asn Ile Asn Phe Ser Glu Ser Phe Ile Glu Ser Phe Lys Asn Asp  
 340 345 350  
 Asn Ile Lys Ile Lys Thr Asn Leu Pro Ile Thr Asp Ile Asn Asn Tyr  
 355 360 365  
 Asn Ile Leu Tyr Gln Tyr Ser Cys Lys Tyr Ile Glu Thr Cys Lys Ile  
 370 375 380  
 Ile Leu Asn Ser Arg Gly Lys Val Leu Ile Phe Glu Pro Leu Val Asn  
 385 390 395 400  
 Phe Glu Gly Ile Ser Leu Lys Cys Tyr Phe Asn Cys Phe Asn Ile  
 405 410 415  
 Ser Tyr Ile Glu Tyr Ser Ser Lys Thr Leu Lys Thr Arg Asp Asn Glu  
 420 425 430  
 Leu Asn Glu Tyr Asn Asn Tyr Glu Asn Asn Asn Gly Lys Lys Val Lys  
 435 440 445  
 Val Cys Ile Phe Ser Tyr Ala Gly Ser Glu Gly Ile Ser Phe Lys Cys  
 450 455 460  
 Ile Asn Asp Ile Ile Ile Leu Asp Met Pro Trp Asn Glu Ser Glu Leu  
 465 470 475 480  
 Lys Gln Ile Ile Gly  
 485

## (2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 235 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

```

AAAAGTTTGA TAAATCACAT TTAAAAATTG TTATGCATAA TAGAGGAAGT GGTAATGTAT      60
TCCCTATTAG ATCACTATAT TTGGAATTAT TGAACGTCAA AGGTTATCCT GTAAAAGCAT      120
CCGATACGTC TAGGTTAGAT GTTGGTGTTT ATAAACTAAA TAAAATATAT ATTGATAATG      180
ATGAAAATAA AATAATTTTA GAAGAAATTG AAACCGATTA TAGATGTGGA AGAGA          235

```

## (2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 77 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

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

```

## (2) INFORMATION FOR SEQ ID NO:45:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 235 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

```

AAAAGTTTGA TAAATCACAT TTAAAAATCG TTATGCACAA TAGAGGAAGC GGTAATGTAT      60
TCCCTATTAG ATCACTATAT TTGGAATTAT TGAACGTCAA AGGTTATCCT GTTAAAGCAT      120
CCGATACGTC TAGGTTAGAC GTTGGTGTTT ATAAACTAAA TAAAATATAT ATTGATAATG      180
ATGAAAATAA AATAATTTTA GAAGAAATCG AAACCGATTA TAGATGTGGA AGAGA          235

```

## (2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 77 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

Lys Phe Asp Lys Ser His Leu Lys Ile Val Met His Asn Arg Gly Ser
 1           5          10

```

88

Gly Asn Val Phe Pro Ile Arg Ser Leu Tyr Leu Glu Leu Leu Asn Val  
                   20                                  25                                  30  
 Lys Gly Tyr Pro Val Lys Ala Ser Asp Thr Ser Arg Leu Asp Val Gly  
                   35                                  40                                  45  
 Val Tyr Lys Leu Asn Lys Ile Tyr Ile Asp Asn Asp Glu Asn Lys Ile  
                   50                                  55                                  60  
 Ile Leu Glu Glu Ile Glu Thr Asp Tyr Arg Cys Gly Arg  
                   65                                  70                                  75

## (2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

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

Lys Phe Lys Tyr Leu Phe Leu Lys Asn Lys  
   1                  5                                  10

## (2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

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

Lys Ser Val Asn Ile Ala Val Ser Phe Leu Asp  
   1                  5                                  10

## (2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

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

Lys Tyr Leu Val Asp Ser Ser Val Gln Ser Gln  
   1                  5                                  10

## (2) INFORMATION FOR SEQ ID NO:50:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

AGATGATGAT TAAAGTGTGG

20

## (2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single

89

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:  
GATAATGATA CTCCGGTTGC 20
- (2) INFORMATION FOR SEQ ID NO:52:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:  
GAAGTNGATC CNGAATATGT 20
- (2) INFORMATION FOR SEQ ID NO:53:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:  
GAAAATAAAA TTATATTGGA 20
- (2) INFORMATION FOR SEQ ID NO:54:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:  
AGACAATTCC AGATATAATG 20
- (2) INFORMATION FOR SEQ ID NO:55:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:  
CCGCATCTAT ATTCTGCTTC 20
- (2) INFORMATION FOR SEQ ID NO:56:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:  
GTTTAAAACC TAAAGTACCC 20

## (2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

TTTCAAATTA ACTGGCAACC

20

## (2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GGGATGGATT TTAGATTGCG

20

## (2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GTTGCATCTG TAGTTACATC

20

## (2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

TCTAGCAATA ATCGACTTAC

20

## (2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GCCTGGTTGG GTAACAACCTC

20

## (2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:  
CATTCTATT AAGCCTAACG 20
- (2) INFORMATION FOR SEQ ID NO:63:  
(i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 20 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: DNA (genomic)  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:  
GTACCTTTAG CAACCAAAC 20
- (2) INFORMATION FOR SEQ ID NO:64:  
(i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 20 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: DNA (genomic)  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:  
CTGCTAGATT ATCTACTCCG 20
- (2) INFORMATION FOR SEQ ID NO:65:  
(i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 20 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: DNA (genomic)  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:  
AATTGCACAT TATCATTGGG 20
- (2) INFORMATION FOR SEQ ID NO:66:  
(i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 20 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: DNA (genomic)  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:  
ATTACCCAAT GATAATGTGC 20

Claims

1           1. A polynucleotide molecule which comprises an Entomopoxvirus  
2 spheroidin polynucleotide sequence; or a fragment or variant of said spheroidin  
3 polynucleotide sequence having sufficient length and homology with an  
4 Entomopoxvirus polynucleotide sequence to hybridize with said Entomopoxvirus  
5 spheroidin polynucleotide sequence under stringent conditions; wherein said  
6 spheroidin polynucleotide sequence is free from association with other viral  
7 nucleotide sequences with which it is associated in nature.

1           2. The polynucleotide molecule, according to claim 1, comprising a  
2 polynucleotide sequence selected from the group consisting of a spheroidin gene  
3 coding sequence, a spheroidin gene regulatory sequence, a spheroidin gene  
4 promoter sequence, and variants or fragments of these spheroidin sequences.

1           3. The polynucleotide molecule, according to claim 1, wherein said  
2 spheroidin polynucleotide sequence is obtainable from the *Amsacta moorei*  
3 Entomopoxvirus.

1           4. The polynucleotide molecule, according to claim 1, wherein said  
2 spheroidin polynucleotide sequence is obtainable from the *Choristoneura*  
3 *fumiferana* Entomopoxvirus.

1           5. The polynucleotide molecule, according to claim 1, wherein said  
2 spheroidin polynucleotide sequence is obtainable from the *Choristoneura biennis*  
3 Entomopoxvirus.

1           6. The polynucleotide molecule, according to claim 3, comprising a  
2 spheroidin polynucleotide nucleotide sequence selected from the group consisting  
3 of SEQ ID NO. 1 and variants or fragments thereof.

1           7. The polynucleotide molecule, according to claim 1, wherein said  
2 spheroidin sequence is further characterized by the ability to direct the expression  
3 of a heterologous gene to which said sequence is operably linked in a selected  
4 host cell or virus.

1           8. The polynucleotide molecule, according to claim 1, comprising said  
2 spheroidin polynucleotide sequence and further comprising a second  
3 polynucleotide sequence encoding a heterologous gene.

1           9. A polynucleotide molecule which comprises an Entomopoxvirus  
2 thymidine kinase polynucleotide sequence; or a fragment or variant of said  
3 thymidine kinase polynucleotide sequence having sufficient length and homology  
4 with an Entomopoxvirus thymidine kinase polynucleotide sequence to hybridize  
5 with said Entomopoxvirus thymidine kinase polynucleotide sequence under  
6 stringent conditions; wherein said polynucleotide sequence is free from association  
7 with other viral nucleotide sequences with which it is associated in nature.

1           10. The polynucleotide molecule, according to claim 9, comprising a  
2 sequence selected from the group consisting of a thymidine kinase gene coding  
3 sequence, a thymidine kinase gene regulatory sequence, a thymidine kinase gene  
4 promoter sequence, and variants or fragments of these thymidine kinase  
5 sequences.

1           11. The polynucleotide molecule, according to claim 9, wherein said  
2 thymidine kinase sequence is obtainable from the *Amsacta moorei*  
3 Entomopoxvirus.

1           12. The polynucleotide molecule, according to claim 9, comprising a  
2 thymidine kinase polynucleotide sequence selected from the group consisting of  
3 SEQ ID NO. 8 and variants or fragments thereof.

1           13. The polynucleotide molecule, according to claim 9, wherein said  
2           thymidine kinase sequence is further characterized by the ability to direct the  
3           expression of a heterologous gene to which said sequence is operably linked in a  
4           selected host cell or virus.

1           14. The polynucleotide molecule, according to claim 9, comprising said  
2           thymidine kinase polynucleotide sequence and further comprising a second  
3           polynucleotide sequence encoding a heterologous gene.

1           15. A substantially pure Entomopoxvirus spheroidin polypeptide, or a  
2           fragment or variant thereof.

1           16. The polypeptide, according to claim 15, fused to a heterologous  
2           protein or peptide.

1           17. A substantially pure Entomopoxvirus thymidine kinase polypeptide or  
2           a fragment or variant thereof.

1           18. The polypeptide, according to claim 17, fused to a heterologous  
2           protein or peptide.

1           19. A recombinant molecule comprising a polynucleotide sequence  
2           encoding an Entomopoxvirus spheroidin promoter sequence, or a variant or  
3           fragment of said promoter sequence, wherein said sequence is operably linked to  
4           a selected heterologous gene sequence, said sequence being capable of directing  
5           the expression of said gene in a selected host cell or virus.

1           20. A recombinant molecule comprising a polynucleotide sequence  
2           encoding an Entomopoxvirus thymidine kinase promoter sequence, or a variant  
3           or fragment of said promoter sequence, wherein said sequence is operably linked  
4           to a selected heterologous gene sequence, said sequence being capable of  
5           directing the expression of said gene in a selected host cell.

1           21. A cell infected with a recombinant virus comprising an  
2 Entomopoxvirus spheroidin gene polynucleotide sequence, or a variant or  
3 fragment thereof, optionally linked to a selected heterologous gene sequence.

1           22. The cell, according to claim 21, selected from the group consisting of  
2 insect cells and mammalian cells.

1           23. A cell infected with a recombinant virus comprising an  
2 Entomopoxvirus thymidine kinase gene polynucleotide sequence, or a variant or  
3 fragment thereof, optionally linked to a selected heterologous gene sequence.

1           24. The cell, according to claim 23, selected from the group consisting of  
2 insect cells and mammalian cells.

1           25. A method for producing a selected polypeptide comprising culturing  
2 a selected host cell transformed with a polynucleotide sequence comprising an  
3 Entomopoxvirus thymidine kinase gene polynucleotide sequence, or a variant or  
4 fragment thereof, operably linked to a heterologous gene sequence encoding said  
5 selected polypeptide, and recovering said polypeptide from the culture medium.

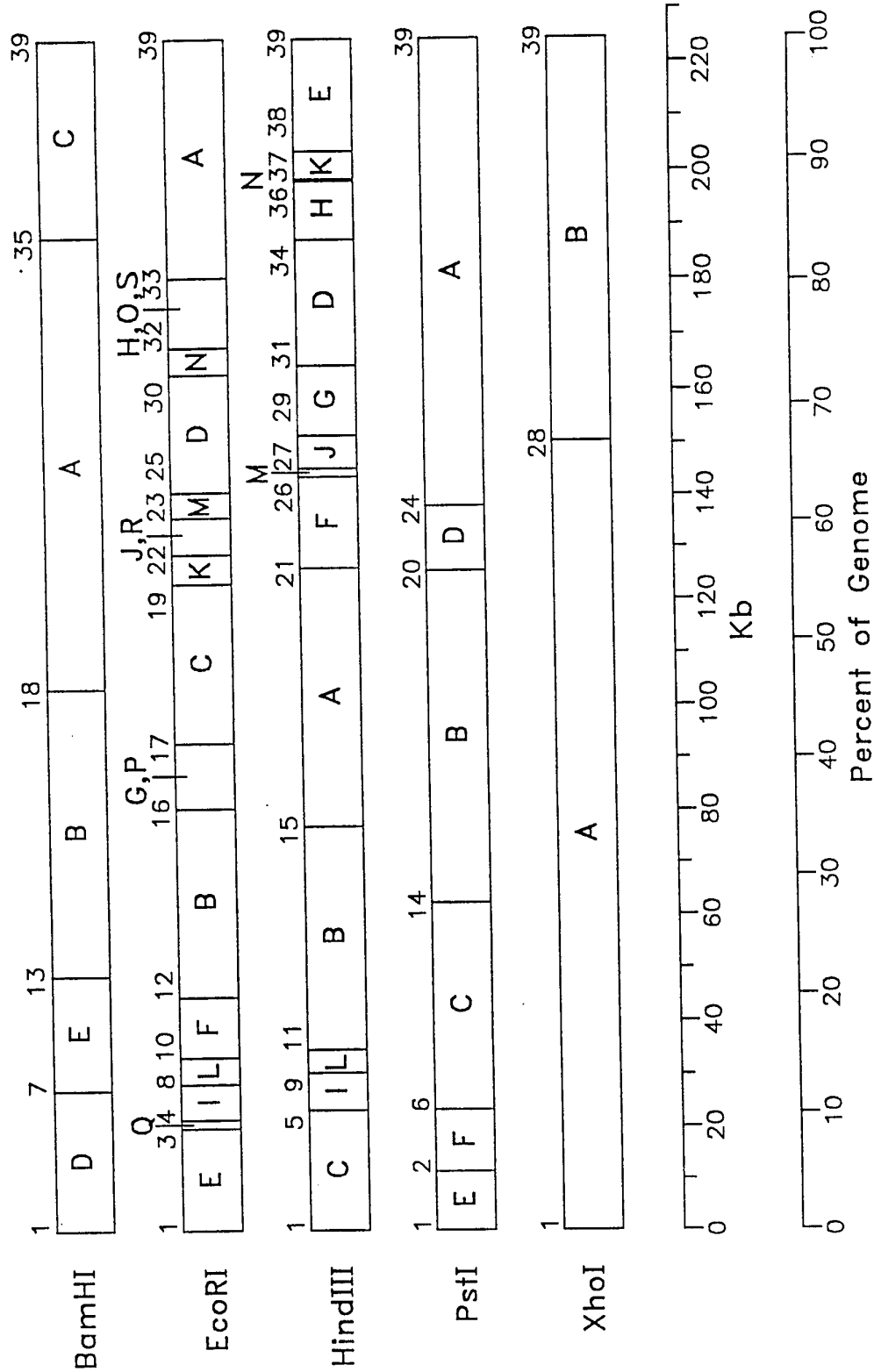
1           26. A method for producing a selected polypeptide comprising culturing  
2 a selected host cell transformed with a polynucleotide sequence comprising an  
3 Entomopoxvirus spheroidin gene polynucleotide sequence, or a variant or  
4 fragment thereof, operably linked to a selected heterologous gene sequence  
5 encoding said selected polypeptide, and recovering said polypeptide from the  
6 culture medium.

1           27. A method for screening recombinant host cells for insertion of  
2 heterologous genes comprising transforming said cells with a polynucleotide  
3 molecule comprising the selected heterologous gene sequence inserted into the  
4 polynucleotide sequence encoding entomopoxvirus spheroidin, wherein the

5       absence of occlusion bodies normally formed by the expression of the spheroidin  
6       protein indicates the integration of the heterologous gene.

1               28. A method for screening recombinant host cells for insertion of  
2       heterologous genes comprising infecting said cells with a polynucleotide molecule  
3       comprising the selected heterologous gene sequence inserted into the  
4       polynucleotide sequence encoding entomopoxvirus thymidine kinase, wherein the  
5       absence of thymidine kinase activity due to the integration of the inactive  
6       thymidine kinase sequence indicates the insertion of the heterologous gene.

Figure 1



2/20

Figure 2A

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| AGATCTGATG TTCTATATAT AGTACAAATT TGTATGATTA ATTGATATTT TAAAAATTC    | 60  |
| GATA TTA AAT ATT AGA TTC TAA ACT ATT CTT CTC ATT ATC AAT ATA ACT    | 109 |
| Ile Asn Ser Glu Leu Ser Asn Lys Glu Asn Asp Ile Tyr Ser             |     |
| 1 5 10                                                              |     |
| ATC ATA ATC ATT TTT TAT TTT ACT ACA TAC ATT CAT AAT TCT ATT ACT ATT | 160 |
| Asp Tyr Asp Asn Lys Ile Lys Ser Cys Val Asn Met Ile Arg Asn Ser Asn |     |
| 15 20 25 30                                                         |     |
| TTT TTT ATA CAT ATC TAT TAA TTC CAT AAA CTT TTT ATT TTT TAT ATT AAA | 211 |
| Lys Lys Tyr Met Asp Ile Leu Glu Met Phe Lys Lys Asn Lys Ile Asn Phe |     |
| 35 40 45                                                            |     |
| TAT TTC TAA TGT ATT TTT AAA TTC GTC AAT ACT ATT AAT ATC ATA TCT AGA | 262 |
| Ile Glu Leu Thr Asn Lys Phe Glu Asp Ile Ser Asn Ile Asp Tyr Arg Ser |     |
| 50 55 60 65                                                         |     |
| AAT AAA TAA TGC ACC TCT ATA ACT ACT AGC CAA TAA ATC ACC AAT AAA ACT | 313 |
| Ile Phe Leu Ala Gly Arg Tyr Ser Ser Ala Leu Leu Asp Gly Ile Phe Ser |     |
| 70 75 80                                                            |     |
| CAT AGA ATA ATA TAA TTT TTT AAA TTC AAA TTT AGA TTT TAT GTT GAA ATA | 364 |
| Met Ser Tyr Tyr Leu Lys Lys Phe Glu Phe Lys Ser Lys Ile Asn Phe Tyr |     |
| 85 90 95                                                            |     |
| AAC TAT ATA ATA TAA AAA TAT TAT ATT AAA CAT ACC ACA ATC GGG ACT ATC | 415 |
| Val Ile Tyr Tyr Leu Phe Ile Ile Asn Phe Met Gly Cys Asp Pro Ser Asp |     |
| 100 105 110 115                                                     |     |
| ATA TTG TAA TTC AAA AGT ATT AAA AAA GTA ATA ATT TAC ATT TTT AAA TAT | 466 |
| Tyr Gln Leu Glu Phe Thr Asn Phe Phe Tyr Tyr Asn Val Asn Lys Phe Ile |     |
| 120 125 130                                                         |     |
| ATC ATT TAA ATA TTC TGA TAG TAC ATC AAT GTA TAA ATA AGC ATA ATT AGT | 517 |
| Asp Asn Leu Tyr Glu Ser Leu Val Asp Ile Tyr Leu Tyr Ala Tyr Asn Thr |     |
| 135 140 145 150                                                     |     |
| ATT AGG AGT ACT ATT GTA GTG TTT ATG GCT TTT TAT AGT CAT ATC AGA TTC | 568 |
| Asn Pro Thr Ser Asn Tyr His Lys His Ser Lys Ile Thr Met Asp Ser Glu |     |
| 155 160 165                                                         |     |
| AAT AAA CAT ATA TTT TTT ATT TTG TTT TAT AAG TTC TGG TAT ATA ACC ACT | 619 |
| Ile Phe Met Tyr Lys Lys Asn Gln Lys Ile Leu Glu Pro Ile Tyr Gly Ser |     |
| 170 175 180                                                         |     |
| ACT ATT AAA AAA GTA TGC AGC TTT TTT ATC TTT ATC AAA GTG TTT ATC TAT | 670 |
| Ser Asn Phe Phe Tyr Ala Ala Lys Lys Asp Lys Asp Phe His Lys Asp Ile |     |
| 185 190 195 200                                                     |     |
| TAC GCA ACA AGT AAA ATG ATC ATT ATA AAT TAT AGG AAA CAT AAA AAA TCT | 721 |
| Val Cys Cys Thr Phe His Asp Asn Tyr Ile Ile Pro Phe Met Phe Phe Arg |     |
| 205 210 215                                                         |     |
| TTT TTT ATC ATT CAT TAA AAA AAA TTT TAC TCT ATC TTC AAG TTT ATA GCA | 772 |
| Lys Lys Asp Asn Met Leu Phe Phe Lys Val Arg Asp Glu Leu Lys Tyr Cys |     |
| 220 225 230 235                                                     |     |
| TCT CAT AGA TGA AGC TAC TGT AGC AAT ATT TTT ATC AGT TTT TTC AAA TAA | 823 |
| Arg Met Ser Ser Ala Val Thr Ala Ile Asn Lys Asp Thr Lys Glu Phe Leu |     |
| 240 245 250                                                         |     |

3/20

Figure 2B

|                                                                     |      |
|---------------------------------------------------------------------|------|
| AAT CAA ATG AAA ATA ATC ATA ATC TGT ATT AAT CAT AGT TAA TGG ATA TAT | 874  |
| Ile Leu His Phe Tyr Asp Tyr Asp Thr Asn Ile Met Thr Leu Pro Tyr Ile |      |
| 255 260 265                                                         |      |
| ACA ATT ATA TAT ATC TCC CGA ACT TAA CCA TGT AGA TTT ATC ATG TTT TCT | 925  |
| Cys Asn Tyr Ile Asp Gly Ser Ser Leu Trp Thr Ser Lys Asp His Lys Arg |      |
| 270 275 280 285                                                     |      |
| TGG GTA AGC TTT AGG TTT AGG ATT AAA TCC CAA AGG CGG TAT TCC TAT TTG | 976  |
| Pro Tyr Ala Lys Pro Lys Pro Asn Phe Gly Leu Pro Pro Ile Gly Ile Gln |      |
| 290 295 300                                                         |      |
| AGC ATC CAA ATC ATC ATA AAT TGT GGC AAA TGT AGA AAA ATC TCT TGT TTT | 1027 |
| Ala Asp Leu Asp Asp Tyr Ile Thr Ala Phe Thr Ser Phe Asp Arg Thr Lys |      |
| 305 310 315 320                                                     |      |
| GGA TAA TTC TGA TTT TAG AAA AGA CTT TCT CAT ATA TAC TAA TGG AAT GCC | 1078 |
| Ser Leu Glu Ser Lys Leu Phe Ser Lys Arg Met Tyr Val Leu Pro Ile Gly |      |
| 325 330 335                                                         |      |
| TTT ATA TTT TTT AGA TGT AAT AAA AGT ATT AAT ATT TAT ATT TTT ATC TTG | 1129 |
| Lys Tyr Lys Lys Ser Thr Ile Phe Thr Asn Ile Asn Ile Asn Lys Asp Gln |      |
| 340 345 350                                                         |      |
| TAA ATA TTT TTT TAT AGT CCA AAA TAG AAA AAA TTT TCT TTT AAT ATT ATT | 1180 |
| Leu Tyr Lys Lys Ile Thr Trp Phe Leu Phe Phe Lys Arg Lys Ile Asn Asn |      |
| 355 360 365 370                                                     |      |
| TTC AAA ATT AAT ATT ATT AAT ATG ATT TGG ATC TAA AAC TAA TTC ATT ATA | 1231 |
| Glu Phe Asn Ile Asn Asn Ile His Asn Pro Asp Leu Val Leu Glu Asn Tyr |      |
| 375 380 385                                                         |      |
| TAA TAT TTC CAA GTA TTT TAT AGG TAT AAA TGT TAC TTT ACC TCT TGT TTC | 1282 |
| Leu Ile Glu Leu Tyr Lys Ile Pro Ile Phe Thr Val Lys Gly Arg Thr Glu |      |
| 390 395 400 405                                                     |      |
| ATC ATC ATC ATC TAT TTT TTC TAA TAT AGC TAT ATT TGC ATT AGT ATT ATA | 1333 |
| Asp Asp Asp Asp Ile Lys Glu Leu Ile Ala Ile Asn Ala Asn Thr Asn Tyr |      |
| 410 415 420                                                         |      |
| TTT AAT AGG ATT TAT AAA ATA TAC CAT ATT ATC TAT TTT ACT AAA AAA TAA | 1384 |
| Lys Ile Pro Asn Ile Phe Tyr Val Met Asn Asp Ile Lys Ser Phe Phe Leu |      |
| 425 430 435                                                         |      |
| CAT AGA CAT AAA ATT AAT ACC AGA TTC TGG CAT TTT TAA ATT TTT ATT TGG | 1435 |
| Met Ser Met Phe Asn Ile Gly Ser Glu Pro Met Lys Leu Asn Lys Asn Pro |      |
| 440 445 450 455                                                     |      |
| AAA TCT TCT AAT TTT ATT ATT CAT TATTTATTTA ATAA ATG TTT CTA GTT TAT | 1488 |
| Phe Arg Arg Ile Lys Asn Asn Met Met Phe Leu Val Tyr                 |      |
| 460 465                                                             |      |
| TTC AAT ACA TTT TTA ATA ATA ATT TTA TTA TTT GGT ATT ATA GGT ATT TAT | 1539 |
| Phe Asn Thr Phe Leu Ile Ile Ile Leu Leu Phe Gly Ile Ile Gly Ile Tyr |      |
| 470 475 480 485                                                     |      |
| ATA TTA ACA TTT GTG TTT AAT ATA GAT TTT TTA ATA AAT AAT AAT AAA ATA | 1590 |
| Ile Leu Thr Phe Val Phe Asn Ile Asp Phe Leu Ile Asn Asn Asn Lys Ile |      |
| 490 495 500                                                         |      |

4/20

Figure 2C

|                                                                     |      |
|---------------------------------------------------------------------|------|
| TAT ATA TTA TCA TAT AAC GCA ACT AAT ATA AAC AAT ATA AAT AAT TTA AAT | 1641 |
| Tyr Ile Leu Ser Tyr Asn Ala Thr Asn Ile Asn Asn Ile Asn Asn Leu Asn |      |
| 505 510 515 520                                                     |      |
| TTA TAC GAT TAT TCA GAT ATT ATA TTT TTG ACA AAT TTT AAC ATA AAT AAT | 1692 |
| Leu Tyr Asp Tyr Ser Asp Ile Ile Phe Leu Thr Asn Phe Asn Ile Asn Asn |      |
| 525 530 535                                                         |      |
| AAT CTT TTA GTA ACA CAA GCT AAT AAT TTA CAA GAT ATA CCA ATA TTT AAT | 1743 |
| Asn Leu Leu Val Thr Gln Ala Asn Asn Leu Gln Asp Ile Pro Ile Phe Asn |      |
| 540 545 550                                                         |      |
| GTA AAT AAT ATT ATA TCT AAT CAA TAT AAT TTT TAT TCA GCG TCT AGT AAT | 1794 |
| Val Asn Asn Ile Ile Ser Asn Gln Tyr Asn Phe Tyr Ser Ala Ser Ser Asn |      |
| 555 560 565 570                                                     |      |
| AAT GTA AAT ATA TTA TTA GGA TTA AGA AAA ACA TTA AAT ATA AAT AGA AAT | 1845 |
| Asn Val Asn Ile Leu Leu Gly Leu Arg Lys Thr Leu Asn Ile Asn Arg Asn |      |
| 575 580 585                                                         |      |
| CCA TTT TTA TTA TTT AGA AAT ACA TCT CTA GCT ATA GTT TTC AAT AAT AAT | 1896 |
| Pro Phe Leu Leu Phe Arg Asn Thr Ser Leu Ala Ile Val Phe Asn Asn Asn |      |
| 590 595 600 605                                                     |      |
| GAA ACT TTT CAC TGT TAT ATA AGT TCA AAT CAA AAT AGT GAT GTA TTA GAT | 1947 |
| Glu Thr Phe His Cys Tyr Ile Ser Ser Asn Gln Asn Ser Asp Val Leu Asp |      |
| 610 615 620                                                         |      |
| ATA GTA TCA CAT ATA GAA TTT ATG AAA TCT AGA TAT AAT AAA TAT GTA ATT | 1998 |
| Ile Val Ser His Ile Glu Phe Met Lys Ser Arg Tyr Asn Lys Tyr Val Ile |      |
| 625 630 635                                                         |      |
| ATA GGA GAA ATA CCC GTA AAT AAT AAT ATA TCT ATT AAT AAT ATA TTA AAT | 2049 |
| Ile Gly Glu Ile Pro Val Asn Asn Asn Ile Ser Ile Asn Asn Ile Leu Asn |      |
| 640 645 650 655                                                     |      |
| AAT TTT GCT ATT ATA ACT AAT GTG AGA TTA ATA GAT AAA TAT AAC TCT ATA | 2100 |
| Asn Phe Ala Ile Ile Thr Asn Val Arg Leu Ile Asp Lys Tyr Asn Ser Ile |      |
| 660 665 670                                                         |      |
| ATA TCA TTT TTA AAT ATC AAC GTA GGA ACA CTT TTT GTC ATA AAT CCA TAA | 2151 |
| Ile Ser Phe Leu Asn Ile Asn Val Gly Thr Leu Phe Val Ile Asn Pro     |      |
| 675 680 685                                                         |      |
| TATTTAGTAA TAATCACTAA CATATTTTTT ATTAAATGA ATAAATATA TATTGTTATT     | 2211 |
| GTCAATATTT TATATCATTT TACAGTC TTA TTT TTT TTT TTT GCT TTT AGG TAT   | 2265 |
| Lys Lys Lys Lys Ser Lys Pro Ile                                     |      |
| 690 695                                                             |      |
| AAT TTT ACC TTC TAA ACG TTT ATC TCC CCA AAC ATC TAC AGT AGA TGG TTT | 2316 |
| Ile Lys Gly Glu Leu Arg Lys Asp Gly Trp Val Asp Val Thr Ser Pro Lys |      |
| 700 705 710                                                         |      |
| ATT AGA TTC TGT GTT ATA CAC ATC TGC TGG ATT TGC GGC ATT TGT ATC CAA | 2367 |
| Asn Ser Glu Thr Asn Tyr Val Asp Ala Pro Asn Ala Ala Asn Thr Asp Leu |      |
| 715 720 725 730                                                     |      |
| ACC ATA ATA TCC AGG TCT ATA ATT ATC TTT AAA AAC TTG GGA TTG AGA TAC | 2418 |
| Gly Tyr Tyr Gly Pro Arg Tyr Asn Asp Lys Phe Val Gln Ser Gln Ser Val |      |
| 735 740 745                                                         |      |

5/20

Figure 2D

```

TTC TTC AGT TTT TAA ATT ATT AAA ATA TCC AAG ATT ATT TTT TTT TGA TGA 2469
Glu Glu Thr Lys Leu Asn Asn Phe Tyr Gly Leu Asn Asn Lys Lys Ser Ser
750 755 760 765

< G3L G4R >
AGA CAT AATTGATATT ATAATACTTT ATAGAT ATG TCA ATA TTT ATC TAC TAT 2522
Ser Met Met Ser Ile Phe Ile Tyr Tyr
770

ATT TTC AAC AAT AGA TTT TAT ATA TAT AAA AGA ATG AAT ACT GTA CAA ATT 2573
Ile Phe Asn Asn Arg Phe Tyr Ile Tyr Lys Arg Met Asn Thr Val Gln Ile
775 780 785 790

TTA GTT GTC ATA TTA ATA ACA ACA GCA TTA TCT TTT CTA GTT TTT CAA TTA 2624
Leu Val Val Ile Leu Ile Thr Thr Ala Leu Ser Phe Leu Val Phe Gln Leu
795 800 805

TGG TAT TAT GCC GAA AAT TAC GAA TAT ATA TTA AGA TAT AAT GAT ACA TAT 2675
Trp Tyr Tyr Ala Glu Asn Tyr Glu Tyr Ile Leu Arg Tyr Asn Asp Thr Tyr
810 815 820 825

TCA AAT TTA CAA TTT GCG AGA AGC GCA AAT ATA AAT TTT GAT GAT TTA ACT 2726
Ser Asn Leu Gln Phe Ala Arg Ser Ala Asn Ile Asn Phe Asp Asp Leu Thr
830 835 840

GTT TTT GAT CCC AAC GAT AAT GTT TTT AAT GTT GAA GAA AAA TGG CGC TGT 2777
Val Phe Asp Pro Asn Asp Asn Val Phe Asn Val Glu Glu Lys Trp Arg Cys
845 850 855

GCT TCA ACT AAT AAT AAT ATA TTT TAT GCA GTT TCA ACT TTT GGA TTT TTA 2828
Ala Ser Thr Asn Asn Asn Ile Phe Tyr Ala Val Ser Thr Phe Gly Phe Leu
860 865 870 875

AGT ACA GAA AGT ACT GGT ATT AAT TTA ACA TAT ACA AAT TCT AGA GAT TGT 2879
Ser Thr Glu Ser Thr Gly Ile Asn Leu Thr Tyr Thr Asn Ser Arg Asp Cys
880 885 890

ATT ATA GAT TTA TTT TCT AGA ATT ATA AAA ATA GTA TAT GAT CCT TGT ACT 2930
Ile Ile Asp Leu Phe Ser Arg Ile Ile Lys Ile Val Tyr Asp Pro Cys Thr
895 900 905 910

GTC GAA ACA TCT AAC GAT TGT AGA TTA TTA AGA TTA TTG ATG GCC AAT ACA 2981
Val Glu Thr Ser Asn Asp Cys Arg Leu Leu Arg Leu Leu Met Ala Asn Thr
915 920 925

TCA TAA ATACATTATA ATATTATTAT AATATCAATC ATAATTTTTA TATATATTTT 3037
Ser

ATCTAAAAGG ACTTTTTTATT TTTTATATAT TAATAATAAT AA ATG AGT AAC GTA CCT 3094
Met Ser Asn Val Pro
930

TTA GCA ACC AAA ACA ATA AGA AAA TTA TCA AAT CGA AAA TAT GAA ATA AAG 3145
Leu Ala Thr Lys Thr Ile Arg Lys Leu Ser Asn Arg Lys Tyr Glu Ile Lys
935 940 945 950

ATT TAT TTA AAA GAT GAA AAT ACT TGT TTC GAA CGT GTA GTA GAT ATG GTA 3196
Ile Tyr Leu Lys Asp Glu Asn Thr Cys Phe Glu Arg Val Val Asp Met Val
955 960 965

```

6/20

Figure 2E

|                                                                     |      |
|---------------------------------------------------------------------|------|
| GTT CCA TTA TAT GAT GTG TGT AAT GAA ACT TCT GGT GTT ACT TTA GAA TCA | 3247 |
| Val Pro Leu Tyr Asp Val Cys Asn Glu Thr Ser Gly Val Thr Leu Glu Ser |      |
| 970 975 980                                                         |      |
| TGT AGT CCA AAT ATA GAA GTA ATT GAA TTA GAC AAT ACT CAT GTT AGA ATC | 3298 |
| Cys Ser Pro Asn Ile Glu Val Ile Glu Leu Asp Asn Thr His Val Arg Ile |      |
| 985 990 995 1000                                                    |      |
| AAA GTT CAC GGC GAT ACA TTA AAA GAA ATG TGT TTT GAA TTA TTG TTC CCG | 3349 |
| Lys Val His Gly Asp Thr Leu Lys Glu Met Cys Phe Glu Leu Leu Phe Pro |      |
| 1005 1010 1015                                                      |      |
| TGT AAT GTA AAC GAA GCC CAA GTA TGG AAA TAT GTA AGT CGA TTA TTG CTA | 3400 |
| Cys Asn Val Asn Glu Ala Gln Val Trp Lys Tyr Val Ser Arg Leu Leu     |      |
| 1020 1025 1030 1035                                                 |      |
| GAT AAT GTA TCA CAT AAT GAC GTA AAA TAT AAA TTA GCT AAT TTT AGA CTG | 3451 |
| Asp Asn Val Ser His Asn Asp Val Lys Tyr Lys Leu Ala Asn Phe Arg Leu |      |
| 1040 1045 1050                                                      |      |
| ACT CTT AAT GGA AAA CAT TTA AAA TTA AAA GAA ATC GAT CAA CCG CTA TTT | 3502 |
| Thr Leu Asn Gly Lys His Leu Lys Leu Lys Glu Ile Asp Gln Pro Leu Phe |      |
| 1055 1060 1065                                                      |      |
| ATT TAT TTT GTC GAT GAT TTG GGA AAT TAT GGA TTA ATT ACT AAG GAA AAT | 3553 |
| Ile Tyr Phe Val Asp Asp Leu Gly Asn Tyr Gly Leu Ile Thr Lys Glu Asn |      |
| 1070 1075 1080 1085                                                 |      |
| ATT CAA AAT AAT AAT TTA CAA GTT AAC AAA GAT GCA TCA TTT ATT ACT ATA | 3604 |
| Ile Gln Asn Asn Asn Leu Gln Val Asn Lys Asp Ala Ser Phe Ile Thr Ile |      |
| 1090 1095 1100                                                      |      |
| TTT CCA CAA TAT GCG TAT ATT TGT TTA GGT AGA AAA GTA TAT TTA AAT GAA | 3655 |
| Phe Pro Gln Tyr Ala Tyr Ile Cys Leu Gly Arg Lys Val Tyr Leu Asn Glu |      |
| 1105 1110 1115 1120                                                 |      |
| AAA GTA ACT TTT GAT GTA ACT ACA GAT GCA ACT AAT ATT ACT TTA GAT TTT | 3706 |
| Lys Val Thr Phe Asp Val Thr Thr Asp Ala Thr Asn Ile Thr Leu Asp Phe |      |
| 1125 1130 1135                                                      |      |
| AAT AAA TCT GTT AAT ATC GCA GTA TCA TTC CTT GAT ATA TAT TAC GAA GTT | 3757 |
| Asn Lys Ser Val Asn Ile Ala Val Ser Phe Leu Asp Ile Tyr Tyr Glu Val |      |
| 1140 1145 1150                                                      |      |
| AAT AAT AAT GAA CAA AAA GAT TTA TTA AAA GAT TTA CTT AAG AGA TAC GGT | 3808 |
| Asn Asn Asn Glu Gln Lys Asp Leu Leu Lys Asp Leu Leu Lys Arg Tyr Gly |      |
| 1155 1160 1165 1170                                                 |      |
| GAA TTT GAA GTC TAT AAC GCA GAT ACT GGA TTA ATT TAT GCT AAA AAT CTA | 3859 |
| Glu Phe Glu Val Tyr Asn Ala Asp Thr Gly Leu Ile Tyr Ala Lys Asn Leu |      |
| 1175 1180 1185                                                      |      |
| AGT ATT AAA AAT TAT GAT ACT GTG ATT CAA GTA GAA AGG TTG CCA GTT AAT | 3910 |
| Ser Ile Lys Asn Tyr Asp Thr Val Ile Gln Val Glu Arg Leu Pro Val Asn |      |
| 1190 1195 1200 1205                                                 |      |
| TTG AAA GTT AGA GCA TAT ACT AAG GAT GAA AAT GGT CGC AAT CTA TGT TTG | 3961 |
| Leu Lys Val Arg Ala Tyr Thr Lys Asp Glu Asn Gly Arg Asn Leu Cys Leu |      |
| 1210 1215 1220                                                      |      |

7/20

Figure 2F

RM58

|                                                                     |                                         |      |
|---------------------------------------------------------------------|-----------------------------------------|------|
| ATG AAA ATA ACA TCT AGT ACA                                         | GAA GTA GAC CCC GAG TAT GTA ACT AGT AAT | 4012 |
| Met Lys Ile Thr Ser Ser Thr                                         | Glu Val Asp Pro Glu Tyr Val Thr Ser Asn |      |
| 1225                                                                | 1230 1235                               |      |
| AAT GCT TTA TTG GGT ACG CTC AGA GTA TAT AAA AAG TTT GAT AAA TCT CAT | 4063                                    |      |
| Asn Ala Leu Leu Gly Thr Leu Arg Val Tyr Lys Lys Phe Asp Lys Ser His |                                         |      |
| 1240                                                                | 1245 1250 1255                          |      |
| TTA AAA ATT GTA ATG CAT AAC AGA GGA AGT GGT AAT GTA TTT CCA TTA AGA | 4114                                    |      |
| Leu Lys Ile Val Met His Asn Arg Gly Ser Gly Asn Val Phe Pro Leu Arg |                                         |      |
| 1260                                                                | 1265 1270                               |      |
| TCA TTA TAT CTG GAA TTG TCT AAT GTA AAA GGA TAT CCA GTT AAA GCA TCT | 4165                                    |      |
| Ser Leu Tyr Leu Glu Leu Ser Asn Val Lys Gly Tyr Pro Val Lys Ala Ser |                                         |      |
| 1275                                                                | 1280 1285 1290                          |      |
| GAT ACT TCG AGA TTA GAT GTT GGT ATT TAC AAA TTA AAT AAA ATT TAT GTA | 4216                                    |      |
| Asp Thr Ser Arg Leu Asp Val Gly Ile Tyr Lys Leu Asn Lys Ile Tyr Val |                                         |      |
| 1295                                                                | 1300 1305                               |      |
| GAT AAC GAC GAA AAT AAA ATT ATA TTG GAA GAA ATT GAA GCA GAA TAT AGA | 4267                                    |      |
| Asp Asn Asp Glu Asn Lys Ile Ile Leu Glu Glu Ile Glu Ala Glu Tyr Arg |                                         |      |
| 1310                                                                | 1315 1320                               |      |
| TGC GGA AGA CAA GTA TTC CAC GAA CGT GTA AAA CTT AAT AAA CAC CAA TGT | 4318                                    |      |
| Cys Gly Arg Gln Val Phe His Glu Arg Val Lys Leu Asn Lys His Gln Cys |                                         |      |
| 1325                                                                | 1330 1335 1340                          |      |
| AAA TAT ACT CCC AAA TGT CCA TTC CAA TTT GTT GTA AAC AGC CCA GAT ACT | 4369                                    |      |
| Lys Tyr Thr Pro Lys Cys Pro Phe Gln Phe Val Val Asn Ser Pro Asp Thr |                                         |      |
| 1345                                                                | 1350 1355                               |      |
| ACG ATT CAC TTA TAT GGT ATT TCT AAT GTT TGT TTA AAA CCT AAA GTA CCC | 4420                                    |      |
| Thr Ile His Leu Tyr Gly Ile Ser Asn Val Cys Leu Lys Pro Lys Val Pro |                                         |      |
| 1360                                                                | 1365 1370 1375                          |      |
| AAA AAT TTA AGA CTT TGG GGA TGG ATT TTA GAT TGC GAT ACT TCT AGA TTT | 4471                                    |      |
| Lys Asn Leu Arg Leu Trp Gly Trp Ile Leu Asp Cys Asp Thr Ser Arg Phe |                                         |      |
| 1380                                                                | 1385 1390                               |      |
| ATT AAA CAT ATG GCT GAT GGA TCT GAT GAT TTA GAT CTT GAC GTT AGG CTT | 4522                                    |      |
| Ile Lys His Met Ala Asp Gly Ser Asp Asp Leu Asp Leu Asp Val Arg Leu |                                         |      |
| 1395                                                                | 1400 1405                               |      |
| AAT AGA AAT GAT ATA TGT TTA AAA CAA GCC ATA AAA CAA CAT TAT ACT AAT | 4573                                    |      |
| Asn Arg Asn Asp Ile Cys Leu Lys Gln Ala Ile Lys Gln His Tyr Thr Asn |                                         |      |
| 1410                                                                | 1415 1420 1425                          |      |
| GTA ATT ATA TTA GAG TAC GCA AAT ACA TAT CCA AAT TGC ACA TTA TCA TTG | 4624                                    |      |
| Val Ile Ile Leu Glu Tyr Ala Asn Thr Tyr Pro Asn Cys Thr Leu Ser Leu |                                         |      |
| 1430                                                                | 1435 1440                               |      |
| GGT AAT AAT AGA TTT AAT AAT GTA TTT GAT ATG AAT GAT AAC AAA ACT ATA | 4675                                    |      |
| Gly Asn Asn Arg Phe Asn Asn Val Phe Asp Met Asn Asp Asn Lys Thr Ile |                                         |      |
| 1445                                                                | 1450 1455 1460                          |      |
| TCT GAG TAT ACT AAC TTT ACA AAA AGT AGA CAA GAC CTT AAT AAC ATG TCA | 4726                                    |      |
| Ser Glu Tyr Thr Asn Phe Thr Lys Ser Arg Gln Asp Leu Asn Asn Met Ser |                                         |      |
| 1465                                                                | 1470 1475                               |      |

8/20

Figure 2G

|                                                                     |      |
|---------------------------------------------------------------------|------|
| TGT ATA TTA GGA ATA AAC ATA GGT AAT TCC GTA AAT ATT AGT AGT TTG CCT | 4777 |
| Cys Ile Leu Gly Ile Asn Ile Gly Asn Ser Val Asn Ile Ser Ser Leu Pro |      |
| 1480 1485 1490                                                      |      |
| GGT TGG GTA ACA CCT CAC GAA GCT AAA ATT CTA AGA TCT GGT TGT GCT AGA | 4828 |
| Gly Trp Val Thr Pro His Glu Ala Lys Ile Leu Arg Ser Gly Cys Ala Arg |      |
| 1495 1500 1505 1510                                                 |      |
| GTT AGA GAA TTT TGT AAA TCA TTC TGT GAT CTT TCT AAT AAG AGA TTC TAT | 4879 |
| Val Arg Glu Phe Cys Lys Ser Phe Cys Asp Leu Ser Asn Lys Arg Phe Tyr |      |
| 1515 1520 1525                                                      |      |
| GCT ATG GCT AGA GAT CTC GTA AGT TTA CTA TTT ATG TGT AAC TAT GTT AAT | 4930 |
| Ala Met Ala Arg Asp Leu Val Ser Leu Leu Phe Met Cys Asn Tyr Val Asn |      |
| 1530 1535 1540 1545                                                 |      |
| ATT GAA ATT AAC GAA GCA GTA TGC GAA TAT CCT GGA TAT GTC ATA TTA TTC | 4981 |
| Ile Glu Ile Asn Glu Ala Val Cys Glu Tyr Pro Gly Tyr Val Ile Leu Phe |      |
| 1550 1555 1560                                                      |      |
| GCA AGA GCT ATT AAA GTA ATT AAT GAT TTA TTA TTA ATT AAC GGA GTA GAT | 5032 |
| Ala Arg Ala Ile Lys Val Ile Asn Asp Leu Leu Leu Ile Asn Gly Val Asp |      |
| 1565 1570 1575                                                      |      |
| AAT CTA GCA GGA TAT TCA ATT TCC TTA CCT ATA CAT TAT GGA TCT ACT GAA | 5083 |
| Asn Leu Ala Gly Tyr Ser Ile Ser Leu Pro Ile His Tyr Gly Ser Thr Glu |      |
| 1580 1585 1590 1595                                                 |      |
| AAG ACT CTA CCA AAT GAA AAG TAT GGT GGT GTT GAT AAG AAA TTT AAA TAT | 5134 |
| Lys Thr Leu Pro Asn Glu Lys Tyr Gly Gly Val Asp Lys Lys Phe Lys Tyr |      |
| 1600 1605 1610                                                      |      |
| CTA TTC TTA AAG AAT AAA CTA AAA GAT TTA ATG CGT GAT GCT GAT TTT GTC | 5185 |
| Leu Phe Leu Lys Asn Lys Leu Lys Asp Leu Met Arg Asp Ala Asp Phe Val |      |
| 1615 1620 1625 1630                                                 |      |
| CAA CCT CCA TTA TAT ATT TCT ACT TAC TTT AGA ACT TTA TTG GAT GCT CCA | 5236 |
| Gln Pro Pro Leu Tyr Ile Ser Thr Tyr Phe Arg Thr Leu Leu Asp Ala Pro |      |
| 1635 1640 1645                                                      |      |
| CCA ACT GAT AAT TAT GAA AAA TAT TTG GTT GAT TCG TCC GTA CAA TCA CAA | 5287 |
| Pro Thr Asp Asn Tyr Glu Lys Tyr Leu Val Asp Ser Ser Val Gln Ser Gln |      |
| 1650 1655 1660                                                      |      |
| GAT GTT CTA CAG GGT CTG TTG AAT ACA TGT AAT ACT ATT GAT ACT AAT GCT | 5338 |
| Asp Val Leu Gln Gly Leu Leu Asn Thr Cys Asn Thr Ile Asp Thr Asn Ala |      |
| 1665 1670 1675 1680                                                 |      |
| AGA GTT GCA TCA AGT GTT ATT GGA TAT GTT TAT GAA CCA TGC GGA ACA TCA | 5389 |
| Arg Val Ala Ser Ser Val Ile Gly Tyr Val Tyr Glu Pro Cys Gly Thr Ser |      |
| 1685 1690 1695                                                      |      |
| GAA CAT AAA ATT GGT TCA GAA GCA TTG TGT AAA ATG GCT AAA GAA GCA TCT | 5440 |
| Glu His Lys Ile Gly Ser Glu Ala Leu Cys Lys Met Ala Lys Glu Ala Ser |      |
| 1700 1705 1710 1715                                                 |      |
| AGA TTA GGA AAT CTA GGT TTA GTA AAT CGT ATT AAT GAA AGT AAT TAC AAC | 5491 |
| Arg Leu Gly Asn Leu Gly Leu Val Asn Arg Ile Asn Glu Ser Asn Tyr Asn |      |
| 1720 1725 1730                                                      |      |

9/20

## Figure 2H

AAA TGT AAT AAA TAT GGT TAT AGA GGA GTA TAC GAA AAT AAC AAA CTA AAA 5542  
 Lys Cys Asn Lys Tyr Gly Tyr Arg Gly Val Tyr Glu Asn Asn Lys Leu Lys  
 1735 1740 1745

ACA AAA TAT TAT AGA GAA ATA TTT GAT TGT AAT CCT AAT AAT AAT AAT GAA 5593  
 Thr Lys Tyr Tyr Arg Glu Ile Phe Asp Cys Asn Pro Asn Asn Asn Asn Glu  
 1750 1755 1760 1765

TTA ATA TCC AGA TAT GGA TAT AGA ATA ATG GAT TTA CAT AAA ATT GGA GAA 5644  
 Leu Ile Ser Arg Tyr Gly Tyr Arg Ile Met Asp Leu His Lys Ile Gly Glu  
 1770 1775 1780

ATT TTT GCA AAT TAC GAT GAA AGT GAA TCT CCT TGC GAA CGA AGA TGT CAT 5695  
 Ile Phe Ala Asn Tyr Asp Glu Ser Glu Ser Pro Cys Glu Arg Arg Cys His  
 1785 1790 1795 1800

TAC TTG GAA GAT AGA GGT CTT TTA TAT GGT CCT GAA TAT GTA CAT CAC AGA 5746  
 Tyr Leu Glu Asp Arg Gly Leu Leu Tyr Gly Pro Glu Tyr Val His His Arg  
 1805 1810 1815

TAT CAA GAA TCA TGT ACG CCT AAT ACG TTT GGA AAT AAC ACA AAT TGT GTA 5797  
 Tyr Gln Glu Ser Cys Thr Pro Asn Thr Phe Gly Asn Asn Thr Asn Cys Val  
 1820 1825 1830

ACA AGA AAT GGT GAA CAA CAC GTA TAC GAA AAT AGT TGT GGA GAT AAT GCA 5848  
 Thr Arg Asn Gly Glu Gln His Val Tyr Glu Asn Ser Cys Gly Asp Asn Ala  
 1835 1840 1845 1850

ACA TGT GGA AGA AGA ACA GGA TAT GGA AGA AGA AGT AGG GAT GAA TGG AAT 5899  
 Thr Cys Gly Arg Arg Thr Gly Tyr Gly Arg Arg Ser Arg Asp Glu Trp Asn  
 1855 1860 1865

GAC TAT AGA AAA CCC CAC GTT TAT GAC AAT TGT GCC GAT GCA AAT AGT TCA 5950  
 Asp Tyr Arg Lys Pro His Val Tyr Asp Asn Cys Ala Asp Ala Asn Ser Ser  
 1870 1875 1880 1885

TCT TCA GAT AGC TGT TCA GAC AGT AGT AGT AGT AGT GAA TCT GAA TCT GAT 6001  
 Ser Ser Asp Ser Cys Ser Asp Ser Ser Ser Ser Ser Glu Ser Glu Ser Asp  
 1890 1895 1900

TCA GAT GGA TGT TGC GAC ACA GAT GCT AGT TTA GAT TCT GAT ATT GAA AAT 6052  
 Ser Asp Gly Cys Cys Asp Thr Asp Ala Ser Leu Asp Ser Asp Ile Glu Asn  
 1905 1910 1915

TGT TAT CAA AAT CCA TCA AAA TGT GAT GCA GGA TGC TAA ATGAAATTTA 6101  
 Cys Tyr Gln Asn Pro Ser Lys Cys Asp Ala Gly Cys  
 1920 1925 1930

ATATTATATA ATATTAACCTT ACAAGTTATA AAAATCATTAA AAATGATTTT TTAAAATGAT 6161

ATTATCGATA GTTGTGATAA TGTGCTCTTT TATTTTATTA ATTGCGATGA TTATAATATT 6221

ATCTTTTAGA TATATTTAAT ATTAATTATA AATCGACTGA CAATAATATT TATTC CTA 6279

TTC ATA ATA ATC ATC TGC TAT ATA TAT TAA TGT ATC ATT CTC TAT TAT AAA 6330  
 Glu Tyr Tyr Asp Asp Ala Ile Tyr Ile Leu Thr Asp Asn Glu Ile Ile Phe  
 1935 1940 1945

TAT AGG TAT ATT GTC TTT ATC AAT CAT TAA TTT TGC TAC AGC TGT ATT ATC 6381  
 Ile Pro Ile Asn Asp Lys Asp Ile Met Leu Lys Ala Val Ala Thr Asn Asp  
 1950 1955 1960 1965

10/20

Figure 2I

|                                                                     |      |
|---------------------------------------------------------------------|------|
| TTT ATA TAC TAT ATT TGT GTC TTT GTT TAA TAA ACC TTT TAA TAT AGT GGC | 6432 |
| Lys Tyr Val Ile Asn Thr Asp Lys Asn Leu Leu Gly Lys Leu Ile Thr Ala |      |
| 1970 1975 1980                                                      |      |
| TCT ATC ATA ATC TTT ACA ATA TGA TAT GGG ATA TAA TTT TAT ATT AAT AAT | 6483 |
| Arg Asp Tyr Asp Lys Cys Tyr Ser Ile Pro Tyr Leu Lys Ile Asn Ile Ile |      |
| 1985 1990 1995                                                      |      |
| AAC ATT AGA TAC GTT CAT TTC TTT CAT TCT AGT TTT ACG TAT TGT GTC AAA | 6534 |
| Val Asn Ser Val Asn Met Glu Lys Met Arg Thr Lys Arg Ile Thr Asp Phe |      |
| 2000 2005 2010 2015                                                 |      |
| AAT TAT TTC ATT TTC TGC TGG TTC TAT ATA TTT ATA TGT GTT ATG AAT AGA | 6585 |
| Ile Ile Glu Asn Glu Ala Pro Glu Ile Tyr Lys Tyr Thr Asn His Ile Ser |      |
| 2020 2025 2030                                                      |      |
| TTC GAT AGA TGA TGA TTT TAA TAA ATC AAA TAT AAC ATT TAT TTT ACC TTG | 6636 |
| Glu Ile Ser Ser Ser Lys Leu Leu Asp Phe Ile Val Asn Ile Lys Gly Gln |      |
| 2035 2040 2045 2050                                                 |      |
| TTT ATC TTT TAT AAT ATC TAA TAT TTC TTT ATC TAC AGA TTT TCT GTT GTT | 6687 |
| Lys Asp Lys Ile Ile Asp Leu Ile Glu Lys Asp Val Ser Lys Arg Asn Asn |      |
| 2055 2060 2065                                                      |      |
| GGT ATA TGA TAT TAA AAA ATG AAC GTT AAC ATA TCT ATA TTC TTG TGG TAA | 6738 |
| Thr Tyr Ser Ile Leu Phe His Val Asn Val Tyr Arg Tyr Glu Gln Pro Leu |      |
| 2070 2075 2080                                                      |      |
| ATC TTT ATG AGA ATT TAA TCT TAT AGA TCT                             | 6768 |
| Asp Lys His Ser Asn Leu Arg Ile Ser Arg                             |      |
| 2085 2090 2094                                                      |      |

11/20

## Figure 3A

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| GAATTCAAGT TAAATAT TTA TAA ACA ACA ATC ATA TTT TTT TAA AGA ATC TAA  | 53  |
| Leu Cys Cys Asp Tyr Lys Lys Leu Ser Asp Leu                         |     |
| 1 5 10                                                              |     |
| TAA ATT TTT TAA CAT TTT ATT ATT ATT TGA TAA TTG TTT ATT TAA TTC GTT | 104 |
| Leu Asn Lys Leu Met Lys Asn Asn Asn Ser Leu Gln Lys Asn Leu Glu Asn |     |
| 15 20 25                                                            |     |
| ATT GAT ATT AAC AAT ATT ATT TAT CAT TTT ACC TAT TTT TTT TCT ATC     | 155 |
| Asn Ile Asn Val Ile Asn Asn Ile Met Lys Gly Ile Lys Lys Lys Arg Asp |     |
| 30 35 40 45                                                         |     |
| RM129                                                               |     |
| TAC TAA CGA AAT ATC AGA TTT TGC ACC TTC AAT ATC AGA ATA ATA ATT ATC | 206 |
| Val Leu Ser Ile Asp Ser Lys Ala Gly Glu Ile Asp Ser Tyr Tyr Asn Asp |     |
| 50 55 60                                                            |     |
| < ORF Q1                                                            |     |
| ATT ATT TTG CAT TTATGAATAA AAATA TTA ATA TGA ATT ATT ATA ACA TAA    | 257 |
| Asn Asn Gln Met Tyr Ser Asn Asn Tyr Cys Leu                         |     |
| 65 70                                                               |     |
| TCT ACA CAC AGG AAC ATA TAA ATC TTG TCC ACC TAT TTC AAT TAT TTG ATT | 308 |
| Arg Cys Val Pro Val Tyr Leu Asp Gln Gly Gly Ile Glu Ile Ile Gln Asn |     |
| 75 80 85 90                                                         |     |
| TTT ATT ATG TTT TTT AAT TGT AAA AGA AGC ATC TTT ATA ACA AAA TTG ACA | 359 |
| Lys Asn His Lys Lys Ile Thr Phe Ser Ala Asp Lys Tyr Cys Phe Gln Cys |     |
| 95 100 105                                                          |     |
| TAT AGC TTG TAA TTT TTT TAT TTT TTC TAC TTT AGG AAT TAA TTT TGA TAT | 410 |
| Ile Ala Gln Leu Lys Lys Ile Lys Glu Val Lys Pro Ile Leu Lys Ser Ile |     |
| 110 115 120                                                         |     |
| RM03                                                                |     |
| AGA ATT AAA TAT ATT TCT GTT AAA GTC ACA ATT TAA TCC AGC AAC AAT AAC | 461 |
| Ser Asn Phe Ile Asn Arg Asn Phe Asp Cys Asn Leu Gly Ala Val Ile Val |     |
| 125 130 135 140                                                     |     |
| TTT TTT TTT ATT ATT AGC CAT TTT ATC ACA AAA TTG TTC TAA ATC ATT TTC | 512 |
| Lys Lys Lys Asn Asn Ala Met Lys Asp Cys Phe Gln Glu Leu Asp Asn Glu |     |
| 145 150 155                                                         |     |
| TTC AAA AAA TTG ACA CTC ATC TAT GCC AAT AAT ATC ATA ATT ATC TAC GAT | 563 |
| Glu Phe Phe Gln Cys Glu Asp Ile Gly Ile Ile Asp Tyr Asn Asp Val Ile |     |
| 160 165 170 175                                                     |     |
| ATT GAT TTC ATT AAT TAA ATT ATT TGT TTT AAT GTA TAA ATA TTC TTT ATT | 614 |
| Asn Ile Glu Asn Ile Leu Asn Asn Thr Lys Ile Tyr Leu Tyr Glu Lys Asn |     |
| 180 185 190                                                         |     |
| TAA TAT ATT TCC GTC ATG ATT TAT TAT ATT TTT ATT TAT AAA TCT ATT ATC | 665 |
| Leu Ile Asn Gly Asp His Asn Ile Ile Asn Lys Asn Ile Phe Arg Asn Asp |     |
| 195 200 205                                                         |     |
| TAT ATT ATG AGT TAT AAT TAC ACA TTT TTG ATT AGA TAA AAT ATA TCT ATT | 716 |
| Ile Asn His Thr Ile Ile Val Cys Lys Gln Asn Ser Leu Ile Tyr Arg Asn |     |
| 210 215 220 225                                                     |     |

SUBSTITUTE SHEET

12/20

Figure 3B

|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
|------------|------------|-----------|-----|-----|-------------|------------|------------|------------|-----|-----|-----|-----|-----|-----|----------|-----|------|--|
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     | RMO4     |     |      |  |
| AAT        | TTT        | TCG       | CAT | CAA | TTC         | TGT        | TGT        | TTT        | GCC | AGA | AAA | CAT | AGG | ACC | AAT      | TAT | 767  |  |
| Ile        | Lys        | Arg       | Met | Leu | Glu         | Thr        | Thr        | Lys        | Gly | Ser | Phe | Met | Pro | Gly | Ile      | Ile |      |  |
|            |            |           |     | 230 |             |            |            |            | 235 |     |     |     |     | 240 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     | < ORF Q2 |     |      |  |
| TAA        | TTC        | TAT       | CGA | CAT | TTTTTTTTTAT | TATTTGATAT | ATTTTTTCAA | AAAAAAATTA |     |     |     |     |     |     |          |     | 822  |  |
| Leu        | Glu        | Ile       | Ser | Met |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
|            |            |           |     | 245 |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     | ORF Q3 > |     |      |  |
| ATCAATGAAA | AAAAAATAAA | ATTATCAAA | ATG | GAT | TTA         | CTA        | AAT        | TCT        | GAT | ATA | ATT |     |     |     |          |     | 878  |  |
|            |            |           | Met | Asp | Leu         | Leu        | Asn        | Ser        | Asp | Ile | Ile |     |     |     |          |     |      |  |
|            |            |           |     | 250 |             |            |            |            | 255 |     |     |     |     |     |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| TTA        | ATA        | AAT       | ATT | TTA | AAA         | TAT        | TAT        | AAT        | TTA | AAA | AAA | ATA | ATA | ATA | AAC      | AGA | 929  |  |
| Leu        | Ile        | Asn       | Ile | Leu | Lys         | Tyr        | Tyr        | Asn        | Leu | Lys | Lys | Ile | Ile | Ile | Asn      | Arg |      |  |
|            |            |           |     | 260 |             |            |            |            | 265 |     |     |     |     | 270 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| GAT        | AAT        | GTT       | ATT | AAT | ATT         | AAT        | ATA        | TTA        | AAA | AAA | TTA | GTT | AAT | TTA | GAA      | GAA | 980  |  |
| Asp        | Asn        | Val       | Ile | Asn | Ile         | Asn        | Ile        | Leu        | Lys | Lys | Leu | Val | Asn | Leu | Glu      | Glu |      |  |
|            |            |           |     | 275 |             |            |            |            | 280 |     |     |     |     | 285 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| TTG        | CAT        | ATA       | ATA | TAT | TAT         | GAT        | AAT        | AAT        | ATT | TTA | AAT | AAT | ATT | CCA | GAA      | AAT | 1031 |  |
| Leu        | His        | Ile       | Ile | Tyr | Tyr         | Asp        | Asn        | Asn        | Ile | Leu | Asn | Asn | Ile | Pro | Glu      | Asn |      |  |
|            |            |           |     | 295 |             |            |            |            | 300 |     |     |     |     | 305 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| ATT        | AAA        | AGT       | TTA | TAT | ATT         | TCA        | AAT        | TTA        | AAT | ATT | ATT | AAT | TTA | AAT | TTT      | ATA | 1082 |  |
| Ile        | Lys        | Ser       | Leu | Tyr | Ile         | Ser        | Asn        | Leu        | Asn | Ile | Ile | Asn | Leu | Asn | Phe      | Ile |      |  |
|            |            |           |     | 310 |             |            |            |            | 315 |     |     |     |     | 320 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| ACA        | AAA        | TTA       | AAA | AAT | ATA         | ACA        | TAT        | TTA        | GAT | ATA | TCT | TAT | AAC | AAA | AAT      | AGC | 1133 |  |
| Thr        | Lys        | Leu       | Lys | Asn | Ile         | Thr        | Tyr        | Leu        | Asp | Ile | Ser | Tyr | Asn | Lys | Asn      | Ser |      |  |
|            |            |           |     | 330 |             |            |            |            | 335 |     |     |     |     | 340 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| AAT        | ATA        | AGT       | AAT | ATT | ATA         | CTA        | CCA        | CAT        | TCT | ATA | GAA | TTT | TTA | AAT | TGT      | GAA | 1184 |  |
| Asn        | Ile        | Ser       | Asn | Ile | Ile         | Leu        | Pro        | His        | Ser | Ile | Glu | Phe | Leu | Asn | Cys      | Glu |      |  |
|            |            |           |     | 345 |             |            |            |            | 350 |     |     |     |     | 355 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| TCA        | TGT        | AAT       | ATA | AAT | GAC         | TAT        | AAT        | TTT        | ATT | AAT | AAT | TTA | GTA | AAT | TTA      | AAA | 1235 |  |
| Ser        | Cys        | Asn       | Ile | Asn | Asp         | Tyr        | Asn        | Phe        | Ile | Asn | Asn | Leu | Val | Asn | Leu      | Lys |      |  |
|            |            |           |     | 360 |             |            |            |            | 365 |     |     |     |     | 370 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| AAA        | TTA        | ATA       | ATA | TCT | AAA         | AAT        | AAA        | TTT        | GGT | AAC | TTT | AAT | AAT | GTT | TTT      | CCT | 1286 |  |
| Lys        | Leu        | Ile       | Ile | Ser | Lys         | Asn        | Lys        | Phe        | Gly | Asn | Phe | Asn | Asn | Val | Phe      | Pro |      |  |
|            |            |           |     | 380 |             |            |            |            | 385 |     |     |     |     | 390 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| ATT        | AGT        | ATA       | GTT | GAG | TTA         | AAT        | ATG        | GAA        | TCA | ATA | CAA | ATA | AAA | GAT | TAT      | AAA | 1337 |  |
| Ile        | Ser        | Ile       | Val | Glu | Leu         | Asn        | Met        | Glu        | Ser | Ile | Gln | Ile | Lys | Asp | Tyr      | Lys |      |  |
|            |            |           |     | 395 |             |            |            |            | 400 |     |     |     |     | 405 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| TTT        | ATA        | GAA       | AAA | TTA | ATT         | AAT        | TTA        | AAA        | AAA | TTA | GAT | ATA | TCT | TTC | AAT      | GTT | 1388 |  |
| Phe        | Ile        | Glu       | Lys | Leu | Ile         | Asn        | Leu        | Lys        | Lys | Leu | Asp | Ile | Ser | Phe | Asn      | Val |      |  |
|            |            |           |     | 415 |             |            |            |            | 420 |     |     |     |     | 425 |          |     |      |  |
|            |            |           |     |     |             |            |            |            |     |     |     |     |     |     |          |     |      |  |
| AAA        | AAA        | AAT       | AAT | ATA | CAT         | TTG        | ATA        | AAA        | TTT | CCA | AAA | AGT | ATA | ACT | CAT      | TTA | 1439 |  |
| Lys        | Lys        | Asn       | Asn | Ile | His         | Leu        | Ile        | Lys        | Phe | Pro | Lys | Ser | Ile | Thr | His      | Leu |      |  |
|            |            |           |     | 430 |             |            |            |            | 435 |     |     |     |     | 440 |          |     |      |  |

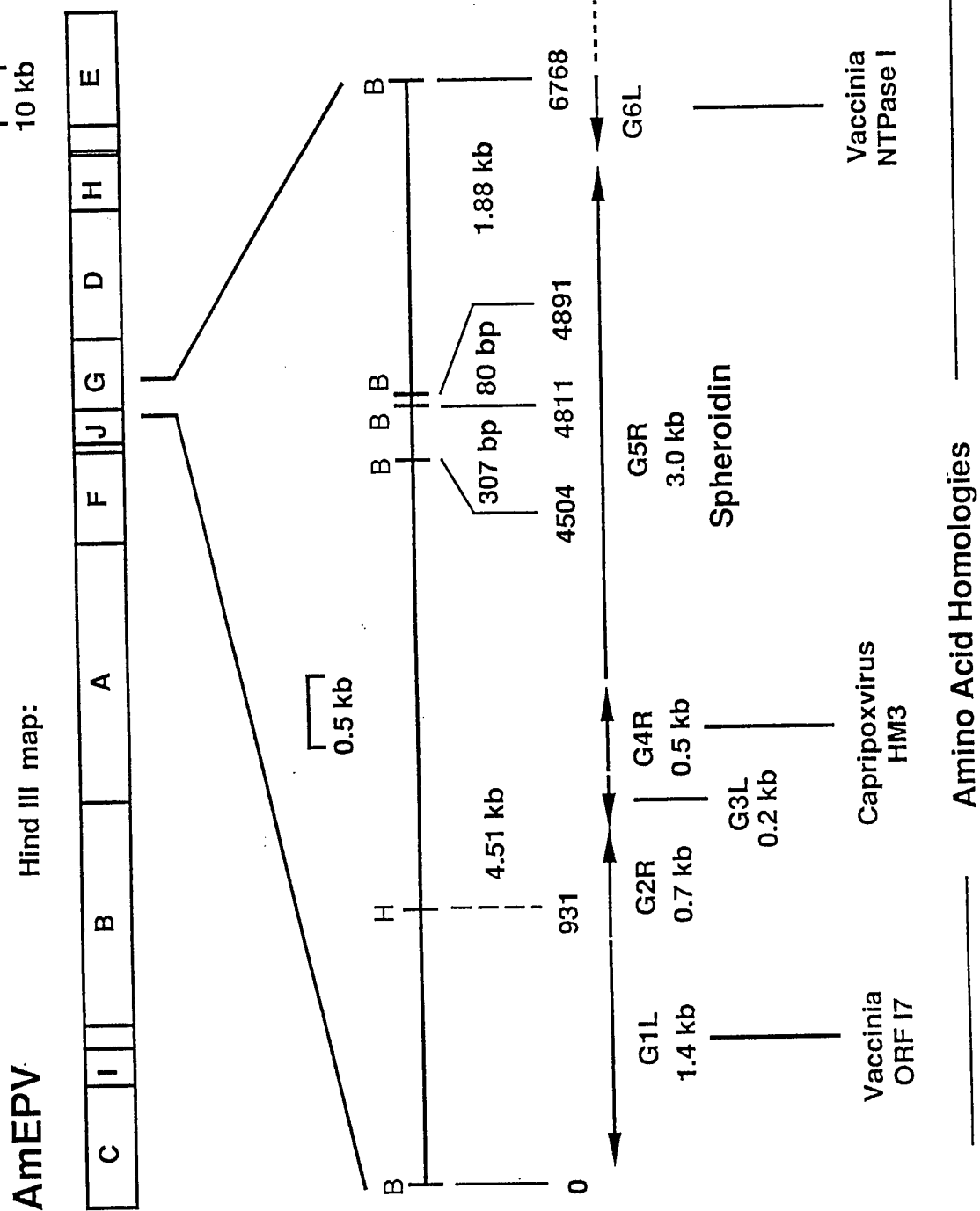
13/20

## Figure 3C

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| TGT | GAT | TAT | CAA | TCA | TAT | AAA | GAA | AAT | TAT | AAT | TAT | TTA | AAA | AAT | TTA | TCA | 1490 |
| Cys | Asp | Tyr | Gln | Ser | Tyr | Lys | Glu | Asn | Tyr | Asn | Tyr | Leu | Lys | Asn | Leu | Ser |      |
| 445 |     |     |     |     | 450 |     |     |     | 455 |     |     |     |     | 460 |     |     |      |
| AAT | ATA | ATT | GAA | TAT | GAA | TTC |     |     |     |     |     |     |     |     |     |     | 1511 |
| Asn | Ile | Ile | Glu | Tyr | Glu | Phe |     |     |     |     |     |     |     |     |     |     |      |
|     |     |     | 465 |     |     |     |     |     |     |     |     |     |     |     |     |     |      |

14/20

Figure 4



15/20

Figure 5

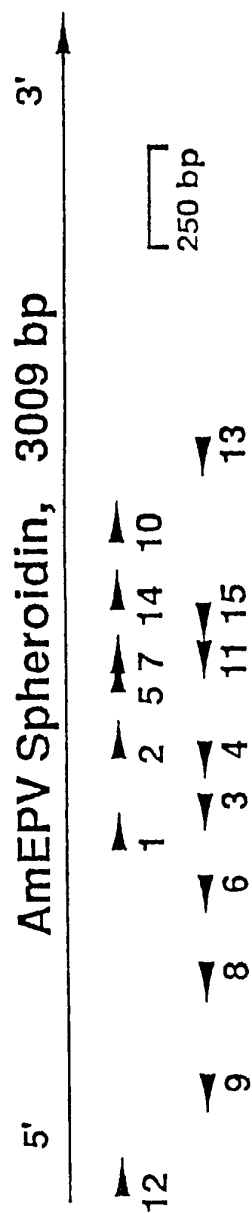




Figure 6B

|       |     |           |              |      |        |         |      |      |     |       |       |      |    |      |      |       |     |
|-------|-----|-----------|--------------|------|--------|---------|------|------|-----|-------|-------|------|----|------|------|-------|-----|
| AmEPV | 323 | KFDKSHLKI | VMHNRGSGNVFP | PLRS | LYLELS | NVKGYPV | KASD | TSRL | DVG | IYKLN | KIYVD | NDEN | KI | ILEE | IEAE | YRCGR | 399 |
|       |     |           |              |      |        |         |      |      |     |       |       |      |    |      |      |       |     |
| CbEPV | 1   | KFDKSHLKI | VMHNRGSGNVFP | PLRS | LYLELS | NVKGYPV | KASD | TSRL | DVG | IYKLN | KIYVD | NDEN | KI | ILEE | IEAE | YRCGR | 77  |
|       |     |           |              |      |        |         |      |      |     |       |       |      |    |      |      |       |     |
| CfEPV | 1   | KFDKSHLKI | VMHNRGSGNVFP | PLRS | LYLELS | NVKGYPV | KASD | TSRL | DVG | IYKLN | KIYVD | NDEN | KI | ILEE | IEAE | YRCGR | 77  |

Figure 6C

|       |     |            |            |     |             |             |     |             |             |
|-------|-----|------------|------------|-----|-------------|-------------|-----|-------------|-------------|
| AmEPV | 211 | 221        | KFKYLFLKNK | 682 | 691         | KSVNIAVSFLD | 726 | 736         | KYLVDSSVQSQ |
|       |     |            |            |     |             |             |     |             |             |
| CbEPV | 1   | KFKYLFLKNK | 10         | 1   | KSVNIAVSFLD | 11          | 1   | KYLVDSSVQSQ | 11          |

18/20

Figure 7A

6769 G I I Q K L E S E N W P M D L I I I D N I C K F S I G E S G A Y S  
T C C T A T T A T T G T T T A A T T C T G A T T C C A C G G C A T A T C T A A T T A T A T A T C A T T G A A T G A T A T G C C T T C A G A T C C A G G T A A G A A

6869 F I C V K V K K G N N E Y N N Y E N L E N D R T K L T K S S Y E  
A A T A T G C A A A C T T T T A C T T T T T A C C A T T A T T A T T A T T C G T T T A A T T C A T T A T C T A G T T T T T A A A G T T T T G C T A G A A T A T T C A A

6969 I Y S I N F C N F Y C K L S S I G E F N V L P E F I L V K G R S N L  
T A T A A G A A A T A T T A A A C A A T T A A A A C A T T T T A A A C T T G A T A T T C C T T C A A A A T T A A C T A A A G T T C A A A T A T T A A T A C T T T T C C T C T C G A A I T T A A

7069 I I K C T E I Y K C S Y Q Y L I N Y N N I D T I P L N T K I K I N  
A A T T A T T T T A C A A G T T T C T A T A T A T T T A C A C G A A T T G A T A T A T A T A T T A T A A T T A T T A T A T C A G T G A T T G G T A A A T T A G T T T T T A T T T T A T A T A T T A

7169 D N K F S E I F S E S F N I N K T F E S F E A L K R K I M D N Y E  
T C A T T T T T A A A C T T T C A T A A A G A T T C A G A G A A T T A T A T T T T T G T A A A C T C G G A A A T T C A G C A A G T T T T C T T T T A A I C A T A T C A T T A T A T T C T A

7269 I N D L D G K I K L D Y F A F S S I L R R M T K L G G L E T K Y D Y  
T A T T A T C T A A A T C C T T T T A T T T A A G A T C A T A A A A A G C A A A T G A A G A T A T T A A T C T T C T C A T A G T T T T T A A A C C A C C T A A T T C A G T T T T A T A T A T C A T A

7369 K E A M N Y L K S Q E D S M I I N H Y F I N K K A Y G D I Y N T E  
T T T T C T G C C A T A T T A T A A T T T A G A T T G C T C A T C G A C A T A A T T A T A T T A T A T A T A T A T A T A T T T T T T T G C A T A T C C A T C T A T A T A A T T T G T T T C T

7469 T L S D A E I L R K Y S C I A L L T E R L E K F N I L K G N N I Y  
G T T A A C T A T C T G C T T C T A T A T C T T T T A T A A G A C A T A T A G C T A A T A A T G T T T C T T A A T T C C T T A A A A T T A A T T A A C T T T C C A T T A T T A T T A T A T T

7569 E E K I N M V N P R L L G I L N N F E S I N N T V P T A S M C L I K  
C T T C T T T A T A T T C A T A A C A T T T G G T C T A A G T A A C C T A T T A A A T T A T T A A A T T C A G A A A T A T T A T T A G T T A C T G G A G T A G C G G A C A T A C A T A A T A T T T T

7669 N N E F N A L K I L K K Y I P T F N R E N N D K K V T R S I L K H  
A T T A T T T C G A A A T T T G C T A A T T T T T A T A A A T A G G A G T A A A A A T T T C T T T C G T T A T T A T C T T T T T T A A C A G T T C T T G A T A T A T T A A T T A T G A

19/20

Figure 7B

V E D I I I L L R S K K N L S S E I S R Y I N N F K D L S S S D  
7769 ACTTCGCTATTATTAGTAATCTACTTTTTTTATTAGAGAACTTTCTATAGATCTATATATATTATAAATTTATCTAACTAGATGACGAATCAT

Y Y I F K I N S T D S I Y S R I T N L W P. D I Y L S K K I F I L I I  
7869 AATATATAAATTTTATATTACTGGTATCTGATATATAATGAICTTATAGTATTTAACCAAGGATCTATGTATAATGATTTTTTAAATAAATATTAAATAT

W R P F L E K I Y K I I Y V A T L T K G M G T D W F L L M S N L N  
7969 CCATCTTGGAAATAATCTTTTATATATTTTATAATATACACAGCAGTTAATGTTTTTCCCATACCAGTATCCCAAATAATAACATACTATTCAAATTT

K L G I F I R S V F Y Q Y D Q L T I E T N T I N N M I K N P L H Q  
8069 TTTAATCCTATGAATATCTACTTACAAAATATTGATAATCTTGTAAATGTAATTTTCAGTATTTGTAAATATTATTCATAAATTTTATTAGGCAATGTTGTG

T K D L A Y N I H K G V I S D L A F M  
8169 TTTTATCAAGTGCATAATTTATATGTTTACCACAAATAGAAATCTAATGCAAAACATTTAGTTATATAAAAAATAATATTATTAATTAAGATGTTTCA

T T A A T T T A T G T C T G T G A T G T G G A G T T A A A C C C A A G A T A T T G A T A T A T C T A T A T C A T T A A T T C T T T T T G A A T C T A T G T C T A T C A A T C G C A A A T T A T  
8269

C C C A G T A T A A T T T C G A G T T T G T T T G C A G C A T A T A A C C A A C A T A C A T A A T G T G G A G T T T T G T G G T T C G G A T G A A A A G C G T A C T T T T 8457  
8369

Figure 8

