



(22) **Date de dépôt/Filing Date:** 1998/09/18  
(41) **Mise à la disp. pub./Open to Public Insp.:** 1999/03/25  
(45) **Date de délivrance/Issue Date:** 2018/11/13  
(62) **Demande originale/Original Application:** 2 304 256  
(30) **Priorité/Priority:** 1997/09/19 (US60/059,353)

(51) **Cl.Int./Int.Cl. C07K 16/12** (2006.01),  
**C07K 14/29** (2006.01)  
(72) **Inventeurs/Inventors:**  
RIKIHISA, YASUKO, US;  
OHASHI, NORIO, US  
(73) **Propriétaire/Owner:**  
OHIO STATE RESEARCH FOUNDATION, US  
(74) **Agent:** NORTON ROSE FULBRIGHT CANADA  
LLP/S.E.N.C.R.L., S.R.L.

(54) **Titre : PROTEINE DE MEMBRANE EXTERNE DE EHRlichIA CANIS ET EHRlichIA CHAFFEENSIS**  
(54) **Title: OUTER MEMBRANE PROTEIN OF EHRlichIA CANIS AND EHRlichIA CHAFFEENSIS**

(57) **Abrégé/Abstract:**

The present invention relates to diagnostic tools for veterinary and human use which are used for serodiagnosing ehrlichiosis in mammals, particularly in members of the Canidae family and in humans. The present invention also provides polynucleotides which encode the outer membrane proteins of *E. chaffeensis*. The polynucleotides encode an OMP-1 family of proteins of *E. chaffeensis* and P30 family of proteins of *E. canis*. The present invention also provides the following isolated proteins of *E. chaffeensis* OMP-1, OMP-1A, OMP-1B, OMP-1C, OMP-1D, OMP-1E, OMP-1F, OMP-1R, OMP-1S, OMP-1T, OMP-1V, OMP-1W, and OMP-1Z, referred to hereinafter collectively as the "OMP family". The present invention also provides the following isolated proteins of *E. canis* P30, P30-1, P30-2, P30-3, P30-4, P30-5, P30-6, P30-7, P30-8, P30-9, and P30-10, referred to hereinafter as the P30 family. The present invention also relates to an assay for diagnosing ehrlichiosis in humans using a recombinant outer membrane protein of *E. chaffeensis*, particularly OMP-1. The present invention also relates to an assay for diagnosing ehrlichiosis in humans and members of the family Canidae using a recombinant outer membrane protein of *E. canis*, particularly P30.

## ABSTRACT

The present invention relates to diagnostic tools for veterinary and human use which are used for serodiagnosing ehrlichiosis in mammals, particularly in members of the Canidae family and in humans. The present invention also provides polynucleotides which encode the  
5 outer membrane proteins of *E. chafeensis*. The polynucleotides encode an OMP-1 family of proteins of *E. chafeensis* and P30 family of proteins of *E. canis*. The present invention also provides the following isolated proteins of *E. chafeensis* OMP-1, OMP-1A, OMP-1B, OMP-1C, OMP-1D, OMP-1E, OMP-1F, OMP-1R, OMP-1S, OMP-1T, OMP-1V, OMP-1W, and OMP-1Z, referred to hereinafter collectively as the "OMP family". The present invention also  
10 provides the following isolated proteins of *E. canis* P30, P30-1, P30-1 P30-2, P30-3, P30-4, P30-5, P30-6 P30-7, P30-8, P30-9, and P30-10, referred to hereinafter as the P30 family. The present invention also relates to an assay for diagnosing ehrlichiosis in humans using a recombinant outer membrane protein of *E. chafeensis*, particularly OMP-1. The present invention also relates to an assay for diagnosing ehrlichiosis in humans and members of the  
15 family Canidae using a recombinant outer membrane protein of *E. canis*, particularly P30.

## OUTER MEMBRANE PROTEIN OF EHRLICHIA CANIS AND EHRLICHIA CHAFFEENSIS

### BACKGROUND OF THE INVENTION

The ehrlichiae are obligate intracellular bacteria that infect circulating leucocytes. *Ehrlichia chaffeensis* infects the monocytes and macrophages in humans and causes human monocytic ehrlichiosis. The clinical manifestations of ehrlichiosis in humans are nonspecific and similar to Rocky Mountain spotted fever. The clinical manifestations include fever, chills, headache myalgia or vomiting and weight loss. Most patients have a history of tick exposure.

*Ehrlichia canis* infects and causes ehrlichiosis in animals belonging to the family Canidae. Canine ehrlichiosis consists of an acute and a chronic phase. The acute phase is characterized by fever, serous nasal and ocular discharges, anorexia, depression, and loss of weight. The chronic phase is characterized by severe pancytopenia, epistaxis, hematuria, blood in feces in addition to more severe clinical signs of the acute disease. If treated early during the course of the disease, dogs respond well to doxycycline. However, chronically infected dogs do not respond well to the antibiotic. Therefore, early diagnosis is very important for treating canine ehrlichiosis.

The primary diagnostic test for diagnosing canine ehrlichiosis and human ehrlichiosis is the indirect fluorescent antibody (IFA) test. This test uses the etiologic agent *Ehrlichia canis* to diagnose canine ehrlichiosis. The IFA test uses *Ehrlichia chaffeensis* as antigen for diagnosing human ehrlichiosis. The IFA test has, however, serious limitations. The IFA test is subject to false positives because the antigens are made of whole infected cells which comprise many nonspecific proteins which will cross-react with sera from some patients. The IFA test is also subject to false negatives because IFA antigens are unstable and may become inactivated during storage. In addition the IFA test requires a special equipment to perform the test. For example, the IFA test requires a tissue culture system for growing the bacterium that are used to prepare the antigen slides, a fluorescent microscope, and trained persons to evaluate the serum reactivity to the bacterial antigen on the slide.

Tools which permit simpler, more rapid, and objective serodiagnosis of canine ehrlichiosis or human ehrlichiosis are desirable.

### SUMMARY OF THE INVENTION

The present invention relates to improved diagnostic tools for veterinary and human use which are used for serodiagnosing ehrlichiosis in mammals, particularly in members of the Canidae family and in humans.

The present invention also provides polynucleotides or nucleic acids which encode the outer membrane proteins of *E. chaffeensis*. The OMP-1 polynucleotide encodes an OMP-1 protein of *E. chaffeensis* having a molecular weight of about 27.7 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 3B. SEQ ID NO: 2. The OMP-1B polynucleotide encodes an OMP-1B protein of *E. chaffeensis* having a molecular weight of about 28.2 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 48. SEQ ID NO: 4. The OMP-1C polynucleotide encodes an OMP-1C protein of *E. chaffeensis* having a molecular weight of about 27.6 kDa and an amino acid

sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 5B, SEQ ID NO: 6. The OMP-1D polynucleotide encodes an OMP-1D protein of *E. chaffeensis* having a molecular weight of about 28.7 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 6B, SEQ ID NO: 8. The OMP-1E polynucleotide encodes an OMP-1E protein of *E. chaffeensis* having a molecular weight of about 27.8 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 7B, SEQ ID NO: 10. The OMP-1F polynucleotide encodes an OMP-1F protein of *E. chaffeensis* having a molecular weight of about 27.9 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 8B, SEQ ID NO: 12. The OMP-1A polynucleotide encodes an OMP-1A protein of *E. chaffeensis* having a molecular weight of about 29.6 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 9B, SEQ ID NO: 14. The OMP-1R polynucleotide encodes an OMP-1R protein of *E. chaffeensis* having a molecular weight of at least 23 kDa and comprising an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 10B, SEQ ID NO: 16. The OMP-1S polynucleotide encodes an OMP-1S protein of *E. chaffeensis* comprising an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 11B, SEQ ID NO: 18. The OMP-1T polynucleotide encodes an OMP-1T protein of *E. chaffeensis* comprising an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 12B, SEQ ID NO: 20. The OMP-1U polynucleotide encodes an OMP-1U protein of *E. chaffeensis* having a molecular weight of about 30.6 kDa and an amino acid sequence which is at least 85% homologous to amino acid sequence shown in FIG. 13B, SEQ ID NO: 22. The OMP-1V polynucleotide encodes an OMP-1V protein of *E. chaffeensis* having a molecular weight of about 28.0 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 14B, SEQ ID NO: 24. The OMP-1W polynucleotide encodes an OMP-1W protein of *E. chaffeensis* having a molecular weight of about 28.8 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 15B, SEQ ID NO: 26. The OMP-1X polynucleotide encodes an OMP-1S protein of *E. chaffeensis* having a molecular weight of about 27.8 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 16B, SEQ ID NO: 28. The OMP-1Y polynucleotide encodes an OMP-1Y protein of *E. chaffeensis* having a molecular weight about 28.8 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 17B, SEQ ID NO: 30. The OMP-1Z polynucleotide encodes an OMP-1Z protein of *E. chaffeensis* comprising an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 18B, SEQ ID NO: 54.

The outer membrane proteins from *E. chaffeensis*, particularly a recombinant form of OMP-1, are immunogenic and, thus are useful for preparing antibodies. Such antibodies are useful for immunolabeling isolates of *E. chaffeensis* and for detecting the presence of *E. chaffeensis* in body fluids, tissues, and particularly in monocytes and macrophages. The isolated outer membrane proteins, particularly OMP-1, are also useful for detecting antibodies to *E. chaffeensis* in the blood of patients with clinical signs of ehrlichiosis. The isolated outer membrane protein, particularly OMP-1, are also useful immunogens for raising antibodies that are capable of reducing the level of infection in an immunized mammal that has been infected with *E. chaffeensis*. The isolated membrane proteins are also useful in a vaccine for protecting against infection with *E. chaffeensis*.

The present invention also relates to isolated polynucleotides which encode 30 kDa outer membrane proteins from *Ehrlichia canis*. The proteins are designated P30 and P30a. The proteins, particularly P30, are

immunogenic and are, thus, useful for preparing antibodies that are useful for immunolabeling isolates of *E. canis*. The P30 protein is also useful for diagnosing canine ehrlichiosis in mammals, particularly in members of the family Canidae, most particularly in dogs and for diagnosing infections with *E. chaffeensis* in humans. The P30 protein is also a useful immunogen for raising antibodies that reduce the level of infection in an immunized mammal that has been infected with *E. canis*. The P30 protein is also useful in a vaccine for protecting animals against infection with *E. canis*.

The present invention also provides the following isolated proteins of *E. chaffeensis* OMP-1 (also known as p28), OMP-1A, OMP-1B, OMP-1C, OMP-1D, OMP-1E, OMP-1F, OMP-1R, OMP-1S, OMP-1T, OMP-1U, OMP-1V, OMP-1W, OMP-1X, and OMP-1Z, referred to hereinafter collectively as the "OMP family". The present invention also provides the following isolated proteins of *E. canis* P30, P30-a, P30-1, P30-2, P30-3, P30-4, P30-5, P30-6, P30-7, P30-8, P30-9, and P30-10, referred to hereinafter as the P30 family.

The present invention also relates to an assay for diagnosing ehrlichiosis in humans using a recombinant outer membrane protein of *E. chaffeensis*, particularly OMP-1. The present invention also relates to an assay for diagnosing ehrlichiosis in humans and members of the family Canidae using a recombinant outer membrane protein of *E. canis*, particularly P30.

#### Brief Description of the Figures

FIG. 1 shows the DNA sequence (SEQ ID NO: 51) and the amino acid sequence (SEQ ID NO: 52) encoded by the *E. chaffeensis* (p28) gene cloned in pCRIIp28. The N-terminal amino acid sequence of native *omp-1* protein (P28) determined chemically is underlined. Five amino acid residues at the N terminus of P28 which were not included in the p28 gene, are indicated by boldface. Arrows indicate annealing positions of the primer pair designed for PCR.

FIG. 2. shows the restriction map of 6.3-kb genomic DNA including the *omp-1* gene copies in *E. chaffeensis*. The four DNA fragments were cloned from the genomic DNA (pPS2.6, pPS3.6, pEC2.6, and pEC3.6). A recombinant plasmid pPS2.6 has an overlapping sequence with that of pEC3.6. The closed boxes at the bottom show PCR-amplified fragments from the genomic DNA for confirmation of the overlapping area. Open boxes at the top indicate open reading frames (ORF) of *omp-1* gene copies with direction by arrows. Open boxes at the bottom show DNA fragments subcloned for DNA sequencing.

FIG. 3B shows one embodiment of the OMP-1 protein; FIG. 3A shows one embodiment of the OMP-1 polynucleotide.

FIG. 4B shows one embodiment of the OMP-1B protein, FIG. 4A shows one embodiment of the OMP-1B polynucleotide

FIG. 5A shows one embodiment of the OMP-1C polynucleotide; FIG 5B shows one embodiment of the OMP-1C protein.

FIG. 6B shows one embodiment of the OMP-1D protein; FIG. 6A shows one embodiment of the OMP-1D polynucleotide.

FIG. 7B shows one embodiment of the OMP-1E protein; FIG 7A shows one embodiment of the OMP-1E polynucleotide.

FIG. 8B shows one embodiment of the OMP-1F protein; FIG 8A shows one embodiment of the OMP-1F polynucleotide.

FIG. 9B shows one embodiment of the OMP-1A protein, FIG 9A shows one embodiment of the OMP-1A polynucleotide.

FIG. 10 B shows one embodiment of a portion of the OMP-1R protein. FIG 10A shows one embodiment of an OMP-1R polynucleotide encoding such polypeptide.

FIG. 11 B shows one embodiment of a portion of the OMP-1S protein. FIG 11A shows one embodiment of the OMP-1S polynucleotide encoding such polypeptide.

FIG. 12 B shows one embodiment of a portion of the OMP-1T protein. FIG 12A shows one embodiment of the OMP-1T polynucleotide encoding such polypeptide.

FIG. 13 B shows one embodiment of the OMP-1U protein. FIG 13A shows one embodiment of the OMP-1U polynucleotide.

FIG. 14 B shows one embodiment of the OMP-1V protein. FIG 14A shows one embodiment of the OMP-1V polynucleotide.

FIG. 15 B shows one embodiment of the OMP-1W protein. FIG 15A shows one embodiment of the OMP-1W polynucleotide.

FIG. 16 B shows one embodiment of the OMP-1X protein. FIG 16A shows one embodiment of the OMP-1X polynucleotide.

FIG. 17 B shows one embodiment of the OMP-1Y protein. FIG 17A shows one embodiment of the OMP-1Y polynucleotide.

FIG. 18 B shows one embodiment of the OMP-1Z protein. FIG 18A shows one embodiment of the OMP-1Z polynucleotide.

FIG. 19 B shows one embodiment of the P30 protein. FIG 19A shows one embodiment of the P30 polynucleotide.

FIG. 20 B shows one embodiment of the P30a protein. FIG 20A shows one embodiment of the p30A polynucleotide.

FIG. 21 B shows one embodiment of the P30-1 protein. FIG 21A shows one embodiment of the p30-1 polynucleotide.

FIG. 22 B shows one embodiment of the P30-2 protein. FIG 22 A shows one embodiment of the p30-2 polynucleotide.

FIG. 23 B shows one embodiment of the P30-3 protein. FIG 23 A shows one embodiment of the p30-3 polynucleotide.

FIG. 24 B shows one embodiment of the P30-4 protein. FIG 22 A shows one embodiment of the p30-4 polynucleotide.

FIG. 25 B shows one embodiment of the P30-5 protein. FIG 22 A shows one embodiment of the p30-5 polynucleotide.

FIG. 26 B shows one embodiment of the P30-6 protein. FIG 26 A shows one embodiment of the p30-6 polynucleotide.

FIG. 27 B shows one embodiment of the P30-7 protein. FIG 27 A shows one embodiment of the p30-7 polynucleotide.

FIG. 28 B shows one embodiment of the P30-8 protein. FIG 28 A shows one embodiment of the p30-8 polynucleotide.

FIG. 29 B shows one embodiment of a portion of the P30-9 protein. FIG 29 A shows one embodiment of the p30-9 polynucleotide.

FIG. 30 B shows one embodiment of a portion of the P30-10 protein, FIG 30 A shows one embodiment of the p30-10 polynucleotide encoding such protein.

FIG. 31 depicts the amino acid sequences alignment of seven *E. chaffeensis* OMP-1s and *Cowdria ruminantium* MAP-1. Aligned positions of identical amino acids with OMP-1F are shown with dots. The sequence of *C. ruminantium* MAP-1 is from the report of Van Vliet et al (1994) Molecular cloning, sequence analysis, and expression of the gene encoding the immunodominant 32-kilodalton protein of *Cowdria ruminantium*. Infect. Immun. 62:1451-1456. Gaps indicated by dashes were introduced for optimal alignment of all proteins. Bars indicates semivariable region (SV) and three hypervariable regions (HV1, HV2, and HV3).

#### DETAILED DESCRIPTION OF THE INVENTION

##### Isolated Polynucleotides Encoding OMP-1, OMP-1A, OMP-1B, OMP-1C, OMP-1D, OMP-1E, OMP-1F and the OMP from *E. canis*

In one aspect, the present invention, provides isolated polynucleotides that encode the outer ...inbrane proteins, OMP-1 (or p28), OMP-1B, OMP-1C, OMP-1D, OMP- 1E, OMP-1F, OMP-1A, OMP-1R, OMP-1S, OMP-1T, OMP -1U, OMP-1V, OMP-1W, OMP-1X, OMP-1Y and OMP-1Z from *E. chaffeensis* and the outer membrane proteins P30, P30-a, P-30-1, P30-2, P30-3, P30-4, P30-5, P30-6, P30-7, P30-8, P30-9, and P30-10 from *E. canis* or an immunogenic fragment thereof.

The polynucleotide is single stranded or double stranded. The polynucleotide may be a DNA or RNA molecule, preferably a DNA molecule, and comprises a sequence which codes for the respective outer membrane protein. Preferably, the polynucleotide encodes at least the mature form of outer membrane protein. The polynucleotide optionally further comprises a leader sequence and encode an outer membrane preprotein that is processed in the cell to form the mature protein. The polynucleotide of the present invention may also be fused in frame to a marker sequence which allows for purification of the corresponding outer membrane protein.

The OMP-1 polynucleotide encodes an OMP-1 protein of *E. chaffeensis* having a molecular weight of about 27.7 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 3B SEQ ID NO: 2; Figure 3B shows one embodiment of the OMP-1 protein, Figure 3A shows one embodiment of the OMP-1 polynucleotide. The OMP-1B polynucleotide encodes an OMP-1B protein of *E. chaffeensis* having a molecular weight of about 28.2 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 4B SEQ ID NO: 4; Figure 4B shows one embodiment of the OMP-1B protein, Figure 4A shows one embodiment of the OMP-1B polynucleotide. The OMP-1C polynucleotide encodes an OMP-1C protein of *E. chaffeensis* having a molecular weight of about 27.6 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 5B SEQ ID NO: 6; Figure 5B shows one embodiment of the OMP-1C protein, Figure 5A shows one embodiment of the OMP-1C polynucleotide. The OMP-1D polynucleotide encodes an OMP-1D protein of *E. chaffeensis* having a molecular weight of about 28.7 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 6B SEQ ID NO: 8; Figure 6B shows one embodiment of the OMP-1D protein, Figure 6A shows one embodiment of the OMP-1D polynucleotide. The OMP-1E polynucleotide encodes an OMP-1E protein of *E. chaffeensis* having a molecular weight of about 27.8 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 7B SEQ ID NO: 10; Figure 7B shows one embodiment of the OMP-1E protein, Figure 7A shows one embodiment of the OMP-1E

polynucleotide. The OMP-1F polynucleotide encodes an OMP-1F protein of *E. chaffeensis* having a molecular weight of about 27.9 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 8B, SEQ ID NO: 12; Figure 8B shows one embodiment of the OMP-1F protein, Figure 8A shows one embodiment of the OMP-1F polynucleotide. The OMP-1A polynucleotide encodes an OMP-1A protein of *E. chaffeensis* having a molecular weight of about 29.6 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 9B, SEQ ID NO: 14; Figure 9B shows one embodiment of the OMP-1A protein, Figure 9A shows one embodiment of the OMP-1A polynucleotide. The OMP-1R polynucleotide encodes an OMP-1R protein of *E. chaffeensis* having a molecular weight of at least 23 kDa and comprising an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 10B, SEQ ID NO: 16; Figure 10B shows one embodiment of a portion of the OMP-1R protein, Figure 10A shows one embodiment of the OMP-1R polynucleotide encoding such polynucleotide. The OMP-1S polynucleotide encodes an OMP-1S protein of *E. chaffeensis* comprising an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 11B, SEQ ID NO: 18; Figure 11B shows one embodiment of a portion of the OMP-1S protein, Figure 11A shows one embodiment of the OMP-1S polynucleotide encoding such polypeptide. The OMP-1T polynucleotide encodes an OMP-1T protein of *E. chaffeensis* comprising an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 12B, SEQ ID NO: 20; Figure 12B shows one embodiment of a portion of the OMP-1T protein, Figure 12A shows one embodiment of a polynucleotide encoding such polypeptide. The OMP-1U polynucleotide encodes an OMP-1U protein of *E. chaffeensis* having a molecular weight of about 30.6 kDa and an amino acid sequence which is at least 85% homologous to amino acid sequence shown in FIG. 13B, SEQ ID NO: 22; Figure 13B shows one embodiment of the OMP-1U protein, Figure 13A shows one embodiment of the OMP-1U polynucleotide. The OMP-1V polynucleotide encodes an OMP-1V protein of *E. chaffeensis* having a molecular weight of about 28.0 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 14B, SEQ ID NO: 24; Figure 14B shows one embodiment of the OMP-1V protein, Figure 14A shows one embodiment of the OMP-1V polynucleotide. The OMP-1W polynucleotide encodes an OMP-1W protein of *E. chaffeensis* having a molecular weight of about 28.8 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 15B, SEQ ID NO: 26; Figure 15B shows one embodiment of the OMP-1W protein, Figure 15A shows one embodiment of the OMP-1W polynucleotide. The OMP-1X polynucleotide encodes an OMP-1X protein of *E. chaffeensis* having a molecular weight of about 27.8 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 16B, SEQ ID NO: 28; Figure 16B shows one embodiment of the OMP-1X protein, Figure 16A shows one embodiment of the OMP-1X polynucleotide. The OMP-1Y polynucleotide encodes an OMP-1Y protein of *E. chaffeensis* having a molecular weight about 28.8 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 17B, SEQ ID NO: 30; Figure 17B shows one embodiment of the OMP-1Y protein, Figure 17A shows one embodiment of the OMP-1Y polynucleotide. The OMP-1Z polynucleotide encodes an OMP-1Z protein of *E. chaffeensis* comprising an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 18B, SEQ ID NO: 54; Figure 18B shows one embodiment of a portion of the OMP-1Z protein, Figure 18A shows one embodiment of an OMP-1Z polynucleotide encoding such polypeptide.

The p30 polynucleotide encodes a P30 protein of *E. canis* having a molecular weight of about 28.8 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 19B, SEQ ID NO: 32; Figure 19B shows one embodiment of the P30 protein, Figure 19A shows one embodiment of the p30 polynucleotide. The p30A polynucleotide encodes a P30a protein of *E. canis* having a molecular weight of about 29.1 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 20B, SEQ ID NO: 34; Figure 20B shows one embodiment of the P30a protein, Figure 20A shows one embodiment of the p30A polynucleotide. The p30-1 polynucleotide encodes a P30-1 protein of *E. canis* having a molecular weight of about 27.7 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 21B, SEQ ID NO: 36; Figure 21B shows one embodiment of the P30-1 protein, Figure 21A shows one embodiment of the p30-1 polynucleotide. The p30-2 polynucleotide encodes a P30-2 protein of *E. canis* having a molecular weight of about 28.0 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 22B, SEQ ID NO: 38; Figure 22B shows one embodiment of the P30-2 protein, Figure 22A shows one embodiment of the p30-2 polynucleotide. The p30-3 polynucleotide encodes a P30-3 protein of *E. canis* having a molecular weight of about 28.7 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 23B, SEQ ID NO: 40; Figure 23B shows one embodiment of the P30-3 protein, Figure 23A shows one embodiment of the p30-3 polynucleotide. The p30-4 polynucleotide encodes a P30-4 protein of *E. canis* having a molecular weight of about 28.0 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 24B, SEQ ID NO: 42; Figure 24B shows one embodiment of the P30-4 protein, Figure 24A shows one embodiment of the p30-4 polynucleotide. The p30-5 polynucleotide encodes a P30-5 protein of *E. canis* having a molecular weight of about 29.4 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 25B, SEQ ID NO: 44; Figure 25B shows one embodiment of the P30-5a protein, Figure 25A shows one embodiment of the p30-5a polynucleotide. The p30-6 polynucleotide encodes a P30-6 protein of *E. canis* having a molecular weight of about 29.5 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 26B, SEQ ID NO: 56; Figure 26B shows one embodiment of the P30-6 protein, Figure 26A shows one embodiment of the p30-6 polynucleotide. The p30-7 polynucleotide encodes a P30-7 protein of *E. canis* having a molecular weight of about 29.9 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 27, SEQ ID NO: 58; Figure 27 shows one embodiment of the P30-7 protein, Figure 27 shows one embodiment of the p30-7 polynucleotide. The p30-8 polynucleotide encodes a P30-8 protein of *E. canis* having a molecular weight of about 30.3 kDa and an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 28B, SEQ ID NO: 46; Figure 28B shows one embodiment of the P30-8 protein, Figure 28A shows one embodiment of the p30-8 polynucleotide. The p30-9 polynucleotide encodes a P30-9 protein of *E. canis* comprising an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 29B, SEQ ID NO: 60; Figure 29B shows one embodiment of a portion of the P30-9 protein, Figure 29A shows one embodiment of the p30-9 polynucleotide encoding such polypeptide. The p30-10 polynucleotide encodes a P30-10 protein of *E. canis* comprising an amino acid sequence which is at least 85% homologous to the amino acid sequence shown in FIG. 30B, SEQ ID NO: 48; Figure 30B shows one embodiment of a portion of the P30-10 protein, Figure 30A shows one embodiment of the p30-10 polynucleotide encoding such polypeptide.

The polynucleotides encoding an *E. chaffeensis* outer membrane protein or an *E. canis* outer membrane protein have a sequence that is at least 85%, preferably at least 90%, more preferably at least 95% homologous to or similar to the amino acid sequences shown in Figures 3B through 30B, and thus embrace polynucleotides encoding outer membrane proteins from different strains of *E. chaffeensis* and *E. canis*. The polynucleotides encode an outer membrane protein whose conserved regions collectively are at least 90%, preferably at 95%, more preferably at least 97% homologous to the conserved regions of the amino acid sequences of the present invention. The outer membrane proteins of *E. chaffeensis* and *E. canis* have six conserved regions, which are separated by one semivariable region and three hypervariable regions. The conserved regions of the outer membrane proteins OMP-1, OMP-1A, OMP-1B, OMP1-C, OMP-1D, OMP1-F are depicted in Fig. 31. Preferably, the amino acid sequence of the outer membrane proteins of *E. chaffeensis* and *E. canis* are at least 30% divergent from the amino acid sequence of MAP-1. Such sequences include allelic, strain variants and other amino acid sequence variants (e.g., including "mteins" or "mutant proteins"), whether naturally-occurring or biosynthetically produced. As used herein, "amino acid sequence homology" is understood to mean amino acid sequence similarity, and homologous sequences share identical or similar amino acids, where similar amino acids are conserved amino acids as defined by Dayoff et al., *Atlas of Protein Sequence and Structure*; vol. 5, Supp. 3, pp. 345-362 (M. O. Dayoff, ed., Nat'l BioMed. Research Fdn., Washington D.C. 1978.) Thus, a candidate sequence sharing 85% amino acid sequence homology with a reference sequence requires that, following alignment of the candidate sequence with the reference sequence, 85% of the amino acids in the candidate sequence are identical to the corresponding amino acid in the reference sequence, or constitute a conservative amino acid change thereto. "Amino acid sequence identity" is understood to require identical amino acids between two aligned sequences. Thus, a candidate sequence sharing 85% amino acid identity with a reference sequence requires that, following alignment of the candidate sequence with the reference sequence, 85% of the amino acids in the candidate sequence are identical to the corresponding amino acid in the reference sequence.

As used herein, all homologies and identities are calculated using the amino acid sequences shown in the cited Figure or SEQ ID NO as the reference sequence. Thus, to determine whether an amino acid sequence is 85% homologous to OMP-1, one uses the amino acid sequence shown in Fig. 3B, SEQ ID NO: 2 as a reference.

Also as used herein, sequences are aligned for homology and identity calculations using the method of the software basic local alignment search tool in the BLAST network service (the National Center for Biotechnology Information, Bethesda, MD) which employs the method of Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* 215, 403-410. Identities are calculated by the Align program (DNASTar, Inc.) In all cases, internal gaps and amino acid insertions in the candidate sequence as aligned are ignored when making the homology/identity calculation.

In another aspect, the present invention provides a nucleotide sequence encoding a polypeptide which comprises a fragment of the OMP1 protein, hereinafter referred to as "rP28". The rP28 polypeptide weighs approximately 31 kDa and comprises all but of the first 5 amino acids of mature OMP-1 protein. The rP28 polypeptide comprises the amino acid sequence extending from amino acid 6 through amino acid 256 of the amino acid sequence shown in Fig. 1, SEQ ID NO. 52. The present invention also embraces polypeptides where one or more of the amino acids in the sequence extending from amino acid 1 or 6 through amino acid 256 Fig. 1

are replaced by conservative amino acid residues. The present invention also relates to derivatives of rP28 that have an amino acid sequence identity of at least 85%, more preferably at least 90%, and most preferably of at least 95% with the amino acid sequence extending from amino acid 1 or 6 through amino acid 251 of the protein and which derivative binds to antibodies in sera from humans infected with *E. chaffeensis*.

The polynucleotides are useful for producing the outer membrane proteins of *E. chaffeensis* and *E. canis*. For example, an RNA molecule encoding the outer membrane protein OMP-1 is used in a cell-free translation systems to prepare OMP-1. Alternatively, a DNA molecule encoding the outer membrane protein is introduced into an expression vector and used to transform cells. Suitable expression vectors include for example chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40, bacterial plasmids, phage DNAs; yeast plasmids, vectors derived from combinations of plasmids and phage DNAs, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. The DNA sequence is introduced into the expression vector by conventional procedures.

Accordingly, the present invention also relates to recombinant constructs comprising one or more of the polynucleotide sequences. Suitable constructs include, for example, vectors, such as a plasmid, phagemid, or viral vector, into which a sequence that encodes the outer membrane protein has been inserted. In the expression vector, the DNA sequence which encodes the outer membrane protein is operatively linked to an expression control sequence, i.e., a promoter, which directs mRNA synthesis. Representative examples of such promoters, include the LTR or SV40 promoter, the *E. coli* lac or trp, the phage lambda PL promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or in viruses. The promoter may also be the natural promoter of the outer membrane protein coding sequence. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. Preferably, the recombinant expression vectors also include an origin of replication and a selectable marker, such as for example, the ampicillin resistance gene of *E. coli* to permit selection of transformed cells, i.e. cells that are expressing the heterologous DNA sequences. The polynucleotide sequence encoding the outer membrane protein is incorporated into the vector in frame with translation initiation and termination sequences. Optionally, the sequence encodes a fusion outer membrane protein which includes an N-terminal or C-terminal peptide or tag that stabilizes or simplifies purification of the expressed recombinant product. Representative examples of such tags include sequences which encode a series of histidine residues, the Herpes simplex glycoprotein D, or glutathione S-transferase.

Polynucleotides which encode portions of the outer membrane proteins of *E. chaffeensis* and *E. canis* are useful as probes for isolating and identifying *E. chaffeensis* genes and *E. canis* genes, particularly full-length genes from new strains or isolates of *E. chaffeensis* and *E. canis*.

#### The Outer Membrane Proteins of *E. chaffeensis* and *E. canis*

In addition to the outer membrane proteins OMP-1, OMP-1B, OMP-1C, OMP-1D, OMP-1E, and OMP-1F, OMP-1R, OMP-1S, OMP-1T, OMP-1U, OMP-1V, OMP-1W, OMP-1X, OMP-1Y, and OMP-1Z from *E. chaffeensis* and the proteins P30, P30A, P30-1, P30-2, P30-3, P30-4, P30-5, P30-6, P30-7, P30-8, P30-9, and P30-10 from *E. canis*, the present inventions embraces non-naturally occurring allelic forms or derivatives of the outer membrane proteins, where one or more of the amino acids have been replaced by conservative amino acid residues, typically by using direct synthesis or recombinant techniques.

### Preparing the Outer Membrane Proteins

The outer membrane proteins of the present invention are synthetically produced by conventional peptide synthesizers. The outer membrane proteins are also produced using cell-free translation systems and RNA molecules derived from DNA constructs that encode the outer membrane protein. Alternatively, the outer membrane protein is made by transfecting host cells with expression vectors that comprise a DNA sequence which encodes the outer membrane protein and then inducing expression of the outer membrane protein in the host cells.

The outer membrane protein is expressed in suitable host cells, preferably bacteria, under the control of suitable promoters. Host cells are transformed with the expression vectors of this invention and cultured in conventional nutrient media. Such media optionally contains additional compounds, such as for example compounds that induce promoters, such as for example isopropyl- $\beta$ -D-thiogalactoside which induces the Lac promoter, or compounds, such as for example, ampicillin, which allows for selection of transformants.

Following transformation of the suitable host strain and growth of the host strain to an appropriate cell density, the cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification of the outer membrane protein. Such purification usually involves salting-out of the protein fraction, and one or more chromatography steps, including aqueous ion exchange chromatography, size exclusion chromatography steps, and high performance liquid chromatography (HPLC).

### Preparation of Antibodies

The isolated outer membrane proteins, particularly the recombinant forms of the outer membrane proteins, are used as immunogens to produce antibodies immunospecific for the corresponding protein. The term "immunospecific" means the antibodies have substantially greater affinity for the protein used as an immunogen than for other proteins. Such antibodies are generated using conventional techniques by administering the respective outer membrane protein or a portion thereof, i.e., the recombinant polypeptide, to an animal, preferably a nonhuman, collecting blood from the immunized animals and isolating the serum and/or the IgG fraction from the blood. Monoclonal antibodies are prepared by injecting animals with the immunogens, extracting antibody-producing B cells from the animal, fusing the B cells with a myeloma cells to produce hybridomas, obtaining the monoclonal antibodies from the hybridomas.

Antibodies to the outer membrane proteins of *E. chaffeensis* and *E. canis* are useful research tools for identifying cells, particularly monocytes, infected with *E. chaffeensis* or *E. canis* and for purifying the corresponding outer membrane protein of *E. chaffeensis* or *E. canis* from partially purified preparations by affinity chromatography. Such antibodies are also useful for identifying bacterial colonies, particularly colonies of genetically-engineered bacteria, that are expressing the major outer membrane protein.

### Diagnostic Method

The present invention also provides a method for detecting antibodies to the *E. chaffeensis* or *E. canis* in a sample of a bodily fluid from a patient. The method comprises providing an isolated outer membrane protein of *E. chaffeensis* or *E. canis*, particularly a recombinant form of the isolated protein, contacting the outer membrane protein or polypeptide with a sample taken from the patient; and assaying for the formation of a complex between the outer membrane protein or polypeptide and antibodies in the sample. For ease of detection, it is preferred that the isolated protein or polypeptide be attached to a substrate such as a column,

plastic dish, matrix, or membrane, preferably nitrocellulose. The sample may be a tissue or a biological fluid, including urine, whole blood, or exudate, preferably serum. The sample may be untreated, subjected to precipitation, fractionation, separation, or purification before combining with the isolated protein or peptide. Interactions between antibodies in the sample and the isolated protein or peptide are detected by radiometric, colorimetric, or fluorometric means, size-separation, or precipitation. Preferably, detection of the antibody-outer membrane protein complex is by addition of a secondary antibody that is coupled to a detectable tag, such as for example, an enzyme, fluorophore, or chromophore. Formation of the complex is indicative of the presence of anti-*E. chaffeensis* or anti-*E. canis* antibodies, either IgM or IgG, in the patient. Thus, the method is used to determine whether a patient is infected with *E. chaffeensis* or *E. canis*.

Preferably, the method employs an enzyme-linked immunosorbent assay (ELISA) or a Western immunoblot procedure. Such methods are relatively simple to perform and do not require special equipment as long as membrane strips are coated with a high quality antigen. Accordingly, it is more advantageous to use a recombinant form of the outer membrane protein of *E. chaffeensis* or *E. canis* since such proteins, typically, are more pure and consistent in quality than a purified form of such protein.

#### Immunogenic Composition

The present invention also relates to immunogenic compositions comprising one or more of the isolated outer membrane proteins of *E. chaffeensis* and a pharmaceutically acceptable adjuvant and to immunogenic compositions comprising an isolated P30 protein of *E. canis* and a pharmaceutically acceptable adjuvant, which, preferably, enhances the immunogenic activity of the outer membrane protein in the host animal.

#### Preparation of a Polynucleotide which Encodes OMP-1(P28)

##### A. Isolation of the Outer Membrane Proteins

*E. chaffeensis* Arkansas strain and *E. canis* Oklahoma strain were cultivated in the DH82 dog macrophage cell line and purified by Percoll\* density gradient centrifugation. Purified ehrlichiae (100 µg) were suspended with 10 mM sodium phosphate buffer, pH 7.4, containing 0.1% Sodium *N*-lauroyl sarcosine (Sarkosyl\*) [Sigma, St. Louis, MO], 50 µg/ml each Dnase I (Sigma) and Rnase A (Sigma), and 2.5 mM MgCl<sub>2</sub>. After incubation at 37° for 30 min, the sample was separated by centrifugation at 10,000 x g for 1 h into the soluble supernatant and the insoluble precipitate. The insoluble pellet was resuspended 2 to 3 times with 0.1% Sarkosyl and centrifuged. The final pellet was analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and by electron microscopy.

Transmission electron microscopy revealed that the purified ehrlichial fraction consists of a mixture of electron dense and light forms of *E. chaffeensis* with slight disintegration of inner membrane. Ehrlichiae were not surrounded with the host inclusion membrane. Various sizes of membrane vesicles (< 1 µm) without significant ribosomes or nuclear materials were observed in the Sarkosyl-insoluble fraction from the organism. Succinic dehydrogenase (inner membrane marker enzyme of gram negative bacteria) activities were at less than the detection limit (1 n moles / min / mg of protein) in the Sarkosyl-insoluble fraction compared to approximately 10 n moles / min / mg of protein in the Percoll-purified organisms, suggesting that the insoluble fraction primarily consisted of the outer membrane of *E. chaffeensis*.

\* Trademark

Analysis of the Sarkosyl-soluble, and insoluble fraction of *E. chaffeensis* by SDS-PAGE suggested that proteins of 30-kDa range in the insoluble fraction represent the major outer membrane proteins of this organism. Analysis of the Sarkosyl-soluble, and insoluble fraction of *E. canis* by SDS-PAGE suggested that proteins of 30-kDa range in the insoluble fraction represent the major outer membrane proteins of this organism also. *E. canis* was antigenically cross reactive with *E. chaffeensis*. These findings indicate that the 30-kDa range proteins represent the major outer membrane proteins of these two *Ehrlichia* spp.

To improve resolution of the outer membrane proteins, proteins in the Sarkosyl-insoluble pellet prepared from 400 µg of purified *E. chaffeensis* were separated by a reversed-discontinuous (Rd) SDS-PAGE (2.5-cm-long 17% gel on top of 11-cm-long 12% gel). At least five proteins of 30-kDa range in *E. chaffeensis* (P23, P25, P27, P28, and P29) were resolved from the Sarkosyl-insoluble proteins.

#### B. Cloning and sequencing of the *p28* gene

The portion of the membrane containing bound proteins was excised and analyzed with an Applied Biosystems protein sequencer (Model 470). The N-terminal amino acid sequence of P28 was determined as D P A G S G I N G N F Y S G K Y M P, SEQ ID NO 61. Based on 6th to 12th amino acids of this sequence, a forward primer, FECH1, having the sequence: 5'-CGGGATCCGAATTCGG(A/T/G/C)AT(A/T/C)AA(T/C)GG(A/T/G/C)AA(T/C)TT(T/C)TA-3', SEQ ID NO 49 was designed. Amino acids at the 1 to 5 positions of the N terminus of P28 were not included in this primer design. For insertion into an expression vector, a 14-bp sequence (underlined) was added at the 5' end of primer to create an *EcoRI* and a *BamHI* site. The reverse primer, RECH2, which includes a *NorI* site at the 5' end for ligation into an expression vector had the sequence: 5'-AGCGGCCGCTTA(A/G)AA(T/C)A(C/G) (A/G)AA (C/T)CT T(C/G)C TCC-3'. SEQ ID NO 50.

Genomic DNA of *E. chaffeensis* was isolated from purified organisms. PCR amplification with FECH1 and RECH2 primers was performed using a Perkin-Elmer Cetus DNA Thermal Cycler (model 480). A 0.8-kb amplified product was cloned in the pCRII vector of a TA cloning kit, as described by the manufacturer (Invitrogen Co., San Diego, CA). The clone obtained was designated pCRIIp28. Both strands of the inserted DNA were sequenced by a dideoxy-termination method with an Applied Biosystems 373A DNA sequencer.

The 0.8-kb DNA fragment, cloned in pCRIIp28, had an open reading frame (ORF) of 756 bp encoding a 251-amino acid recombinant protein (including both PCR primer regions) with a molecular mass of 27,685 Da. The nucleotide sequence of the open reading frame, SEQ ID NO: 1, and the amino acid sequence of the polypeptide of the OMP-I protein, SEQ ID NO 2, are shown in Figs 3A and 3B, respectively.

A DNA fragment comprising the *p30* gene was prepared in a similar manner, i.e., by PCR amplification of genomic DNA of *E. canis* using primers FECH1 and RECH1.

#### Preparation of Polynucleotides which encode OMP-1A, OMP-1B, OMP-1C, OMP-1D, OMP-1E, and OMP-1F

A. Southern blot analysis. Genomic DNA extracted from the purified *E. chaffeensis* (200 ng each) was digested with restriction endonucleases, electrophoresed, and transferred to Hybond-N\* nylon membrane (Amersham, Arlington Heights, IL), by a standard method. The 0.8-kb *p28* gene fragment from the clone pCRIIp28 was labeled with [ $\alpha$ -<sup>32</sup>P]dATP by the random primer method using a kit (Boehringer Mannheim, Indianapolis, IN) and the labeled fragment was used as a DNA probe. Hybridization was performed at 60°C in rapid hybridization buffer (Amersham) for 20 h. The nylon sheet was washed in 0.1 x SSC (1 x SSC containing

\* Trademark

0.15M sodium chloride and 0.015M sodium citrate) with 1% SDS at 55°C and the hybridized probes were exposed to Hyperfilm\* (Amersham) at -80°C.

Genomic Southern blot analysis with several restriction enzymes resulted in one or more DNA fragment(s) of *E. chaffeensis* which hybridized to <sup>32</sup>P-labeled *p28* gene probe. The restriction enzymes used did not cut within the *p28* gene portion of the pCRJ1*p28* insert. *Xba* I, *Bgl* II, and *Kpn* I produced two bands, *Spe* I generated three bands, and *EcoR* V and *Pst* I produced multiple bands with different densities. *EcoR* I generated a broad band of 2.5 to 4kb. These *p28* homologous genes are designated as *omp-1* (outer membrane protein-1) family.

B. Cloning and sequencing of genomic copies of *E. chaffeensis p28* gene. The *EcoR* I and *Pst* I fragments of DNA, detected by genomic Southern blot analysis as described above, were inserted into pBluescript II KS (+) vectors, and the recombinant plasmids were introduced into *E. coli* DH5 $\alpha$ . Using the colony hybridization method with the <sup>32</sup>P-labeled *p28* gene probe, four positive clones were isolated from the transformant. The positive clones were designated pEC2.6, pEC3.6, pPS2.6, and pPS3.6. These contained the ehrlichial DNA fragments of 2.6-kb (*EcoR* I), 3.6 kb (*EcoR* I), 2.6 kb (*Pst* I), and 3.6 kb (*Pst* I), respectively. The inserts of the clones pEC3.6 and pPS2.6 overlapped as shown in Fig. 2. The overlapping area was further confirmed by PCR of *E. chaffeensis* genomic DNA with two pairs of primer sets interposing the junctions of the four clones. The 1.1- to 1.6-kb DNA fragments of *Hind*III-*Hind*III, *Hind*III-*EcoR*I, or *Nho*I-*EcoR*I in the pEC2.6 and pEC3.6 were subcloned for sequencing. DNA sequencing was performed with suitable synthetic primers by dideoxy-termination method as described above.

Four DNA fragments from 2.6 to 3.6 kb were cloned from the *EcoR*I-digested and the *Pst*I-digested genomic DNA of *E. chaffeensis* by colony hybridization with radiolabeled *p28* gene probe. The inserted DNA of the two recombinant clones, pEC3.6 and pPS2.6, were overlapped. Sequencing revealed one 5'-truncated ORF of 243 bp (designated *omp-1A*) and five complete ORF of 836-861 bp (designated *omp-1B* to *omp-1F*), which are tandemly-arrayed and are homologous to the *p28* gene (but are not identical), in the ehrlichial genomic DNA of 6.292 bp. The intergenic spaces were 581 bp between *omp-1A* and *omp-1B* and 260-308 bp among others. Putative promoter regions and ribosome-binding sites were identified in the noncoding regions.

#### Sequence analysis and GenBank accession number.

Nucleotide sequences were analyzed with the DNASIS program (Hitachi Software Engineering Co., Ltd., Yokohama, Japan). A homology search was carried out with databases of the GenBank, Swiss Plot, PDB and PIR by using the software basic local alignment search tool in the BLAST network service (the National Center for Biotechnology Information, Bethesda, MD). Phylogenetic analysis was performed by using the PHYLIP software package (version 3.5). An evolutionary distance matrix, generated by using the Kimura formula in the PROTDIST, was used for construction of a phylogenetic tree by using the unweighted pair-group method analysis (UPGMA) (Felsenstein, J. 1989. PHYLIP-phylogeny inference package (version 3.3). *Cladistics* 5:164-166). The data were also examined using parsimony analysis (PROTPARS in PHYLIP). A bootstrap analysis was carried out to investigate the stability of randomly generated trees by using SEQBOOT and CONSENSE in the same package. The nucleotide sequence of the *p28* gene and its gene copies has been assigned GenBank accession numbers U72291 and AF021338, respectively.

#### Proteins of the *E. chaffeensis omp-1* Family.

\* Trademark

Five complete *omp-1* gene copies (*omp-1B* to *omp-1F*) encode 279 to 287-amino acid proteins with molecular masses of 30.320 - 31.508 Da. *Omp-1A* encodes an 82-amino acid partial protein (9.243 Da) which lacks the N-terminal region. The 25-amino acid sequence at the N-terminus of OMP-1B to OMP-1F (encoded in *omp-1B* to *omp-1F*) is predicted to be a signal peptide because three carboxyl-terminal amino acids of the signal peptides (Ser-X-Ala in OMP-1B, Leu-X-Ser for OMP-1C, and Ser-X-Ser for OMP-1D and OMP-1F) are included in the preferred amino acid sequence of signal peptidase at the processing sites proposed by Oliver. The calculated molecular masses of the mature OMP-1B to OMP-1F from the predicted amino acid sequences are 28,181 Da for OMP-1B, 27,581 Da for OMP-1C, 28,747 Da for OMP-1D, 27,776 Da for OMP-1E, and 27,933 Da for OMP-1F. The estimated isoelectric points are 4.76-5.76 in the mature OMP-1B to OMP-1F. An amino acid sequence in *omp-1F* gene (the 80th to 94th amino acids) was identical to the N-terminal amino acid sequences of *E. chaffeensis* native P23 protein as determined chemically, which indicates that P23 is derived from the *omp-1F* gene. Amino acid sequences identical to the N-terminal sequences of P25, P27, and P29 were not found in those from *omp-1* gene copies cloned in this study.

Alignment of predicted amino acid sequences of the *E. chaffeensis* OMP-1 family and *Cowdria ruminantium*, revealed substitutions or deletions of one or several contiguous amino acid residues throughout the molecules. The significant differences in sequences among the aligned proteins are seen in the regions indicated SV (semivariable region) and HV (hypervariable region) 1 to 3 in Fig 31. Computer analysis for hydropathy revealed that protein molecules predicted from all *omp-1* gene copies contain alternative hydrophilic and hydrophobic motifs which are characteristic of transmembrane proteins. The HV1 and HV2 were found to locate in the hydrophilic regions.

The amino acid sequences of 5 mature proteins without signal peptides (OMP-1C to OMP-1F and P28) were similar to one another (71-83%) but the sequence of OMP-1B was dissimilar to those of the 5 proteins (45-48%). The amino acid sequences of the 5 proteins showed an intermediate degree of similarity with that of *C. ruminantium* MAP-1 (59-63%), but the similarity between that of the OMP-1B and the *C. ruminantium* MAP-1 was low (45%). These relations are shown in a phylogenetic tree which was obtained based on the amino acid sequence alignment by UPGMA method in the PHYLIP software package. Three proteins (P28, OMP-1D, and OMP-1F) and two proteins (OMP-1C and OMP-1E) formed two separate clusters. The OMP-1B was located distantly from these two clusters. The *C. ruminantium* MAP-1 was positioned between the OMP-1B and other members in the OMP-1 family.

#### Preparation of a Recombinant form of OMP-1 and P30

The 0.8-kb *p28* gene was excised from the clone pCRII*p28* by *EcoRI-NotI* double-digestion, ligated into *EcoRI-NotI* sites of a pET 29a expression vector, and amplified in *Escherichia coli* BL21 (DE3)pLysS (Novagen, Inc., Madison, WI). The clone (designated pET29*p28*) produced a fusion protein with a 35-amino acid sequence carried from the vector at the N terminus. The amino acid sequence of the OMP-1 portion of the fusion protein is depicted in Fig. 1, amino acid 6 through 256.

An expression vector comprising the *p30* gene was used to prepare the recombinant form of P30.

### Preparation of anti-rP28 (anti-OMP1) antibody

The (r) PS antigen was prepared by exercising the gel band corresponding to the rP28 in SDS-PAGE, mincing the band in phosphate-buffered saline (PBS). pH 7.4, and mixing with an equal volume of Freund's incomplete adjuvant (Sigma). The rP28 mixture (1 mg of protein each time) was subcutaneously injected into a rabbit every 2 weeks four times. A serum sample was collected from the rabbit to provide the anti-rP28 antibody.

The anti-rP28 antibody was examined by western immunoblots analysis. The results indicated that the rabbit anti-rP28 antibody recognized not only rP28 (31 kDa) and P28, but also P29 and P25 of *E. chaffeensis* and P30 of *E. canis*. These results indicate that P28 shares antigenic epitopes with P25 and P29 in *E. chaffeensis* and P30 of *E. canis*.

The following examples are for purposes of illustration only and are not intended to limit the scope of the claims which are appended hereto.

### Example 1. Assaying for the presence of anti-OMP-1 antibody in a Patient

Convalescent-phase serum from a patient with clinical signs of human ehrlichiosis was used. Western blot analyses using the rP28 protein as antigen was performed with 1:1,000 dilutions of this serum. Alkaline phosphatase-conjugated affinity-purified anti-human immunoglobulin G (Kirkegaard & Perry Laboratories, Inc., Gaithersburg, MD) was used at a 1:1,000 or 1:2,000 dilution as secondary antibodies. Results indicated that serum from a patient with clinical signs of human ehrlichiosis reacted strongly to rP28 (31 kDa).

### Example 2 Assaying for the presence of anti-P30 antibody in a patient

Convalescent-phase serum from a patient with clinical signs of human ehrlichiosis was reacted with the rP30 protein of *E. canis* as described in Example 1. The serum reacted strongly to rP30. These results indicate the rP30 is useful for diagnosing an infection with *E. chaffeensis* in human patients.

### Example 3. Identifying *E. chaffeensis*-infected cells using anti-rP28 antibody

*E. chaffeensis*-infected DH82 cells were sonicated and centrifuged at 400 x g for 10 min. The supernatant was then centrifuged at 10,000 x g for 10 min to obtain ehrlichia-enriched pellet. The pellet was resuspended and incubated with rabbit anti-rP28 antibody or normal rabbit serum (1:100 dilution) at 37°C for 1h in PBS containing 1% bovine serum albumin (BSA-PBS). After washing, the ehrlichiae was incubated with gold-conjugated protein G (20 nm). Sigma) at 1:30 dilution for 1h at room temperature in BSA-PBS. After washing again, the specimen was fixed with 1.25% formaldehyde, 2.5% glutaraldehyde, and 0.03% trinitrophenol in 0.1 M cacodylate buffer (pH 7.4) for 24h and postfixed in 1% osmium-1.5% potassium ferricyanide for 1h (34). The section was then embedded in PolyBed 812 (Polysciences, Warrington, Pa). The specimen

was ultrathin sectioned at 60 nm, stained with uranyl acetate and lead citrate, and observed with a Philips 300 transmission electron microscope at 60 kV.

Transmission immunoelectron microscopy with colloidal gold-conjugated protein G and rabbit anti-rP28 antibody revealed gold particles bound to *E. chaffeensis* surface. The distribution of the particles was random, close to the surface, and appeared as if almost embedded in the membrane, suggesting that the antigenic epitope protrudes very little from the lipid bilayer. Nonetheless, the antigenic epitope was surface-exposed, and thus, could be recognized by rabbit anti-rP28 antibody. No gold particles were observed on host cytoplasmic membrane or *E. chaffeensis* incubated with normal rabbit serum.

Example 4. Immunization of mice and *E. chaffeensis* challenge.

The rP28 band in SDS- PAGE was excised, minced, and mixed with an equal volume of Freund's incomplete or complete adjuvant. Nine BALB/c male mice (6 weeks old) were divided into two groups. Five mice were intraperitoneally immunized a total of four times at 10-day intervals; twice with a mixture of the minced gel with the rP28 (30 to 40 µg of protein per mouse each time) and incomplete adjuvant, and twice with a mixture of the recombinant protein (the same amount as before) and complete adjuvant. Four mice were intraperitoneally injected with a mixture of the minced gel without protein and the respective adjuvants. For ehrlichia-challenge, approximately  $1 \times 10^7$  DH82 cells heavily-infected with *E. chaffeensis* were disrupted by sonication in serum-free DMEM (GIBCO-BRL) and centrifuged at  $200 \times g$  for 5 min. The supernatant was diluted to a final volume of 5 ml, and 0.3 ml was inoculated intraperitoneally into each mouse 10 days after the last immunization. Before challenge, all 5-immunized mice had a titer of 1:160 against *E. chaffeensis* antigen by IFA and all 4-nonimmunized mice were negative.

At day 5 post-challenge, approximately 1 ml of blood was collected in an EDTA tube from each mouse and protection was assessed by PCR detection of *E. chaffeensis* 16S rDNA in the buffy coat of the collected blood. *E. chaffeensis* could not be reisolated in cell culture at day 10 postinfection. Day 5 post challenge is the optimum time at which establishment of ehrlichial infection can be examined by PCR without the influence of residual DNA from the ehrlichiae used as the challenge before the spontaneous clearance of organisms take place. The *E. chaffeensis*-specific DNA fragment was observed in all nonimmunized mice but not in any immunized mice, indicating that immunization of rP28 apparently protects mice from ehrlichial infection and indicating that the P28 is a potential protective antigen.

#### Example 5 Assaying for the presence of anti-P30 antibody in Dogs

The rP30 protein was used as an antigen in a Western immunoblot analysis and dot blot analysis to detect the presence of antibody to *E. canis* in serum from E-canis infected dogs. The results of the Western immunoblot analysis indicated that reactivity of the sera with rP30 was stronger than the reactivity that was observed when purified *E. canis* was used as antigen. The results of the dot blot assay indicated that rP30 is a useful and sensitive tool for serodiagnosis of canine ehrlichiosis.

What is claimed is:

1. A method for diagnosing an infection with *E. chaffeensis* in a human patient comprising:
  - a) providing a serum sample from the patient;
  - b) providing an isolated or purified P30 protein of *E. canis*, or an immunoreactive fragment thereof, wherein the P30 protein or immunoreactive fragment thereof comprises an amino acid sequence at least 95% identical to amino acid 26 through amino acid 281 of SEQ ID NO: 32;
  - c) contacting the serum sample with the P30 protein or immunoreactive fragment thereof; and
  - d) assaying for the formation of a complex between antibodies in the serum sample and the P30 protein or immunoreactive fragment thereof, wherein formation of said complex is indicative of infection with *E. chaffeensis*.
  
2. The method of claim 1, wherein the P30 protein of *E. canis* comprises an amino acid sequence or immunoreactive fragment thereof that is at least 95% identical to amino acid 33 through amino acid 224 of SEQ ID NO: 32.
  
3. A method for diagnosing an infection with *E. canis* in an animal belonging to the family *Canidae* comprising:
  - a) providing a serum sample from the animal;
  - b) providing an isolated or purified P30 protein of *E. canis*, or an immunoreactive fragment thereof, wherein the P30 protein or immunoreactive fragment thereof comprises an amino acid sequence at least 95% identical to amino acid 26 through amino acid 281 of SEQ ID NO: 32;
  - c) contacting the serum sample with the P30 protein or an immunoreactive fragment thereof; and
  - d) assaying for the formation of a complex between antibodies in the serum sample and the P30 protein or immunoreactive fragment thereof, wherein formation of said complex is indicative of infection with *E. canis*.

4. A method for diagnosing an *E. canis* infection in an animal belonging to the family *Canidae* comprising:
- a) contacting a serum sample from the animal with and *E. canis* P30 protein or an antigenic fragment of the *E. canis* P30 protein, wherein the P30 protein or antigenic fragment thereof comprises an amino acid sequence at least 95% identical to amino acid 26 through amino acid 281 of SEQ ID NO: 32; and
  - b) assaying for the formation of complex between antibodies in the serum sample and the *E. canis* P30 protein or the antigenic fragment of the *E. canis* P30 protein, wherein formation of said complex is indicative of infection with *E. canis*.
5. The method of claim 4, wherein said antigenic fragment comprises amino acid 33 through amino acid 224 of SEQ ID NO: 32.
6. A kit for diagnosing *E. canis* in an animal belonging to the family *Canidae*, said kit comprising i) *E. canis* P30 protein, an antigenic fragment of the *E. canis* P30 protein, or both, and ii) a reagent for assaying for the formation of a complex between antibodies in an animal's serum sample and the *E. canis* P30 protein or the antigenic fragment of the *E. canis* P30 protein, wherein the P30 protein or antigenic fragment thereof comprises an amino acid sequence at least 95% identical to amino acid 26 through amino acid 281 of SEQ ID NO: 32.
7. The kit of claim 6, wherein said antigenic fragment comprises amino acid 33 through amino acid 224 of SEQ ID NO: 32.

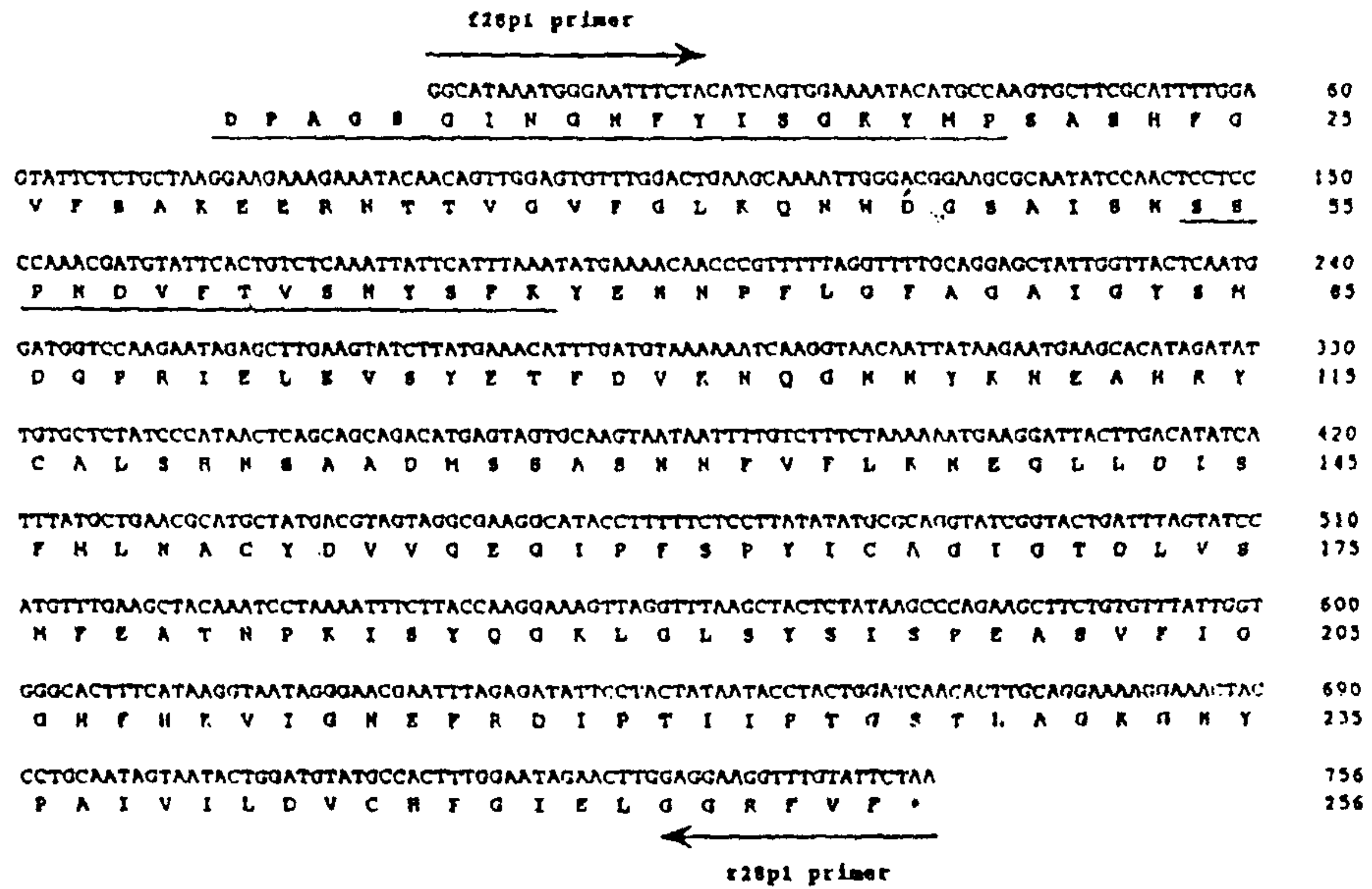


Fig. 1

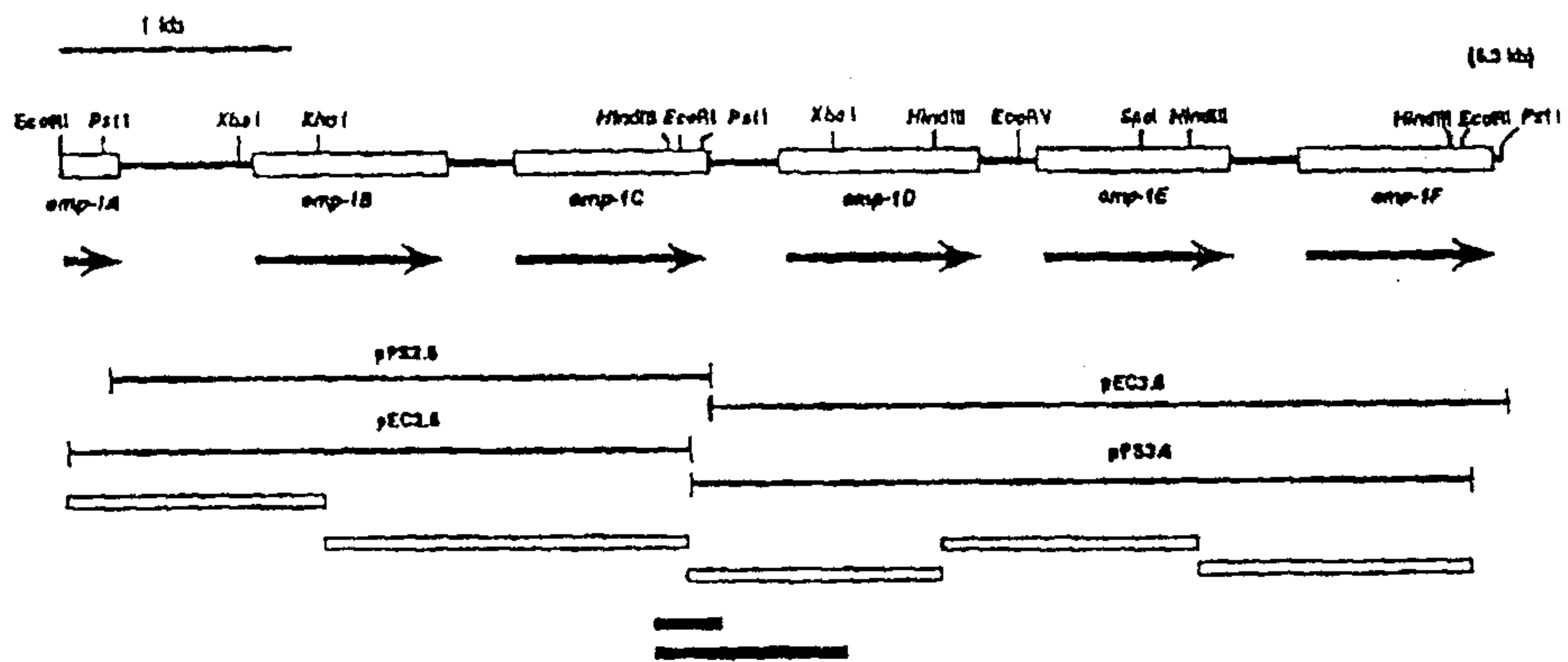


Fig. 2

10	20	30	40	50	60
ATGAATTACA	AAAAAGTTTT	CATAACAAGT	GCATTGATAT	CATTAATATC	TTCTCTACCT
70	80	90	100	110	120
GGAGTATCAT	TTTCCGACCC	AGCAGGTAGT	GGTATTAACG	GTAATTTCTA	CATCAGTGGA
130	140	150	160	170	180
AAATACATGC	CAAGTGCTTC	GCATTTTGGG	GTAATTTCTA	CTAAGGAAGA	AAGAAATACA
190	200	210	220	230	240
ACAGTTGGAG	TGTTTGGACT	GAAGCAAAT	TGGGACGGAA	GCGCAATATC	CAACTCCTCC
250	260	270	280	290	300
CCAAACGATG	TATTCACTGT	CTCAAATTAT	TCATTTAAAT	ATGAAAACAA	CCCGTTTTTA
310	320	330	340	350	360
GGTTTTGCAG	GAGCTATTGG	TTACTCAATG	GATGGTCCAA	GAATAGAGCT	TGAAGTATCT
370	380	390	400	410	420
TATGAAACAT	TTGATGTAAA	AAATCAAGGT	AACAATTATA	AGAATGAAGC	ACATAGATAT
430	440	450	460	470	480
TGTGCTCTAT	CCCATAACTC	AGCAGCAGAC	ATGAGTAGTG	CAAGTAATAA	TTTTGTCTTT
490	500	510	520	530	540
CTAAAAAATG	AAGGATTACT	TGACATATCA	TTTATGCTGA	ACGCATGCTA	TGACGTAGTA
550	560	570	580	590	600
GGCGAAGGCA	TACCTTTTTC	TCCTTATATA	TGCGCAGGTA	TCGGTACTGA	TTTAGTATCC
610	620	630	640	650	660
ATGTTTGAAG	CTACAAATCC	TAAAATTTCT	TACCAAGGAA	AGTTAGGTTT	AAGCTACTCT
670	680	690	700	710	720
ATAAGCCCAG	AAGCTTCTGT	GTTTATTGGT	GGGCACTTTC	ATAAGGTAAT	AGGGAACGAA
730	740	750	760	770	780
TTTAGAGATA	TTCCTACTAT	AATACCTACT	GGATCAACAC	TTGCAGGAAA	AGGAAACTAC
790	800	810	820	830	840
CCTGCAATAG	TAATACTGGA	TGTATGCCAC	TTTGGAAATAG	AACTTGGAGG	AAGGTTTGTA
850	860	870	880	890	900
TTCTAA....	.....	.....	.....	.....	.....

Fig. 3A

10	20	30	40	50	60
MNYKKVFITS	ALISLISSLP	GVSFSDPAGS	GINGNFYISG	KYMPSASHFG	VFSAKEERNT
70	80	90	100	110	120
TVGVFGLKQN	WDGSAISNSS	PNDVFTVSNY	SFKYENNPFL	GFAGAIGYSM	DGPRIELEVS
130	140	150	160	170	180
YETFDVKNQG	NNYKNEAHRY	CALSHNSAAD	MSSASNNEVF	LKNEGLLDIS	FMLNACYDVV
190	200	210	220	230	240
GEGIPFSPYI	CAGIGTDLVS	MFEATNPKIS	YQGKLGLSYS	ISPEASVFIG	GHFHKVIGNE
250	260	270	280	290	300
FRDIPTIIPT	GSTLAGKGNV	PAIVILDVCH	FGIEJ-GGREF	F.....	.....

Fig. 3B

10	20	30	40	50	60
ATGAATTACA	AGAAAATTTT	TGTAAGCAGT	GCATTAATTT	CATTAATGTC	AATCTTACCT
70	80	90	100	110	120
TACCAATCTT	TTGCAGATCC	TGTAACTTCA	AATGATACAG	GAATCAACGA	CAGCAGAGAA
130	140	150	160	170	180
GGCTTCTACA	TTAGTGTAAG	GTATAATCCA	AGCATATCAC	ACTTCAGAAA	ATTCTCAGCT
190	200	210	220	230	240
GAAGAAGCTC	CCATCAATGG	AAATACTTCT	ATCACTAAAA	AGGTTTTTCGG	GCTGAAAAAA
250	260	270	280	290	300
GACGGAGATA	TAGCACAATC	TGCCAATTTT	AACAGGACAG	ATCCAGCCCT	CGAGTTTCAG
310	320	330	340	350	360
AATAACCTAA	TATCAGGATT	CTCAGGAAGT	ATTGGTTATG	CTATGGATGG	GCCAAGAATA
370	380	390	400	410	420
GAACTTGAAG	CTGCATACCA	AAAATTTGAT	GCAAAAAATC	CTGACAACAA	TGACACTAAT
430	440	450	460	470	480
AGCGGTGACT	ACTATAAATA	CTTTGGACTA	TCTCGTGAAG	ACGCAATAGC	AGATAAGAAA
490	500	510	520	530	540
TATGTTGTCC	TTAAAAATGA	AGGCATCACT	TTTATGTTCAT	TAATGGTTAA	CACTTGCTAT
550	560	570	580	590	600
GACATTACAG	CTGAAGGAGT	ACCTTTCATA	CCGTATGCAT	GTGCAGGTGT	AGGAGCAGAC
610	620	630	640	650	660
CTTATAAACG	TATTTAAGGA	TTTTAATTTA	AAATTCTCAT	ACCAAGGGAA	AATAGGTATT
670	680	690	700	710	720
AGCTATCCAA	TCACACCAGA	AGTTTCCGCT	TTTATTGGAG	GATACTACCA	CGGAGTTATA
730	740	750	760	770	780
GGAAATAATT	TTAACAAAAT	ACCTGTAATA	ACACCTGTAG	TATTAGAAGG	AGCTCCTCAA
790	800	810	820	830	840
ACCACATCTG	CGCTAGTAAC	TATTGACACT	GGATACTTTG	GCGGAGAAGT	TGGAGTAAGG
850	860	870	880	890	900
TTCACCTTCT	AG.....	.....	.....	.....	.....

Fig. 4A

10	20	30	40	50	60
MNYKKIFVSS	ALISLMSILP	YQSEADPVT	NDTGINDSRE	GFYISVKYNP	SISHFRKESA
70	80	90	100	110	120
EEAPINGNTS	ITKKVFGLKK	DGDIAQSANF	NRTDPALEFQ	NNLISGFSGS	IGYAMDGPRI
130	140	150	160	170	180
ELEAAYQKFD	AKNPDNNDTN	SGDYKYFGL	SREDAIADKK	YVVLKNEGIT	FMSLMVNTCY
190	200	210	220	230	240
DITAEGVPFI	PYACAGVGAD	LINVKDFNL	KFSYQKGIGI	SYFITPEVSA	FIGGYHGVVI
250	260	270	280	290	300
GNNFNKIPVI	TPVVLEGAPQ	TTSALVTIDT	GYFGGEVGVV	FTF.....	.....

Fig. 4B

10	20	30	40	50	60
ATGAACTGCA	AAAAATTTT	TATAACAAC	GCATTGGCAT	TGCCAATGTC	TTTCTTACCT
70	80	90	100	110	120
GGAATATTAC	TTTCTGAACC	AGTACAAGAT	GACAGTGTGA	GTGGCAATTT	CTATATTAGT
130	140	150	160	170	180
GGCAAGTACA	TGCCAAGTGC	TTCTCATTTT	GGAGTTTTCT	CTGCCAAAGA	AGAAAAAAT
190	200	210	220	230	240
CCTACTGTGC	CGTTGTATGG	TTTGAAACAA	GATTGGAACG	GTGTTAGTGC	TTCAAGTCAT
250	260	270	280	290	300
GCTGATGCGG	ACTTTAATAA	CAAAGGTTAT	TCTTTTAAAT	ACGAAAACAA	TCCATTTCTA
310	320	330	340	350	360
GGTTTTGCAG	GAGCTATTGG	TTATTCAATG	GGTGGTCCAA	GAATAGAGTT	TGAAGTGTCC
370	380	390	400	410	420
TATGAAACAT	TTGACGTGAA	AAATCAAGGT	GGTAATTACA	AAAATGATGC	TCACAGATAC
430	440	450	460	470	480
TGTGCCTTAG	ATCGTAAAGC	AAGCAGCACT	AATGCCACAG	CTAGTCACTA	CGTGCTACTA
490	500	510	520	530	540
AAAAATGAAG	GACTACTTGA	TATATCACTT	ATGTTGAATG	CATGCTATGA	CGTAGTAAGT
550	560	570	580	590	600
GAAGGAATAC	CTTTCTCTCC	TTACATATGT	GCAGGTGTTG	GTACCGATTT	AATATCCATG
610	620	630	640	650	660
TTTGAAGCTA	TAAACCCTAA	AATTTCTTAT	CAAGGAAAGT	TAGGTTTGAG	TTACTCTATA
670	680	690	700	710	720
AACCCAGAAG	CTTCTGTCTT	TGTTGGTGGA	CATTTTCATA	AAGTTGCAGG	TAATGAATTC
730	740	750	760	770	780
AGGGACATTT	CTACTCTTAA	AGCGTTTGCT	ACACCATCAT	CTGCAGCTAC	TCCAGACTTA
790	800	810	820	830	840
GCAACAGTAA	CACTGAGTGT	GTGTCACTTT	GGAGTAGAAC	TTGGAGGAAG	ATTTAACTTC
850	860	870	880	890	900
TAA.....	.....	.....	.....	.....	.....

Fig. 5A

10	20	30	40	50	60
MNCKKFFITT	ALALPMSFLP	GILLSEPVDQ	DSVSGNFYIS	GKYMPSASHF	GVFSAKEEKN
70	80	90	100	110	120
PTVALYGLKQ	DWNGVSASSH	ADADFNKGY	SFKYENNPFL	GFAGAIGYSM	GGPRIEFEVS
130	140	150	160	170	180
YETFDVKNQG	GNYKNDHRY	CALDRKASST	NATASHYVLL	KNEGLLDISL	MLNACYDVVS
190	200	210	220	230	240
EGIPFSPYIC	AGVGTDLISM	FEAINPKISY	QGKLGLSYSI	NPEASVFEVGG	HFHKVAGNEF
250	260	270	280	290	300
RDISTLKAF	TPSSAATPDL	ATVTLVCHF	GVELGGRENF	.....	.....

Fig. 5B

10	20	30	40	50	60
ATGAACTGCG	AAAAATTTT	TATAACAAC	GCATTAACAT	TACTAATGTC	CTTCTTACCT
70	80	90	100	110	120
GGAATATCAC	TTTCTGATCC	AGTACAGGAT	GACAACATTA	GTGGTAATTT	CTACATCAGT
130	140	150	160	170	180
GGAAAGTATA	TGCCAAGCGC	TTCGCATTTT	GGAGTTTTTT	CTGCCAAGGA	AGAAAGAAAT
190	200	210	220	230	240
ACAACAGTTG	GAGTATTTGG	AATAGAGCAA	GATTGGGATA	GATGTGTAAT	ATCTAGAACC
250	260	270	280	290	300
ACTTTAAGCG	ATATATTCAC	CGTTCCAAAT	TATTCATTTA	AGTATGAAA	TAATCTATTT
310	320	330	340	350	360
TCAGGATTTG	CAGGAGCTAT	TGGCTACTCA	ATGGATGGCC	CAAGAATAGA	GCTTGAAGTA
370	380	390	400	410	420
TCTTATGAAG	CATTCGATGT	TAAAAATCAA	GGTAACAATT	ATAAGAACGA	AGCACATAGA
430	440	450	460	470	480
TATTATGCTC	TGTCCCATCT	TCTCGGCACA	GAGACACAGA	TAGATGGTGC	AGGCAGTGCG
490	500	510	520	530	540
TCTGTCTTTC	TAATAAATGA	AGGACTACTT	GATAAATCAT	TTATGCTGAA	CGCATGTTAT
550	560	570	580	590	600
GATGTAATAA	GTGAAGGCAT	ACCTTTTTCT	CCTTATATAT	GTGCAGGTAT	TGGTATTGAT
610	620	630	640	650	660
TTAGTATCCA	TGTTTGAAGC	TATAAATCCT	AAAATTTCTT	ATCAAGGAAA	ATTAGGCTTA
670	680	690	700	710	720
AGTTACCCTA	TAAGCCCAGA	AGCTTCTGTG	TTTATTGGTG	GACATTTTCA	TAAGGTGATA
730	740	750	760	770	780
GGAAACGAAT	TTAGAGATAT	TCCTACTATG	ATACCTAGTG	AATCAGCGCT	TGCAGGAAAA
790	800	810	820	830	840
GGAAACTACC	CTGCAATAGT	AACACTGGAC	GTGTTCTACT	TTGGCATAGA	ACTTGGAGGA
850	860	870	880	890	900
AGGTTTAACT	TCCAACCTTG	A.....	.....	.....	.....

Fig. 6A

10	20	30	40	50	60
MNCEKFFITT	ALTLMSFLP	GISLSDPVQD	DNISGNFYIS	GKYMPSASHE	GVFSAKEERN
70	80	90	100	110	120
TTVGVFGIEQ	DWDRCVISRT	TLSDIFTVPN	YSEKYENNLF	SGFAGAIGYS	MDGPRIELEV
130	140	150	160	170	180
SYEAFDVKNQ	GNNYKNEAHR	YYALSHLLGT	ETQIDGAGSA	SVFLINEGLL	DKSEMLNACY
190	200	210	220	230	240
DVISEGIPFS	PYICAGIGID	LVSMFEAINP	KISYQGKGLG	SYPIspeasv	FIGGHFHKVI
250	260	270	280	290	300
GNEFRDIPTM	IPSESALAGK	GNYPaiVtLD	Vfyfgielgg	Rfnfql....	.....

Fig. 6B

10	20	30	40	50	60
ATGAATTGCA	AAAAATTTT	TATAACAAC	GCATTAGTAT	CACTAATGTC	CTTTCTACCT
70	80	90	100	110	120
GGAATATCAT	TTTCTGATCC	AGTGCAAGGT	GACAATATTA	GTGGTAATTT	CTATGTTAGT
130	140	150	160	170	180
GGCAAGTATA	TGCCAAGTGC	TTCGCATTTT	GGCATGTTTT	CTGCCAAAGA	AGAAAAAAT
190	200	210	220	230	240
CCTACTGTTG	CATTGTATGG	CTTAAAACAA	GATTGGGAAG	GGATTAGCTC	ATCAAGTCAC
250	260	270	280	290	300
AATGATAATC	ATTTCAATAA	CAAGGGTTAT	TCATTTAAAT	ATGAAAATAA	CCCATTTTTA
310	320	330	340	350	360
GGGTTTGCAG	GAGCTATTGG	TTATTCAATG	GGTGGTCCAA	GAGTAGAGTT	TGAAGTGTCC
370	380	390	400	410	420
TATGAAACAT	TTGACGTAA	AAATCAGGGT	AATAACTATA	AAAATGATGC	TCACAGATAC
430	440	450	460	470	480
TGTGCTTTAG	GTCAACAAGA	CAACAGCGGA	ATACCTAAA	CTAGTAAATA	CGTACTGTTA
490	500	510	520	530	540
AAAAGCGAAG	GATTGCTTGA	CATATCATT	ATGCTAAATG	CATGCTATGA	TATAATAAAC
550	560	570	580	590	600
GAGAGCATA	CTTTGTCTCC	TTACATATGT	GCAGGTGTTG	GTACTGATTT	AATATCCATG
610	620	630	640	650	660
TTTGAAGCTA	CAAATCCTAA	AATTTCTTAC	CAAGGGAAGT	TAGGTCTAAG	TTACTCTATA
670	680	690	700	710	720
AACCCAGAAG	CTTCTGTATT	TATTGGTGG	CATTTTCATA	AGGTGATAGG	AAACGAATTT
730	740	750	760	770	780
AGGGACATTC	CTACTCTGAA	AGCATTGTG	ACGTCATCAG	CTACTCCAGA	TCTAGCAATA
790	800	810	820	830	840
GTAACACTAA	GTGTATGTCA	TTTTGGAATA	GAACTTGGAG	GAAGGTTTAA	CTTCTAA...

Fig. 7A

10	20	30	40	50	60
MNCKKFFITT	ALVSLMSFLP	GISFSDPVQG	DNISGNFYVS	GKYMPSASHF	GMFSAKEEKN
70	80	90	100	110	120
PTVALYGLKQ	DWEGISSSSH	NDNHFNKGY	SFKYENNPFL	GFAGAIGYSM	GGPRVEFEVS
130	140	150	160	170	180
YETFDVKNQG	NNYKNDAHRY	CALGQQDNSG	IPKTSKYVLL	KSEGLLDISE	MLNACYDIIN
190	200	210	220	230	240
ESIPLSPYIC	AGVGTDLISM	FEATNPKISY	QGKLGLSYSI	NPEASVFIGG	HFHKVIGNEF
250	260	270	280	290	300
RDIPFLKAFV	TSSATPDLAI	VTLSVCHFGI	ELGGRFNF..	.....	.....

Fig. 7B

10	20	30	40	50	60
ATGAAATTGCA	AAAAATTTT	TATAACAACT	ACATTAGTAT	CGCTAATGTC	CTTCTTACCT
70	80	90	100	110	120
GGAATATCAT	TTTCTGATGC	AGTACAGAAC	GACAATGTTG	GTGGTAATTT	CTATATCAGT
130	140	150	160	170	180
GGGAAATATG	TACCAAGTGT	TTCACATTTT	GGCGTATTCT	CTGCTAAACA	GGAAAGAAAT
190	200	210	220	230	240
ACAACAACCG	GAGTATTTGG	ATTAAAGCAA	GATTGGGATG	GCAGCACAAT	ATCTAAAAAT
250	260	270	280	290	300
TCTCCAGAAA	ATACATTTAA	CGTTCCAAAT	TATTCATTTA	AATATGAAAA	TAATCCATTT
310	320	330	340	350	360
CTAGGTTTTG	CAGGAGCTGT	TGGTTATTTA	ATGAATGGTC	CAAGAATAGA	GTTAGAAATG
370	380	390	400	410	420
TCCTATGAAA	CATTTGATGT	GAAAAACCAG	GGTAATAACT	ATAAGAACGA	TGCTCACAAA
430	440	450	460	470	480
TATTATGCTT	TAACCCATAA	CAGTGGGGGA	AAGCTAAGCA	ATGCAGGTGA	TAAGTTTGTT
490	500	510	520	530	540
TTTCTAAAAA	ATGAAGGACT	ACTTGATATA	TCACTTATGT	TGAATGCATG	CTATGATGTA
550	560	570	580	590	600
ATAAGTGAAG	GAATACCTTT	CTCTCCTTAC	ATATGTGCAG	GTGTTGGTAC	TGATTTAATA
610	620	630	640	650	660
TCCATGTTTG	AAGCTATAAA	CCCTAAAATT	TCTTATCAAG	GAAAGTTAGG	TTTGAGTTAC
670	680	690	700	710	720
TCCATAAGCC	CAGAAGCTTC	TGTTTTTGT	GGTGGACATT	TTCATAAGGT	GATAGGGAAT
730	740	750	760	770	780
GAATTCAGAG	ATATTCCTGC	TATGATACCC	AGTACCTCAA	CTCTCACAGG	TAATCACTTT
790	800	810	820	830	840
ACTATAGTAA	CACTAAGTGT	ATGCCACTTT	GGAGTGGAAC	TTGGAGGAAG	GTTTAACTTT
850	860	870	880	890	900
TAA.....	.....	.....	.....	.....	.....

Fig. 8A

10	20	30	40	50	60
MNCKKFFITT	TLVSLMSFLP	GISFSDAVQN	DNVGGNFYIS	GKYVPSVSHF	GVFSAKQERN
70	80	90	100	110	120
TTTGVFGLKQ	DWDGSTISKN	SPENTFNVPN	YSFKYENNPF	LGFAGAVGYL	MNGPRIELEM
130	140	150	160	170	180
SYETFVKNQ	GNNYKNDHK	YYALTHNSGG	KLSNAGDKFV	FLKNEGLLDI	SLMLNACYDV
190	200	210	220	230	240
ISEGIPFSPY	ICAGVGTDLI	SMFEAINPKI	SYQGKLGLSY	SISPEASVFF	GGHFHKVIGN
250	260	270	280	290	300
EFRDIPAMIP	STSTLTGNHF	TIVTSLVCHF	GVELGGRFNF	.....	.....

Fig. 8B

10	20	30	40	50	60
ATGGAAAATC	TCATGAATAA	GAAAAACAAA	TTCTTTACAA	TAAGTACAGC	AATGGTATGC
70	80	90	100	110	120
TTATTGTTAT	TACCTGGTAT	ATCATTTTCA	GAAACTATAA	ACAACAGTGC	TAAAAAACAG
130	140	150	160	170	180
CCTGGGTTAT	ATATCAGTGG	GCAGTACAAA	CCTAGTGTTT	CAGTTTTTAG	TAATTTTTCA
190	200	210	220	230	240
GTAAAAGAAA	CTAATGTTCC	CACAAAGCAG	TTAATAGCAC	TTAAAAAAGA	CATTAATTCT
250	260	270	280	290	300
GTTGCAGTTG	GTAGTAATGC	TACTACAGGT	ATTAGCAATC	CAGGTAATTT	CACAATTCCT
310	320	330	340	350	360
TATACTGCAG	AATTTCAAGA	TAATGTTGCC	AATTTCAATG	GGGCTGTTGG	TTACTCTTTT
370	380	390	400	410	420
CCTGATAGTC	TAAGAATTGA	AATAGAGGGA	TTTCATGAAA	AATTTGATGT	CAAAAACCCT
430	440	450	460	470	480
GGAGGTACA	CACAAGTAAA	AGATGCGTAC	CGTTATTTTG	CACTAGCAGC	TGATTTAAAA
490	500	510	520	530	540
GATGGCTTCT	TTGAACCTAA	AGCGGAAGAT	ACAGGTGTTT	ATCATACTGT	TATGAAAAAT
550	560	570	580	590	600
GATGGATTAT	CTATTTTATC	TACTATGGTT	AACGTCTGTT	ACGATTTTTT	TGTAGATGAA
610	620	630	640	650	660
TTACCAGTCT	TACCTTATAT	ATGTGCAGGT	ATGGGTATAA	ACGCCATAGA	ATTCTTCGAC
670	680	690	700	710	720
GCTTTACATG	TAAAATTTGC	TTACCAAGGC	AAACTAGGTA	TTAGCTATCA	ACTATTTACT
730	740	750	760	770	780
AAAGTAAATT	TATTCCTTGA	TGGGTATTAC	CATCAAGTAA	TAGGCAATCA	ATTCAAAAAC
790	800	810	820	830	840
TTAAACGTAA	ACCATGTTTA	CACACTTAAA	GAATCTCCTA	AAGTCACATC	TGCAGTAGCT
850	860	870	880	890	900
ACACTTGACA	TTGCATACTT	TGGTGGCGAA	GTTGGAATAA	GATTCACATT	TTAA.....

Fig. 9A

10	20	30	40	50	60
MVCLLLLPGI	SFSETINNSA	KKQPGLYISG	QYKPSVSVFS	NFSVKETNVP	TKQLIALKKD
70	80	90	100	110	120
INSVAVGSNA	TTGISNPGNF	TIPYTAEFQD	NVANFNGAVG	YSFPDSLRIE	IEGFHEKFDV
130	140	150	160	170	180
KNPGGYTQVK	DAYRYFALAR	DLKDGFFEPK	AEDTGVYHTV	MKNDGLSILS	TMVNVCYDFS
190	200	210	220	230	240
VDELPVLPYI	CAGMGINAIE	FFDALHVKEA	YQKLGISYQ	LETKVNLELD	GYHQVIGNQ
250	260	270	280	290	300
EKNLNVNHVY	TLKESPKVTS	AVATLDIAYF	GGEVGIRFTE	.....	.....

Fig. 9B

10	20	30	40	50	60
ATGATATATA	AAGAAAAACT	TACTAGAGTG	GGAGAATATA	TCTTAGCATA	TTTATCATTT
70	80	90	100	110	120
ATTCTTTCTA	CTTATATCTT	TCTAGTGCTG	GTAAATATTA	TTAGATATAA	CAGCCTTGCT
130	140	150	160	170	180
ATATGTGTTA	TCAGTCTACT	AAGAACTAAT	ATCTTTAACG	TTAGCACAAA	AAAATTAATA
190	200	210	220	230	240
AAAGATAAAT	GTCGTGATAC	TAAGTTTAGT	AACATGAATT	GTTATTTGTA	CGGTAAACCG
250	260	270	280	290	300
TTAAATTTAC	AAATTTTTTA	TGGAATATTT	TCCTTTATTA	GAACTTTCA	AAATAACACA
310	320	330	340	350	360
CTAATAATTC	CTAATGATAG	TAAATGCGGC	TTCTATACCA	CGTTATGGGA	TAATCCAGCA
370	380	390	400	410	420
CTACATTATA	CATATACACT	TACTGGCAGT	GAGTACCGTA	ATTTTTTTGA	CATTCTATAT
430	440	450	460	470	480
GAAAACATTA	TCTGTCAATG	TAAATTAATT	ATTAACTATA	ACCGTTCTGT	ATTAAACCAA
490	500	510	520	530	540
CATAATAAAA	ATACTCTCGT	AATAATACCA	ATACCTAATG	CTAGAGAGTT	CAGTAATGAA
550	560	570	580	590	600
ATTCGAGTAA	GGAATATATC	AATAAATAAG	GAAAGTTCTT	ATGAGTGCTA	A.....

Fig. 10A

10	20	30	40	50	60
MIYKEKLTRV	GEYILAYLSF	ILSTYIFLVL	VNIIRYNSLA	ICVISLLRTN	IFNVSTKKLI
70	80	90	100	110	120
KDKCRDTKFS	NMNCYLYGKP	LNLOIFYGIF	SFIRNFQNNT	LIIPNSKCG	FYTTLWDNPA
130	140	150	160	170	180
LHYTYTLTGS	EYRNFFDILY	ENIICQCKLL	INYNRSVLNQ	HNKNTLVIIP	IPNAREFSNE
190	200	210	220	230	240
IRVRNISINK	ESSYEC....	.....	.....	.....	.....

Fig. 10B

10	20	30	40	50	60
ATGAATAAAA	AAAACAAGTT	TATTATAGCT	ACAGCATTGG	TATATTTACT	GTCATTACCT
70	80	90	100	110	120
AGTGTATCGT	TTTCAGAGGT	TACAAACAGC	AGTATTAAAA	AACACTCTGG	GTTATATATT
130	140	150	160	170	180
AGTGGACAAT	ACAAACCAAG	TGTTTCTGTT	TTTAGTAGTT	TCTCAATTAA	AGAAACTAAC
190	200	210	220	230	240
ACTATCACAA	AAAATCTTAT	AGCGTTAAAA	AAAGATATTA	ACTCTCTTGA	AGTTAACGCC
250	260	270	280	290	300
GATGCTAGTC	AAGGTATTAG	TCATCCAGGA	AATTTTACTA	TACCTTATAT	AGCAGCATTT
310	320	330	340	350	360
GAAGATAATG	CTTTTAATTT	CAACGGTGCT	ATTGGTTACA	TTACTGAAGG	TCTAAGGATT
370	380	390	400	410	420
GAAATAGAAG	G TTCCTATGA	AGAATTTGAT	GCTGAAAACC	CTGGAGGTTA	TGGTCTAAAT
430	440	450	460	470	480
GATGCCTTTC	GGTACTTTGC	TTTAGCACGT	GATATGGAAA	GCAACAAGTT	CCTACCAAAA
490	500	510	520	530	540
GCACAAAGCT	CAC.....	.....	.....	.....	.....

Fig. 11A

10	20	30	40	50	60
MNKKNKFIIA	TALVYLLSLP	SVSFSEVTNS	SIKKHSGLYI	SGQYKPSVSV	FSSFSIKETN
70	80	90	100	110	120
TITKNLIALK	KDINSLEVNA	DASQGISHPG	NFTIPYIAAF	EDNAFNENGA	IGYITEGLRI
130	140	150	160	170	180
EIEGSYEED	AENPGGYGLN	DAFRYFALAR	DMESNKFLPK	AQSS.....	.....

Fig. 11B

10	20	30	40	50	60
TCTAGAATAC	ATGATGAAAA	TTATGCTATT	ACAACAAATA	ATAAATTATC	CATCGCATCT
70	80	90	100	110	120
ATTATGGTTA	ACACCTGCTA	TGATATTTCA	ATTAATAATA	CATCAATAGT	ACCGTATTTA
130	140	150	160	170	180
TGCACAGGCA	TTGGTGAAGA	TCTTGTAGGG	CTTTTTAATA	CAATACATTT	TAAACTTGCA
190	200	210	220	230	240
TATCAAGGGA	AAGTTGGAAT	GAGTTATTTG	ATAAATAACA	ATATCCTATT	ATTTTCTGAC
250	260	270	280	290	300
ATATATTATC	ATAAAGTCAT	GGGTAACAGA	TTTAAAATT	TGTACATGCA	ATATGTAGCT
310	320	330	340	350	360
GATCCTAATA	TTTCTGAAGA	AACTATACCT	ATATTAGCAA	AACTTGATAT	TGGTTATTTT
370	380	390	400	410	420
GGAAGTGAAA	TTGGAATAAG	GTTTATGTTT	AACTAA....	.....	.....

Fig. 12A

10	20	30	40	50	60
SRIHDENYAI	TTNNKLSIAS	IMVNTCYDIS	INNTSIVPYL	CTGIGEDLVG	LFNTIHFKLA
70	80	90	100	110	120
YQKVGMSYL	INNNILFSD	IYYHKVMGNR	FKNLYMQYVA	DPNISEETIP	ILAKLDIGYF
130	140	150	160	170	180
GSEIGIREMF	N.....	.....	.....	.....	.....

Fig. 12B

10	20	30	40	50	60
ATGACAAAGA	AATTTAATTT	TGTAAATGTT	ATATTAACAT	TTTTGTTATT	TCTTTTCCCA
70	80	90	100	110	120
CTTAAGTCAT	TTACAACATA	TGCAAATAAT	AACACAATCA	CTCAAAAAGT	TGGATTGTAC
130	140	150	160	170	180
ATAAGTGGTC	AATATAAGCC	AAGTATTCCT	CATTTCAAGA	ATTTTTCAGT	AGAAGAAAAT
190	200	210	220	230	240
GACAAAGTAG	TAGATTTGAT	AGGTCTTACA	ACTGATGTTA	CATATATCAC	AGAACATATA
250	260	270	280	290	300
TTACGAGATA	ATACAAAATT	CAACACTCAT	TATATTGCAA	AGTTCAAGAA	CAATTTTATA
310	320	330	340	350	360
AATTCAGCA	GTGCAATTGG	TTATTATTCT	GGGCAAGGAC	CAAGGTTAGA	AATAGAAAGC
370	380	390	400	410	420
TCTTATGGGG	ATTTTGATGT	TGTAAATTAT	AAAAATTATG	CAGTACAAGA	TGTTAATAGA
430	440	450	460	470	480
TATTTTGCTT	TAGTACGTGA	AAAAAATGGT	TCAAATTTCT	CTCCAAAACC	ACATGAAACT
490	500	510	520	530	540
AGTCAACCCT	CTGACAGTAA	TCCTAAAAAG	TCTTTTTATA	CTTTAATGAA	GAATAATGGG
550	560	570	580	590	600
GTATTTGTTG	CATCAGTAAT	AATCAACGGT	TGTTATGATT	TTTCTTTTAA	TAACACAACA
610	620	630	640	650	660
ATATCACCTT	ACGTATGTAT	AGGAGTTGGA	GGAGATTTTA	TAGAGTTTTT	TGAAGTAATG
670	680	690	700	710	720
CATATCAAGT	TTGCTTGCCA	AAGTAAGGTT	GGTATTAGCT	ATCCAATATC	TCCCTCTATT
730	740	750	760	770	780
ACTATTTTTG	CTGATGCACA	TTATCACAAAG	GTCATAAATA	ATAAATTTAA	CAACCTACAT
790	800	810	820	830	840
GTTAAGTATT	CATATGAACT	TAAAAACTCA	CCTACCATTA	CCTCTGCAAC	AGCCAAACTA
850	860	870	880	890	900
AACATTGAAT	ATTTTGGTGG	TGAAGTTGGG	ATGAGATTTA	TATTTTAA..	.....

Fig. 13A

10	20	30	40	50	60
MTKKFNFVNV	ILTFLLFLFP	LKSFTTYANN	NTITQKVGLY	ISGQYKPSIP	HFKNFSVEEN
70	80	90	100	110	120
DKVVDLIGLT	TDVTYITEHI	LRDNTKFENTH	YIAKFKNNEI	NFSSAIGYYS	GQGPRLEIES
130	140	150	160	170	180
SYGDEFVVNY	KNYAVQDVNR	YFALVREKNG	SNFSPKPHET	SQPSDSNPKK	SFYTLMKNG
190	200	210	220	230	240
VFVASVIING	CYDFEFNNTT	ISP <sub>Y</sub> VCIGVG	GDFIEFFEVM	HIKFACQSKV	GISYPISPSI
250	260	270	280	290	300
TIFADAHYHK	VINNKFNHLH	VKYSYELKNS	PTITSATAKL	NIEYFGGEVG	MRFIF.....

Fig. 13B

10	20	30	40	50	60
ATGAGCAAAA	AAAAGTTTAT	TACAATAGGA	ACAGTACTTG	CATCTCTATT	ATCATTCTTA
70	80	90	100	110	120
TCTATTGAAT	CCTTTTCAGC	TATAAATCAT	AATCATAACAG	GAAATAACAC	TAGTGGTATA
130	140	150	160	170	180
TATATTACAG	GGCAGTATAG	ACCAGGAGTA	TCCCATTTTA	GCAATTTCTC	AGTAAAAGAA
190	200	210	220	230	240
ACTAATGTTG	ATACAATACA	ACTAGTAGGA	TATAAAAAAA	GTGCGTCTTC	TATCGATCCT
250	260	270	280	290	300
AACACTTATT	CAAACCTTCA	AGGTCCATAT	ACTGTTACAT	TTCAAGATAA	TGCTGCTAGT
310	320	330	340	350	360
TTCAGTGGAG	CAATTGGATA	TTCTTACCCC	GAAAGTCTAA	GACTTGAACT	TGAAGGTTCT
370	380	390	400	410	420
TACGAAAAAT	TTGATGTCAA	AGATCCTAAA	GACTACTCAG	CAAAAGATGC	TTTTAGGTTT
430	440	450	460	470	480
TTTGCTCTAG	CACGTAATAC	GTCTACTACT	GTTCCCTGATG	CTCAAAAATA	TACAGTTATG
490	500	510	520	530	540
AAGAATAATG	GCTTATCTGT	TGCATCAATC	ATGATCAATG	GTTGTTATGA	TCTATCTTTT
550	560	570	580	590	600
AATAATTTAG	TCGTATCACC	TTATATATGT	GCAGGTATTG	GTGAAGATTT	CATTGAATTT
610	620	630	640	650	660
TTTGATACTT	TGCACATTAA	ACTTGCTTAT	CAAGGAAAAC	TAGGTATTAG	TTATTACTTC
670	680	690	700	710	720
TTTCCTAAGA	TTAATGTATT	TGCTGGTGGG	TACTATCATA	GAGTTATAGG	GAATAAATTT
730	740	750	760	770	780
AAAAATTTAA	ATGTTAACCA	TGTTGTTACA	CTTGATGAAT	TTCCTAAAGC	AACTTCTGCA
790	800	810	820	830	840
GTAGCTACAC	TTAATGTTGC	TTATTTTGGT	GGTGAAGCTG	GAGTAAAGTT	TACATTTTAA
850	860	870	880	890	900
.....	.....	.....	.....	.....	.....

Fig. 14A

10	20	30	40	50	60
MSKKKFITIG	TVLASLLSFL	SIESFSAINH	NHTGNNTSGI	YITGQYRPGV	SHFSNFSVKE
70	80	90	100	110	120
TNVDTIQLVG	YKKSASSIDP	NTYSNEQGPY	TVTFQDNAAS	FSGAIGYSYP	ESLRLELEGS
130	140	150	160	170	180
YEKFDVKDPK	DYSAKDAFRF	FALARNTSTT	VPDAQKYTVM	KNNGLSVASI	MINGCYDLSF
190	200	210	220	230	240
NNLVVSPYIC	AGIGEDFIEF	FDTLHIKLAY	QGKLGISYF	FPKINVFAGG	YXHRVIGNKF
250	260	270	280	290	300
KNLNVNHVVT	LDEFPKATSA	VATLNVAYFG	GEAGVKFTF.	.....	.....

Fig. 14B

10	20	30	40	50	60
ATGAGTGCTA	AAAAAAGCT	TTTTATAATA	GGGTCAGTGT	TAGTATGTTT	AGTGTCATAC
70	80	90	100	110	120
TTACCTACTA	AATCTTTGTC	AAACTTAAAT	AATATTAATA	ATAACACTAA	GTGCACTGGG
130	140	150	160	170	180
CTATATGTCA	GTGGACAATA	TAAACCTACT	GTTTCTCACT	TTAGTAATTT	TTCACTTAAA
190	200	210	220	230	240
GAAACTTATA	CTGACACTAA	AGAGTTATTA	GGACTAGCAA	AAGATATTAA	GTCTATTACA
250	260	270	280	290	300
GATATAACAA	CAAATAAAAA	ATTCAACATT	CCTTATAACA	CAAATTTTCA	AGATAATGCT
310	320	330	340	350	360
GTTAGCTTCA	GTGCAGCTGT	TGGATATATT	TCCCAAGACA	GTCCAAGGGT	TGAGGTAGAA
370	380	390	400	410	420
TGGTCTTATG	AAGAATTTGA	CGTTAAAAAT	CCTGGTAATT	ACGTAGTAAG	TGAAGCCTTC
430	440	450	460	470	480
AGGTATATTG	CTTTAGCAAG	AGGAATTGAT	AATCTTCAA	AATATCCTGA	AACAAATAAG
490	500	510	520	530	540
TATGTTGTTA	TAAAGAACAA	TGGCTTATCT	GTCGCATCCA	TTATAATCAA	TGGCTGTTAT
550	560	570	580	590	600
GATTTTTCTT	TAAACAATTT	AAAAGTATCA	CCTTACATAT	GCGTAGGGTT	TGGTGGGGAC
610	620	630	640	650	660
ATTATAGAAT	TTTTTAGTGC	TGTAAGTTTT	AAATTTGCTT	ATCAAGGTAA	GGTAGGTATC
670	680	690	700	710	720
AGTTATCCAT	TATTCTCTAA	TATGATTATA	TTTGCTGACG	GATATTACCA	TAAGGTCATA
730	740	750	760	770	780
GGAAATAAAT	TTAACAATTT	AAATGTTCAA	CACGTTGTTA	GTCTTAACAG	TCATCCTAAG
790	800	810	820	830	840
TCTACTTTTG	CAGTAGCTAC	TCTTAATGTT	GAGTATTTTC	GTAGTGAATT	TGGGTAAAAA
850	860	870	880	890	900
TTTATATTTT	AA.....	.....	.....	.....	.....

Fig. 15A

10	20	30	40	50	60
MSAKKKLEFII	GSVLVCLVSY	LPTKSLSNLN	NINNNTKCTG	LYVSGQYKPT	VSHFSNFSLK
70	80	90	100	110	120
ETYTDTKELL	GLAKDIKSIT	DITTNKKFNI	PYNTKEQDNA	VSFSAAVGYI	SQDSPRVEVE
130	140	150	160	170	180
WSYEEFDVKN	PGNYVVSEAF	RYIALARGID	NLQKYPETNK	YVVIKNNGLS	VASIIINGCY
190	200	210	220	230	240
DFSLNNLKVS	PYICVGFEGD	IEFFSAVSE	KFAYQKVGVI	SYPLESNMII	FADGYXHKVI
250	260	270	280	290	300
GNKFNNLNVQ	HVVSLNSHPK	STFAVATLNV	EYFGSEFGLK	FIF.....	.....

Fig. 15B

10	20	30	40	50	60
ATGAGTAAAA	AAAATTTTAT	TACAATAGGA	GCAACACTTA	TTCATATGTT	GTTACCTAAC
70	80	90	100	110	120
ATATCTTTTC	CAGAAACTAT	TAACAATAAC	ACTGATAAAC	TTTCTGGGTT	ATATATAAGT
130	140	150	160	170	180
GGGCAATATA	AACCAGGGAT	TTCTCATTTC	AGCAAATTTT	CAGTCAAAGA	AATCTATAAT
190	200	210	220	230	240
GATAACATTC	AACTAATTGG	GTTAAGACAC	AACGCAATTT	CTACTAGTAC	CCTTAATATT
250	260	270	280	290	300
AATACAGATT	TTAATATCCC	CTATAAAGTA	ACATTTCAA	ATAACATTAC	CAGCTTTAGT
310	320	330	340	350	360
GGAGCTATTG	GTTATTCTGA	TCCCACAGGG	GCAAGATTTG	AGCTTGAAGG	TTCTTATGAA
370	380	390	400	410	420
GAATTTGATG	TGACAGATCC	TGGAGACTGC	TTAATAAAAG	ATACCTATAG	ATATTTGCGT
430	440	450	460	470	480
TTAGCTAGAA	ACCCATCAGG	TTCTAGCCCT	ACCTCAAACA	ACTATACTGT	TATGAGAAAT
490	500	510	520	530	540
GATGGTGTTT	CCATTACTTC	TGTTATATTT	AATGGCTGTT	ATGACATCTT	TTTAAAGGAT
550	560	570	580	590	600
TTAGAAGTAT	CACCTTATGT	ATGTGTTGGT	GTAGGTGGAG	ATTTTATAGA	ATTTTTTGAC
610	620	630	640	650	660
GCATTACACA	TTAAATTAGC	ATACCAAGGC	AAGTTAGGTA	TCAATTATCA	CTTATCGACT
670	680	690	700	710	720
CAAGCAAGCG	TATTTATTGA	TGGATATTAT	CATAAGGTTA	TAGGAAATCA	ATTCAACAAT
730	740	750	760	770	780
CTAAATGTTC	AACACGTGGC	TAGTACAGAT	TTTGGACCTG	TATACGCAGT	AGCCACACTT
790	800	810	820	830	840
AACATTGGTT	ATTTTGGTGG	TGAAATCGGA	ATTAGACTTA	CATTTTAA..	.....

Fig. 16A

10	20	30	40	50	60
MSKKNFITIG	ATLIHMLLPN	ISEPETINNN	TDKLSGLYIS	GQYKPGISHE	SKFSVKEIYN
70	80	90	100	110	120
DNIQLIGLRH	NAISTSTLNI	NTDFNIPYKV	TFQNNITSFS	GAIGYSDPTG	ARFELEGSYE
130	140	150	160	170	180
EFDVTDPGDC	LIKDTYRYEA	LARNPSGSSP	TSNNYTVMRN	DGVSITSVIF	NGCYDIELKD
190	200	210	220	230	240
LEVSPYVCVG	VGGDFIEFFD	ALHIKLAYQG	KLGINYHLST	QASVFIDGY	HKVIGNQFNN
250	260	270	280	290	300
LNVQHVASTD	FGPVYAVATL	NIGYFGGEIG	IRLTF.....	.....	.....

Fig. 16B

10	20	30	40	50	60
ATGAATAATA	GAAAAAGTTT	TTTTATAATA	GGTGCATCAT	TACTAGCAAG	CTTATTATTC
70	80	90	100	110	120
ACATCTGAGG	CCTCTTCTAC	AGGAAATGTA	AGTAACCATA	CTTATTTTAA	ACCTAGGTTA
130	140	150	160	170	180
TATATCAGTG	GACAATATAG	ACCAGGAGTT	TCTCATTTTA	GCAAATTTTC	AGTCAAAGAA
190	200	210	220	230	240
ACCAACTACA	ATACTACTCA	ACTAGTTGGG	CTTAAAAAGG	ACATCAGTGT	CATAGGGAAC
250	260	270	280	290	300
AGTAATATCA	CAACCTACAC	AAATTTCAAC	TTTCCTTACA	TTGCAGAATT	TCAAGACAAT
310	320	330	340	350	360
GCCATAAGTT	TCAGTGGGGC	AATTGGATAC	TTGTATTCCG	AGAATTTTAT	AATTGAAGTA
370	380	390	400	410	420
GAGGCTTCTT	ATGAAGAATT	TGATGTTAAA	AATCCAGAAG	GATCTGCTAC	AGACGCATAC
430	440	450	460	470	480
AGGTATTTTG	CACTAGCACG	TGCTATGGAT	GGCACTAATA	AATCTAGTCC	TGATGACACA
490	500	510	520	530	540
AGAAAATTCA	CTGTCATGAG	AAATGACGGG	TTATCAATTT	CATCAGTAAT	GATAAATGGG
550	560	570	580	590	600
TGTTACAATT	TTACATTAGA	TGATATACCA	GTAGTACCGT	ATGTATGCGC	AGGAATAGGA
610	620	630	640	650	660
GGAGATTTCA	TAGAGTTTTT	TAATGATTTA	CATGTTAAGT	TTGCTCATCA	AGGCAAGGTA
670	680	690	700	710	720
GGTATTAGTT	ATTCTATATC	CCCTGAAGTA	AGTTTATTTT	TTAACGGATA	TTACCATAAA
730	740	750	760	770	780
GTAACAGGTA	ACAGATTTAA	AAACTTACAC	GTTCAACACG	TAAGTGATTT	AAGTGACGCT
790	800	810	820	830	840
CCTAAGTTCA	CATCTGCAGT	TGCTACACTC	AATGTTGGGT	ACTTTGGTGG	CGAAATTGGA
850	860	870	880	890	900
GTAAGATTTA	TATTTTAA..	.....	.....	.....	.....

Fig. 17A

10	20	30	40	50	60
MNNRKSFFII	GASLLASLLF	TSEASSTGNV	SNHTYFKPRL	YISGQYRPGV	SHFSKFSVKE
70	80	90	100	110	120
TNYNTTQLVG	LKKDISVIGN	SNITTYTNFN	FPYLAEFQDN	AISFSGAIGY	LYSENFRIEV
130	140	150	160	170	180
EASYEEFDVK	NPEGSATDAY	RYFALARAMD	GTNKSSPDDT	RKFTVMRNDG	LSISSVMING
190	200	210	220	230	240
CYNFTLDDIP	VVPYVCAGIG	GDFIEFFNDL	HVKEFAHQGV	GISYSISPEV	SLFLNGYYHK
250	260	270	280	290	300
VTGNRFKNLH	VQHVS DLSDA	PKFSAVATL	NVGYFGGEIG	VRFIF.....	.....

Fig. 17B

10	20	30	40	50	60
TAGCAGCACT	AAAAAACAGT	TTGGGTTATA	TGTTAGTGGA	CAACACCAGC	CTAGTGTTTC
70	80	90	100	110	120
TATTTTTAGC	AATTTCTCAG	TAAAGGAAAC	TAATTTTCCT	ACAAAGTATT	CTAGCAGCTT
130	140	150	160	170	180
CTTAAAAAAA	GACATTAATT	CTGTCGAATT	TGACGATAGT	GTTACTGCTG	GCATTAGTTA
190	200	210	220	230	240
CCCACCTAAT	TTCAGTACTC	CTTATATAGC	TGTATTTCAA	GATAATATTT	CTAATTTTAA
250	260	270	280	290	300
TGGCGCTATT	GGGTACACTT	TTGTTGAAGG	CCCAAGAATT	GAAATAGAAG	GTTCTTATGA
310	320	330	340	350	360
AGAATTCGAT	GTCAAAGACC	CTGGAAGATA	TACAGAAATA	CAAGATGCAT	ACCGTTACTT
370	380	390	400	410	420
TGCTTTAGCA	CGTGATATAG	ACTCTATTCC	TACTAGCCCA	AAAAATAGAA	CTTCACATGA
430	440	450	460	470	480
TGGCAACAGT	TCATATAAGG	TATAACCACAC	TGTAATGAAA	AATGAAGGAC	TATCTATAAT
490	500	510	520	530	540
ATCCATTATG	GTCAATGGCT	GCTATGATTT	TTCTTCAGAT	AATTTATCAA	TATTACCTTA
550	560	570	580	590	600
TGTATGTGGT	GGTATAGGTG	TAAATGCTAT	AGAGTTTTTC	GATGCATTAC	ATGTTAAATT
610	620	630	640	650	660
CGCGTGTCAG	GGTAAATTAG	GTATTACTTA	TCCATTATCT	TCCAACGTTA	GTTTATTTGC
670	680	690	700	710	720
TGGTGGATAT	TATCACCAAG	TAATGGGCAA	CCAATTTAAA	AATCTAAATG	TTCAACATGT
730	740	750	760	770	780
AGCTGAACTT	AATGACGCAC	CCAAAGTTAC	ATCTGCAGTA	GCTACACTTG	ACATTGGGTA
790	800	810	820	830	840
TTTTGGTGGT	GAAATTGGAG	CAAGGCTTAT	ATTTTAA...	.....	.....

Fig. 18A

10	20	30	40	50	60
SSTKKQFGLY	VSGQHQPVS	IFSNFSVKET	NFPTKYSSSF	LKKDINSVEF	DDSVTAGISY
70	80	90	100	110	120
PLNFSTPYIA	VFQDNISNFN	GAIGYTFVEG	PRIEIEGSYE	EFDVKDPGRY	TEIQDAYRYF
130	140	150	160	170	180
ALARDIDSIP	TSPKNRTSHD	GNSSYKVYHT	VMKNEGLSII	SIMVNGCYDF	SSDNLSILPY
190	200	210	220	230	240
VCGGIGVNAI	EFFDALHVKE	ACQGKLGITY	PLSSNVSLEA	GGYYHQVMGN	QFKNLNVQHV
250	260	270	280	290	300
AELNDAPKVT	SAVATLDIGY	FGGEIGARLI	F.....	.....	.....

Fig. 18B

10	20	30	40	50	60
ATGAATTGCA	AAAGATTTTT	CATAGCAAGT	GCATTGATAT	CACTAATGTC	TTTCTTACCT
70	80	90	100	110	120
AGCGTATCTT	TTTCTGAATC	AATACATGAA	GATAATATAA	ATGGTAACTT	TTACATTAGT
130	140	150	160	170	180
GCAAAGTATA	TGCCAAGTGC	CTCACACTTT	GGCGTATTTT	CAGTTAAAGA	AGAGAAAAAC
190	200	210	220	230	240
ACAACAACCTG	GAGTTTTTCGG	ATTAAAACAA	GATTGGGACG	GAGCAACAAT	AAAGGATGCA
250	260	270	280	290	300
AGCAGCAGCC	ACACAATAGA	CCCAAGTACA	ATATTCTCCA	TTTCAAATTA	TTCATTTAAA
310	320	330	340	350	360
TATGAAAACA	ATCCATTTTT	AGGGTTTGCA	GGAGCTATTG	GCTACTCAAT	GGGTGGTCCA
370	380	390	400	410	420
AGGGTAGAGT	TTGAAGTGTC	TTACGAAATA	TTTGATGTAA	AAAACCAAGG	TAACAGTTAC
430	440	450	460	470	480
AAGAACGATG	CTCACAAATA	TTGCGCTTTA	TCAAGACACA	CCGGAGGTAT	GCCACAAGCC
490	500	510	520	530	540
GGTCATCAAA	ATAAATTTGT	CTTCCTAAAA	AATGAAGGAT	TACTTGACAT	ATCACTTATG
550	560	570	580	590	600
ATAAACGCAT	GTTATGATAT	AACAATCGAC	AGCATGCCAT	TTTCTCCATA	TATATGTGCA
610	620	630	640	650	660
GGTATTGGTA	GTGACTTAGT	TTCGATGTTT	GAAACTACAA	ATCCTAAAAT	TTCTTATCAA
670	680	690	700	710	720
GGAAAATTAG	GTGTAAGTTA	CTCCATAAGC	CCAGAAGCAT	CTGTTTTTGT	TGGAGGACAC
730	740	750	760	770	780
TTTCACAGAG	TTATAGGTAA	TGAATTTAAA	GACATTCCTG	CAATAACTCC	TGCTGGAGCA
790	800	810	820	830	840
ACAGAAATTA	AAGGCACACA	GTTTACAACA	GTAACATTAA	ACATATGCCA	CTTCGGACTA
850	860	870	880	890	900
GAGCTTGGAG	GCAGGTTTAC	TTTTTAA...	.....	.....	.....

Fig. 19A

10	20	30	40	50	60
MNCKRFFIAS	ALISLMSFLP	SVSFSESIHE	DNINGNFYIS	AKYMPSASHF	GVFSVKEEKN
70	80	90	100	110	120
TTTGVFGLKQ	DWDGATIKDA	SSSHTIDPST	IFSISNYSFK	YENNPFLGFA	GAIGYSMGGP
130	140	150	160	170	180
RVEFEVSYEI	FDVKNQNSY	KNDAHKYCAL	SRHTGGMPOA	GHQNKFEVLK	NEGLLDISLM
190	200	210	220	230	240
INACYDITID	SMPFSPYICA	GIGSDLVSMF	ETTNPKISYQ	GKLGVSYSIS	PEASVFEVGGH
250	260	270	280	290	300
FHRVIGNEFK	DIPAITPAGA	TEIKGTQFTT	VTLNICHFGL	ELGGRFTF..	.....

Fig. 19B

10	20	30	40	50	60
ATGAAATATA	AAAAAACTTT	TACAGTAACT	GCATTAGTAT	TATTAECTTC	CTTTACACAT
70	80	90	100	110	120
TTTATACCTT	TTTATAGTCC	AGCACGTGCC	AGTACAATTC	ACAACTTCTA	CATTAGTGGA
130	140	150	160	170	180
AAATATATGC	CAACAGCGTC	ACATTTTGGA	ATTTTTTCAG	CTAAAGAAGA	ACAAAGTTTT
190	200	210	220	230	240
ACTAAGGTAT	TAGTTGGGTT	AGATCAACGA	TTATCACATA	ATATTATAAA	CAATAATGAT
250	260	270	280	290	300
ACAGCAAAGA	GTCTTAAGGT	TCAAAATTAT	TCATTTAAAT	ACAAAAATAA	CCCATTTCTA
310	320	330	340	350	360
GGATTTGCAA	GAGCTATTGG	TTATTCAATA	GGCAATTCAA	GAATAGAACT	AGAAGTATCA
370	380	390	400	410	420
CATGAAATAT	TTGATACTAA	AAACCCAGGA	AACAATTATT	TAAATGACTC	TCACAAATAT
430	440	450	460	470	480
TGCGCTTTAT	CTCATGGAAG	TCACATATGC	AGTGATGGAA	ATAGCGGAGA	TTGGTACACT
490	500	510	520	530	540
GCAAAAACCTG	ATAAGTTTGT	ACTTCTGAAA	AATGAAGGTT	TACTTGACGT	CTCATTTATG
550	560	570	580	590	600
TTAAACGCAT	GTTATGACAT	AACAACCTGAA	AAAATGCCTT	TTTCACCTTA	TATATGTGCA
610	620	630	640	650	660
GGTATTGGTA	CTGATCTCAT	ATCTATGTTT	GAGACAACAC	AAAACAAAAT	ATCTTATCAA
670	680	690	700	710	720
GGAAAGTTAG	GTTTAAACTA	TACTATAAAC	TCAAGAGTTT	CTGTTTTTGC	AGGTGGGCAC
730	740	750	760	770	780
TTTCATAAAG	TAATAGGTAA	TGAATTTAAA	GGTATTCCTA	CTCTATTACC	TGATGGATCA
790	800	810	820	830	840
AACATTAAG	TACAACAGTC	TGCAACAGTA	ACATTAGATG	TGTGCCATTT	CGGGTTAGAG
850	860	870	880	890	900
ATTGGAAGTA	GATTTTTCTT	TTAA.....	.....	.....	.....

Fig. 20A

10	20	30	40	50	60
MKYKKTFTVT	ALVLLTSFTH	FIPEYSPARA	STIHNEYISG	KYMPTASHEG	IFSAKEEQSF
70	80	90	100	110	120
TKVLVGLDQR	LSHNIINNND	TAKSLRVQNY	SFKYKNNPFL	GFARAIGYSI	GNSRIELEVS
130	140	150	160	170	180
HEIFDTKNPG	NNYLNDSHKY	CALSHGSHIC	SDGNSGDWYT	AKTDKFEVLLK	NEGLLDVSEM
190	200	210	220	230	240
LNACYDITTE	KMPFSPYICA	GIGTDLISME	ETTQNKISYQ	GKLGLNYTIN	SRVSVFAGGH
250	260	270	280	290	300
FHKVIGNEFK	GIPTLLPDGS	NIKVQQSATV	TLDVCHGLE	IGSRFFF...	.....

Fig. 20B

10	20	30	40	50	60
ATGTTTTATA	CTAATATATA	TATTCTGGCT	TGTATTTACT	TTGCACTTCC	ACTATTGTTA
70	80	90	100	110	120
ATTTATTTTC	ACTATTTTAG	GTGTAATATG	AATTGCAAAA	AAATTCTTAT	AACAACGCA
130	140	150	160	170	180
TTAATATCAT	TAATGTACTC	TATTCCAAGC	ATATCTTTTT	CTGATACTAT	ACAAGATGGT
190	200	210	220	230	240
AACATGGGTG	GTAACCTCTA	TATTAGTGGA	AAGTATGTAC	CAAGTGTCTC	ACATTTTGGT
250	260	270	280	290	300
AGCTTCTCAG	CTAAAGAAGA	AAGCAAATCA	ACTGTTGGAG	TTTTTGGATT	AAAACATGAT
310	320	330	340	350	360
TGGGATGGAA	GTCCAATACT	TAAGAATAAA	CACGCTGACT	TTACTGTTCC	AACTATTCCG
370	380	390	400	410	420
TTCAGATACG	AGAACAATCC	ATTTCTAGGG	TTTGCAGGAG	CTATCGGTTA	CTCAATGGGT
430	440	450	460	470	480
GGCCCAAGAA	TAGAATTCGA	AATATCTTAT	GAAGCATTTC	ACGTAAAAAG	TCCTAATATC
490	500	510	520	530	540
AATTATCAAA	ATGACGCGCA	CAGGTACTGC	GCTCTATCTC	ATCACACATC	GGCAGCCATG
550	560	570	580	590	600
GAAGCTGATA	AATTTGTCTT	CTTAAAAAAC	GAAGGGTTAA	TTGACATATC	ACTTGCAATA
610	620	630	640	650	660
AATGCATGTT	ATGATATAAT	AAATGACAAA	GTACCTGTTT	CTCCTTATAT	ATGCGCAGGT
670	680	690	700	710	720
ATTGGTACTG	ATTTGATTTC	TATGTTTGAA	GCTACAAGTC	CTAAAATTTT	CTACCAAGGA
730	740	750	760	770	780
AAACTGGGCA	TTAGTTACTC	TATTAATCCG	GAAACCTCTG	TTTTCATCGG	TGGGCATTTT
790	800	810	820	830	840
CACAGGATCA	TAGGTAATGA	GTTTAGAGAT	ATTCCTGCAA	TAGTACCTAG	TAACTCAACT
850	860	870	880	890	900
ACAATAAGTG	GACCACAATT	TGCAACAGTA	ACACTAAATG	TGTGTCACTT	TGGTTTAGAA
910	920	930	940	950	960
CTTGAGGAA	GATTTAACTT	CTAA.....	.....	.....	.....

Fig. 21A

10	20	30	40	50	60
MEYTNIIYILA	CIYFALPELLL	IYFHYFRONM	NCKKILITTA	LISLMYSIPS	ISFSDTIQDG
70	80	90	100	110	120
NMGGNFYISG	KYVPSVSHFG	SFSAKEESKS	TVGVEGLKHD	WDGSPILKNK	HADFTVPNYS
130	140	150	160	170	180
FRYENNPFLG	FAGAIGYSMG	GPRIEFEISY	EAFDVKSPNI	NYQNDAHRYC	ALSHHTSAAM
190	200	210	220	230	240
EADKFVFLKN	EGLIDISLAI	NACYDIINDK	VPVSPYICAG	IGTDLISMFE	ATSPKISYQG
250	260	270	280	290	300
KLGISYSINP	ETSVFIGGHE	HRIIGNEFRD	IPAIVPSNST	TISGPQFATV	TLNVCHFGLF
310	320	330	340	350	360
LGGRENF...	.....	.....	.....	.....	.....

Fig. 21B

10	20	30	40	50	60
ATGAATTGCA	AAAAAATTCT	TATAACAAC	GCATTAATGT	CATTAATGTA	CTATGCTCCA
70	80	90	100	110	120
AGCATATCTT	TTTCTGATAC	TATAACAAGAC	GATAACACTG	GTAGCTTCTA	CATCAGTGG
130	140	150	160	170	180
AAATATGTAC	CAAGTGTTTC	ACATTTTGGT	GTTTTCTCAG	CTAAAGAAGA	AAGAAACTCA
190	200	210	220	230	240
ACTGTTGGAG	TTTTTGGATT	AAAACATGAT	TGGAATGGAG	GTACAATATC	TAACCTCTCT
250	260	270	280	290	300
CCAGAAAATA	TATTCACAGT	TCAAATTAT	TCGTTTAAAT	ACGAAAACAA	CCCATCTCTTA
310	320	330	340	350	360
GGGTTTGCAG	GAGCTATTGG	TTATTCAATG	GGTGGCCCAA	GAATAGAACT	TGAAGTTCTG
370	380	390	400	410	420
TACGAGACAT	TCGATGTGAA	AAATCAGAAC	AATAATTATA	AGAACGGCGC	ACACAGATAC
430	440	450	460	470	480
TGTGCTTTAT	CTCATCATAG	TTCAGCAACA	AACATGTCCT	CCGCAAGTAA	CAAATTTGTT
490	500	510	520	530	540
TTCTTAAAAA	ATGAAGGGTT	AATTGACTTA	TCATTTATGA	TAAATGCATG	CTATGACATA
550	560	570	580	590	600
ATAATTGAAG	GAATGCCTTT	TTCACCTTAT	ATTTGTGCAG	GTGTTGGTAC	TGATGTTGTT
610	620	630	640	650	660
TCCATGTTTG	AAGCTATAAA	TCCTAAAATT	TCTTACCAAG	GAAACTAGG	ATTAGGTTAT
670	680	690	700	710	720
AGTATAAGTT	CAGAAGCCTC	TGTTTTTATC	GGTGGACACT	TTCACAGAGT	CATAGGTAAT
730	740	750	760	770	780
GAATTTAGAG	ACATCCCTGC	TATGGTTCCT	AGTGGATCAA	ATCTTCCAGA	AAACCAATTT
790	800	810	820	830	840
GCAATAGTAA	CACTAAATGT	GTGTCACTTT	GGTTTAGAAC	TTGGAGGAAG	ATTTAACTTC
850	860	870	880	890	900
TGA.....	.....	.....	.....	.....	.....

Fig. 22A

10	20	30	40	50	60
MNCKKILITT	ALMSLMYYAP	SISFSDTIQD	DNTGSFYISG	KYVPSVSHFG	VFSAKEERNS
70	80	90	100	110	120
TVGVFGLKHD	WNGGTISNSS	PENIFTVQNY	SFKYENNPFL	GFAGAIGYSM	GGPRIELEVEL
130	140	150	160	170	180
YETFDVKNQN	NNYKNGAHRY	CALSHHSSAT	NMSSASNKFV	FLKNEGLIDL	SFMINACYDI
190	200	210	220	230	240
IIEGMPFSPY	ICAGVGTDVV	SMFEAINPKI	SYQGKLGGLY	SISSEASVFI	GGHFHRVIGN
250	260	270	280	290	300
EERDIPAMVP	SGSNLPENQF	AIVTLNVCHF	GLELGGRENF	.....	.....

Fig. 22B

10	20	30	40	50	60
ATGAATTGTA	AAAAAGTTTT	CACAATAAGT	GCATTGATAT	CATCCATATA	CTTCCTACCT
70	80	90	100	110	120
AATGTCTCAT	ACTCTAACCC	AGTATATGGT	AACAGTATGT	ATGGTAATTT	TTACATATCA
130	140	150	160	170	180
GGAAAGTACA	TGCCAAGTGT	TCCTCATTTT	GGAATTTTTT	CAGCTGAAGA	AGAGAAAAAA
190	200	210	220	230	240
AAGACAACCTG	TAGTATATGG	CTTAAAAGGA	AAACTGGCAG	GAGATGCAAT	ATCTAGTCAA
250	260	270	280	290	300
AGTCCAGATG	ATAATTTTAC	CATTTCGAAAT	TACTCATTCA	AGTATGCAAG	CAACAAGTTT
310	320	330	340	350	360
TTAGGGTTTG	CAGTAGCTAT	TGGTTACTCG	ATAGGCAGTC	CAAGAATAGA	AGTTGAGATG
370	380	390	400	410	420
TCTTATGAAG	CATTTGATGT	GAAAAATCCA	GGTGATAATT	ACAAAAACGG	TGCTTACAGG
430	440	450	460	470	480
TATTGTGCTT	TATCTCATCA	AGATGATGCG	GATGATGACA	TGACTAGTGC	AACTGACAAA
490	500	510	520	530	540
TTTGTATATT	TAATTAATGA	AGGATTACTT	AACATATCAT	TTATGACAAA	CATATGTTAT
550	560	570	580	590	600
GAAACAGCAA	GCAAAAATAT	ACCTCTCTCT	CCTTACATAT	GTGCAGGTAT	TGGTACTGAT
610	620	630	640	650	660
TTAATTCACA	TGTTTGAAAC	TACACATCCT	AAAATTTCTT	ATCAAGGAAA	GCTAGGGTTG
670	680	690	700	710	720
GCCTACTTCG	TAAGTGCAGA	GTCTTCGGTT	TCTTTTGGTA	TATATTTTCA	TAAAATTATA
730	740	750	760	770	780
AATAATAAGT	TTAAAAATGT	TCCAGCCATG	GTACCTATTA	ACTCAGACGA	GATAGTAGGA
790	800	810	820	830	840
CCACAGTTTG	CAACAGTAAC	ATTAAATGTA	TGCTACTTTG	GATTAGAACT	TGGATGTAGG
850	860	870	880	890	900
TTCAACTTCT	AA.....	.....	.....	.....	.....

Fig. 23A

10	20	30	40	50	60
MNCKKVFTIS	ALISSIYFLP	NVSYSNPVYG	NSMYGNFYIS	GKYMPSVPHF	GIFSAEEEEK
70	80	90	100	110	120
KTTVVYGLKG	KLAGDAISSQ	SPDDNFTIRN	YSEKYASNKE	LGFVAIGYS	IGSPRIEVEM
130	140	150	160	170	180
SYEAFDVKNP	GDNYKNGAYR	YCALSHQDDA	DDDMTSATDK	FVYLINEGLL	NISEMTNICY
190	200	210	220	230	240
ETASKNIPLS	PYICAGIGTD	LIHMFETTHP	KISYQGLGL	AYFVSAESSV	SFGIYFHKII
250	260	270	280	290	300
NNKFKNVPAM	VPINSDEIVG	PQFATVTLNV	CYFGLELGR	FNF.....	.....

Fig. 23B

10	20	30	40	50	60
ATGAACTGTA	AAAAATTTCT	TATAACAAC	ACATTGGTAT	CACTAACAAT	TCTTTTACCT
70	80	90	100	110	120
GGCATATCTT	TCTCCAAACC	AATACATGAA	AACAATACTA	CAGGAAACTT	TTACATTATT
130	140	150	160	170	180
GGAAAATATG	TACCAAGTAT	TTCACATTTT	GGGAACTTTT	CAGCTAAAGA	AGAAAAAAC
190	200	210	220	230	240
ACAACTACTG	GAATTTTTGG	ATTAAAAGAA	TCATGGACTG	GTGGTATCAT	CCTTGATAAA
250	260	270	280	290	300
GAACATGCAG	CTTTTAATAT	CCCAAATTAT	TCATTTAAAT	ATGAAAATAA	TCCATTTTTA
310	320	330	340	350	360
GGATTTGCAG	GGGTAATTGG	CTATTCAATA	GGTAGTCCAA	GAATAGAATT	TGAAGTATCA
370	380	390	400	410	420
TACGAGACAT	TCGATGTACA	AAATCCAGGA	GATAAGTTTA	ACAATGATGC	ACATAAGTAT
430	440	450	460	470	480
TGTGCTTTAT	CCAATGATTC	CAGTAAAACA	ATGAAAAGTG	GTAAATTCGT	TTTTCTCAA
490	500	510	520	530	540
AATGAAGGAT	TAAGTGACAT	ATCACTCATG	TTAAATGTAT	GTTATGATAT	AATAAACAAA
550	560	570	580	590	600
AGAATGCCTT	TTTCACCTTA	CATATGTGCA	GGCATTGGTA	CTGACTTAAT	ATTCATGTTT
610	620	630	640	650	660
GACGCTATAA	ACCATAAAGC	TGCTTATCAA	GGAAAATTAG	GTTTTAATTA	TCCAATAAGC
670	680	690	700	710	720
CCAGAAGCTA	ACATTTCTAT	GGGTGTGCAC	TTTCACAAAG	TAACAAACAA	CGAGTTTAGA
730	740	750	760	770	780
GTTCCIGTTC	TATTAAGTGC	TGGAGGACTC	GCTCCAGATA	ATCTATTTGC	AATAGTAAAG
790	800	810	820	830	840
TTGAGTATAT	GTCATTTTGG	GTTAGAATTT	GGGTACAGGG	TCAGTTTTTA	A.....

Fig. 24A

10	20	30	40	50	60
MNCKKFLITT	TLVSLTILLP	GISFSKPIHE	NNTTGNFYII	GKYVPSISHF	GNFSAKEEKN
70	80	90	100	110	120
TTTGIFGLKE	SWGGIILDK	EHAAFNIPNY	SFKYENNPFL	GFAGVIGYSI	GSPRIEFEVS
130	140	150	160	170	180
YETFVQNP	DKFNNDHAKY	CALSNDSSKT	MKSGKFVFLK	NEGLSDISLM	LNVCYDIINK
190	200	210	220	230	240
RMPFSPYICA	GIGTDLIFMF	DAINHKAAYQ	GKLGFNYPIS	PEANISMGVH	FHKVTNNEFR
250	260	270	280	290	300
VPVLLTAGGL	APDNLFAIVK	LSICHEGLEF	GYRVSF.....	.....	.....

Fig. 24B

10	20	30	40	50	60
ATGAATAATA	AACTCAAATT	TACTATAATA	AACACAGTAT	TAGTATGCTT	ATTGTCATTA
70	80	90	100	110	120
CCTAATATAT	CTTCCTCAAA	GGCCATAAAC	AATAACGCTA	AAAAGTACTA	CGGATTATAT
130	140	150	160	170	180
ATCAGTGGAC	AATATAAACC	CAGTGTCTT	GTTTTTCAGTA	ATTTTTTCAGT	TAAAGAAACC
190	200	210	220	230	240
AATGTCATAA	CTAAAAACCT	TATAGCTTTA	AAAAAAGATG	TTGACTCTAT	TGAAACCAAG
250	260	270	280	290	300
ACTGATGCCA	GTGTAGGTAT	TAGTAACCCA	TCAAATTTTA	CTATCCCCTA	TACAGCTGTA
310	320	330	340	350	360
TTTCAAGATA	ATTCTGTCAA	TTTCAATGGA	ACTATTGGTT	ACACCTTTGC	TGAAGGTACA
370	380	390	400	410	420
AGAGTTGAAA	TAGAAGGTTT	TTATGAGGAA	TTTGATGTTA	AAAACCCTGG	AGGCTATACA
430	440	450	460	470	480
CTAAGTGATG	CCTATCGCTA	TTTTGCATTA	GCACGTGAAA	TGAAAGGTAA	TAGTTTTTACA
490	500	510	520	530	540
CCTAAAGAAA	AAGTTTCTAA	TAGTATTTTT	CACACTGTAA	TGAGAAATGA	TGGATTATCT
550	560	570	580	590	600
ATAATATCTG	TTATAGTAAA	TGTTTGCTAC	GATTTCTCTT	TGAACAATTT	GTCAATATCG
610	620	630	640	650	660
CCTTACATAT	GTGGAGGAGC	AGGGGTAGAT	GCTATAGAAT	TCTTCGATGT	ATTACACATT
670	680	690	700	710	720
AAGTTTGCAT	ATCAAAGCAA	GCTAGGTATT	GCTTATTCTC	TACCATCTAA	CATTAGTCTC
730	740	750	760	770	780
TTTGCTAGTT	TATATTACCA	TAAAGTAATG	GGCAATCAAT	TTAAAAATTT	AAATGTCCAA
790	800	810	820	830	840
CATGTTGCTG	AACTTGCAAG	TATACCTAAA	ATTACATCCG	CAGTTGCTAC	ACTTAATATT
850	860	870	880	890	900
GGTTATTTTG	GAGGTGAAAT	TGGTGCAAGA	TTGACATTTT	AA.....	.....

Fig. 25A

10	20	30	40	50	60
MNNKLFETII	NTVLVCLLSL	PNISSSKAIN	NNAKYYGLY	ISGQYKPSVS	VFSNFSVKET
70	80	90	100	110	120
NVITKNLIAL	KRDVDSIETK	TDASVGISNP	SNFTIPYTAV	FQDNSVNENG	TIGYTFAEGT
130	140	150	160	170	180
RVEIEGSYEE	FDVKNPGGYT	LSDAYRYFAL	AREMKGNSFT	PKEKVSNSIF	HTVMRNDGLS
190	200	210	220	230	240
IISVIVNVCY	DFSLNLSIS	PYICGGAGVD	AIEFFDVLHI	KFAYQSKLGI	AYSLPSNISL
250	260	270	280	290	300
FASLYYHKVM	GNQFKNLNVQ	HVAELASIPK	ITSAVATLNI	GYFGGEIGAR	LTF.....

Fig. 25B

10	20	30	40	50	60
ATGGCAAATT	TTATGTACAA	AAAATACAAA	CTAATGACAG	CAGGTGTAGT	ATTATTTTCCAC
70	80	90	100	110	120
ATGTTATTTT	TACCTCATGT	TTCTTTTCGCA	AAAAATACAA	ACAGCAATAA	ACTTGGATTA
130	140	150	160	170	180
TACATCAGTG	GACAGTATAA	CCCTAGTGTT	TCTGTTTTTA	GCAATTTTTC	AGCAAAGAA
190	200	210	220	230	240
ACCAATGTTT	ATACAGTACA	ACTCATGGCG	CTTAAAAAAG	ACATTGATTC	TATTGAAGTT
250	260	270	280	290	300
GATACTGGAA	ATAGCGCAGG	TATTAGCAAA	CCACAAAATT	TCACAGTTCT	TTATACTCCA
310	320	330	340	350	360
AAATTTCAAG	ATAATGTTGC	TGGTCTTAGC	GGTGCACCTG	GATTCTTTTA	TTCTAAAGGA
370	380	390	400	410	420
TTAAGGATTG	AAATGGGGTT	TTCTTATGAA	AAATTTGATG	CTAAAGACCT	TGGTGAGTAC
430	440	450	460	470	480
ACCAAATAA	AAGATGCTTA	TAGATATTTT	GCTCTAGTAC	GTGAAATGCA	TGTTAGTCTC
490	500	510	520	530	540
ATTTATCCAA	AAGATAATAA	CACAGGAACA	CATTATACTG	TTATGAGAAA	TGATGGTATA
550	560	570	580	590	600
TCTATTTCTT	CTGCTACAGT	AAATGGCTGC	TATGATTCTT	TTTTCCAGTT	TATCTTTGTC
610	620	630	640	650	660
ACCTATATGT	GTATAGGCAT	CGGTATAGAT	GCTATAGAAT	TTCTTAATGC	ATACATATTA
670	680	690	700	710	720
AGTTTGCTTG	CCAAGGTAGT	TAAGGTGTTA	ACTTATTCTG	TATCTCCCAA	TGTTAATTTA
730	740	750	760	770	780
TTTGCAGATG	GATATTATCA	TAAAGTGATG	GGCAATAAAT	TTAAAAATTT	ACCTGTTCAA
790	800	810	820	830	840
TACGTTAATA	CTTTAGAAGA	GTATCCAAGA	GTTACATCTG	CAATTGCTAC	ACTTGATATT
850	860	870	880	890	900
GGCTACCTCG	GTGGTGAAAT	TGGCATAAGA	TTTATATTTT	AA.....	.....

Fig. 26A

10	20	30	40	50	60
MYKKYKLMTA	GVVLFHMLFL	PHVSEAKNTN	SNKLGGLYISG	QYNPSVSVFS	NFSAKETNVH
70	80	90	100	110	120
TVQLMALKKD	IDSIEVDTGN	SAGISKPQNF	TVLYTPKFQD	NVAGLSGALG	FFYSKGLRIE
130	140	150	160	170	180
MGFSYEKFDA	KDLGEYTKIK	DAYRYFALVR	EMHVS LIYPK	DNNTGTHYTV	MRNDGISISS
190	200	210	220	230	240
ATVNGCYDSF	FQFIFVTYMC	IGIGIDAIEF	LNAYILSLLA	KVVKVLTYSV	SPNVNLEADG
250	260	270	280	290	300
YYHKVMGNKF	KNLPVQYVNT	LEEYPRV TSA	IATLDIGYLG	GEIGIRFIF.	.....

Fig. 26B

10	20	30	40	50	60
ATGGGAAATT	CTATGAATAA	TAAAAGTCAA	TTCTTAATAA	GATTTATATT	TTTAACATGC
70	80	90	100	110	120
ATGCTGTCAT	TACCTAATAT	ATCTCTTTCA	AAAGTAAATA	ACGAAAAACA	TTCTGGTTTG
130	140	150	160	170	180
TATATTAGCG	GGCAATACAA	ACCCAGTGTT	TCTGTTTTCA	GTAATTTTTC	AGTTAAAGAA
190	200	210	220	230	240
ACCAACTTTC	ATACAAAACA	TCTCATAGCT	CTTAAACAAG	ATGTTGATTC	TGTTGAAATT
250	260	270	280	290	300
GATACTGGTA	GTAATACAGC	AGGTATTAGT	AACCCATCTA	ACTTTACAAT	CCCTTATACT
310	320	330	340	350	360
GCAGAATTC	AAGACAACCA	TACTAACTGC	AATGGCTCTA	TTGGTTATGC	TTTTGCTGAA
370	380	390	400	410	420
GGTCCAAGAA	TTGAAATAGA	ATTATCATAT	GAAAAATTTG	ATGTTAAAAA	TCCCACAGGG
430	440	450	460	470	480
TATACTACAG	TAAAAGATGC	TTATAGATAC	TTTGCTTTAG	CACGTGAAAT	AAATATTTCT
490	500	510	520	530	540
CTATTCCAAC	CAAAACAAA	AGAAGGTAGT	GGAATTTACC	ATGTCGTAAT	GAAAAACGAT
550	560	570	580	590	600
GGGTTATCTA	TCTTATCCAA	TATAGTTAAT	ATTTGCTACG	ATTTTCTTT	AAATAATTTA
610	620	630	640	650	660
CCTATATCAC	CTTATTTATG	CGGAGGAATG	GGTATAAATG	CCATAGAATT	CTTTGACGCT
670	680	690	700	710	720
TTACATGTGA	AATTTGCTTA	TCAAAGCAAG	GCAGGAATTA	GTTATCAACT	ATTACGTAAA
730	740	750	760	770	780
ATCAACTTAT	TTATTGATGT	ATATTACTAC	GAAGTAATAA	GTAATAAATT	TAAAAACCTG
790	800	810	820	830	840
AAAGTCCAAC	ATGTACATGA	ACTTAAAGAT	AATCCAAAAG	TCACATCTGC	AGTTGCTACA
850	860	870	880	890	900
CTTGATATAG	CATATTTTGG	TAGTGAAGCT	GGCATAAGAA	TTATATTTTA	A.....

Fig. 27A

10	20	30	40	50	60
MNNKSQFLIR	FIFLTCMLSL	PNISLSKVNN	EKHSGLYISG	QYKPSVSVES	NFSVKETNEH
70	80	90	100	110	120
TKHLIALKQD	VDSVEIDTGS	NTAGISNPSN	FTIPYTAEFQ	DNHTNCNGSI	GYAFAEGPRI
130	140	150	160	170	180
EIELSYEKFD	VKNPTGYTTV	KDAYRYFALA	REINISLFQP	KQKEGSGIYH	VVMKNDGLSI
190	200	210	220	230	240
LSNIVNICYD	FSLNNLPISP	YLCGGMGINA	IEFFDALHVK	FAYQSKAGIS	YQLLRKINLF
250	260	270	280	290	300
IDVYYYEVIS	NKFKNLKVQH	VHELKDNPKV	TSAVATLDIA	YFGSEAGIRI	IF.....

Fig. 27B

10	20	30	40	50	60
ATGAATAGCA	AGAGTAAGTT	CTTTACAATA	TGTACATCGT	TAATATGCTT	ATTATCATCA
70	80	90	100	110	120
CCTAACACAT	CTCTCTCAAA	CTTCATAGGC	AATAGTACAA	AACATTCTGG	ATTATATGTT
130	140	150	160	170	180
AGCGGACAAT	ATAAGCCCAG	CGTTTCCATT	TTTAGCAAAT	TTTCAGTAAA	AGAAACAAAT
190	200	210	220	230	240
ACACATACAG	TACAGTTAGT	AGCTCTTAAA	AAAGATGTTA	ATTCTATTTT	TATGAACATC
250	260	270	280	290	300
AGTAATGGTG	CTACAGGCAT	TAGCAAAGCA	ACAAATTTTA	ATCTTCCTTA	TGTTGCAGAA
310	320	330	340	350	360
TTTCAAGACA	ATGCCTTCAA	CTTCAGTGGA	GCTATTGGTT	ATTCACTTTT	TGAACAATA
370	380	390	400	410	420
AACATTGAAG	TTGAAGGTTT	TTATGAAGAA	TTCGATGCCA	AAAATCCTGG	TGGTTATATT
430	440	450	460	470	480
TTAAATGATG	CATTCCGCTA	TTTTGCATTG	GCACGTGAAA	TGGGACAAGA	AAAAAATGAT
490	500	510	520	530	540
AATAAGCATC	TTAGTCCTAA	GGAGGAGCAT	GATATAAGTA	AAACATATTA	CACAGTCATG
550	560	570	580	590	600
AGAAATAATG	GGTTATCTAT	ATTATCTATT	ATGATAAATG	GCTGCTATAA	TCTACCTCTC
610	620	630	640	650	660
AATGATTTAT	CAATATCACC	TTATTTTTGT	ACAGGAATAG	GTGTAGATGC	TATAGAATTT
670	680	690	700	710	720
TTTGATGCAC	TGCATCTTAA	ACTTGCTTTG	CAAAGTAAAA	TAGGAGCTAC	TTACCAATTA
730	740	750	760	770	780
TCAGACAACA	TTAGTTTATT	TACAAATGGA	TATTACCATC	AAGTAATAGG	TGATCAATTT
790	800	810	820	830	840
AAAAACTTAA	AAGTCCAATA	TATAGGTGAA	CTTAAAGAGA	ACCCGAAAAT	TACATCTGCA
850	860	870	880	890	900
GTTGCTACTC	TCAATGTTGG	ATACTTTGGA	GGTGAAATTG	GAGTAAGACT	CACACTTTAA
910	920	930	940	950	960
.....	.....	.....	.....	.....	.....

Fig. 28A

10	20	30	40	50	60
MNSKSKFFTI	CTSLICLLSS	PNTSLSNFIG	NSTRHSGLYV	SGQYKPSVSI	FSKFSVKETN
70	80	90	100	110	120
THTVQLVALK	KDVNSISMNI	SNGATGISKA	TNFNLPYVAE	FQDNAFNFSG	AIGYSLFEQL
130	140	150	160	170	180
NIEVEGSYEE	FDAKNPGGYI	LNDAFRYFAL	AREMGQEKND	NKHLSPKEEH	DISKTYITVM
190	200	210	220	230	240
RNNGLSILSI	MINGCYNLPL	NDLSISPYFC	TGIGVDAIEF	FDALHLKLAL	QSKIGATYQL
250	260	270	280	290	300
SDNISLFTNG	YYHQVIGDQF	KNLKVQYIGE	LKENPKITSA	VATLNVGYFG	GEIGVRLTL.

Fig. 28B

10	20	30	40	50	60
AAGCTTCTTA	TGAAGAATTT	GACGTTAAAA	ATCCTGAAGG	ATCTACTACA	GACTCCTATA
70	80	90	100	110	120
GATATTTTCGC	GTTAGCACGT	GGCATGGATG	GTAATAATAT	TCCTACAAGT	CAAAAATTTA
130	140	150	160	170	180
CTGTAATGAG	AAACGACGGG	TTATTAATCT	CATCTGTTAT	GATAAATGGC	TGTTACAATG
190	200	210	220	230	240
TCATACTAAA	TGATATACAA	GCAGAACCTT	ACATATGTGC	AGGACTAGGA	GGAGATTTTA
250	260	270	280	290	300
TAGAATTCTT	CAATGGCTTT	CATGTTAAGC	TAGCTTATCA	AGGTAAAGTA	GGCATTAGTT
310	320	330	340	350	360
ATCAAATATT	CCCTGAAGTA	AGATTATTTA	TTGATGGATA	CTACCATAAA	GTAAAAGGCA
370	380	390	400	410	420
ACAAGTTTAA	AAATTTACAC	GTTCAACATG	TAGGTGCACT	TGCAGCACTC	CCTAAAGTTA
430	440	450	460	470	480
CATCTGCAGT	TGCAACACTT	AATATTGGAT	ACTTTGGTTG	TGAAGCTGGA	GTAAGATTCA
490	500	510	520	530	540
TATTTTAA..	.....	.....	.....	.....	.....

Fig. 29A

10	20	30	40	50	60
ASYEEFDVKN	PEGSTTDSYR	YFALARGMDG	NNIPTSQKFT	VMRNDGLLIS	SVMINGCYNV
70	80	90	100	110	120
ILNDIQAEPY	ICAGLGGDFI	EFFNGEHVKL	AYQGKVGISY	QIFPEVRLFI	DGYHVKVGN
130	140	150	160	170	180
KEKNLHVQHV	GALAALPKVT	SAVATLNIGY	FGCEAGVRFI	F.....	.....

Fig. 29B

10	20	30	40	50	60
ATGAATTATA	AGAAAATTCT	AGTAAGAAGC	GCGTTAATCT	CATTAATGTC	AATCTTACCA
70	80	90	100	110	120
TATCAGTCTT	TTGCAGATCC	TGTAGGTTCA	AGAACTAATG	ATAACAAAGA	AGGCTTCTAC
130	140	150	160	170	180
ATTAGTGCAA	AGTACAATCC	AAGTATATCA	CACTTTAGAA	AATTCTCTGC	TGAAGAAACT
190	200	210	220	230	240
CCTATTAATG	GAACAAATTC	TCTCACTAAA	AAAGTTTTTCG	GACTAAAGAA	AGATGGTGAT
250	260	270	280	290	300
ATAACAAAAA	AAGACGATTT	TACAAGAGTA	GCTCCAGGCA	TTGATTTTCA	AAATAACTTA
310	320	330	340	350	360
ATATCAGGAT	TTTCAGGAAG	TATTGGTTAC	TCTATGGACG	GACCAAGAAT	AGAACTTGAA
370	380	390	400	410	420
GCTGCATATC	ACAATTTAAT	CCAAAAACAC	GATAACAATG	ATACTGATAA	TGGTGAATAC
430	440	450	460	470	480
TATAAACATT	TTGCATATCT	CGTAAAGATG	CCATGGAAGA	TCAGCCATAT	GTTGTTCTTA
490	500	510	520	530	540
AAAATGACGG	CATAC.....	.....	.....	.....	.....

Fig. 30A

10	20	30	40	50	60
MNYKKILVRS	ALISLMSILP	YQSFADPVGS	RTNDNKEGEFY	ISAKYNPSIS	HERKFSAEET
70	80	90	100	110	120
PINGTNSLTK	KVFGLKKGDD	ITKKDDEFTRV	APGIDFQNNL	ISGFSGSIGY	SMDGPRIELE
130	140	150	160	170	180
AAYHNLIQKH	DNNDTDNGEY	YKHFAYLVKM	PWKISHMLFL	KMTAY.....	.....

Fig. 30B

	SV	HV1	
OHP-1F	MNCKKFFITL TLVELMSFLP GISFSDAVON DMVG-GH--- -FYISGKYVF SVSHFGVPSA KD-----ERM TITGVFGLKQ DWGOSTISKR SPENTFHYEN		90
OHP-1E	.....A.....P..G ..IS-----V...M. .A...M... E-----K. P.VALY... ..E.-IS.SS FGD.M..NKG		89
OHP-1D	...E.....A.TL.....L..P..D ..IS-----N. .A..... E-----V...IE. ...RCV..RT TLESDI.T...		90
OHP-1C	.....A.ALP.....LL.EP..D .S.S-----M. .A..... E-----K. P.VALY... ..N.-VSASE HADAD..NKG		89
OHP-1B	..Y..I.VSS A.I...I.. YQ..A.P.TS NDT.INDSRE G...V..N. .I...RX... ZEAPINGHTS I.KK...K ..GDI AQSAM..RTD		94
P28	.....FA-G SGIN-----H. .A..... E-----V...N...A..NS ..NDV.T.S.		64
HAP-1	.....I...S T.I..V... .V...VI.E E.MPV.S--- -V...A..K. TA...RH.I E-----DSR D.KA.....K ...VKTFSQ NTNSI.TEKD		91
OHP-1A	-----		
	HV2		
OHP-1F	YSPKYENPF LGFAGAVCYL MNGPRIELM SYETFDVYQO GMYQNDAM- -KYALTH-- NSGGKLSNAG DAPVFLKNEG LLDISIMLNA CYDVISGIP		186
OHP-1E	.....I...S .G...V.F.V .....-R.C..OQ-- -ODNSGIFKT S.Y.L..S... ..F... ..I.N.S..		184
OHP-1D	.....L. S.....I..S .D.....V .....A.....E...-R...S.LL GTETQIDG.. SAG...I... ..K.F... ..V...V...		188
OHP-1C	.....I...S .G...F.V .....G.....-R.C..DR-- -KASSTNATA ERY.L... ..V...V...		184
OHP-1B	PALEFO..LI S..S.SI..A .D.....A AYQK..A..P D..DT.SGDY Y..FG.SR-- ----EDAI.D K.Y.V... ..ITFM...V.T ...ITA..V.		188
P28	.....I..S .D.....V .....E...-R.C..SH-- ..ADM.S.S NN... ..F... ..VO... ..VO...		160
HAP-1	.....S .....F.V .....R.P .G.....-K.C.---L DTASSSTAGA TTS.MV...N .T... ..IMLD.M.		185
OHP-1A	-----		
	HV3		
OHP-1F	FSPYICAGVG TDLISMTEAI NPKISYQGLL GLSYSISPEA SVYVGGHFMK VIGNEFRDIP AMIPSTILT GN-NF---T IVTLSVCHFG VELGGRFNY		280
OHP-1E	L.....T .....N... ..I..... TLKAFVTSS-- -ATPDL---A .....I.....		278
OHP-1D	.....I. I..V...f.....P.....I..... T...E.A.A .KGNYP---A ...D.FY.. I.....QL		286
OHP-1C	.....N.....S TLKAFATPSS AATPDL---A T.....		280
OHP-1B	..I..A... A...NV.KDF .L.F...I .I..P.T.V .A.I..YY.G ...M.WK.. VIT.VVLEGA POTTSE---A L..IDTGY.. G.V.V..T.		283
P28	.....I. ...V.H...T .....I..... TI..TG...A .KGNYP---A .I.D... I...V.		256
HAP-1	V...V...I. ...V.VIN.T ...L...I...N... .I.I...R .....K..A TSKVF..SCN ASSAVSPOFA SAI.D... I.I...V.		284
OHP-1A	-----EF.D.L HV.FA.... .I..QLTKV NL.LD.YY.Q ...Q.KNLN VHWVY.LKES FKVTS---A VA..DIAY.. G.V.I..T.		81

Fig. 31