



(72) KUNSCH, Charles A., US

(72) CHOI, Gil H., US

(72) JOHNSON, L. Sydnor, US

(72) HROMOCKYJ, Alex, US

(71) HUMAN GENOME SCIENCES, INC., US

(51) Int.Cl.<sup>6</sup> C12N 15/31, C12Q 1/68, G01N 33/68, G01N 33/569,  
C07K 14/315, C12N 5/18, A61K 39/09

(30) 1996/10/31 (60/029,960) US

(54) **ANTIGENES ET VACCINS ACTIFS CONTRE  
STREPTOCOCCUS PNEUMONIAE**

(54) **STREPTOCOCCUS PNEUMONIAE ANTIGENS AND VACCINES**

(57) La présente invention concerne des préparations vaccinales destinées à la prévention ou la lutte contre l'infection par *Streptococcus pneumoniae*. L'invention concerne également des molécules d'acide nucléiques qui ont été isolées et qui codent des polypeptides antigènes de *Streptococcus pneumoniae*. L'invention concerne aussi, d'une part des polypeptides antigènes tels que des vecteurs et des cellules hôtes, et d'autre part des procédés de production correspondants. L'invention concerne enfin des techniques de diagnostic permettant de détecter des acides nucléiques, des polypeptides et des anticorps de *Streptococcus* dans un échantillon biologique.

(57) The present invention relates to novel vaccines for the prevention or attenuation of infection by *Streptococcus pneumoniae*. The invention further relates to isolated nucleic acid molecules encoding antigenic polypeptides of *Streptococcus pneumoniae*. Antigenic polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. The invention additionally relates to diagnostic methods for detecting *Streptococcus* nucleic acids, polypeptides and antibodies in a biological sample.

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

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

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>(51) International Patent Classification <sup>6</sup> :</b><br><b>C12N 15/31, 5/18, 1/21, C07K 14/315,</b><br><b>C12Q 1/68, A61K 39/09, G01N 33/569,</b><br><b>33/68</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | <b>A3</b> | <b>(11) International Publication Number:</b> <b>WO 98/18930</b><br><b>(43) International Publication Date:</b> 7 May 1998 (07.05.98)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <b>(21) International Application Number:</b> PCT/US97/19422<br><b>(22) International Filing Date:</b> 30 October 1997 (30.10.97)<br><b>(30) Priority Data:</b><br>60/029,960 31 October 1996 (31.10.96) US<br><b>(71) Applicant (for all designated States except US):</b> HUMAN<br>GENOME SCIENCES, INC. [US/US]; 9410 Key West<br>Avenue, Rockville, MD 20850 (US).<br><b>(72) Inventors; and</b><br><b>(75) Inventors/Applicants (for US only):</b> KUNSCH, Charles, A.<br>[US/US]; 2398B Dunwoody Crossing, Atlanta, GA 30338<br>(US). CHOI, Gil, H. [KR/US]; 11429 Potomac Oaks<br>Drive, Rockville, MD 20850 (US). JOHNSON, L., Sydnor<br>[US/US]; 13545 Ambassador Drive, Germantown, MD<br>20874 (US). HROMOCKYJ, Alex [US/US]; 10003 Sidney<br>Road, Silver Spring, MD 20901 (US).<br><b>(74) Agents:</b> BROOKES, A., Anders et al.; Human Genome<br>Sciences, Inc., 9410 Key West Avenue, Rockville, MD<br>20850 (US). |           | <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR,<br>BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE,<br>GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK,<br>LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO,<br>NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR,<br>TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH,<br>KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ,<br>BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE,<br>CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,<br>PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN,<br>ML, MR, NE, SN, TD, TG).<br><br><b>Published</b><br><i>With international search report.</i><br><i>Before the expiration of the time limit for amending the claims</i><br><i>and to be republished in the event of the receipt of amendments.</i><br><br><b>(88) Date of publication of the international search report:</b><br>8 October 1998 (08.10.98) |
| <b>(54) Title:</b> <i>STREPTOCOCCUS PNEUMONIAE</i> ANTIGENS AND VACCINES<br><br><b>(57) Abstract</b><br><p>The present invention relates to novel vaccines for the prevention or attenuation of infection by <i>Streptococcus pneumoniae</i>. The invention further relates to isolated nucleic acid molecules encoding antigenic polypeptides of <i>Streptococcus pneumoniae</i>. Antigenic polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. The invention additionally relates to diagnostic methods for detecting <i>Streptococcus</i> nucleic acids, polypeptides and antibodies in a biological sample.</p>                                                                                                                                                                                                                                                         |           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |

## ***Streptococcus pneumoniae* Antigens and Vaccines**

### ***Field of the Invention***

The present invention relates to novel *Streptococcus pneumoniae* antigens for the detection of *Streptococcus* and for the prevention or attenuation of disease caused by *Streptococcus*. The invention further relates to isolated nucleic acid molecules encoding antigenic polypeptides of *S. pneumoniae*. Antigenic polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. The invention additionally relates to diagnostic methods for detecting *Streptococcus* gene expression.

### ***Background of the Invention***

*Streptococcus pneumoniae* has been one of the most extensively studied microorganisms since its first isolation in 1881. It was the object of many investigations that led to important scientific discoveries. In 1928, Griffith observed that when heat-killed encapsulated pneumococci and live strains constitutively lacking any capsule were concomitantly injected into mice, the nonencapsulated could be converted into encapsulated pneumococci with the same capsular type as the heat-killed strain. Years later, the nature of this "transforming principle," or carrier of genetic information, was shown to be DNA. (Avery, O.T., *et al.*, *J. Exp. Med.*, 79:137-157 (1944)).

In spite of the vast number of publications on *S. pneumoniae* many questions about its virulence are still unanswered, and this pathogen remains a major causative agent of serious human disease, especially community-acquired pneumonia. (Johnston, R.B., *et al.*, *Rev. Infect. Dis.* 13(Suppl. 6):S509-517 (1991)). In addition, in developing countries, the pneumococcus is responsible for the death of a large number of children under the age of 5 years from pneumococcal pneumonia. The incidence of pneumococcal disease is highest in infants under 2 years of age and in people over 60 years of age. Pneumococci are the second most frequent cause (after *Haemophilus influenzae* type b) of bacterial meningitis and otitis media in children. With the recent introduction of conjugate vaccines for *H. influenzae* type b, pneumococcal meningitis is likely to become increasingly prominent. *S. pneumoniae* is the most important etiologic agent of community-acquired pneumonia in adults and is the second most common cause of bacterial meningitis behind *Neisseria meningitidis*.

The antibiotic generally prescribed to treat *S. pneumoniae* is benzylpenicillin, although resistance to this and to other antibiotics is found occasionally. Pneumococcal resistance to penicillin results from mutations in its

penicillin-binding proteins. In uncomplicated pneumococcal pneumonia caused by a sensitive strain, treatment with penicillin is usually successful unless started too late. Erythromycin or clindamycin can be used to treat pneumonia in patients hypersensitive to penicillin, but resistant strains to these drugs exist. Broad spectrum antibiotics (e.g., the tetracyclines) may also be effective, although tetracycline-resistant strains are not rare. In spite of the availability of antibiotics, the mortality of pneumococcal bacteremia in the last four decades has remained stable between 25 and 29%. (Gillespie, S.H., *et al.*, *J. Med. Microbiol.* 28:237-248 (1989).

*S. pneumoniae* is carried in the upper respiratory tract by many healthy individuals. It has been suggested that attachment of pneumococci is mediated by a disaccharide receptor on fibronectin, present on human pharyngeal epithelial cells. (Anderson, B.J., *et al.*, *J. Immunol.* 142:2464-2468 (1989). The mechanisms by which pneumococci translocate from the nasopharynx to the lung, thereby causing pneumonia, or migrate to the blood, giving rise to bacteremia or septicemia, are poorly understood. (Johnston, R.B., *et al.*, *Rev. Infect. Dis.* 13(Suppl. 6):S509-517 (1991).

Various proteins have been suggested to be involved in the pathogenicity of *S. pneumoniae*, however, only a few of them have actually been confirmed as virulence factors. Pneumococci produce an IgA1 protease that might interfere with host defense at mucosal surfaces. (Kornfield, S.J., *et al.*, *Rev. Inf. Dis.* 3:521-534 (1981). *S. pneumoniae* also produces neuraminidase, an enzyme that may facilitate attachment to epithelial cells by cleaving sialic acid from the host glycolipids and gangliosides. Partially purified neuraminidase was observed to induce meningitis-like symptoms in mice; however, the reliability of this finding has been questioned because the neuraminidase preparations used were probably contaminated with cell wall products. Other pneumococcal proteins besides neuraminidase are involved in the adhesion of pneumococci to epithelial and endothelial cells. These pneumococcal proteins have as yet not been identified. Recently, Cundell *et al.*, reported that peptide permeases can modulate pneumococcal adherence to epithelial and endothelial cells. It was, however, unclear whether these permeases function directly as adhesions or whether they enhance adherence by modulating the expression of pneumococcal adhesions. (DeVelasco, E.A., *et al.*, *Micro. Rev.* 59:591-603 (1995). A better understanding of the virulence factors determining its pathogenicity will need to be developed to cope with the devastating effects of pneumococcal disease in humans.

Ironically, despite the prominent role of *S. pneumoniae* in the discovery of DNA, little is known about the molecular genetics of the organism. The *S. pneumoniae* genome consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory elements, such as prophages, plasmids, transposons and the like. Most physical characteristics and almost all of the genes of *S. pneumoniae* are unknown. Among the few that have been identified, most have not been physically mapped or characterized in detail. Only a few genes of this organism have been sequenced. (See, for instance current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. pneumoniae* such as those set out elsewhere herein.) Identification of *in vivo*-expressed, and broadly protective, antigens of *S. pneumoniae* has remained elusive.

### *Summary of the Invention*

The present invention provides isolated nucleic acid molecules comprising polynucleotides encoding the *S. pneumoniae* polypeptides described in Table 1 and having the amino acid sequences shown as SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and so on through SEQ ID NO:226. Thus, one aspect of the invention provides isolated nucleic acid molecules comprising polynucleotides having a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding any of the amino acid sequences of the polypeptides shown in Table 1; and (b) a nucleotide sequence complementary to any of the nucleotide sequences in (a).

Further embodiments of the invention include isolated nucleic acid molecules that comprise a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical, to any of the nucleotide sequences in (a) or (b) above, or a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide in (a) or (b) above. This polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues. Additional nucleic acid embodiments of the invention relate to isolated nucleic acid molecules comprising polynucleotides which encode the amino acid sequences of epitope-bearing portions of an *S. pneumoniae* polypeptide having an amino acid sequence in (a) above.

The present invention also relates to recombinant vectors, which include the isolated nucleic acid molecules of the present invention, and to host cells containing the recombinant vectors, as well as to methods of making such

vectors and host cells and for using these vectors for the production of *S. pneumoniae* polypeptides or peptides by recombinant techniques.

The invention further provides isolated *S. pneumoniae* polypeptides having an amino acid sequence selected from the group consisting of an amino acid sequence of any of the polypeptides described in Table 1.

The polypeptides of the present invention also include polypeptides having an amino acid sequence with at least 70% similarity, and more preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% similarity to those described in Table 1, as well as polypeptides having an amino acid sequence at least 70% identical, more preferably at least 75% identical, and still more preferably 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to those above; as well as isolated nucleic acid molecules encoding such polypeptides.

The present invention further provides a vaccine, preferably a multi-component vaccine comprising one or more of the *S. pneumoniae* polynucleotides or polypeptides described in Table 1, or fragments thereof, together with a pharmaceutically acceptable diluent, carrier, or excipient, wherein the *S. pneumoniae* polypeptide(s) are present in an amount effective to elicit an immune response to members of the *Streptococcus* genus in an animal. The *S. pneumoniae* polypeptides of the present invention may further be combined with one or more immunogens of one or more other streptococcal or non-streptococcal organisms to produce a multi-component vaccine intended to elicit an immunological response against members of the *Streptococcus* genus and, optionally, one or more non-streptococcal organisms.

The vaccines of the present invention can be administered in a DNA form, *e.g.*, "naked" DNA, wherein the DNA encodes one or more streptococcal polypeptides and, optionally, one or more polypeptides of a non-streptococcal organism. The DNA encoding one or more polypeptides may be constructed such that these polypeptides are expressed fusion proteins.

The vaccines of the present invention may also be administered as a component of a genetically engineered organism. Thus, a genetically engineered organism which expresses one or more *S. pneumoniae* polypeptides may be administered to an animal. For example, such a genetically engineered organism may contain one or more *S. pneumoniae* polypeptides of the present invention intracellularly, on its cell surface, or in its periplasmic space. Further, such a genetically engineered organism may secrete one or more *S. pneumoniae* polypeptides.

The vaccines of the present invention may be co-administered to an animal with an immune system modulator (e.g., CD86 and GM-CSF).

The invention also provides a method of inducing an immunological response in an animal to one or more members of the *Streptococcus* genus, preferably one or more isolates of the *S. pneumoniae* genus, comprising administering to the animal a vaccine as described above.

The invention further provides a method of inducing a protective immune response in an animal, sufficient to prevent or attenuate an infection by members of the *Streptococcus* genus, preferably at least *S. pneumoniae*, comprising administering to the animal a composition comprising one or more of the polynucleotides or polypeptides described in Table 1, or fragments thereof. Further, these polypeptides, or fragments thereof, may be conjugated to another immunogen and/or administered in admixture with an adjuvant.

The invention further relates to antibodies elicited in an animal by the administration of one or more *S. pneumoniae* polypeptides of the present invention and to methods for producing such antibodies.

The invention also provides diagnostic methods for detecting the expression of genes of members of the *Streptococcus* genus in an animal. One such method involves assaying for the expression of a gene encoding *S. pneumoniae* peptides in a sample from an animal. This expression may be assayed either directly (e.g., by assaying polypeptide levels using antibodies elicited in response to amino acid sequences described in Table 1) or indirectly (e.g., by assaying for antibodies having specificity for amino acid sequences described in Table 1). An example of such a method involves the use of the polymerase chain reaction (PCR) to amplify and detect *Streptococcus* nucleic acid sequences.

The present invention also relates to nucleic acid probes having all or part of a nucleotide sequence described in Table 1 (shown as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and so on through SEQ ID NO:225) which are capable of hybridizing under stringent conditions to *Streptococcus* nucleic acids. The invention further relates to a method of detecting one or more *Streptococcus* nucleic acids in a biological sample obtained from an animal, said one or more nucleic acids encoding *Streptococcus* polypeptides, comprising: (a) contacting the sample with one or more of the above-described nucleic acid probes, under conditions such that hybridization occurs, and (b) detecting hybridization of said one or more probes to the *Streptococcus* nucleic acid present in the biological sample.

The invention also includes immunoassays, including an immunoassay for detecting *Streptococcus*, preferably at least isolates of the *S. pneumoniae* genus, comprising incubation of a sample (which is suspected of being infected with *Streptococcus*) with a probe antibody directed against an antigen/epitope of *S. pneumoniae*, to be detected under conditions allowing the formation of an antigen-antibody complex; and detecting the antigen-antibody complex which contains the probe antibody. An immunoassay for the detection of antibodies which are directed against a *Streptococcus* antigen comprising the incubation of a sample (containing antibodies from a mammal suspected of being infected with *Streptococcus*) with a probe polypeptide including an epitope of *S. pneumoniae*, under conditions that allow the formation of antigen-antibody complexes which contain the probe epitope containing antigen.

Some aspects of the invention pertaining to kits are those for: investigating samples for the presence of polynucleotides derived from *Streptococcus* which comprise a polynucleotide probe including a nucleotide sequence selected from Table 1 or a fragment thereof of approximately 15 or more nucleotides, in an appropriate container; analyzing the samples for the presence of antibodies directed against a *Streptococcus* antigen made up of a polypeptide which contains a *S. pneumoniae* epitope present in the polypeptide, in a suitable container; and analyzing samples for the presence of *Streptococcus* antigens made up of an anti-*S. pneumoniae* antibody, in a suitable container.

### **Detailed Description**

The present invention relates to recombinant antigenic *S. pneumoniae* polypeptides and fragments thereof. The invention also relates to methods for using these polypeptides to produce immunological responses and to confer immunological protection to disease caused by members of the genus *Streptococcus*, at least isolates of the *S. pneumoniae* genus. The invention further relates to nucleic acid sequences which encode antigenic *S. pneumoniae* polypeptides and to methods for detecting *S. pneumoniae* nucleic acids and polypeptides in biological samples. The invention also relates to *S. pneumoniae*-specific antibodies and methods for detecting such antibodies produced in a host animal.

### **Definitions**

The following definitions are provided to clarify the subject matter which the inventors consider to be the present invention.

As used herein, the phrase "pathogenic agent" means an agent which causes a disease state or affliction in an animal. Included within this definition, for examples, are bacteria, protozoans, fungi, viruses and metazoan parasites which either produce a disease state or render an animal infected with such an organism susceptible to a disease state (*e.g.*, a secondary infection). Further included are species and strains of the genus *Streptococcus* which produce disease states in animals.

As used herein, the term "organism" means any living biological system, including viruses, regardless of whether it is a pathogenic agent.

As used herein, the term "*Streptococcus*" means any species or strain of bacteria which is members of the genus *Streptococcus*. Such species and strains are known to those of skill in the art, and include those that are pathogenic and those that are not.

As used herein, the phrase "one or more *S. pneumoniae* polypeptides of the present invention" means polypeptides comprising the amino acid sequence of one or more of the *S. pneumoniae* polypeptides described in Table 1 and disclosed as SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and so on through SEQ ID NO:226. These polypeptides may be expressed as fusion proteins wherein the *S. pneumoniae* polypeptides of the present invention are linked to additional amino acid sequences which may be of streptococcal or non-streptococcal origin. This phrase further includes polypeptide comprising fragments of the *S. pneumoniae* polypeptides of the present invention.

Additional definitions are provided throughout the specification.

### ***Explanation of Table 1***

Table 1, below, provides information describing 113 open reading frames (ORFs) which encode potentially antigenic polypeptides of *S. pneumoniae* of the present invention. The table lists the ORF identifier which consists of the letters SP, which denote *S. pneumoniae*, followed immediately by a three digit numeric code, which arbitrarily number the potentially antigenic polypeptides of *S. pneumoniae* of the present invention and the nucleotide or amino acid sequence of each ORF and encoded polypeptide. The table further correlates the ORF identifier with a sequence identification number (SEQ ID NO:). The actual nucleotide or amino acid sequence of each ORF identifier is also shown in the Sequence Listing under the corresponding SEQ ID NO.

Thus, for example, the designation "SP126" refers to both the nucleotide and amino acid sequences of *S. pneumoniae* polypeptide number 126 of the present invention. Further, "SP126" correlates with the nucleotide

sequence shown as SEQ ID NO:223 and with the amino acid sequence shown as SEQ ID NO:224 as is described in Table 1.

The open reading frame within each "ORF" begins with the second nucleotide shown. Thus, the first codon for each nucleotide sequence shown is bases 2-4, the second 5-7, the third 8-10, and so on.

### ***Explanation of Table 2***

Table 2 lists the antigenic epitopes present in each of the *S. pneumoniae* polypeptides described in Table 1 as predicted by the inventors. Each *S. pneumoniae* polypeptide shown in Table 1 has one or more antigenic epitopes described in Table 2. It will be appreciated that depending on the analytical criteria used to predict antigenic determinants, the exact address of the determinant may vary slightly. The exact location of the antigenic determinant may shift by about 1 to 5 residues, more likely 1 to 2 residues, depending on the criteria used. Thus, the first antigenic determinant described in Table 2, "Lys-1 to Ile-10" of SP001, represents a peptide comprising the lysine at position 1 in SEQ ID NO:2 through and including the isoleucine at position 10 in SEQ ID NO:2, but may include more or fewer residues than those 10. It will also be appreciated that, generally speaking, amino acids can be added to either terminus of a peptide or polypeptide containing an antigenic epitope without affecting its activity, whereas removing residues from a peptide or polypeptide containing only the antigenic determinant is much more likely to destroy activity. It will be appreciated that the residues and locations shown described in Table 2 correspond to the amino acid sequences for each ORF shown in Table 1 and in the Sequence Listing.

### ***Explanation of Table 3***

Table 3 shows PCR primers designed by the inventors for the amplification of polynucleotides encoding polypeptides of the present invention according to the method of Example 1. PCR primer design is routine in the art and those shown in Table 3 are provided merely for the convenience of the skilled artisan. It will be appreciated that others can be used with equal success.

For each primer, the table lists the corresponding ORF designation from Table 1 followed by either an "A" or a "B". The "A" primers are the 5' primers and the "B" primers 3'. A restriction enzyme site was built into each primer to allow ease of cloning. The restriction enzyme which will recognize and cleave a sequence within each primer is shown in Table 3, as well, under the heading

"RE" for restriction enzyme. Finally the sequence identifier is shown in Table 3 for each primer for easy correlation with the Sequence Listing.

***Selection of Nucleic Acid Sequences Encoding Antigenic S. pneumoniae Polypeptides***

The present invention provides a select number of ORFs from those presented in the fragments of the *S. pneumoniae* genome which may prove useful for the generation of a protective immune response. The sequenced *S. pneumoniae* genomic DNA was obtained from a sub-cultured isolate of *S. pneumoniae* Strain 7/87 14.8.91, which has been deposited at the American Type Culture Collection, as a convenience to those of skill in the art. The *S. pneumoniae* isolate was deposited on October 10, 1996 at the ATCC, 12301 Park Lawn Drive, Rockville, Maryland 20852, and given accession number 55840. A genomic library constructed from DNA isolated from the *S. pneumoniae* isolate was also deposited at the ATCC on October 11, 1996 and given ATCC Deposit No. 97755. A more complete listing of the sequence obtained from the *S. pneumoniae* genome may be found in co-pending U.S. Provisional Application Serial No. 60/029,960, filed 10/31/96, incorporated herein by reference in its entirety. Some ORFs contained in the subset of fragments of the *S. pneumoniae* genome disclosed herein were derived through the use of a number of screening criteria detailed below.

The selected ORFs do not consist of complete ORFs. Although a polypeptide representing a complete ORF may be the closest approximation of a protein native to an organism, it is not always preferred to express a complete ORF in a heterologous system. It may be challenging to express and purify a highly hydrophobic protein by common laboratory methods. Thus, the polypeptide vaccine candidates described herein may have been modified slightly to simplify the production of recombinant protein. For example, nucleotide sequences which encode highly hydrophobic domains, such as those found at the amino terminal signal sequence, have been excluded from some constructs used for *in vitro* expression of the polypeptides. Furthermore, any highly hydrophobic amino acid sequences occurring at the carboxy terminus have also been excluded from the recombinant expression constructs. Thus, in one embodiment, a polypeptide which represents a truncated or modified ORF may be used as an antigen.

While numerous methods are known in the art for selecting potentially immunogenic polypeptides, many of the ORFs disclosed herein were selected

on the basis of screening all theoretical *S. pneumoniae* ORFs for several aspects of potential immunogenicity. One set of selection criteria are as follows:

1. *Type I signal sequence*: An amino terminal type I signal sequence generally directs a nascent protein across the plasma and outer membranes to the exterior of the bacterial cell. Experimental evidence obtained from studies with *Escherichia coli* suggests that the typical type I signal sequence consists of the following biochemical and physical attributes (Izard, J. W. and Kendall, D. A. *Mol. Microbiol.* **13**:765-773 (1994)). The length of the type I signal sequence is approximately 15 to 25 primarily hydrophobic amino acid residues with a net positive charge in the extreme amino terminus. In addition, the central region of the signal sequence adopts an alpha-helical conformation in a hydrophobic environment. Finally, the region surrounding the actual site of cleavage is ideally six residues long, with small side-chain amino acids in the -1 and -3 positions.

2. *Type IV signal sequence*: The type IV signal sequence is an example of the several types of functional signal sequences which exist in addition to the type I signal sequence detailed above. Although functionally related, the type IV signal sequence possesses a unique set of biochemical and physical attributes (Strom, M. S. and Lory, S., *J. Bacteriol.* **174**:7345-7351 (1992)). These are typically six to eight amino acids with a net basic charge followed by an additional sixteen to thirty primarily hydrophobic residues. The cleavage site of a type IV signal sequence is typically after the initial six to eight amino acids at the extreme amino terminus. In addition, type IV signal sequences generally contain a phenylalanine residue at the +1 site relative to the cleavage site.

3. *Lipoprotein*: Studies of the cleavage sites of twenty-six bacterial lipoprotein precursors has allowed the definition of a consensus amino acid sequence for lipoprotein cleavage. Nearly three-fourths of the bacterial lipoprotein precursors examined contained the sequence L-(A,S)-(G,A)-C at positions -3 to +1, relative to the point of cleavage (Hayashi, S. and Wu, H. C., *J. Bioenerg. Biomembr.* **22**:451-471 (1990)).

4. *LPXTG motif*: It has been experimentally determined that most anchored proteins found on the surface of gram-positive bacteria possess a highly conserved carboxy terminal sequence. More than fifty such proteins from organisms such as *S. pyogenes*, *S. mutans*, *E. faecalis*, *S. pneumoniae*, and others, have been identified based on their extracellular location and carboxy terminal amino acid sequence (Fischetti, V. A., *ASM News* **62**:405-410 (1996)). The conserved region consists of six charged amino acids at the extreme carboxy terminus coupled to 15-20 hydrophobic amino acids

presumed to function as a transmembrane domain. Immediately adjacent to the transmembrane domain is a six amino acid sequence conserved in nearly all proteins examined. The amino acid sequence of this region is L-P-X-T-G-X, where X is any amino acid.

5 An algorithm for selecting antigenic and immunogenic *S. pneumoniae* polypeptides including the foregoing criteria was developed. Use of the algorithm by the inventors to select immunologically useful *S. pneumoniae* polypeptides resulted in the selection of a number of the disclosed ORFs. Polypeptides comprising the polypeptides identified in this group may be  
10 produced by techniques standard in the art and as further described herein.

### ***Nucleic Acid Molecules***

The present invention provides isolated nucleic acid molecules comprising polynucleotides encoding the *S. pneumoniae* polypeptides having  
15 the amino acid sequences described in Table 1 and shown as SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and so on through SEQ ID NO:226, which were determined by sequencing the genome of *S. pneumoniae* and selected as putative immunogens.

Unless otherwise indicated, all nucleotide sequences determined by  
20 sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the Model 373 from Applied Biosystems, Inc.), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of DNA sequences determined as above. Therefore, as is known in the art for any DNA sequence determined by this  
25 automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other  
30 approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid  
35 sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

Unless otherwise indicated, each "nucleotide sequence" set forth herein is presented as a sequence of deoxyribonucleotides (abbreviated A, G, C and

T). However, by "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides, and for an RNA molecule or polynucleotide, the corresponding sequence of ribonucleotides (A, G, C and U), where each thymidine deoxyribonucleotide (T) in the specified deoxyribonucleotide sequence is replaced by the ribonucleotide uridine (U). For instance, reference to an RNA molecule having a sequence described in Table 1 set forth using deoxyribonucleotide abbreviations is intended to indicate an RNA molecule having a sequence in which each deoxyribonucleotide A, G or C described in Table 1 has been replaced by the corresponding ribonucleotide A, G or C, and each deoxyribonucleotide T has been replaced by a ribonucleotide U.

Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

By "isolated" nucleic acid molecule(s) is intended a nucleic acid molecule, DNA or RNA, which has been removed from its native environment. For example, recombinant DNA molecules contained in a vector are considered isolated for the purposes of the present invention. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Isolated nucleic acid molecules of the present invention include DNA molecules comprising a nucleotide sequence described in Table 1 and shown as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and so on through SEQ ID NO:225; DNA molecules comprising the coding sequences for the polypeptides described in Table 1 and shown as SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and so on through SEQ ID NO:226; and DNA molecules which comprise sequences substantially different from those described above but which, due to the degeneracy of the genetic code, still encode the *S. pneumoniae* polypeptides described in Table 1. Of course, the genetic code is well known in the art. Thus, it would be routine for one skilled in the art to generate such degenerate variants.

The invention also provides nucleic acid molecules having sequences complementary to any one of those described in Table 1. Such isolated molecules, particularly DNA molecules, are useful as probes for detecting expression of *Streptococcal* genes, for instance, by Northern blot analysis or the polymerase chain reaction (PCR).

The present invention is further directed to fragments of the isolated nucleic acid molecules described herein. By a fragment of an isolated nucleic acid molecule having a nucleotide sequence described in Table 1, is intended fragments at least about 15 nt, and more preferably at least about 17 nt, still more preferably at least about 20 nt, and even more preferably, at least about 25 nt in length which are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments 50-100 nt in length are also useful according to the present invention as are fragments corresponding to most, if not all, of a nucleotide sequence described in Table 1. By a fragment at least 20 nt in length, for example, is intended fragments which include 20 or more contiguous bases of a nucleotide sequence as described in Table 1. Since the nucleotide sequences identified in Table 1 are provided as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and so on through SEQ ID NO:225, generating such DNA fragments would be routine to the skilled artisan. For example, such fragments could be generated synthetically.

Preferred nucleic acid fragments of the present invention also include nucleic acid molecules comprising nucleotide sequences encoding epitope-bearing portions of the *S. pneumoniae* polypeptides identified in Table 1. Such nucleic acid fragments of the present invention include, for example, nucleotide sequences encoding polypeptide fragments comprising from about the amino terminal residue to about the carboxy terminal residue of each fragment shown in Table 2. The above referred to polypeptide fragments are antigenic regions of the *S. pneumoniae* polypeptides identified in Table 1.

In another aspect, the invention provides isolated nucleic acid molecules comprising polynucleotides which hybridize under stringent hybridization conditions to a portion of a polynucleotide in a nucleic acid molecule of the invention described above, for instance, a nucleic acid sequence identified in Table 1. By "stringent hybridization conditions" is intended overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 g/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

By polynucleotides which hybridize to a "portion" of a polynucleotide is intended polynucleotides (either DNA or RNA) which hybridize to at least about 15 nucleotides (nt), and more preferably at least about 17 nt, still more preferably at least about 20 nt, and even more preferably about 25-70 nt of the reference polynucleotide. These are useful as diagnostic probes and primers as discussed above and in more detail below.

Of course, polynucleotides hybridizing to a larger portion of the reference polynucleotide, for instance, a portion 50-100 nt in length, or even to the entire length of the reference polynucleotide, are also useful as probes according to the present invention, as are polynucleotides corresponding to most, if not all, of a nucleotide sequence as identified in Table 1. By a portion of a polynucleotide of "at least 20 nt in length," for example, is intended 20 or more contiguous nucleotides from the nucleotide sequence of the reference polynucleotide (*e.g.*, a nucleotide sequences as described in Table 1). As noted above, such portions are useful diagnostically either as probes according to conventional DNA hybridization techniques or as primers for amplification of a target sequence by PCR, as described in the literature (for instance, in *Molecular Cloning, A Laboratory Manual*, 2nd. edition, Sambrook, J., Fritsch, E. F. and Maniatis, T., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), the entire disclosure of which is hereby incorporated herein by reference).

Since nucleic acid sequences encoding the *S. pneumoniae* polypeptides of the present invention are identified in Table 1 and provided as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, and so on through SEQ ID NO:225, generating polynucleotides which hybridize to portions of these sequences would be routine to the skilled artisan. For example, the hybridizing polynucleotides of the present invention could be generated synthetically according to known techniques.

As indicated, nucleic acid molecules of the present invention which encode *S. pneumoniae* polypeptides of the present invention may include, but are not limited to those encoding the amino acid sequences of the polypeptides by themselves; and additional coding sequences which code for additional amino acids, such as those which provide additional functionalities. Thus, the sequences encoding these polypeptides may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (Qiagen, Inc.), among others, many of which are

commercially available. As described by Gentz and colleagues (*Proc. Natl. Acad. Sci. USA* **86**:821-824 (1989)), for instance, hexa-histidine provides for convenient purification of the resulting fusion protein.

Thus, the present invention also includes genetic fusions wherein the *S. pneumoniae* nucleic acid sequences coding sequences identified in Table 1 are linked to additional nucleic acid sequences to produce fusion proteins. These fusion proteins may include epitopes of streptococcal or non-streptococcal origin designed to produce proteins having enhanced immunogenicity. Further, the fusion proteins of the present invention may contain antigenic determinants known to provide helper T-cell stimulation, peptides encoding sites for post-translational modifications which enhance immunogenicity (*e.g.*, acylation), peptides which facilitate purification (*e.g.*, histidine "tag"), or amino acid sequences which target the fusion protein to a desired location (*e.g.*, a heterologous leader sequence).

In all cases of bacterial expression, an N-terminal methionine residues is added. In many cases, however, the N-terminal methionine residues is cleaved off post-translationally. Thus, the invention includes polypeptides shown in Table 1 with, and without an N-terminal methionine.

The present invention thus includes nucleic acid molecules and sequences which encode fusion proteins comprising one or more *S. pneumoniae* polypeptides of the present invention fused to an amino acid sequence which allows for post-translational modification to enhance immunogenicity. This post-translational modification may occur either *in vitro* or when the fusion protein is expressed *in vivo* in a host cell. An example of such a modification is the introduction of an amino acid sequence which results in the attachment of a lipid moiety.

Thus, as indicated above, the present invention includes genetic fusions wherein a *S. pneumoniae* nucleic acid sequence identified in Table 1 is linked to a nucleotide sequence encoding another amino acid sequence. These other amino acid sequences may be of streptococcal origin (*e.g.*, another sequence selected from Table 1) or non-streptococcal origin.

The present invention further relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs or derivatives of the *S. pneumoniae* polypeptides described in Table 1. Variants may occur naturally, such as a natural allelic variant. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism (*Genes II*, Lewin, B., ed., John Wiley & Sons,

New York (1985)). Non-naturally occurring variants may be produced using art-known mutagenesis techniques.

Such variants include those produced by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. These variants may be altered in coding regions, non-coding regions, or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the *S. pneumoniae* polypeptides disclosed herein or portions thereof. Silent substitution are most likely to be made in non-epitopic regions. Guidance regarding those regions containing epitopes is provided herein, for example, in Table 2. Also especially preferred in this regard are conservative substitutions.

Further embodiments of the invention include isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical to: (a) a nucleotide sequence encoding any of the amino acid sequences of the polypeptides identified in Table 1; and (b) a nucleotide sequence complementary to any of the nucleotide sequences in (a) above.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence encoding a *S. pneumoniae* polypeptide described in Table 1, is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the subject *S. pneumoniae* polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

Certain nucleotides within some of the nucleic acid sequences shown in Table 1 were ambiguous upon sequencing. Completely unknown sequences are shown as an "N". Other unresolved nucleotides are known to be either a

purine, shown as "R", or a pyrimidine, shown as "Y". Accordingly, when determining identity between two nucleotide sequences, identity is met where any nucleotide, including an "R", "Y" or "N", is found in a test sequence and at the corresponding position in the reference sequence (from Table 1). Likewise, an A, G or "R" in a test sequence is identical to an "R" in the reference sequence; and a T, C or "Y" in a test sequence is identical to a "Y" in the reference sequence.

As a practical matter, whether any particular nucleic acid molecule is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, a nucleotide sequence described in Table 1 can be determined conventionally using known computer programs such as the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). Bestfit uses the local homology algorithm of Smith and Waterman (*Advances in Applied Mathematics* 2:482-489 (1981)), to find the best segment of homology between two sequences. When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference nucleotide sequence and that gaps in homology of up to 5% of the total number of nucleotides in the reference sequence are allowed.

The present application is directed to nucleic acid molecules at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleic acid sequences described in Table 1. One of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe or a polymerase chain reaction (PCR) primer. Uses of the nucleic acid molecules of the present invention include, *inter alia*, (1) isolating *Streptococcal* genes or allelic variants thereof from either a genomic or cDNA library and (2) Northern Blot or PCR analysis for detecting *Streptococcal* mRNA expression.

Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of nucleic acid molecules having a sequence at least 90%, 95%, 96%, 97%, 98%, or 99% identical to a nucleic acid sequence identified in Table 1 will encode the same polypeptide. In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even without performing the above described comparison assay.

It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode

proteins having antigenic epitopes of the *S. pneumoniae* polypeptides of the present invention. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect the antigenicity of a polypeptide (*e.g.*, replacement of an amino acid in a region which is not believed to form an antigenic epitope). For example, since antigenic epitopes have been identified which contain as few as six amino acids (see Harlow, *et al.*, *Antibodies: A Laboratory Manual*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988), page 76), in instances where a polypeptide has multiple antigenic epitopes the alteration of several amino acid residues would often not be expected to eliminate all of the antigenic epitopes of that polypeptide. This is especially so when the alterations are in regions believed to not constitute antigenic epitopes.

### ***Vectors and Host Cells***

The present invention also relates to vectors which include the isolated DNA molecules of the present invention, host cells which are genetically engineered with the recombinant vectors, and the production of *S. pneumoniae* polypeptides or fragments thereof by recombinant techniques.

Recombinant constructs may be introduced into host cells using well known techniques such as infection, transduction, transfection, transvection, electroporation and transformation. The vector may be, for example, a phage, plasmid, viral or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line and then transduced into host cells.

Preferred are vectors comprising *cis*-acting control regions to the polynucleotide of interest. Appropriate *trans*-acting factors may be supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression, which may be inducible and/or cell type-specific. Particularly preferred among such vectors are those inducible by environmental factors that are easy to manipulate, such as temperature and nutrient additives.

Expression vectors useful in the present invention include chromosomal-, episomal- and virus-derived vectors, *e.g.*, vectors derived from bacterial plasmids, bacteriophage, yeast episomes, yeast chromosomal elements, viruses such as baculoviruses, papova viruses, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as cosmids and phagemids.

The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli lac*, *trp* and *tac* promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will preferably include a translation initiating site at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture and tetracycline or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A available from Stratagene; pET series of vectors available from Novagen; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Among known bacterial promoters suitable for use in the present invention include the *E. coli lacI* and *lacZ* promoters, the T3 and T7 promoters, the *gpt* promoter, the lambda PR and PL promoters and the *trp* promoter. Suitable eukaryotic promoters include the CMV immediate early promoter, the

HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus (RSV), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals (for example, Davis, *et al.*, *Basic Methods In Molecular Biology* (1986)).

Transcription of DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are *cis*-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

For secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals, but also additional heterologous functional regions. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to solubilize proteins. For example, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is thoroughly advantageous for use in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262).

On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified in the advantageous manner described. This is the case when Fc portion proves to be a hindrance to use in therapy and diagnosis, for example when the fusion protein is to be used as antigen for immunizations. In drug discovery, for example, human proteins, such as, hIL-5-receptor has been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. See Bennett, D. *et al.*, *J. Molec. Recogn.* 8:52-58 (1995) and Johanson, K. *et al.*, *J. Biol. Chem.* 270 (16):9459-9471 (1995).

The *S. pneumoniae* polypeptides can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography and high performance liquid chromatography ("HPLC") is employed for purification. Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells.

### ***Polypeptides and Fragments***

The invention further provides isolated polypeptides having the amino acid sequences described in Table 1, and shown as SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, and so on through SEQ ID NO:226, and peptides or polypeptides comprising portions of the above polypeptides. The terms "peptide" and "oligopeptide" are considered synonymous (as is commonly recognized) and each term can be used interchangeably as the context requires to indicate a chain of at least two amino acids coupled by peptidyl linkages. The word "polypeptide" is used herein for chains containing more than ten amino acid residues. All oligopeptide and polypeptide formulas or sequences herein are written from left to right and in the direction from amino terminus to carboxy terminus.

Some amino acid sequences of the *S. pneumoniae* polypeptides described in Table 1 can be varied without significantly effecting the antigenicity of the polypeptides. If such differences in sequence are contemplated, it should be remembered that there will be critical areas on the polypeptide which determine antigenicity. In general, it is possible to replace residues which do

not form part of an antigenic epitope without significantly effecting the antigenicity of a polypeptide. Guidance for such alterations is given in Table 2 wherein epitopes for each polypeptide is delineated.

The polypeptides of the present invention are preferably provided in an isolated form. By "isolated polypeptide" is intended a polypeptide removed from its native environment. Thus, a polypeptide produced and/or contained within a recombinant host cell is considered isolated for purposes of the present invention. Also intended as an "isolated polypeptide" is a polypeptide that has been purified, partially or substantially, from a recombinant host cell. For example, recombinantly produced versions of the *S. pneumoniae* polypeptides described in Table 1 can be substantially purified by the one-step method described by Smith and Johnson (*Gene* 67:31-40 (1988)).

The polypeptides of the present invention include: (a) an amino acid sequence of any of the polypeptides described in Table 1; and (b) an amino acid sequence of an epitope-bearing portion of any one of the polypeptides of (a); as well as polypeptides with at least 70% similarity, and more preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% similarity to those described in (a) or (b) above, as well as polypeptides having an amino acid sequence at least 70% identical, more preferably at least 75% identical, and still more preferably 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to those above.

By "% similarity" for two polypeptides is intended a similarity score produced by comparing the amino acid sequences of the two polypeptides using the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) and the default settings for determining similarity. Bestfit uses the local homology algorithm of Smith and Waterman (*Advances in Applied Mathematics* 2:482-489 (1981)) to find the best segment of similarity between two sequences.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a reference amino acid sequence of a *S. pneumoniae* polypeptide is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to

5 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

10 The amino acid sequences shown in Table 1 may have on or more "X" residues. "X" represents unknown. Thus, for purposes of defining identity, if any amino acid is present at the same position in a reference amino acid sequence (shown in Table 1) where an X is shown, the two sequences are identical at that position.

15 As a practical matter, whether any particular polypeptide is at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to, for instance, an amino acid sequence shown in Table 1, can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference amino acid sequence and that gaps in homology of up to 5% of the total number of amino acid residues in the reference sequence are allowed.

25 As described below, the polypeptides of the present invention can also be used to raise polyclonal and monoclonal antibodies, which are useful in assays for detecting *Streptococcal* protein expression.

30 In another aspect, the invention provides peptides and polypeptides comprising epitope-bearing portions of the *S. pneumoniae* polypeptides of the invention. These epitopes are immunogenic or antigenic epitopes of the polypeptides of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein or polypeptide is the immunogen. These immunogenic epitopes are believed to be confined to a few loci on the molecule. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic determinant" or "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes (Geysen, *et al.*, *Proc. Natl. Acad. Sci. USA* **81**:3998- 4002 (1983)). Predicted antigenic epitopes are shown in Table 2, below.

As to the selection of peptides or polypeptides bearing an antigenic epitope (*i.e.*, that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein (for instance, Sutcliffe, J., *et al.*, *Science* **219**:660-666 (1983)). Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (*i.e.*, immunogenic epitopes) nor to the amino or carboxyl terminals. Peptides that are extremely hydrophobic and those of six or fewer residues generally are ineffective at inducing antibodies that bind to the mimicked protein; longer, peptides, especially those containing proline residues, usually are effective (Sutcliffe, *et al.*, *supra*, p. 661). For instance, 18 of 20 peptides designed according to these guidelines, containing 8-39 residues covering 75% of the sequence of the influenza virus hemagglutinin HA1 polypeptide chain, induced antibodies that reacted with the HA1 protein or intact virus; and 12/12 peptides from the MuLV polymerase and 18/18 from the rabies glycoprotein induced antibodies that precipitated the respective proteins.

Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. Thus, a high proportion of hybridomas obtained by fusion of spleen cells from donors immunized with an antigen epitope-bearing peptide generally secrete antibody reactive with the native protein (Sutcliffe, *et al.*, *supra*, p. 663). The antibodies raised by antigenic epitope-bearing peptides or polypeptides are useful to detect the mimicked protein, and antibodies to different peptides may be used for tracking the fate of various regions of a protein precursor which undergoes post-translational processing. The peptides and anti-peptide antibodies may be used in a variety of qualitative or quantitative assays for the mimicked protein, for instance in competition assays since it has been shown that even short peptides (*e.g.*, about 9 amino acids) can bind and displace the larger peptides in immunoprecipitation assays (for instance, Wilson, *et al.*, *Cell* **37**:767-778 (1984) p. 777). The anti-peptide antibodies of the invention also are useful for purification of the mimicked protein, for instance, by adsorption chromatography using methods well known in the art.

Antigenic epitope-bearing peptides and polypeptides of the invention designed according to the above guidelines preferably contain a sequence of at

least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. However, peptides or polypeptides comprising a larger portion of an amino acid sequence of a polypeptide of the invention, containing about 30 to about 50 amino acids, or any length up to and including the entire amino acid sequence of a polypeptide of the invention, also are considered epitope-bearing peptides or polypeptides of the invention and also are useful for inducing antibodies that react with the mimicked protein. Preferably, the amino acid sequence of the epitope-bearing peptide is selected to provide substantial solubility in aqueous solvents (*i.e.*, the sequence includes relatively hydrophilic residues and highly hydrophobic sequences are preferably avoided); and sequences containing proline residues are particularly preferred.

Non-limiting examples of antigenic polypeptides or peptides that can be used to generate *Streptococcal*-specific antibodies include portions of the amino acid sequences identified in Table 1. More specifically, Table 2 discloses antigenic fragments of polypeptides of the present invention, which antigenic fragments comprise amino acid sequences from about the first amino acid residues indicated to about the last amino acid residue indicated for each fragment. The polypeptide fragments disclosed in Table 2 are believed to be antigenic regions of the *S. pneumoniae* polypeptides described in Table 1. Thus the invention further includes isolated peptides and polypeptides comprising an amino acid sequence of an epitope shown in Table 2 and polynucleotides encoding said polypeptides.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means for making peptides or polypeptides including recombinant means using nucleic acid molecules of the invention. For instance, an epitope-bearing amino acid sequence of the present invention may be fused to a larger polypeptide which acts as a carrier during recombinant production and purification, as well as during immunization to produce anti-peptide antibodies. Epitope-bearing peptides also may be synthesized using known methods of chemical synthesis. For instance, Houghten has described a simple method for synthesis of large numbers of peptides, such as 10-20 mg of 248 different 13 residue peptides representing single amino acid variants of a segment of the HA1 polypeptide which were prepared and characterized (by ELISA-type binding studies) in less than four weeks (Houghten, R. A. Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985)). This "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten and coworkers (1986). In this procedure the individual

resins for the solid-phase synthesis of various peptides are contained in separate solvent-permeable packets, enabling the optimal use of the many identical repetitive steps involved in solid-phase methods. A completely manual procedure allows 500-1000 or more syntheses to be conducted simultaneously (Houghten, *et al.*, *supra*, p. 5134).

Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art (for instance, Sutcliffe, *et al.*, *supra*; Wilson, *et al.*, *supra*; Chow, M., *et al.*, *Proc. Natl. Acad. Sci. USA* **82**:910-914; and Bittle, F. J., *et al.*, *J. Gen. Virol.* **66**:2347-2354 (1985)). Generally, animals may be immunized with free peptide; however, anti-peptide antibody titer may be boosted by coupling of the peptide to a macromolecular carrier, such as keyhole limpet hemacyanin (KLH) or tetanus toxoid. For instance, peptides containing cysteine may be coupled to carrier using a linker such as m-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS), while other peptides may be coupled to carrier using a more general linking agent such as glutaraldehyde. Animals such as rabbits, rats and mice are immunized with either free or carrier-coupled peptides, for instance, by intraperitoneal and/or intradermal injection of emulsions containing about 100 µg peptide or carrier protein and Freund's adjuvant. Several booster injections may be needed, for instance, at intervals of about two weeks, to provide a useful titer of anti-peptide antibody which can be detected, for example, by ELISA assay using free peptide adsorbed to a solid surface. The titer of anti-peptide antibodies in serum from an immunized animal may be increased by selection of anti-peptide antibodies, for instance, by adsorption to the peptide on a solid support and elution of the selected antibodies according to methods well known in the art.

Immunogenic epitope-bearing peptides of the invention, *i.e.*, those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. For instance, Geysen, *et al.*, *supra*, discloses a procedure for rapid concurrent synthesis on solid supports of hundreds of peptides of sufficient purity to react in an enzyme-linked immunosorbent assay. Interaction of synthesized peptides with antibodies is then easily detected without removing them from the support. In this manner a peptide bearing an immunogenic epitope of a desired protein may be identified routinely by one of ordinary skill in the art. For instance, the immunologically important epitope in the coat protein of foot-and-mouth disease virus was located by Geysen *et al. supra* with a resolution of seven amino acids by synthesis of an overlapping set of all 208 possible hexapeptides covering the

entire 213 amino acid sequence of the protein. Then, a complete replacement set of peptides in which all 20 amino acids were substituted in turn at every position within the epitope were synthesized, and the particular amino acids conferring specificity for the reaction with antibody were determined. Thus, peptide  
5 analogs of the epitope-bearing peptides of the invention can be made routinely by this method. U.S. Patent No. 4,708,781 to Geysen (1987) further describes this method of identifying a peptide bearing an immunogenic epitope of a desired protein.

Further still, U.S. Patent No. 5,194,392, to Geysen (1990), describes a  
10 general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (*i.e.*, a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092, also to Geysen (1989), describes a method of detecting or determining a  
15 sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. *et al.* (1996) discloses linear C<sub>1</sub>-C<sub>7</sub>-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries  
20 for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

The entire disclosure of each document cited in this section on  
25 "Polypeptides and Fragments" is hereby incorporated herein by reference.

As one of skill in the art will appreciate, the polypeptides of the present invention and the epitope-bearing fragments thereof described above can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification  
30 and show an increased half-life *in vivo*. This has been shown, *e.g.*, for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (EPA 0,394,827; Traunecker *et al.*, *Nature* 331:84-86 (1988)). Fusion proteins that have a disulfide-linked  
35 dimeric structure due to the IgG part can also be more efficient in binding and neutralizing other molecules than a monomeric *S. pneumoniae* polypeptide or

fragment thereof alone (Fountoulakis *et al.*, *J. Biochem.* 270:3958-3964 (1995)).

### ***Diagnostic Assays***

5           The present invention further relates to a method for assaying for *Streptococcal* infection in an animal *via* detecting the expression of genes encoding *Streptococcal* polypeptides (*e.g.*, the polypeptides described Table 1). This method comprises analyzing tissue or body fluid from the animal for *Streptococcus*-specific antibodies or *Streptococcal* nucleic acids or proteins.

10          Analysis of nucleic acid specific to *Streptococcus* can be done by PCR or hybridization techniques using nucleic acid sequences of the present invention as either hybridization probes or primers (*cf. Molecular Cloning: A Laboratory Manual, second edition*, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989; Ereemeeva *et al.*, *J. Clin. Microbiol.* 32:803-810

15          (1994) which describes differentiation among spotted fever group *Rickettsiae* species by analysis of restriction fragment length polymorphism of PCR-amplified DNA). Methods for detecting *B. burgdorferi* nucleic acids *via* PCR are described, for example, in Chen *et al.*, *J. Clin. Microbiol.* 32:589-595 (1994).

20          Where diagnosis of a disease state related to infection with *Streptococcus* has already been made, the present invention is useful for monitoring progression or regression of the disease state whereby patients exhibiting enhanced *Streptococcus* gene expression will experience a worse clinical outcome relative to patients expressing these gene(s) at a lower level.

25          By "assaying for *Streptococcal* infection in an animal *via* detection of genes encoding *Streptococcal* polypeptides" is intended qualitatively or quantitatively measuring or estimating the level of one or more *Streptococcus* polypeptides or the level of nucleic acid encoding *Streptococcus* polypeptides in a first biological sample either directly (*e.g.*, by determining or estimating

30          absolute protein level or nucleic level) or relatively (*e.g.*, by comparing to the *Streptococcus* polypeptide level or mRNA level in a second biological sample). The *Streptococcus* polypeptide level or nucleic acid level in the second sample used for a relative comparison may be undetectable if obtained from an animal which is not infected with *Streptococcus*. When monitoring the progression or

35          regression of a disease state, the *Streptococcus* polypeptide level or nucleic acid level may be compared to a second sample obtained from either an animal infected with *Streptococcus* or the same animal from which the first sample was obtained but taken from that animal at a different time than the first. As will be

appreciated in the art, once a standard *Streptococcus* polypeptide level or nucleic acid level which corresponds to a particular stage of a *Streptococcus* infection is known, it can be used repeatedly as a standard for comparison.

By "biological sample" is intended any biological sample obtained from an animal, cell line, tissue culture, or other source which contains *Streptococcus* polypeptide, mRNA, or DNA. Biological samples include body fluids (such as plasma and synovial fluid) which contain *Streptococcus* polypeptides, and muscle, skin, and cartilage tissues. Methods for obtaining tissue biopsies and body fluids are well known in the art.

The present invention is useful for detecting diseases related to *Streptococcus* infections in animals. Preferred animals include monkeys, apes, cats, dogs, cows, pigs, mice, horses, rabbits and humans. Particularly preferred are humans.

Total RNA can be isolated from a biological sample using any suitable technique such as the single-step guanidinium-thiocyanate-phenol-chloroform method described in Chomczynski and Sacchi, *Anal. Biochem.* 162:156-159 (1987). mRNA encoding *Streptococcus* polypeptides having sufficient homology to the nucleic acid sequences identified in Table 1 to allow for hybridization between complementary sequences are then assayed using any appropriate method. These include Northern blot analysis, S1 nuclease mapping, the polymerase chain reaction (PCR), reverse transcription in combination with the polymerase chain reaction (RT-PCR), and reverse transcription in combination with the ligase chain reaction (RT-LCR).

Northern blot analysis can be performed as described in Harada *et al.*, *Cell* 63:303-312 (1990). Briefly, total RNA is prepared from a biological sample as described above. For the Northern blot, the RNA is denatured in an appropriate buffer (such as glyoxal/dimethyl sulfoxide/sodium phosphate buffer), subjected to agarose gel electrophoresis, and transferred onto a nitrocellulose filter. After the RNAs have been linked to the filter by a UV linker, the filter is prehybridized in a solution containing formamide, SSC, Denhardt's solution, denatured salmon sperm, SDS, and sodium phosphate buffer. A *S. pneumoniae* polypeptide DNA sequence shown in Table 1 labeled according to any appropriate method (such as the <sup>32</sup>P-multiprimered DNA labeling system (Amersham)) is used as probe. After hybridization overnight, the filter is washed and exposed to x-ray film. DNA for use as probe according to the present invention is described in the sections above and will preferably at least 15 bp in length.

S1 mapping can be performed as described in Fujita *et al.*, *Cell* 49:357-367 (1987). To prepare probe DNA for use in S1 mapping, the sense strand of an above-described *S. pneumoniae* DNA sequence of the present invention is used as a template to synthesize labeled antisense DNA. The antisense DNA can then be digested using an appropriate restriction endonuclease to generate further DNA probes of a desired length. Such antisense probes are useful for visualizing protected bands corresponding to the target mRNA (*i.e.*, mRNA encoding *Streptococcus* polypeptides).

Preferably, levels of mRNA encoding *Streptococcus* polypeptides are assayed using the RT-PCR method described in Makino *et al.*, *Technique* 2:295-301 (1990). By this method, the radioactivities of the "amplicons" in the polyacrylamide gel bands are linearly related to the initial concentration of the target mRNA. Briefly, this method involves adding total RNA isolated from a biological sample in a reaction mixture containing a RT primer and appropriate buffer. After incubating for primer annealing, the mixture can be supplemented with a RT buffer, dNTPs, DTT, RNase inhibitor and reverse transcriptase. After incubation to achieve reverse transcription of the RNA, the RT products are then subject to PCR using labeled primers. Alternatively, rather than labeling the primers, a labeled dNTP can be included in the PCR reaction mixture. PCR amplification can be performed in a DNA thermal cycler according to conventional techniques. After a suitable number of rounds to achieve amplification, the PCR reaction mixture is electrophoresed on a polyacrylamide gel. After drying the gel, the radioactivity of the appropriate bands (corresponding to the mRNA encoding the *Streptococcus* polypeptides)) is quantified using an imaging analyzer. RT and PCR reaction ingredients and conditions, reagent and gel concentrations, and labeling methods are well known in the art. Variations on the RT-PCR method will be apparent to the skilled artisan.

Assaying *Streptococcus* polypeptide levels in a biological sample can occur using any art-known method. Preferred for assaying *Streptococcus* polypeptide levels in a biological sample are antibody-based techniques. For example, *Streptococcus* polypeptide expression in tissues can be studied with classical immunohistological methods. In these, the specific recognition is provided by the primary antibody (polyclonal or monoclonal) but the secondary detection system can utilize fluorescent, enzyme, or other conjugated secondary antibodies. As a result, an immunohistological staining of tissue section for pathological examination is obtained. Tissues can also be extracted, *e.g.*, with urea and neutral detergent, for the liberation of *Streptococcus* polypeptides for

Western-blot or dot/slot assay (Jalkanen, M., *et al.*, *J. Cell. Biol.* 101:976-985 (1985); Jalkanen, M., *et al.*, *J. Cell. Biol.* 105:3087-3096 (1987)). In this technique, which is based on the use of cationic solid phases, quantitation of a *Streptococcus* polypeptide can be accomplished using an isolated *Streptococcus* polypeptide as a standard. This technique can also be applied to body fluids.

Other antibody-based methods useful for detecting *Streptococcus* polypeptide gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). For example, a *Streptococcus* polypeptide-specific monoclonal antibodies can be used both as an immunoabsorbent and as an enzyme-labeled probe to detect and quantify a *Streptococcus* polypeptide. The amount of a *Streptococcus* polypeptide present in the sample can be calculated by reference to the amount present in a standard preparation using a linear regression computer algorithm. Such an ELISA for detecting a tumor antigen is described in Iacobelli *et al.*, *Breast Cancer Research and Treatment* 11:19-30 (1988). In another ELISA assay, two distinct specific monoclonal antibodies can be used to detect *Streptococcus* polypeptides in a body fluid. In this assay, one of the antibodies is used as the immunoabsorbent and the other as the enzyme-labeled probe.

The above techniques may be conducted essentially as a "one-step" or "two-step" assay. The "one-step" assay involves contacting the *Streptococcus* polypeptide with immobilized antibody and, without washing, contacting the mixture with the labeled antibody. The "two-step" assay involves washing before contacting the mixture with the labeled antibody. Other conventional methods may also be employed as suitable. It is usually desirable to immobilize one component of the assay system on a support, thereby allowing other components of the system to be brought into contact with the component and readily removed from the sample.

*Streptococcus* polypeptide-specific antibodies for use in the present invention can be raised against an intact *S. pneumoize* polypeptide of the present invention or fragment thereof. These polypeptides and fragments may be administered to an animal (*e.g.*, rabbit or mouse) either with a carrier protein (*e.g.*, albumin) or, if long enough (*e.g.*, at least about 25 amino acids), without a carrier.

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')<sub>2</sub> fragments) which are capable of specifically binding to a *Streptococcus* polypeptide. Fab and F(ab')<sub>2</sub> fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may

have less non-specific tissue binding of an intact antibody (Wahl *et al.*, *J. Nucl. Med.* 24:316-325 (1983)). Thus, these fragments are preferred.

The antibodies of the present invention may be prepared by any of a variety of methods. For example, the *S. pneumoniae* polypeptides identified in Table 1, or fragments thereof, can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of a *S. pneumoniae* polypeptide of the present invention is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of high specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies. Such monoclonal antibodies can be prepared using hybridoma technology (Kohler *et al.*, *Nature* 256:495 (1975); Kohler *et al.*, *Eur. J. Immunol.* 6:511 (1976); Kohler *et al.*, *Eur. J. Immunol.* 6:292 (1976); Hammerling *et al.*, In: *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y., (1981) pp. 563-681 ). In general, such procedures involve immunizing an animal (preferably a mouse) with a *S. pneumoniae* polypeptide antigen of the present invention. Suitable cells can be recognized by their capacity to bind anti-*Streptococcus* polypeptide antibody. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP<sub>2</sub>O), available from the American Type Culture Collection, Rockville, Maryland. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.* (*Gastroenterology* 80:225-232 (1981)). The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the *Streptococcus* polypeptide antigen administered to immunized animal.

Alternatively, additional antibodies capable of binding to *Streptococcus* polypeptide antigens may be produced in a two-step procedure through the use of anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and that, therefore, it is possible to obtain an antibody

which binds to a second antibody. In accordance with this method, *Streptococcus* polypeptide-specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the *Streptococcus* polypeptide-specific antibody can be blocked by a *Streptococcus* polypeptide antigen. Such antibodies comprise anti-idiotypic antibodies to the *Streptococcus* polypeptide-specific antibody and can be used to immunize an animal to induce formation of further *Streptococcus* polypeptide-specific antibodies.

It will be appreciated that Fab and  $F(ab')_2$  and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce  $F(ab')_2$  fragments). Alternatively, *Streptococcus* polypeptide-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

Of special interest to the present invention are antibodies to *Streptococcus* polypeptide antigens which are produced in humans, or are "humanized" (*i.e.*, non-immunogenic in a human) by recombinant or other technology. Humanized antibodies may be produced, for example by replacing an immunogenic portion of an antibody with a corresponding, but non-immunogenic portion (*i.e.*, chimeric antibodies) (Robinson, R.R. *et al.*, International Patent Publication PCT/US86/02269; Akira, K. *et al.*, European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison, S.L. *et al.*, European Patent Application 173,494; Neuberger, M.S. *et al.*, PCT Application WO 86/01533; Cabilly, S. *et al.*, European Patent Application 125,023; Better, M. *et al.*, *Science* 240:1041-1043 (1988); Liu, A.Y. *et al.*, *Proc. Natl. Acad. Sci. USA* 84:3439-3443 (1987); Liu, A.Y. *et al.*, *J. Immunol.* 139:3521-3526 (1987); Sun, L.K. *et al.*, *Proc. Natl. Acad. Sci. USA* 84:214-218 (1987); Nishimura, Y. *et al.*, *Canc. Res.* 47:999-1005 (1987); Wood, C.R. *et al.*, *Nature* 314:446-449 (1985); Shaw *et al.*, *J. Natl. Cancer Inst.* 80:1553-1559 (1988). General reviews of "humanized" chimeric antibodies are provided by Morrison, S.L. (*Science*, 229:1202-1207 (1985)) and by Oi, V.T. *et al.*, *BioTechniques* 4:214 (1986)). Suitable "humanized" antibodies can be alternatively produced by CDR or CEA substitution (Jones, P.T. *et al.*, *Nature* 321:552-525 (1986);

Verhoeyan *et al.*, *Science* 239:1534 (1988); Beidler, C.B. *et al.*, *J. Immunol.* 141:4053-4060 (1988)).

Suitable enzyme labels include, for example, those from the oxidase group, which catalyze the production of hydrogen peroxide by reacting with substrate. Glucose oxidase is particularly preferred as it has good stability and its substrate (glucose) is readily available. Activity of an oxidase label may be assayed by measuring the concentration of hydrogen peroxide formed by the enzyme-labeled antibody/substrate reaction. Besides enzymes, other suitable labels include radioisotopes, such as iodine ( $^{125}\text{I}$ ,  $^{121}\text{I}$ ), carbon ( $^{14}\text{C}$ ), sulphur ( $^{35}\text{S}$ ), tritium ( $^3\text{H}$ ), indium ( $^{112}\text{In}$ ), and technetium ( $^{99\text{m}}\text{Tc}$ ), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

Further suitable labels for the *Streptococcus* polypeptide-specific antibodies of the present invention are provided below. Examples of suitable enzyme labels include malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast-alcohol dehydrogenase, alpha-glycerol phosphate dehydrogenase, triose phosphate isomerase, peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase, and acetylcholine esterase.

Examples of suitable radioisotopic labels include  $^3\text{H}$ ,  $^{111}\text{In}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ ,  $^{51}\text{Cr}$ ,  $^{57}\text{Co}$ ,  $^{58}\text{Co}$ ,  $^{59}\text{Fe}$ ,  $^{75}\text{Se}$ ,  $^{152}\text{Eu}$ ,  $^{90}\text{Y}$ ,  $^{67}\text{Cu}$ ,  $^{217}\text{Bi}$ ,  $^{211}\text{At}$ ,  $^{212}\text{Pb}$ ,  $^{47}\text{Sc}$ ,  $^{109}\text{Pd}$ , etc.  $^{111}\text{In}$  is a preferred isotope where *in vivo* imaging is used since it avoids the problem of dehalogenation of the  $^{125}\text{I}$  or  $^{131}\text{I}$ -labeled monoclonal antibody by the liver. In addition, this radionuclide has a more favorable gamma emission energy for imaging (Perkins *et al.*, *Eur. J. Nucl. Med.* 10:296-301 (1985); Carasquillo *et al.*, *J. Nucl. Med.* 28:281-287 (1987)). For example,  $^{111}\text{In}$  coupled to monoclonal antibodies with 1-(P-isothiocyanatobenzyl)-DPTA has shown little uptake in non-tumorous tissues, particularly the liver, and therefore enhances specificity of tumor localization (Esteban *et al.*, *J. Nucl. Med.* 28:861-870 (1987)).

Examples of suitable non-radioactive isotopic labels include  $^{157}\text{Gd}$ ,  $^{55}\text{Mn}$ ,  $^{162}\text{Dy}$ ,  $^{52}\text{Tr}$ , and  $^{56}\text{Fe}$ .

Examples of suitable fluorescent labels include an  $^{152}\text{Eu}$  label, a fluorescein label, an isothiocyanate label, a rhodamine label, a phycoerythrin label, a phycocyanin label, an allophycocyanin label, an o-phthaldehyde label, and a fluorescamine label.

Examples of suitable toxin labels include diphtheria toxin, ricin, and cholera toxin.

Examples of chemiluminescent labels include a luminal label, an isoluminal label, an aromatic acridinium ester label, an imidazole label, an acridinium salt label, an oxalate ester label, a luciferin label, a luciferase label, and an aequorin label.

5        Examples of nuclear magnetic resonance contrasting agents include heavy metal nuclei such as Gd, Mn, and iron.

Typical techniques for binding the above-described labels to antibodies are provided by Kennedy *et al.*, *Clin. Chim. Acta* 70:1-31 (1976), and Schurs *et al.*, *Clin. Chim. Acta* 81:1-40 (1977). Coupling techniques mentioned in the  
10        latter are the glutaraldehyde method, the periodate method, the dimaleimide method, the m-maleimidobenzyl-N-hydroxy-succinimide ester method, all of which methods are incorporated by reference herein.

In a related aspect, the invention includes a diagnostic kit for use in screening serum containing antibodies specific against *S. pneumoniae*  
15        infection. Such a kit may include an isolated *S. pneumoniae* antigen comprising an epitope which is specifically immunoreactive with at least one anti-*S. pneumoniae* antibody. Such a kit also includes means for detecting the binding of said antibody to the antigen. In specific embodiments, the kit may include a recombinantly produced or chemically synthesized peptide or polypeptide  
20        antigen. The peptide or polypeptide antigen may be attached to a solid support.

In a more specific embodiment, the detecting means of the above-described kit includes a solid support to which said peptide or polypeptide antigen is attached. Such a kit may also include a non-attached reporter-labelled anti-human antibody. In this embodiment, binding of the antibody to the *S.*  
25        *pneumoniae* antigen can be detected by binding of the reporter labelled antibody to the anti-*S. pneumoniae* antibody.

In a related aspect, the invention includes a method of detecting *S. pneumoniae* infection in a subject. This detection method includes reacting a body fluid, preferably serum, from the subject with an isolated *S. pneumoniae*  
30        antigen, and examining the antigen for the presence of bound antibody. In a specific embodiment, the method includes a polypeptide antigen attached to a solid support, and serum is reacted with the support. Subsequently, the support is reacted with a reporter-labelled anti-human antibody. The support is then examined for the presence of reporter-labelled antibody.

35        The solid surface reagent employed in the above assays and kits is prepared by known techniques for attaching protein material to solid support material, such as polymeric beads, dip sticks, 96-well plates or filter material. These attachment methods generally include non-specific adsorption of the

protein to the support or covalent attachment of the protein, typically through a free amine group, to a chemically reactive group on the solid support, such as an activated carboxyl, hydroxyl, or aldehyde group. Alternatively, streptavidin coated plates can be used in conjunction with biotinylated antigen(s).

### *Therapeutics and Modes of Administration*

The present invention also provides vaccines comprising one or more polypeptides of the present invention. Heterogeneity in the composition of a vaccine may be provided by combining *S. pneumoniae* polypeptides of the present invention. Multi-component vaccines of this type are desirable because they are likely to be more effective in eliciting protective immune responses against multiple species and strains of the *Streptococcus* genus than single polypeptide vaccines. Thus, as discussed in detail below, a multi-component vaccine of the present invention may contain one or more, preferably 2 to about 20, more preferably 2 to about 15, and most preferably 3 to about 8, of the *S. pneumoniae* polypeptides identified in Table 1, or fragments thereof.

Multi-component vaccines are known in the art to elicit antibody production to numerous immunogenic components. Decker, M. and Edwards, K., *J. Infect. Dis.* 174:S270-275 (1996). In addition, a hepatitis B, diphtheria, tetanus, pertussis tetravalent vaccine has recently been demonstrated to elicit protective levels of antibodies in human infants against all four pathogenic agents. Aristegui, J. *et al.*, *Vaccine* 15:7-9 (1997).

The present invention thus also includes multi-component vaccines. These vaccines comprise more than one polypeptide, immunogen or antigen. An example of such a multi-component vaccine would be a vaccine comprising more than one of the *S. pneumoniae* polypeptides described in Table 1. A second example is a vaccine comprising one or more, for example 2 to 10, of the *S. pneumoniae* polypeptides identified in Table 1 and one or more, for example 2 to 10, additional polypeptides of either streptococcal or non-streptococcal origin. Thus, a multi-component vaccine which confers protective immunity to both a Streptococcal infection and infection by another pathogenic agent is also within the scope of the invention.

As indicated above, the vaccines of the present invention are expected to elicit a protective immune response against infections caused by species and strains of *Streptococcus* other than strain of *S. pneumoniae* deposited with that ATCC.

Further within the scope of the invention are whole cell and whole viral vaccines. Such vaccines may be produced recombinantly and involve the

expression of one or more of the *S. pneumoniae* polypeptides described in Table 1. For example, the *S. pneumoniae* polypeptides of the present invention may be either secreted or localized intracellular, on the cell surface, or in the periplasmic space. Further, when a recombinant virus is used, the *S. pneumoniae* polypeptides of the present invention may, for example, be localized in the viral envelope, on the surface of the capsid, or internally within the capsid. Whole cells vaccines which employ cells expressing heterologous proteins are known in the art. See, e.g., Robinson, K. *et al.*, *Nature Biotech.* 15:653-657 (1997); Sirard, J. *et al.*, *Infect. Immun.* 65:2029-2033 (1997); Chabalgoity, J. *et al.*, *Infect. Immun.* 65:2402-2412 (1997). These cells may be administered live or may be killed prior to administration. Chabalgoity, J. *et al.*, *supra*, for example, report the successful use in mice of a live attenuated *Salmonella* vaccine strain which expresses a portion of a platyhelminth fatty acid-binding protein as a fusion protein on its cells surface.

A multi-component vaccine can also be prepared using techniques known in the art by combining one or more *S. pneumoniae* polypeptides of the present invention, or fragments thereof, with additional non-streptococcal components (e.g., diphtheria toxin or tetanus toxin, and/or other compounds known to elicit an immune response). Such vaccines are useful for eliciting protective immune responses to both members of the *Streptococcus* genus and non-streptococcal pathogenic agents.

The vaccines of the present invention also include DNA vaccines. DNA vaccines are currently being developed for a number of infectious diseases. Boyer, J *et al.*, *Nat. Med.* 3:526-532 (1997); reviewed in Spier, R., *Vaccine* 14:1285-1288 (1996). Such DNA vaccines contain a nucleotide sequence encoding one or more *S. pneumoniae* polypeptides of the present invention oriented in a manner that allows for expression of the subject polypeptide. The direct administration of plasmid DNA encoding *B. burgdorferi* OspA has been shown to elicit protective immunity in mice against borrelial challenge. Luke, C. *et al.*, *J. Infect. Dis.* 175:91-97 (1997).

The present invention also relates to the administration of a vaccine which is co-administered with a molecule capable of modulating immune responses. Kim, J. *et al.*, *Nature Biotech.* 15:641-646 (1997), for example, report the enhancement of immune responses produced by DNA immunizations when DNA sequences encoding molecules which stimulate the immune response are co-administered. In a similar fashion, the vaccines of the present invention may be co-administered with either nucleic acids encoding immune modulators or the immune modulators themselves. These immune modulators

include granulocyte macrophage colony stimulating factor (GM-CSF) and CD86.

The vaccines of the present invention may be used to confer resistance to streptococcal infection by either passive or active immunization. When the vaccines of the present invention are used to confer resistance to streptococcal infection through active immunization, a vaccine of the present invention is administered to an animal to elicit a protective immune response which either prevents or attenuates a streptococcal infection. When the vaccines of the present invention are used to confer resistance to streptococcal infection through passive immunization, the vaccine is provided to a host animal (*e.g.*, human, dog, or mouse), and the antisera elicited by this antisera is recovered and directly provided to a recipient suspected of having an infection caused by a member of the *Streptococcus* genus.

The ability to label antibodies, or fragments of antibodies, with toxin molecules provides an additional method for treating streptococcal infections when passive immunization is conducted. In this embodiment, antibodies, or fragments of antibodies, capable of recognizing the *S. pneumoniae* polypeptides disclosed herein, or fragments thereof, as well as other *Streptococcus* proteins, are labeled with toxin molecules prior to their administration to the patient. When such toxin derivatized antibodies bind to *Streptococcus* cells, toxin moieties will be localized to these cells and will cause their death.

The present invention thus concerns and provides a means for preventing or attenuating a streptococcal infection resulting from organisms which have antigens that are recognized and bound by antisera produced in response to the polypeptides of the present invention. As used herein, a vaccine is said to prevent or attenuate a disease if its administration to an animal results either in the total or partial attenuation (*i.e.*, suppression) of a symptom or condition of the disease, or in the total or partial immunity of the animal to the disease.

The administration of the vaccine (or the antisera which it elicits) may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the compound(s) are provided in advance of any symptoms of streptococcal infection. The prophylactic administration of the compound(s) serves to prevent or attenuate any subsequent infection. When provided therapeutically, the compound(s) is provided upon or after the detection of symptoms which indicate that an animal may be infected with a member of the *Streptococcus* genus. The therapeutic administration of the compound(s) serves to attenuate any actual infection. Thus, the *S. pneumoniae* polypeptides, and

fragments thereof, of the present invention may be provided either prior to the onset of infection (so as to prevent or attenuate an anticipated infection) or after the initiation of an actual infection.

5 The polypeptides of the invention, whether encoding a portion of a native protein or a functional derivative thereof, may be administered in pure form or may be coupled to a macromolecular carrier. Example of such carriers are proteins and carbohydrates. Suitable proteins which may act as macromolecular carrier for enhancing the immunogenicity of the polypeptides of the present invention include keyhole limpet hemacyanin (KLH) tetanus toxoid, pertussis toxin, bovine serum albumin, and ovalbumin. Methods for coupling the polypeptides of the present invention to such macromolecular carriers are disclosed in Harlow *et al.*, *Antibodies: A Laboratory Manual*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988), the entire disclosure of which is incorporated by reference herein.

15 A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient animal and is otherwise suitable for administration to that animal. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

20 While in all instances the vaccine of the present invention is administered as a pharmacologically acceptable compound, one skilled in the art would recognize that the composition of a pharmacologically acceptable compound varies with the animal to which it is administered. For example, a vaccine intended for human use will generally not be co-administered with Freund's adjuvant. Further, the level of purity of the *S. pneumoniae* polypeptides of the present invention will normally be higher when administered to a human than when administered to a non-human animal.

25 As would be understood by one of ordinary skill in the art, when the vaccine of the present invention is provided to an animal, it may be in a composition which may contain salts, buffers, adjuvants, or other substances which are desirable for improving the efficacy of the composition. Adjuvants are substances that can be used to specifically augment a specific immune response. These substances generally perform two functions: (1) they protect the antigen(s) from being rapidly catabolized after administration and (2) they nonspecifically stimulate immune responses.

30 Normally, the adjuvant and the composition are mixed prior to presentation to the immune system, or presented separately, but into the same

40

site of the animal being immunized. Adjuvants can be loosely divided into several groups based upon their composition. These groups include oil adjuvants (for example, Freund's complete and incomplete), mineral salts (for example,  $\text{AlK}(\text{SO}_4)_2$ ,  $\text{AlNa}(\text{SO}_4)_2$ ,  $\text{AlNH}_4(\text{SO}_4)$ , silica, kaolin, and carbon),  
5 polynucleotides (for example, poly IC and poly AU acids), and certain natural substances (for example, wax D from *Mycobacterium tuberculosis*, as well as substances found in *Corynebacterium parvum*, or *Bordetella pertussis*, and members of the genus *Brucella*. Other substances useful as adjuvants are the saponins such as, for example, Quil A. (Superfos A/S, Denmark). Preferred  
10 adjuvants for use in the present invention include aluminum salts, such as  $\text{AlK}(\text{SO}_4)_2$ ,  $\text{AlNa}(\text{SO}_4)_2$ , and  $\text{AlNH}_4(\text{SO}_4)$ . Examples of materials suitable for use in vaccine compositions are provided in *Remington's Pharmaceutical Sciences* (Osol, A, Ed, Mack Publishing Co, Easton, PA, pp. 1324-1341 (1980), which reference is incorporated herein by reference).

15 The therapeutic compositions of the present invention can be administered parenterally by injection, rapid infusion, nasopharyngeal absorption (intranasopharangeally), dermoabsorption, or orally. The compositions may alternatively be administered intramuscularly, or intravenously. Compositions for parenteral administration include sterile  
20 aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Carriers or occlusive dressings can be used to increase skin permeability and enhance antigen absorption. Liquid dosage forms for oral administration may generally  
25 comprise a liposome solution containing the liquid dosage form. Suitable forms for suspending liposomes include emulsions, suspensions, solutions, syrups, and elixirs containing inert diluents commonly used in the art, such as purified water. Besides the inert diluents, such compositions can also include adjuvants, wetting agents, emulsifying and suspending agents, or sweetening, flavoring,  
30 or perfuming agents.

Therapeutic compositions of the present invention can also be administered in encapsulated form. For example, intranasal immunization of mice against *Bordetella pertussis* infection using vaccines encapsulated in biodegradable microsphere composed of poly(DL-lactide-co-glycolide) has been  
35 shown to stimulate protective immune responses. Shahin, R. *et al.*, *Infect. Immun.* 63:1195-1200 (1995). Similarly, orally administered encapsulated *Salmonella typhimurium* antigens have also been shown to elicit protective

immunity in mice. Allaoui-Attarki, K. *et al.*, *Infect. Immun.* 65:853-857 (1997). Encapsulated vaccines of the present invention can be administered by a variety of routes including those involving contacting the vaccine with mucous membranes (*e.g.*, intranasally, intracolonicly, intraduodenally).

Many different techniques exist for the timing of the immunizations when a multiple administration regimen is utilized. It is possible to use the compositions of the invention more than once to increase the levels and diversities of expression of the immunoglobulin repertoire expressed by the immunized animal. Typically, if multiple immunizations are given, they will be given one to two months apart.

According to the present invention, an "effective amount" of a therapeutic composition is one which is sufficient to achieve a desired biological effect. Generally, the dosage needed to provide an effective amount of the composition will vary depending upon such factors as the animal's or human's age, condition, sex, and extent of disease, if any, and other variables which can be adjusted by one of ordinary skill in the art.

The antigenic preparations of the invention can be administered by either single or multiple dosages of an effective amount. Effective amounts of the compositions of the invention can vary from 0.01-1,000  $\mu\text{g/ml}$  per dose, more preferably 0.1-500  $\mu\text{g/ml}$  per dose, and most preferably 10-300  $\mu\text{g/ml}$  per dose.

Having now generally described the invention, the same will be more readily understood through reference to the following example which is provided by way of illustration, and is not intended to be limiting of the present invention, unless specified.

### **Examples**

#### ***Example 1: Expression and Purification of S. pneumoniae Polypeptides in E. coli***

The bacterial expression vector pQE10 (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311) is used in this example for cloning of the nucleotide sequences shown in Table 1 and for expressing the polypeptides identified in Table 1. The components of the pQE10 plasmid are arranged such that the inserted DNA sequence encoding a polypeptide of the present invention expresses the polypeptide with the six His residues (*i.e.*, a "6 X His tag") covalently linked to the amino terminus.

The DNA sequences encoding the desired portions of the polypeptides of Table 1 are amplified using PCR oligonucleotide primers from either a DNA

library constructed from *S. pneumoniae*, such as the one deposited by the inventors at the ATCC for convenience, ATCC Deposit No. 97755, or from DNA isolated from the same organism such as the *S. pneumoniae* strain deposited with the ATCC as Deposit No. 55840. A list of PCR primers which can be used for this purpose is provided in Table 3, below. The PCR primers anneal to the nucleotide sequences encoding both the amino terminal and carboxy terminal amino acid sequences of the desired portion of the polypeptides of Table 1. Additional nucleotides containing restriction sites to facilitate cloning in the pQE10 vector were added to the 5' and 3' primer sequences, respectively. Such restriction sites are listed in Table 3 for each primer. In each case, the primer comprises, from the 5' end, 4 random nucleotides to prevent "breathing" during the annealing process, a restriction site (shown in Table 3), and approximately 15 nucleotides of *S. pneumoniae* ORF sequence (the complete sequence of each cloning primer is shown as SEQ ID NO:227 through SEQ ID NO:452).

For cloning the polypeptides of Table 1, the 5' and 3' primers were selected to amplify their respective nucleotide coding sequences. One of ordinary skill in the art would appreciate that the point in the protein coding sequence where the 5' primer begins may be varied to amplify a DNA segment encoding any desired portion of the complete amino acid sequences described in Table 1. Similarly, one of ordinary skill in the art would further appreciate that the point in the protein coding sequence where the 3' primer begins may also be varied to amplify a DNA segment encoding any desired portion of the complete amino acid sequences described in Table 1.

The amplified DNA fragment and the pQE10 vector are digested with the appropriate restriction enzyme(s) and the digested DNAs are then ligated together. The ligation mixture is transformed into competent *E. coli* cells using standard procedures such as those described in Sambrook *et al.*, *Molecular Cloning: a Laboratory Manual, 2nd Ed.*; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Transformants are identified by their ability to grow under selective pressure on LB plates. Plasmid DNA is isolated from resistant colonies and the identity of the cloned DNA confirmed by restriction analysis, PCR and DNA sequencing.

Clones containing the desired constructs are grown overnight ("O/N") in liquid culture under selection. The O/N culture is used to inoculate a large culture, at a dilution of approximately 1:25 to 1:250. The cells are grown to an optical density at 600 nm ("OD600") of between 0.4 and 0.6. Isopropyl-b-D-thiogalactopyranoside ("IPTG") is then added to a final concentration of 1 mM

to induce transcription from the *lac* repressor sensitive promoter, by inactivating the *lacI* repressor. Cells subsequently are incubated further for 3 to 4 hours. Cells are then harvested by centrifugation.

The cells are stirred for 3-4 hours at 4 C in 6M guanidine-HCl, pH 8. The cell debris is removed by centrifugation, and the supernatant containing the protein of interest is loaded onto a nickel-nitrilo-tri-acetic acid ("NiNTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6x His tag bind to the NI-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist, 1995, QIAGEN, Inc., *supra*). Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH8, then washed with 10 volumes of 6 M guanidine-HCl pH6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.0.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins can be eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

The DNA sequences encoding the amino acid sequences of Table 1 may also be cloned and expressed as fusion proteins by a protocol similar to that described directly above, wherein the pET-32b(+) vector (Novagen, 601 Science Drive, Madison, WI 53711) is preferentially used in place of pQE10.

Each of the polynucleotides shown in Table 1, was successfully amplified and subcloned into pQE10 as described above using the PCR primers shown in Table 3. These pQE10 plasmids containing the DNAs of Table 1, except SP023, SP042, SP054, SP063, SP081, SP092, SP114, SP122, SP123, SP126, and SP127, were deposited with the ATCC as a pooled deposit as a convenience to those of skill in the art. This pooled deposit was deposited on October 16, 1997 and given ATCC Deposit No. 209369. Those of ordinary skill in the art appreciate that isolating an individual plasmid from the pooled deposit is trivial provided the information and reagents described herein. Each of the deposited clones is capable of expressing its encoded *S. pneumoniae* polypeptide.

**Example 2: Immunization and Detection of Immune Responses****Methods****Growth of bacterial inoculum, immunization of Mice and Challenge with *S. pneumoniae*.**

Propagation and storage of, and challenge by *S. pneumoniae* are preformed essentially as described in Aaberge, I.S. et al., Virulence of *Streptococcus pneumoniae* in mice: a standardized method for preparation and frozen storage of the experimental bacterial inoculum, *Microbial Pathogenesis*, 18:141 (1995), incorporated herein by reference.

Briefly, Todd Hewitt (TH) broth (Difco laboratories, Detroit, MI) with 17% FCS, and horse blood agar plates are used for culturing the bacteria. Both broth and blood plates are incubated at 37°C in a 5% CO<sub>2</sub> atmosphere. Blood plates are incubated for 18 hr. The culture broth is regularly 10-fold serially diluted in TH broth kept at room temperature and bacterial suspensions are kept at room temperature until challenge of mice.

For active immunizations C3H/HeJ mice (The Jackson Laboratory, Bar Harbor, ME) are injected intraperitoneally (i.p.) at week 0 with 20 g of recombinant streptococcal protein, or phosphate-buffered saline (PBS), emulsified with complete Freund's adjuvant (CFA), given a similar booster immunization in incomplete Freund's adjuvant (IFA) at week 4, and challenged at week 6. For challenge *S. pneumoniae* are diluted in TH broth from exponentially-growing cultures and mice are injected subcutaneously (s.c.) at the base of the tail with 0.1 ml of these dilutions (serial dilutions are used to find medium infectious dose). Streptococci used for challenge are passaged fewer than six times *in vitro*. To assess infection, blood samples are obtained from the distal part of the lateral femoral vein into heparinized capillary tubes. A 25 ul blood sample is serially 10-fold diluted in TH broth, and 25 ul of diluted and undiluted blood is plated onto blood agar plates. The plates are incubated for 18 hr. and colonies are counted.

Other methods are known in the art, for example, see Langermann, S. et al., *J. Exp. Med.*, 180:2277 (1994), incorporated herein by reference.

### *Immunoassays*

Several immunoassay formats are used to quantify levels of streptococcal-specific antibodies (ELISA and immunoblot), and to evaluate the functional properties of these antibodies (growth inhibition assay). The ELISA and immunoblot assays are also used to detect and quantify antibodies elicited in response to streptococcal infection that react with specific streptococcal antigens. Where antibodies to certain streptococcal antigens are elicited by infection this is taken as evidence that the streptococcal proteins in question are expressed *in vivo*. Absence of infection-derived antibodies (seroconversion) following streptococcal challenge is evidence that infection is prevented or suppressed. The immunoblot assay is also used to ascertain whether antibodies raised against recombinant streptococcal antigens recognize a protein of similar size in extracts of whole streptococci. Where the natural protein is of similar, or identical, size in the immunoblot assay to the recombinant version of the same protein, this is taken as evidence that the recombinant protein is the product of a full-length clone of the respective gene.

### *Enzyme-Linked Immunosorbant Assay (ELISA).*

The ELISA is used to quantify levels of antibodies reactive with streptococcus antigens elicited in response to immunization with these streptococcal antigens. Wells of 96 well microtiter plates (Immunlon 4, Dynatech, Chantilly, Virginia, or equivalent) are coated with antigen by incubating 50  $\mu$ l of 1  $\mu$ g/ml protein antigen solution in a suitable buffer, typically 0.1 M sodium carbonate buffer at pH 9.6. After decanting unbound antigen, additional binding sites are blocked by incubating 100  $\mu$ l of 3% nonfat milk in wash buffer (PBS, 0.2% Tween 20, pH 7.4). After washing, duplicate serial two-fold dilutions of sera in PBS, Tween 20, 1% fetal bovine serum, are incubated for 1 hr, removed, wells are washed three times, and incubated with horseradish peroxidase-conjugated goat anti-mouse IgG. After three washes, bound antibodies are detected with H<sub>2</sub>O<sub>2</sub> and 2,2'-azino-di-(3-ethylbenzthiazoline sulfonate) (Schwan, T.G., *et al.*, *Proc. Natl. Acad. Sci. USA* 92:2909-2913 (1985)) (ABTS®, Kirkegaard & Perry Labs., Gaithersburg, MD) and A<sub>405</sub> is quantified with a Molecular Devices, Corp. (Menlo Park, California) Vmax™ plate reader. IgG levels twice the background level in serum from naive mice are assigned the minimum titer of 1:100.

***Sodiumdodecylsulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Immunoblotting***

Using a single well format, total streptococcal protein extracts or recombinant streptococcal antigen are boiled in SDS/2-ME sample buffer before electrophoresis through 3% acrylamide stacking gels, and resolving gels of higher acrylamide concentration, typically 10-15% acrylamide monomer. Gels are electro-blotted to nitrocellulose membranes and lanes are probed with dilutions of antibody to be tested for reactivity with specific streptococcal antigens, followed by the appropriate secondary antibody-enzyme (horseradish peroxidase) conjugate. When it is desirable to confirm that the protein had transferred following electro-blotting, membranes are stained with Ponceau S. Immunoblot signals from bound antibodies are detected on x-ray film as chemiluminescence using ECL™ reagents (Amersham Corp., Arlington Heights, Illinois).

***Example 3: Detection of Streptococcus mRNA expression***

Northern blot analysis is carried out using methods described by, among others, Sambrook *et al.*, *supra*, to detect the expression of the *S. pneumoniae* nucleotide sequences of the present invention in animal tissues. A cDNA probe containing an entire nucleotide sequence shown in Table 1 is labeled with <sup>32</sup>P using the *rediprime*™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using a CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to detect the expression of *Streptococcus* mRNA in an animal tissue sample.

Animal tissues, such as blood or spinal fluid, are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70 C overnight, and films developed according to standard procedures.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of all publications (including patents, patent applications, journal articles, laboratory manuals, books, or other documents) cited herein are hereby incorporated by reference.

Table 1

**SP001 nucleotide (SEQ ID NO:1)**

TAAATCTACGACAATAAAATCAACTCATTGCTGACTTGGGTTCTGAACGCCGCGTCAATGCCCAAGC  
 TAATGATATTCCACAGATTGGTTAAGGCAATCGTTTCTATCGAAGACCATCGCTTCTTCGACCACAG  
 GGGGATTGATACCATCCGTATCCTGGGAGCTTTCTTGCAGCAATCTGCAAAGCAATTCCCTCCAAGGTGG  
 ATCAACTCTCACCCAACAGTTGATTAGTTGACTTACTTTTCAACTTCGACTTCCGACCAGACTATTTTC  
 TCGTAAGGCTCAGGAAGCTTGGTTAGCGATTAGTTAGAACAAAAAGCAACCAAGCAAGAAATCTTGAC  
 CTACTATATAAATAAGGTTACATGTCTAATGGGAACATATGGAAATGCAGACAGCAGCTCAAAACTACTA  
 TGGTAAAGACCTCAATAATTTAAGTTTACCTCAGTTAGCCTTGCTGGCTGGAATGCCTCAGGCACCAAA  
 CCAATATGACCCCTATTACATCCAGAAGCAGCCCAAGACCGCCGAAACTTGGTCTTATCTGAAATGAA  
 AAATCAAGGCTACATCTCTGCTGAACAGTATGAGAAAGCAGTCAATACACCAATTACTGATGGACTACA  
 AAGTCTCAAATCAGCAAGTAATTACCCTGCTTACATGGATAATTACCTCAAGGAAGTCATCAATCAAGT  
 TGAAGAAGAAACAGGCTATAACCTACTCACAACCTGGGATGGATGTCTACACAAATGTAGACCAAGAAGC  
 TCAAAAACATCTGTGGGATATTTACAATACAGACGAATACGTTGCCTATCCAGACGATGAATTGCAAGT  
 CGCTTCTACCATTGTTGATGTTTCTAACGGTAAAGTCATTGCCAGCTAGGAGCAGCCATCAGTCAAG  
 TAATGTTTCTTCCGAATTAACCAAGCAGTAGAAACAAACCGCGACTGGGGATCAACTATGAAACCGAT  
 CACAGACTATGCTCCTGCCTTGGAGTACGGTGTCTACGATTCAACTGCTACTATCGTTCACGATGAGCC  
 CTATAACTACCTGGGACAAATACTCCTGTTTATACTGGGATAGGGGCTACTTTGGCAACATCACCTT  
 GCAATACGCCCTGCAACAATCGCGAAACGTCCCAGCCGTGGAACTCTAAACAAGGTCCGACTCAACCG  
 CGCCAAGACTTTCCTAAATGGTCTAGGAATCGACTACCCAAGTATTCACTACTCAAATGCCATTTCAAG  
 TAACACAACCGAATCAGACAAAAATATGGAGCAAGTAGTGAAAAGATGGCTGCTGCTTACGCTGCCTT  
 TGCAAATGGTGGAATCTACTATAAACCAATGTATATCCATAAAGTCGCTCTTAGTGATGGGAGTGAAAA  
 AGAGTTCTCTAATGTCGGAACCTCGTGCCATGAAGGAAACGACAGCCTATATGATGACCGACATGATGAA  
 AACAGTCTTGACTTATGGAACCTGGACGAAATGCCTATCTTGCTTGGCTCCCTCAGGCTGGTAAAAACAGG  
 AACCTCTAACTATACAGACGAGGAAATTGAAACCACATCAAGACCTCTCAATTTGTAGCACCTGATGA  
 ACTATTTGCTGGCTATACGCGTAAATATTCAATGGCTGTATGGACAGGCTATTCTAACCGTCTGACACC  
 ACTTGTAGGCAATGGCCTTACGGTGCCTGCCAAAGTTTACCGCTCTATGATGACCTACCTGTCTGAAGG  
 AAGCAATCCAGAAGATTGGAATATACCAGAGGGGCTCTACAGAAATGGAGAATTCGTATTTAAAAATGG  
 TGCTCGTTCTACGTGGAATCACCTGCTCCACAACAACCCCATCAACTGAAAGTTCAAGCTCATCATC  
 AGATAGTTCAACTTCACAGTCTAGCTCAACCACTCCAAGCACAAATAATAGTACGACTACCAATCTTAA  
 CAATAATACGCAACAATCAAAATACAACCCCTGATCAACAAAATCAGAATCCTCAACCAGCACAAACA

**SP001 AMINO ACID (SEQ ID NO:2)**

KIYDNKNQLIADLGSERRVNAQANDIPTDLVKAIVSIEDHRFFDHRGIDTIRILGAFLRNLOQSNLSLQGG  
 STLTLQQLIKLTYFSTSTSDQTI SRKAQEAWLAIQLEQKATKQEI LTYIYINKVYMSNGNYGMQTAQNY  
 GKDLNLSLPQLALLAGMPQAPNQYDPYSHPEAAQDRRLVLSMKNQGYISAEQYKAVNTPITDGLQ  
 SLKSASNYPAYMVDIYQLKEVINQVEETGYNLLTTGMDVYTNVDQEAQKHLWDIYNTDEYVAYPDDELQV  
 ASTIVDVSNKGVIAQLGARHQSSNVSFGINQAVETNRDWGSTMKPI TDYAPALEYGVYDSTATIVHDEP  
 YNYPGTNTPVYNWDRGYFGNITLQYALQOSRNPVAVETLNKVLNRAKTFNLGLIDYPSIHYSNAISS  
 NTTESDKKYGASSEKMAAAAYAAFANGGTYYKPMYIHKVVFSDGSEKEFSNVGTRAMKETTA YMMTDMMK  
 TVLTYGTGRNAYLAWLPQAGKTGTSNYTDEEIEENHIKTSQFVAPDELFAGYTRKYSMAVWTGYSNRLTP  
 LVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFVFKNGARSTWNSPAPQPPSTESSSSSS  
 DSSTSQSSSTTPSTNNSTTNPNNNTQQSNTTPDQONQNPPQPAQP

**SP004 nucleotide (SEQ ID NO:3)**

AAATTACAATACGGACTATGAATTGACCTCTGGAGAAAAATTACCTCTTCTTAAAGAGATTTTCAAGTTA  
 CACTTATATTGGATATATCAAAGAGGGAAAAACGACTTCTGAGTCTGAAGTAAGTAATCAAAGAGTTT  
 AGTTGCCACTCTACAAAACAACAAAAGGTGGATTATAATGTTACACCGAATTTTGTAGACCATCCATC  
 AACAGTACAAGCTATTCAAGAACAAACACCTGTTTCTTCAACTAAGCCGACAGAAGTTCAAGTAGTTGA  
 AAAACCTTTCTCTACTGAATTAATCAATCCAAGAAAAGAAGAGAAACAATCTTCAGATTCTCAAGAACA  
 ATTAGCCGAACATAAGAATCTAGAAACGAAGAAAGAGGAGAAGATTTCTCCAAAAGAAAAGACTGGGGT  
 AAATACATTAAATCCACAGGATGAAGTTTATCAGGTCAATTGAACAAACCTGAACCTCTATATCGTGA  
 GGAAACTATGGAGACAAAAATAGATTTTCAAGAAAGAAATTCAGAAAATCCTGATTTAGCTGAAGGAAC  
 TCTAAGAGTAAAACAAGAAGGTAAATTAGGTAAGAAAAGTTGAAATCGTCAGAATATTCTCTGTAAACAA  
 GGAAGAAGTTTCGCGAGAAATTGTTTCAACTTCAACGACTGCGCCTAGTCCAAGAATAGTCGAAAAAGG  
 TACTAAAAAACTCAAGTTATAAAGGAACAACCTGAGACTGGTGTAGAACATAAGGACGTACAGTCTGG  
 AGCTATTGTTGAACCCGCAATTCAGCCTGAGTTGCCCCAAGCTGTAGTAAGTGACAAAGGCGAACCAGA  
 AGTTCAACCTACATTACCCGAAGCAGTTGTGACCGACAAAGGTGAGACTGAGGTTCAACCAGAGTCGCC  
 AGATACTGTGGTAAGTGATAAAGGTGAACCAGAGCAGGTAGCACCGCTTCCAGAATATAAGGGTAATAT

Table 1

TGAGCAAGTAAAACCTGAAACTCCGGTTGAGAAGACCAAAGAACAAGGTCCAGAAAAAACTGAAGAAGT  
 TCCAGTAAAACCAACAGAGAAACACCAGTAAATCCAAATGAAGGTACTACAGAAAGGAACCTCAATTCA  
 AGAAGCAGAAAAATCCAGTTCAACCTGCAGAAGAATCAACAACGAATTCAGAGAAAGTATCACCAGATAC  
 ATCTAGCAAAAAATACTGGGGAAGTGTCCAGTAATCTTAGTGATTTCGACAACCTCAGTTGGAGAATCAAA  
 TAAACCAGAACATAATGACTCTAAAAATGAAAATTCAGAAAAAACTGTAGAAGAAGTTCCAGTAAATCC  
 AAATGAAGGCACAGTAGAAGGTACCTCAAATCAAGAAACAGAAAAACCAGTTCAACCTGCAGAGAAAC  
 ACAAACTCTGGGAAAAATAGCTAACGAAAAATACTGGAGAAGTATCCAATAAACCTAGTGATTCAAAA  
 ACCACCACTTGAAGAATCAAATCAACCAGAAAAAACCGGAACCTGCAACAAAACAGAAAAATTCAGGTAA  
 TACAACATCAGAGAATGGACAAACAGAACAGAACCATCAAACGGAAATTCAACTGAGGATGTTTCAAC  
 CGAATCAAACACATCCAATTCAAATGGAAACGAAGAAATTAACAAGAAAAATGAACTAGACCCCTGATAA  
 AAAGGTAGAAGAACCAGAGAAAAACACTTGAATTAAGAAATGTTTCCGACCTAGAGTTA

**SP004 amino acid (SEQ ID NO:4)**

NYNTDYELTSGEKLPKEISGYTYIGYIKEGKTTSESEVSNQKSSVATPTKQKQVDYNTPNFVDHPS  
 TVQAIQEQTVPVSSTKPTQVVEKPFSTELINPRKEEKQSSDSQEQLAHKNLETKKEEKISPKEKTGV  
 NTLNPQDEVLSGQLNKPELLYREETMETKIDFQEEIQENPDLAEGTVRVKQEGKLGKKVEIVRIFSVNK  
 EEVSREIVSTSTTAPSPRIVEKGTKKTKQVIKEQPETGVEHKDVQSGAIVEPAIQPELPEAVVSDKGEPE  
 VQPTLPEAVVTDKGETEVQPEPDTVVSDKGEPEQVAPLPEYKGNIEQVKPETPVEKTEQGEPEKTEEV  
 PVKPTETPVNPNEGTTEGTSIQEAENPVQPAEESTTNSEKVSPTSSKNTGEVSSNPSTSTSVGESN  
 KPEHNSKNENSEKTVEVPVNPNEGTVEGTSNQETEKPVQPAEETQTNSGKIANENTGEVSNKPSDSK  
 PPVEESNQPEKNGTATKPNESGNTTSENGQTEPEPSNGNSTEDVSTESNTSNSNGNEEIKQENELDPDK  
 KVEEPEKTLERNVSDLEL

**SP006 nucleotide (SEQ ID NO:5)**

TGAGAAATCAAGCTACACCCAAAGAGACTAGCGCTCAAAAGACAATCGTCCTTGCTACAGCTGGCGACGT  
 GCCACCATTTGACTACGAAGACAAGGGCAATCTGACAGGCTTTGATATCGAAGTTTTAAAGGCAGTAGA  
 TGAAAACTCAGCGACTACGAGATTCAATTCCAAAGAACCGCCTGGGAGAGCATCTTCCAGGACTTGA  
 TTCTGGTCACTATCAGGCTGCGGCCAATAAATTGAGTTACACAAAAGAGCGTGCTGAAAAATACCTTTA  
 CTCGCTTCCAATTTCCAACAATCCCTCGTCCTTGTCAGCAACAAGAAAAATCCTTTGACTTCTCTTGA  
 CCAGATCGCTGGTAAAAACAACAAGAGGATACCGGAACCTCTAACGCTCAATTCATCAATAACTGGAA  
 TCAGAAACACACTGATAATCCCGCTACAATTAATTTTTCTGGTGAGGATATTGGTAAACGAATCCTAGA  
 CCTTGCTAACGGAGAGTTTGATTTCTAGTTTTTGACAAGGTATCCGTTCAAAAGATTATCAAGGACCG  
 TGGTTTAGACCTCTCAGTCGTTGATTTACCTTCTGCAGATAGCCCCAGCAATTATATCATTTTCTCAAG  
 CGACCAAAAGAGTTTAAAGAGCAATTTGATAAAGCGCTCAAAGAACTCTATCAAGACGGAACCCCTGA  
 AAAACTCAGCAATACCTATCTAGGTGGTTCTTACCTCCCAGATCAATCTCAGTTACAA

**SP006 amino acid (SEQ ID NO:6)**

ENQATPKETSAQKTIVLATAGDVPPFDYEDKGNLTGFDIEVLKAVDEKLSDYEIQFQRTAWESIFPGLD  
 SGHYQAAANNLSYTKERAELYLSLPISNNPLVLVSNKKNPLTSLDQIAGKTTQEDTGTSNAQFINNWN  
 QKHTDNPATINFSGEDIGKRILDLANGEFDLFLVDKVSQKIIKDRGLDLSVDLPSADSPSNYIIFSS  
 DQKEFKEQFDKALKELYQDGTLEKLSNTYLGGSYLPDQSQLQ

**SP007 nucleotide (SEQ ID NO:7)**

TGGTAACCGCTCTTCTCGTAACGCAGCTTCATCTTCTGATGTGAAGACAAAAGCAGCAATCGTCACTGA  
 TACTGGTGGTGTGATGACAAATCATTCAACCAATCAGCTTGGGAAGGTTTGCAGGCTTGGGGTAAAGA  
 ACACAACTTTTCAAAAGATAACGGTTTCACTTACTTCCAATCAACAAGTGAAGCTGACTACGCTAACAA  
 CTTGCAACAAGCGGCTGGAAGTTACAACCTAATCTTCGGTGTGGTTTTGCCCTTAATAATGCAGTTAA  
 AGATGCAGCAAAAGAACACACTGACTTGAAGTATGTCTTGATTGATGATGTGATTAAAGACCAAAAGAA  
 TGTTGCGAGCGTAACCTTCGCTGATAATGAGTCAGGTTACCTTGCAGGTGTGGCTGCAGCAAAAACAAC  
 TAAGACAAAACAAGTTGGTTTTGTAGGTGGTATCGAATCTGAAGTTATCTCTCGTTTTGAGCAGGATT  
 CAAGGCTGGTGTGCGTCACTAGACCCATCTATCAAAGTCCAAGTTGACTACGCTGGTTTCAATTTGGTGA  
 TGCGGCTAAAGGTAAAAACAATTGCAGCCGCACAATACGCAGCCGGTGCAGATATTGTTTACCAAGTAGC  
 TGGTGGTACAGGTGCAGGTGTCTTTGCAGAGGCAAAATCTCTCAACGAAAGCCGTCCTGAAAAATGAAAA  
 AGTTTGGGTTATCGGTGTTGATCGTGACCAAGAAGCAGAAGGTAAATACACTTCTAAAGATGGCAAAAGA  
 ATCAAACCTTTGTTCTTGATCTACTTTGAAACAAGTTGGTACAACGTGTAAGATATTTCTAAACAAGGC  
 AGAAAGAGGAGAATTCCTTGGCGGTCAAGTGATCGTTTACTCATTGAAGGATAAAGGGGTTGACTTGGC  
 AGTAACAAACCTTTCAGAAAGAAGGTAAAAAAGCTGTGCAAGATGCAAAAGCTAAAATCCTTGATGGAAG  
 CGTAAAGTTCTTGAAAAA

Table 1

**SP007 amino acid (SEQ ID NO:8)**

GNRSSRNAASSSDVKTAAIVTDTGGVDDKSFNQSAWEGLQAWGKEHNLSKDNNGFTYFQSTSEADYANN  
 LQQAAGSYNLIFGVGFALNNNAVKDAAKEHTDLNVLIDDVIKDQKNVASVTFADNESGYLAGVAAAKTT  
 KTKQVGVFVGGIESEVISRFEAGFKAGVASVDPSIKVQVDYAGSFGDAAKGKTIAAAQYAAGADIVYQVA  
 GGTGAGVFAEAKSLNESRPENЕКVWVIGVDRDQEAEGKYTSKDGKESNFVLVSTLKQVGTTVKDISNKA  
 ERGEFPGGQVIVYSLKDKGVDLAVTNLSEEGKKAVEDAKAKILDGSVKVPEK

**SP008 nucleotide (SEQ ID NO:9)**

TGTGGAAATTTGACAGGTAACAGCAAAAAAGCTGCTGATTGAGGTGACAAACCTGTTATCAAAATGTAC  
 CAAATCGGTGACAAACCAGACAACCTTGGATGAATTGTTAGCAAATGCCAACAAAATCATTGAAGAAAAA  
 GTTGGTGCCAAAATGGATATCCAATACCTTGGCTGGGGTGAATGTTAGTAAGAAAATGTCAGTTATCACA  
 TCATCTGGTGAAAACATGATATTGCCTTTGCAGATAACTATATTGTAAATGCTCAAAAAGGTGCTTAC  
 GCTGACTTGACAGAATTGTACAAAAAGAAGGTAAAGACCTTTACAAAGCACTTGACCCAGCTTACATC  
 AAGGGTAATACTGTAAATGGTAAGATTACGCTGTTCCAGTTGCAGCCAACGTTGCATCATCTCAAAAC  
 TTTGCCTTCAACGGAACTCTCCTTGCTAAATATGGTATCGATATTTGAGGTGTTACTTCTTACGAACT  
 CTTGAGCCAGTCTTGAAACAAATCAAAGAAAAAGCTCCAGACGTAGTACCATTGCTATTGGTAAAGTT  
 TTCATCCCATCTGATAATTTGACTACCCAGTAGCAAACGGTCTTCCATTGCTTATCGACCTTGAAGGC  
 GATACTACTAAAGTTGTAAACCGTTACGAAGTGCCTCGTTTCAAAGAACACTTGAAGACTCTTCACAAA  
 TTCTATGAAGCTGGCTACATTCCAAAAGACGTGCGCAACAAGCGATACTTCCCTTTGACCTTCAACAAGAT  
 ACTTGGTTTCGTTTCGTGAAGAAACAGTAGGACCAGCTGACTACGGTAACAGCTTGCTTTTACGTTGTTGCC  
 AACAAAGATATCCAAATCAAACCAATTACTAAGTTCATCAAGNAAAACCAACAACACAAGTTGCTAAC  
 TTTGTCTATCTCAACAACCTCTAAGAACAAGAAAAATCAATGGAAATCTTGAACCTCTTGAATACGAAC  
 CCAGAACTCTTGAACGGTCTTGTGTTACGGTCCAGAAGGCAAGAAGTGGGAAAAAATTGAAGGTAAAGAA  
 AACCGTGTTCGCGTTCTTGTGATGGCTACAAAGGAAACACTCACATGGGTGGATGGAACACTGGTAACAAC  
 TGGATCCTTTACATCAACGAAAACGTTACAGACCAACAAATCGAAAATCTAAGAAAGAATTGGCAGAA  
 GCTAAAGAATCTCCAGCGCTTGGATTTATCTTCAATACTGACAATGTGAAATCTGAAATCTCAGCTATT  
 GCTAACACAATGCAACAATTTGATACAGCTATCAACACTGGTACTGTAGACCCAGATAAAGCGATTCCA  
 GAATTGATGGAAAAATTGAAATCTGAAGGTGCCTACGAAAAAGTATTGAACGAAATGCAAAAACAATAC  
 GATGAATCTTGAAAAACAAAAAA

**SP008 amino acid (SEQ ID NO:10)**

CGNLTGNSKKAADSGDKPVIMYQIGDKPDNLDELLANANKIIIEKVGAKLDIQYLGWGDYGKKMSVIT  
 SSGENYDIAFADNYIVNAQKAYADLTELYKKEGKDLKALDPAYIKGNTVNGKIYAVPVAANVASSQN  
 FAFNGTLLAKYIGIDISGVTSEYETLEPVLKQIKEKAPDVVPFAIGKVFIPSDNFDYPVANGLPFVIDLEG  
 DTTKVVNRVEVPRFKEHLKTLHKFYEAGYIPKDVATSDTSFDLQQDTWVREETVGPADYGNLSLSRVA  
 NKDIQIKPITNFIFKXNQTTQVANFVISNNSKNKEKSMEILNLLNTNPPELLNGLVYGPEGKNWEKIEGKE  
 NRVRLVDGYKGNTHMGWNTGNNWILYINENVTDQQIENSKKELAEAKESPALGFIFNTDNVKSEISAI  
 ANTMQQFDTAINTGTVPDKAIPELMEKLKSEGAYEKVLNEMQKQYDEFLKNKK

**SP009 nucleotide (SEQ ID NO:11)**

TGGTCAAGGAAGTCTTCTAAAGACAACAAAGAGGCAGAACTTAAGAAGGTTGACTTTATCCTAGACTG  
 GACACCAAAATACCAACCACACAGGGCTTTATGTTGCCAAGGAAAAAGGTTATTTCAAAGAAGCTGGAGT  
 GGATGTTGATTTGAAATTGCCACCAGAAGAAAGTTCTTCTGACTTGGTTATCAACGGAAAGGCACCATT  
 TGCAGTGTATTTCCAAGACTACATGGCTAAGAAATTGAAAAAGGAGCAGGAATCACTGCCGTTGCAGC  
 TATTGTTGAACACAATACATCAGGAATCATCTCTCGTAAATCTGATAATGTAAGCAGTCCAAAAGACTT  
 GGTGTTGTAAGAAATATGGGACATGGAATGACCCAACCTGAACTTGTCTATGTTGAAAACCTTGGTAGAATC  
 TCAAGGTGGAGACTTTGAGAAGGTTGAAAAAGTACCAATAACGACTCAAACCTCAATCACACCGATTGC  
 CAATGGCGTCTTTGATACTGCTTGGATTTACTACGGTTGGGATGGTATCCTTGCTAAATCTCAAGGTGT  
 AGATGCTAACTTCATGTACTTGAAGACTATGTCAAGGAGTTTGAAGTACTATTCCACCAGTTATCATCGC  
 AAACAACGACTATCTGAAAGATAACAAAGAAAGAGCTCGCAAGTCAATCAAGCCATCAAAAAAGGCTA  
 CCAATATGCCATGGAACATCCAGAAGAGCTGCAGATATTCTCATCAAGAATGCACCTGAACTCAAGGA  
 AAAACGTGACTTTGTCTCATGAATCTCAAAAAACTTGTCAAAAAGAAATACGCAAGCGACAAGGAAAAATG  
 GGGTCAATTTGACGCAGCTCGCTGGAATGCTTTCTACAAATGGGATAAAGAAAAATGGTATCCTTAAAGA  
 AGACTTGACAGACAAAGGCTTACCAACGAATTTGTGAAA

**SP009 amino acid (SEQ ID NO:12)**

Table 1

GQGTASKDNKEAELKKVDFILDWTPNTNHTGLYVAKEKGYFKEAGVDVDLKLPPPESSSDLVINGKAPF  
AVYFQDYMAKKLEKGAGITAVAAIVEHNTSGIISRKSDNVSSPKDLVGKKYGTWNDPTELAMLKTLVES  
QGGDFEKVEKVPNNDSNSITPIANGVFDTAWIYYGWDGILAKSQGV DANFMYLKDYVKEFDYSPVIIA  
NNDYLDKNKEEARKVIOAIKKGYQYAMEHP EEAADILIKNAPELKEKRDFVIESQKYL SKEYASDKEKW  
GQFDAARWNAFYKWDKENGILKEDLTDKGFTNEFVK

**SP010 nucleotide (SEQ ID NO:13)**

TAGCTCAGGTGGAAACGCTGGTTCATCCTCTGGAAAAACAAC TGCCAAAGCTCGCACTATCGATGAAAT  
CAAAAAAAGCGGTGAAC TCGAATCGCCGTGTTTGGAGATAAAAAACCGTTTGGCTACGTTGACAATGA  
TGGTTCTACCAAGGTACGCTACGATATTGAAC TAGGGAACCAACTAGCTCAAGACCTTGGTGTCAAGGT  
TAAATACATTTTCAGTCGATGCTGCCAACC GTGCGGAATACTTGATTTCAAACAAGGTAGATATTACTCT  
TGCTAACTTTACAGTAACTGACGAACGTAAGAAAACAAGTTGATTTTGCCCTTCCATATATGAAAGTTTC  
TCTGGGTGTCGTATCACCTAAGACTGGTCTCATTACAGACGTC AAAACAACCTGAAGGTAAAACCTTAAT  
TGTCACAAAAGGAACGACTGCTGAGACTTATTTTGAAGAATCATCCAGAAATCAAAC TCCAAAAATA  
CGACCAATACAGTGACTCTTACCAAGCTCTTCTTGACGGACGTGGAGATGCCTTTTCAACTGACAATAC  
GGAAGTTCTAGCTTGGGCGCTTGA AAAATAAAGGATTTGAAGTAGGAATTACTTCCCTCGGTGATCCCGA  
TACCATTGCGGCAGCAGTTCAAAAAGGCAACCAAGAATTGCTAGACTTCATCAATAAAGATATTGAAAA  
ATTAGGCAAGGAAAAC TCTTCCACAAGGCCTATGAAAAGACACTTCACCCAACCTACGGTGACGCTGC  
TAAAGCAGATGACCTGGTTGTTGAAGGTGGAAAAGTTGAT

**SP010 amino acid (SEQ ID NO:14)**

SSGGNAGSSSGKTTAKARTIDEIKSGELRIAVFGDKKPFYVDNDGSTKVRDYDIELGNQLAQDLGVKV  
KYISVDAANRAEY LISNKVDITLANFTVTDERKKQVDFALPYMKVSLGVVSPKTGLITDVKQLEGKTLI  
VTKGTTAET YFEKNHPEIKLQKYDQYSDSYQALLDGRGDAFSTDNTEVLAWALENKGFEVGITSLGDPD  
TIAAAVQKGNQELLD FINKDIEKLGKENFFHKAYEKT LHPTYGDAAKADDLVVEGGKVD

**SP011 nucleotide (SEQ ID NO:15)**

CTCCAAC TATGGTAAATCTGCGGATGGCACAGTGACCATCGAGTATTTCAACCAGAAAAAGAAATGAC  
CAAAACCTTGAAGAAATCACTCGTGATTTT GAGAAGGAAAAACCC TAAGATCAAGGTCAAAGTCGTCAA  
TGTACCAAATGCTGGTGAAGTATTGAAGACACGCGTTCTCGCAGGAGATGTGCCTGATGTGGTCAATAT  
TTACCCACAGTCCATCGA ACTGCAAGAATGGGCAAAAGCAGGTGTTTTTGAAGATTTGAGCAACAAAGA  
CTACCTGAAACGCGTGAAAAATGGCTACGCTGAAAAATATGCTGTAAACGAAAAAGTTTACAACGTTCC  
TTTTACAGCTAATGCCTTATGGAATTTACTACAACAAAGATAAATTCGAAGAACTGGGCTTGAAGGTTCC  
TGAAACCTGGGATGAATTTGAACAGTTAGTCAAAGATATCGTTGCTAAAGGACAAACACCATTTGGAAT  
TGCAGGTGCAGATGCTTGGACACTCAATGGTTACAATCAATTAGCCTTTGCGACAGCAACAGGTGGAGG  
AAAAGAAGCAAATCAATACCTTCGTTATTCTCAACCAAATGCCATTAAATTGTCGGATCCGATTATGAA  
AGATGATATCAAGGTCATGGACATCCTTCGCATCAATGGATCTAAGCAAAAGAACTGGGAAGGTGCTGG  
CTATACCGATGTTATCGGAGCCTTCGCACGTGGGGATGTCCCTCATGACACCAAATGGGTCTTGGGCGAT  
CACAGCGATTAATGAACAAAAACCGAAC TTTAAGATTGGGACCTTCATGATTCCAGGAAAAAGAAAAAGG  
ACAAAGCTTAACCGTTGGTGC GGGAGACTTGGCATGGTCTATCTCAGCCACCACCAAACATCCAAAAGA  
AGCCAATGCCTTTGTGGAATATATGACCCGTCCAGAAGTCATGCAAAAATACTACGATGTGGACGGATC  
TCCAACAGCGATCGAAGGGGTCAAACAAGCAGGAGAAGATTCACCGCTTGCTGGTATGACCGAATATGC  
CTTTACGGATCGTCACTTGGTCTGGTTGCAACAATACTGGACCAGTGAAGCAGACTTCCATACCTTGAC  
CATGAACTATGTCTTGACCGGTGATAAACAAGGCATGGTCAATGATTTGAATGCCTTCTTTAACCCGAT  
GAAAGCGGATGTGGAT

**SP011 amino acid (SEQ ID NO:16)**

SNYGKSADGTVTIEYFNQKKEMTKTLEEITRDFEKENPKIKVKVNVNPNAGEVLKTRVLAGDVPDVVNI  
YPQSIELQEWAKAGVFEDLSNKDY LKRVKNGYAEKYAVNEKVYNVPFTANAYGIYYNKDKFEELGLKVP  
ETWDEFELVKDIVAKGQTPFGIAGADAWTLNGYNQLAFATATGGGKEANQYLRYSQPNAIKLSDPIMK  
DDIKVMDILRINGSKQKNWEGAGYTDVIGAFARGDVLMTPNGSWAITAINEQKPNFKIGTFMIPGKEKG  
QSLTVGAGDLAWSISATTKHPKEANAFVEYMTREPEVMQKYVDVGSPTAIEGVKQAGEDSPLAGMTEYA  
FTDRHLVWLQQYWTSEADFHTLTMNYVLTGDKQGMVNDLNAFFNPMKADVD

**SP012 nucleotide (SEQ ID NO:17)**

TGGGAAAAATTCTAGCGAAACTAGTGGAGATAATTGGTCAAAGTACCAGTCTAACAAGTCTATTACTAT  
TGGATTTGATAGTACTTTTGTTC AATGGGATTTGCTCAGAAAGATGGTTCTTATGCAGGATTTGATAT  
TGATTTAGCTACAGCTGTTTTTGA AAAATACGGAATCACGGTAAATTGGCAACCGATTGATTGGGATTT

Table 1

GAAAGAAGCTGAATTGACAAAAGGAACGATTGATCTGATTGGAATGGCTATTCGCTACAGACGAACG  
CCGTGAAAAGGTGGCTTTCAGTAACATCATATATGAAGAATGAGCAGGTATTGGTTACGAAGAAATCATC  
TGGTATCACGACTGCAAAGGATATGACTGGAAAGACATTAGGAGCTCAAGCTGGTTCATCTGGTTATGC  
GGACTTTGAAGCAAATCCAGAAATTTTGAAGAATATTGTCGCTAATAAGGAAGCGAATCAATACCAAAC  
CTTTAATGAAGCCTTGATTGATTTGAAAAACGATCGAATTGATGGTCTATTGATTGACCGTGTCTATGC  
AAACTATTATTTAGAAGCAGAAGGTGTTTTAAACGATTATAATGTCTTTACAGTTGGACTAGAAACAGA  
AGCTTTTGCGGTTGGAGCCCGTAAGGAAGATACAACTTGGTTAAGAAGATAAATGAAGCTTTTTCTAG  
TCTTTACAAGGACGGCAAGTTCCAAGAAATCAGCCAAAAATGGTTTGGAGAAGATGTAGCAACCAAAGA  
AGTAAAGAAGGACAG

**SP012 nucleotide (SEQ ID NO:18)**

GKNSSETSGDNWSKYQSNKSITIGFDSTFVPMGFAQKDGSYAGFDIDLATAVFEKYGITVNWQPIDWDL  
KEAELTKGTIDLIWNGYSATDERREKVAFSNSYMKNEQVLVTKSSGITTAKDMTGKTLGAQAGSSGYA  
DFEANPEILKNIVANKEANQYQTFNEALIDLKNDRIDGLLIDRVYANYYLEAEGVLNDYNVFTVGLETE  
AFAVGARKEDTNLVKKINEAFSSLYKDGKFQEI SQWFGEDVATKEVKEGQ

**SP013 nucleotide (SEQ ID NO:19)**

TGCTAGCGGAAAAAAGATACAACTTCTGGTCAAAAACATAAAGTTGTTGCTACAACTCAATCATCGC  
TGATATTACTAAAAATATTGCTGGTGACAAAATTGACCTTCATAGTATCGTTCGGATTGGGCAAGACCC  
ACACGAATACGAACCACTTCTGAAGACGTTAAGAAAACCTTCTGAGGCTAATTTGATTTTCTATAACGG  
TATCAACCTTGAAACAGGTGGCAATGCTTGGTTTACAAAATTGGTAGAAAATGCCAAGAAAACCTGAAAA  
CAAAGACTACTTTCGAGTCAGCGACGGCGTTGATGTTATCTACCTTGAAGGTCAAAAATGAAAAAGGAAA  
AGAAGACCCACACGCTTGGCTTAACCTTGAAAACGGTATTATTTTTTGCTAAAAATATCGCCAAACAATT  
GAGCGCCAAAGACCTTAACAATAAAGAATTCTATGAAAAAATCTCAAAGAATATACTGATAAGTTAGA  
CAAACCTTGATAAAGAAAGTAAGGATAAATTTAATAAGATCCCTGCTGAAAAGAACTCATTGTAACCAG  
CGAAGGAGCATTCAAATACTTCTCTAAAGCCTATGGTGTCCCAAGTGCTTACATCTGGGAAATCAATAC  
TGAAGAAGAAGGAACCTCTGAACAAATCAAGACCTTGGTTGAAAAACCTTCGCCAAACAAAAGTTCCATC  
ACTCTTTGTAGAATCAAGTGTGGATGACCGTCCAATGAAAACCTGTTTCTCAAGACACAAACATCCCAAT  
CTACGCTCAAACTTTTACTGACTCTATCGCAGAACAAAGGTAAAGAAGGCGACAGCTACTACAGCATGAT  
GAAATACAACCTTGACAAGATTGCTGAAGGATTGGCAAAA

**SP013 amino acid (SEQ ID NO:20)**

ASGKKDTSQGKLKVATNSIIADITKNIAGDKIDLHSIVPIGQDPHEYEPLPEDVKKTSEANLIFYNG  
INLETGGNAWFTKLVENAKKTENKDYFAVSDGVDVIYLEGQNEKGKEDPHAWLNLENGIIFAKNIAKQL  
SAKDPNNKEFYEKNLKEYTDKLDKDKESKDKFNKIPAEEKLIVTSEGAFKYFSKAYGVPSAYIWEINT  
EEEGTPEQIKTLVEKLRQTKVPSLFVSSVDDRPMTVSQDTNPIPIYAQIFTDSIAEQGKEGDSYYSMM  
KYNLDKIAEGLAK

**SP014 nucleotide (SEQ ID NO:21)**

TGGCTCAAAAAATACAGCTTCAAGTCCAGATTATAAGTTGGAAGGTGTAACATTCCCGCTTCAAGAAAA  
GAAAACATTGAAGTTTATGACAGCCAGTTCACCGTTATCTCCTAAAGACCCAAATGAAAAGTTAATTTT  
GCAACGTTTGGAGAAGGAAACTGGCGTTCATATTGACTGGACCAACTACCAATCCGACTTTGCAGAAAA  
ACGTAACCTTGGATATTTCTAGTGGTGATTTACCAGATGCTATCCACAACGACGGAGCTTCAGATGTGGA  
CTTGATGAACTGGGCTAAAAAAGGTGTTATTATTCAGTTGAAGATTTGATTGATAAATACATGCCAAA  
TCTTAAGAAAAATTTGGATGAGAAACCAGAGTACAAGGCCTTGATGACAGCACCTGATGGGCACATTTA  
CTCATTTCCATGGATTGAAGAGCTTGGAGATGGTAAAGAGTCTATTACAGTGTCAACGATATGGCTTG  
GATTAACAAAGATTGGCTTAAGAACTTGGTCTTGAAATGCCAAAAACTACTGATGATTTGATTAAAGT  
CCTAGAAGCTTTCAAAAACGGGGATCCAAATGGAATGGAGAGGCTGATGAAATTCATTTTTCATTTAT  
TAGTGGTAACGGAAACGAAGATTTTAAATTCCTATTTGCTGCATTTGGTATAGGGGATAACGATGATCA  
TTTAGTAGTAGGAAATGATGGCAAAGTTGACTTCACAGCAGATAACGATAACTATAAAGAAGGTGTCAA  
ATTTATCCGTCATTTGCAAGAAAAAGGCCTGATTGATAAAGAAGCTTTTCAACATGATTGGAATAGTTA  
CATTGCTAAAGGTCATGATCAGAAATTTGGTGTCTTACTTTACATGGGATAAGAATAATGTTACTGGAAG  
TAACGAAAGTTATGATGTTTTACCAGTACTTGTCTGGACCAAGTGGTCAAAAACACGTAGCTCGTACAAA  
CGGTATGGGATTTGCACGTGACAAGATGGTTATTACCAGTGTAACAAAAACCTAGAATTGACAGCTAA  
ATGGATTGATGCACAAATACGCTCCACTCCAATCTGTGCAAAAATAACTGGGGAACCTACGGAGATGACAA  
ACAACAAAACATCTTTGAATTGGATCAAGCGTCAAATAGTCTAAAAACCTTACCACTAAACGGAACCTGC  
ACCAGCAGAACTTCGTCAAAAGACTGAAGTAGGAGGACCACTAGCTATCCTAGATTACACTATGGTAA  
AGTAAACAACCATGCCTGATGATGCCAAATGGCGTTTGGATCTTATCAAAGAATATTATGTTCCCTTACAT

WO 98/18930

PCT/US97/19422

Table 1

53

GAGCAATGTCAATAACTATCCAAGAGTCTTTATGACACAGGAAGATTGTTGACAAGATTGCCCATATCGA  
AGCAGATATGAATGACTATATCTACCGTAAACGTGCTGAATGGATTGTAAATGGCAATATTGATACTGA  
GTGGGATGATTACAAGAAAGAACTTGAAAAATACGGACTTTCTGATTACCTCGCTATTAAACAAAAATA  
CTACGACCAATACCAAGCAACAAAAAC

**SP014 amino acid (SEQ ID NO:22)**

GSKNTASSPDYKLEGVTFPLQEKTKLFMTASSPLSPKDPNEKLILQRLKETGVHIDWTNYQSDFAEK  
RNLDISSGDLPDIAHNDGASDVLDLNMWAKKGVII PVEDLIDKYPNLKKILDEKPEYKALMTAPDGHY  
SFPWIEELGDGKESIHSVNDMAWINKDWLKKLGLEMPKTTDDLKLVLEAFKNGDPNGNGEAD EIPFSFI  
SGNGNEDFKFLFAAFGIGDNDHLVVGNDGKVDFTADNDNYKEGVKFI RQLQEKGLIDKEAFEHDWNSY  
IAKGHDQKFGVYFTWDKNNVTGSNESYDVLPLVLAGPSGQKHVARTNGMGFARDKMVITSVNKNLELTAK  
WIDAQYAPLQSVQNNWGTYGDDKQONIFELDQASNSLKHPLNGTAPAE LRQKTEVGGPLAILDSYYGK  
VTTMPDDAKWRLDLIKEYYVPYMSNVNNYPRVFMTQEDLDKIAHIEADMNDYIYRKRAEWIVNGNIDTE  
WDDYKKELEKYGLSDYLAIKQKYDQYQANKN

**SP015 nucleotide (SEQ ID NO:23)**

TAGTACAAACTCAAGCACTAGTCAGACAGAGACCAGTAGCTCTGCTCCAACAGAGGTAACCATTTAAAAG  
TTCAC TGGACGAGGTCAAAC TTTCCAAAGTTCCTGAAAAGATTGTGACCTTTGACCTCGGCGCTGCGGA  
TACTATTCGCGCTTTAGGATTTGAAAAAATATCGTCGGAATGCCTACAAAAACTGTTCCGACTTATCT  
AAAAGACCTAGTGGGAACTGTCAAAAATGTTGGTTCTATGAAAGAACCTGATTTAGAAGCTATCGCCGC  
CCTTGAGCCTGATTTGATTATCGCTTCGCCACGTACACAAAAATTCGTAGACAAATTCAAAGAAATCGC  
CCCAACCGTTCCTCTTCCAAGCAAGCAAGGACGACTACTGGACTTCTACCAAGGCTAATATCGAATCCTT  
AGCAAGTGCCTTCGGCGAAACTGGTACACAGAAAGCCAAGGAAGAATTGACCAAGCTAGACAAGAGCAT  
CCAAGAAGTCGCTACTAAAAATGAAAGCTCTGACAAAAAAGCCCTTGCGATCCTCCTTAATGAAGGAAA  
AATGGCAGCCTTTGGTGCCAAATCTCGTTTCTCTTTCTTGTAACCAACCTTGAAATTCAAACCAACTGA  
TACAAAATTTGAAGACTCAGCCACGGACAAGAAGTCAGCTTTGAAAGTGTCAAAGAAATCAACCTTGA  
CATCCTCTTTGTCTATCAACCGTACCCTTGCCATCGGTGGGGACAACCTAGCAACGACGGTGTCTCTAGA  
AAATGCCCTTATCGCTGAAACACCTGCTGCTAAAAATGGTAAGATTATCCAACCTAACACCAGACCTCTG  
GTATCTAAGCGGAGGCGGACTTGAATCAACAAAACTCATGATTGAAGACATACAAAAAGCTTTTGAAA

**SP015 amino acid (SEQ ID NO:24)**

STNSSTSQTETSSSAPTEVTIKSSLDEVKLSKVPEKIVTFDLGAADTIRALGF EKNIVGMPTKTVPTYL  
KDLVGTVKNVSGMKEPDLEAIAALEPDLIIASPR TQKFVDKFKEIAPT VLFQASKDDYWTSTKANIESL  
ASAFGETGTQAKEELTKLDKSIQEVATKNESDDKALAILLNEGKMAAFGAKSRFSFLYQTLKFKPTD  
TKFEDSRHQEVSFESVKEINPDILFVINRTLAIGDNDSSNDGVLENALIAETPAAKNGKIIQLTPDLW  
YLSGGGLESTKLMIEDIQKALK

**SP016 nucleotide (SEQ ID NO:25)**

TGGCAATTC TGGCGAAGTAAAGATGCTGCCAAATCAGGTGGTGACGGTGCCAAAACAGAAATCACTTG  
GTGGGCATTTCCAGTATTTACCCAAGAAAAAACTGGTGACGGTGTTGGAAC TTTATGAAAAATCAATCAT  
CGAAGCGTTTGAAAAAGCAAACCCAGATATAAAAGTGAAATTGGAACCATCGACTTCAAGTCAGGTCC  
TGAAAAAATCACACAGCCATCGAAGCAGGAACAGCTCCAGACGTACTCTTTGATGCACCAGGACGTAT  
CATCCAATACGGTAAAAACGGTAAATTTGGCTGAGTTGAATGACCTCTTCACAGATGAATTTGTTAAAGA  
TGTCACCAATGAAAAACATCGTACAAGCAAGTAAAGCTGGAGACAAGGCTTATATGTATCCGATTAGTTC  
TGCCCCATTCTACATGGCAATGAACAAGAAAAATGTTAGAAGATGCTGGAGTAGCAAACCTTGTAAGA  
AGGTTGGCAACTGATGATTTTGAAAAAGTATTGAAAGCACTTAAAGACAAGGGTTACACACCAGGTTC  
ATTGTTTCAGTTCTGGTCAAGGGGGAGACCAAGGAACACGTGCCTTTATCTCTAACCTTTATAGCGGTTC  
TGTAACAGATGAAAAAGTTAGCAAAATATACAACTGATGATCCTAAATTCGTCAAAGGTCTTGAAAAAGC  
AACTAGCTGGATTAAAGACAATTTGATCAATAATGGTTTACAATTTGACGGTGGGGCAGATATCCAAAA  
CTTTGCCAACGGTCAAACATCTTACACAATCCTTTGGGCACCAGCTCAAAATGGTATCCAAGCTAAACT  
TTTAGAAGCAAGTAAGGTAGAAGTGGTAGAAGTACCATTCCCATCAGACGAAGGTAAAGCCAGCTCTTGA  
GTACCTTGTAACGGGTTTGTCAGTATTCAACAATAAAGACGACAAGAAAGTCGCTGCATCTAAGAAATT  
CATCCAGTTTATCGCAGATGACAAGGAGTGGGGACCTAAAGACGTAGTTTCGTACAGGTGCTTTCCCACT  
CCGTACTTCATTTGAAAACTTTATGAAGACAAACGCATGGAACAATCAGCGGCTGGACTCAATACTA  
CTCACCATACTACAACACTATTGATGGATTGCTGAAATGAGAACACTTTGGTTCCCAATGTTGCAATC  
TGTATCAAAATGGTGACGAAAAACCAGCAGATGCTTTGAAAGCCTTCAC TGAAAAAGCGAACGAAACAAT  
CAAAAAAGCTATGAAACAA

Table 1

**SP016 amino acid (SEQ ID NO:26)**

GNSGGSKDAASKGGDGAKTEITWWAFPVFTQEKTDGVDGTYEKSIIIEAFEKANPDIKVKLETIDFKSGP  
 EKITTAIEAGTAPDVLFDAPGRRIIQYGKNGKLAELNDLFTDEFVKDVNNENIVQASKAGDKAYMPISS  
 APFYMAMNKKMLEADAGVANLVKEGWTTDDFEKVLKALKDKGYTPGSLFSSGQGGDQGTRAFISNLYSGS  
 VTDEKVSKEYTTDDPKFVKGLEKATSWIKDNLINNGSQFDGGADIQNFANGQTSYTLWAPAQNGIQAKL  
 LEASKVEVVEVPFSPDEGKPALEYLVNGFAVFNNKDDKKVAASKKFIQFIADDKEWGPDKDVVRTGAFPV  
 RTSFGKLYEDKRMETISGWTQYYSPYYNTIDGFAEMRTLWF PMLQSVSNGDEKPADALKAFTEKANETI  
 KKAMKQ

**SP017 nucleotide (SEQ ID NO:27)**

TTCACAAGAAAAACAAAAATGAAGATGGAGAACTAAGACAGAACAGACAGCCAAAGCTGATGGAAC  
 AGTCGGTAGTAAGTCTCAAGGAGCTGCCCAGAAAGCAGAAAGTGGTCAATAAAGGTGATTACTACAG  
 CATTCAAGGGAAATACGATGAAATCATCGTAGCCAACAAACACTATCCATTGCTTAAAGACTATAATCC  
 AGGGGAAAATCCAACAGCCAAGGCAGAGTTGGTCAAACCTCATCAAAGCGATGCAAGAGGCAGGTTTCCC  
 TATTAGTGATCATTACAGTGGTTTTAGAAAGTTATGAACTCAGACCAAGCTCTATCAAGATTATGTCAA  
 CCAAGATGGAAAGGCAGCAGCTGACCGTTACTCTGCCCCGTCCTGGCTATAGCGAACACCAGACAGGCTT  
 GGCTTTTGATGTGATTGGGACTGATGGTGATTTGGTGACAGAAGAAAAAGCAGCCCAATGGCTCTTGGA  
 TCATGCAGCTGATTATGGCTTTGTTGTCCGTTATCTCAAAGGCAAGGAAAAGGAAACAGGCTATATGGC  
 TGAAGAATGGCACCTGCGTTATGTAGGAAAAGAAGCTAAAGAAATTGCTGCAAGTGGTCTCAGTTTGA  
 AGAATACTATGGCTTTGAAGGCGGAGACTACGTCGAT

**SP017 amino acid (SEQ ID NO:28)**

SQEKTKNEDGETKTEQTAKADGTVGSKSQGAAQKKADEVVNKGDYYSIQGKYDEIIVANKHYPLSKDYNP  
 GENPTAKAELVKLIKAMQEAGFPISDHYSGFRSYETQTKLYQDYVNQDGKAAADRY SARPGYSEHQ TGL  
 AFDVIGTDGDLVTEEKAAQWLLDHAADYGFVVRYLKGKEKETGYMAEEWHLRYVGKEAKEIAASGLSLE  
 EYYGFEGGDYVD

**SP019 nucleotide (SEQ ID NO:29)**

GAAAGGTCTGTGGTCAAATAATCTTACCTGCGGTTATGATGAAAAATAATCTTGGAAAAATATAAATAT  
 AAAAAATACCTGAAGAAAAAATATCAGTTATTATTGGGTCAAATGGTTGTGGGAAATCAACACTCATTA  
 AACCTTGTCTCGACTTATAAAGCCATTAGAGGGAGAAGTATTGCTTGATAATAAATCAATTAATTCTTA  
 TAAAGAAAAAGATTTAGCAAAACACATAGCTATATTACCTCAATCTCCAATAATCCCTGAATCAATAAC  
 AGTAGCTGATCTTGTAAGCCGTGGTTCGTTTCCCCTACAGAAAGCCTTTTAAGAGTCTTGGAAAAGATGA  
 CCTTGAAATAATAAACAGATCAATGGTTAAGGCCAATGTTGAAGATCTAGCAAATAACCTAGTTGAAGA  
 ACTTTCTGGGGGTCAAAGGCAAAGAGTATGGATAGCTCTAGCCCTAGCCCAAGATACAAGTATCCTACT  
 TTTAGATGAGCCAACACTACTTACTTGGATATCTCATATCAAATAGAACTATTAGACCTCTTGACTGATCT  
 AAACCAAAAAATATAAGACAACCATTTGCATGATTTTGCACGATATAAATCTAACAGCAAGATACGCTGA  
 TTACCTATTTGCAATTAAGAAGGTAACTTGTGTCAGAGGGAAAGCCTGAAGATATACTAAATGATAA  
 ACTAGTTAAAGATATCTTTAATCTTGAAGCAAAAATTATACGTGACCCTATTTCCAATTGCGCTCTAAT  
 GATTCCTATTGGCAAGCACCATGTTAACTCT

**SP019 amino acid (SEQ ID NO:30)**

KGLWSNNLTCGYDEKIILENINIKIPEEKISVIIIGSNGCGKSTLIKTLRLIKPLEGEVLLDNKSINSY  
 KEKDLAKHIAILPQSPPIIPESITVADLVSRGRFPYRKPFKSLGKDDLEIINRSMVKANVEDLANNLVEE  
 LSGGQRQRVWIALALAQDTSILLLDEPTTYLDISYQIELLDLLTDLNQKYKTTICMILHDINLTARYAD  
 YLFAIKEGKLVAEGKPEDILNDKLVKIDIFNLEAKIIRDPISNSPLMIPIGKHHVS

**SP020 nucleotide (SEQ ID NO:31)**

AAACTCAGAAAAGAAAGCAGACAATGCAACAACATCAAAATCGCAACTGTTAACCCTAGCGGTTCTGA  
 AGAAAAACGTTGGGACAAAAATCCAAGAATTGGTTAAAAAAGACGGAATTACCTTGGAATTTACAGAGTT  
 CACAGACTACTCACAACCAACAAAGCAACTGCTGATGGCGAAGTAGATTGTAACGCTTTCCAACACTA  
 TAACTTCTTGAACAACTGGAACAAAGAAAACGGAAAAGACCTTGTAGCGATTGCAGATACTTACATCTC  
 TCCAATCCGCTTTACTCAGGTTTGAATGGAAGTGCCAACAAGTACACTAAAGTAGAAGACATCCCAGC  
 AAACGGAGAAATCGCTGTACCGAATGACGCTACAACGAAAGCCGTGCGCTTTATTTGCTTCAATCAGC  
 TGGCTTGATTAAATTTGGATGTTTCTGGAACCTGCTTGTGCAACAGTTGCCAACATCAAAGAAAAATCCAAA  
 GAACCTTGAAAACTACTGAATTGGACGCTAGCCAACACGCTCGTTTCATTGTGTCATCAGTTGACGCTGCCGT  
 TGTAAACAATACCTTCGTTACAGAAGCAAAAATTGGACTACAAGAAATCACTTTTCAAAGAACAGCTGA  
 TGAAAAC TCAAAACAATGGTACAACATCATTGTTGCAAAAAAGATTGGGAAACATCACCTAAGGCTGA

Table 1

TGCTATCAAGAAAGTAATCGCAGCTTACCACACAGATGACGTGAAAAAAGTTATCGAAGAATCATCAGA  
TGGTTTGGATCAACCAGTTTGG

**SP020 amino acid (SEQ ID NO:32)**

NSEKKADNATTIKIATVNRSGSEKRWDKIQELVKKDGITLEFTEFTDYSQPNKATADGEVDLNAFQHY  
NFLNNWNKENGKDLVAIADTYISPIRLYSGLNGSANKYTKVEDIPANGEI AVPN DATNESRALYLLQSA  
GLIKLDVSGTALATVANIKENPKNLKITELDASQTARSLSSVDAAVVNNTFVTEAKLDYKKS LFK EQAD  
ENSKQWYNIIVAKKDWETSPKADA IKKVIAAYHTDDVKKVIEESSDGLDQPVW

**SP021 nucleotide (SEQ ID NO:33)**

TTCGAAAGGGTCAGAAGGTGCAGACCTTATCAGCATGAAAGGGGATGTCATTACAGAACATCAATTTTA  
TGAGCAAGTGAAAAGCAACCCCTTCAGCCCAACAAGTCTTGTTAAATATGACCATCCAAAAAGTTTGTGA  
AAAACAATATGGCTCAGAGCTTGATGATAAAGAGGTTGATGATACTATTGCCGAAGAAAAAAAACAATA  
TGGCGAAAAC TACCAACGTGTCTTGTCACAAGCAGGTATGACTCTTGAAACACGTAAAGCTCAAATTCCG  
TACAAGTAAATTAGTTGAGTTGGCAGTTAAGAAGGTAGCAGAAGCTGAATTGACAGATGAAGCCTATAA  
GAAAGCCTTTGATGAGTACACTCCAGATGTAACGGCTCAAATCATCCGCTCTTAATAATGAAGATAAGGC  
CAAAGAAGTTCTCGAAAAAGCCAAGGCAGAAGGTGCTGATTTTGCTCAATTAGCCAAAGATAATTCAAC  
TGATGAAAAAACAAAAGAAAATGGTGGAGAAATTACCTTTGATTCTGCTTCAACAGAAAGTACCTGGAGC  
AAGTCCAAAAAGCCGCTTTTTCGCTTTTAGATGTGGGATGGTGTTTCTGGATGTGGATTACAGCAACTG  
GGGCACACCAAGCCTACAG

**SP021 amino acid (SEQ ID NO:34)**

SKGSEGADLISMKGDVITEHQFYEQVKSNP SAQQVLLNMTIQKVF EKQY GSELDDKEVDDTIAEEKKQY  
GENYQRVLSQAGMTLETRKAQIRTSKLVELAVKKVAEAE LTDEAYKKAFDEYTPDVTAQIIRLNNEDKA  
KEVLEKAKAEGADFAQLAKDNSTDEKTKENGGEITFDSASTEVP GASPKKPLFAFRCGMVFLDV DYSNW  
GTPSLQ

**SP022 nucleotide (SEQ ID NO:35)**

GGGGATGGCAGCTTTTAAAAATCCTAACAATCAATACAAAGCTATTACAATTGCTCAAAC TCTAGGTGA  
TGATGCTTCTTCAGAGGAATTGGCTGGTAGATATGGTCTCTGCTGTT CAGTGTACAGAAGTGACTGCCTC  
AAACCTTTTCAACAGTTAAAACTAAAGCTACGGTTGTAGAAAAACCACTGAAAGATTTT TAGAGCGTCTAC  
GTCTGATCAGTCTGGTTGGGTGGAATCTAATGGTAAATGGTATTTCTATGAGTCTGGTGATGTGAAGAC  
AGGTTGGGTGAAAACAGATGGTAAATGGTACTATTTGAATGACTTAGGTGTCATGCAGACTGGATTTGT  
AAAATTTTCTGGTAGCTGGTATTACTTGAGCAATTCAGGTGCTATGTTTACAGGCTGGGGAACAGATGG  
TAGCAGATGGTTCTACTTTGACGGCTCAGGAGCTATGAAGACAGGCTGGTACAAGGAAAATGGCACTTG  
GTATTACCTTGACGAAGCAGGTATCATGAAGACAGGTTGGTTTAAAGTCGGACCACACTGGTACTATGC  
CTACGGTT CAGGAGCTTTGGCTGTGAGCACAACAACACCAGATGGTTACCGTGTAATGCTAATGGTGA  
ATGGGTAAAC

**SP022 amino acid (SEQ ID NO:36)**

GMAAFKNPNNQYKAITIAQTLGDDASSEELAGRYGSAVQC TEVTASNLS TVKTKATVVEKPLKDFRAST  
SDQSGWVESNGKWYFYESGDVKTGWVKTDGKWYYLNDLGVMQTGFVKFSGSWY YLSNSGAMFTGWGTDG  
SRWFYFDGSGAMKTGWYKENG TWYYLDEAGIMKTGWFKVGP HWYYAYGSGALAVSTTT PDGYRVNNGE  
WVN

**SP023 nucleotide (SEQ ID NO:37)**

AGACGAGCAAAAAATTAAGCAAGCAGAAGCGGAAGTTGAGAGTAAACAAGCTGAGGCTACAAGGTTAAA  
AAAAATCAAGACAGATCGTGAAAGAAGCAGAAGAAGAAGCTAAACGAAGAGCAGATGCTAAAGAGCAAGG  
TAAACCAAAGGGCGGGCAAAACGAGGAGTTCTTGAGAGCTAGCAACACCTGATAAAAAAGAAAAATGA  
TGCGAAGTCTTCAGATTCTAGCGTAGGTGAAGAACTCTTCCAAGCCCATCCCTGAAACCAGAAAAAAA  
GGTAGCAGAAGCTGAGAAGAAGGTTGAAGAAGCTAAGAAAAAGCCGAGGATCAAAAAGAAGAAGATCG  
CCGTAAC TACCCAACCAATACTTACAAAACGCTTGAAC TTGAAATTGCTGAGTCCGATGTGGAAGTTAA  
AAAAGCGGAGCTTGAAGTAGTAAAAGAGGAAGCTAAGGAACCTCGAAACGAGGAAAAAGTTAAGCAAGC  
AAAAGCGGAAGTTGAGAGTAAAAAGCTGAGGCTACAAGGTTAGAAAAAATCAAGACAGATCGTAAAAA  
AGCAGAAGAAGAAGCTAAACGAAAAGCAGCAGAAGAAGATAAAGTTAAAGAAAAACCAGCTGAACAACC  
ACAACCAGCGCCGGCTCCAAAAGCAGAAAAACAGCTCCAGCTCCAAAACCAGAGAAATCCAGCTGAACA  
ACCAAAAGCAGAAAAACCAGCTGATCAACAAGCTGAAGAAGACTATGCTCGTAGATCAGAAGAAGAATA  
TAATCGCTTGACTCAACAGCAACCGCCAAAAACTGAAAAACCAGCACAACCATCTACTCCAAAAACAGG

WO 98/18930

PCT/US97/19422

Table 1

56

CTGGAAACAAGAAAACGGTATGTGGTACTTCTACAATACTGATGGTTCAATGGCGACAGGATGGCTCCA  
 AAACAATGGCTCATGGTACTACCTCAACAGCAATGGCGCTATGGCGACAGGATGGCTCCAAAACAATGG  
 TTCATGGTACTATCTAAACGCTAATGGTTCAATGGCAACAGGATGGCTCCAAAACAATGGTTCATGGTA  
 CTACCTAAACGCTAATGGTTCAATGGCGACAGGATGGCTCCAATACAATGGCTCATGGTACTACCTAAA  
 CGCTAATGGTTCAATGGCGACAGGATGGCTCCAATACAATGGCTCATGGTACTACCTAAACGCTAATGG  
 TGATATGGCGACAGGTTGGGTGAAAGATGGAGATACCTGGTACTATCTTGAAGCATCAGGTGCTATGAA  
 AGCAAGCCAATGGTTCAAAGTATCAGATAAATGGTACTATGTCAATGGCTCAGGTGCCCTTGCAGTCAA  
 CACAACGTAGATGGCTATGGAGTCAATGCCAATGGTGAATGGGTAAAC

**SP023 amino acid (SEQ ID NO:38)**

DEQKIKQAEAEVESKQAEATRLKKIKTDREEAEAEAKRRADAEQKPKGRAKRGVPGELATPDKKEND  
 AKSSDSSVGEETLPSPLKPEKKVAEAEKKVEEAKKKAEDQKEEDRRNYPTNTYKLTLELEIAESDVEVK  
 KAELELVKEEAKPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEEEEAKRKAEEEDKVKEKPAEQP  
 QPAPAPKAEKPAPAPKPNPAEQPKAEKPADQQAEEYARRSEEEYNRLTQQQPPKTEKPAQPSTPKTG  
 WKQENGWYFYNTDGSMTGWLQNNGSWYYLNSNGAMATGWLQNNGSWYYLNANGSMATGWLQNNGSWY  
 YLNANGSMATGWLQYNGSWYYLNANGSMATGWLQYNGSWYYLNANGDMATGWVKDGDWYYLEASGAMK  
 ASQWFKVSDKWYYVNGSGALAVNTTVDGYGVNANGEWVN

**SP025 nucleotide (SEQ ID NO:39)**

CTGTGGTGAGGAAGAACTAAAAAGACTCAAGCAGCACAAACAGCCAAAACAACAAACGACTGTACAACA  
 AATTGCTGTTGGAAGATGCTCCAGACTTCACATTGCAATCCATGGATGGCAAAGAAGTTAAGTTATC  
 TGATTTTAAGGGTAAAAAGGTTTACTTGAAGTTTGGGCTTCATGGTGTGGTCCATGCAAGAAAAGTAT  
 GCCAGAGTTGATGGAAGTAGCGGCGAAACCAGATCGTGATTTGCAAAATCTTACTGTCAATTGCACCAGG  
 AATTCAGGTGAAAAAACTGTTGAGCAATTCCCACAATGGTTCCAGGAACAAGGATATAAGGATATCCC  
 AGTTCTTTATGATACCAAAGCAACCACTTCCAAGCTTATCAAATTCGAAGCATTCTACAGAATATT

**SP025 amino acid (SEQ ID NO:40)**

CGEEETKKTQAAQPKQQTTVQQIAVGKDAPDFTLQSMGKEVKLSDFKGGKVYLKFWASWCGPCKKSM  
 PELMELAAKPRDRDFEILTIVAPIQGEKTVEQFPQWFQEQGYKDI PVLYDTKATTSKLIKFEAFLONI

**SP028 nucleotide (SEQ ID NO:41)**

GACTTTTAAACAATAAACTATGAAGAGTTGCACAATCTCCTTGTCTCTAAGGAAATTTCTGCAACAGA  
 ATTGACCCAAGCAACACTTGAAATATCAAGTCTCGTGAGGAAGCCCTCAATTCATTTGTCACCATCGC  
 TGAGGAGCAAGCTCTTGTCAAGCTAAAGCCATTGATGAAGCTGGAATTGATGCTGACAATGTCCTTTC  
 AGGAATTCCACTTGCTGTTAAGGATAACATCTCTACAGACGGTATTCTCACAACCTGCTGCCTCAAAAAT  
 GCTCTACAACATATGAGCCAATCTTTGATGCGACAGCTGTTGCCAATGCAAAAACCAAGGGCATGATTGT  
 CGTTGGAAAGACCAACATGGACGAATTTGCTATGGGTGGTTCAGGTGAAACTTCACACTACGGAGCAAC  
 TAAAAACGCTTGGAACCAACAGCAAGGTTCTGGTGGGTTCATCAAGTGGTCTGCCGCAGCTGTAGCCTC  
 AGGACAAGTTCGCTTGCTACTTGGTCTGATACTGGTGGTTCATCCGCCAACCTGCTGCCTTCAACGG  
 AATCGTTGGTCTCAAACCAACCTACGGAACAGTTTCACGTTTCGGTCTCATTGCCTTTGGTAGCTCATT  
 AGACCAGATTGGACCTTTGCTCCTACTGTTAAGGAAAATGCCCTCTTGCTCAACGCTATTGCCAGCGA  
 AGATGCTAAAGACTCTACTTCTGCTCCTGTCCGCATCGCCGACTTTACTTCAAAAATCGGCCAAGACAT  
 CAAGGGTATGAAAATCGCTTTGCCTAAGGAATACCTAGGCGAAGGAATTGATCCAGAGGTTAAGGAAAC  
 AATCTTAAACGCGGCCAAACACTTTGAAAAATTGGGTGCTATCGTCGAAGAAGTCAGCCTTCTCTACTC  
 TAAATACGGTGTGCGCTTTATTACATCATCGCTTCATCAGAAGCTTCATCAAACCTTGCAACGCTTCGA  
 CGGTATCCGTTACGGCTATCGCGCAGAAGATGCAACCAACCTTGATGAAATCTATGTAAACAGCCGAAG  
 CCAAGGTTTTTGGTGAAGAGGTAAACGTCGTATCATGCTGGGTACTTTCAGTCTTTCATCAGGTTACTA  
 TGATGCCTACTACAAAAGGCTGGTCAAGTCCGTACCCTCATCATTCAAGATTTGAAAAAGTCTTCGC  
 GGATTACGATTTGATTTTGGGTCCAACCTGCTCCAAGTGTTCCTATGACTTGGATTCTCTCAACCATGA  
 CCCAGTTGCCATGTAAGTACCGGACCTATTGACCATACCTGTAAACTTGGCAGGACTGCCTGGAATTTT  
 GATTCCTGCTGGATTCTCTCAAGGTCTACCTGTGCGACTCCAATTGATTGGTCCCAAGTACTCTGAGGA  
 AACCATTTACCAAGCTGCTGCTGCTTTTGAAGCAACAACAGACTACCACAAACAACAACCCGTGATTTT  
 TGGAGGTGACAAC

**SP028 amino acid (SEQ ID NO:42)**

TFNNKTIIEELHNLVSKEISATELTQATLENIKSREEALNSFVTIAEEQALVQAKAIDEAGIDADNVLS  
 GIPLAVKDNISTDGILTTAASKMLYNYEPIFDATAVANAKTKGMI VVGKTNMDEFAMGSSGETSHYGAT  
 KNAWNHKSVPGGSSSGSAAAVASGQVRLSLGSDTGGSIRQPAAFNGIVGLKPTYGTVSRFGLIAFGSSL

Table 1

57

DQIGPFAPTVKENALLLNIAISEDADKSTAPVRIADFTSKIGQDIKGMKIALPKEYLLEGIDPEVKET  
ILNAAKHFEKLGAIVEEVSLPHSKYGVAVVYIIASSEASSNLQRFDGIRYGYRAEDATNLDEIYVNSRS  
QGFGEVVKRRIMLGTFSLSSGYDDAYYKKAGQVRTLIIQDFEKVFADYDLILGPTAPSVAYDLDSLNDH  
PVAMYLADLLTIPVNLAGLPGISIPAGFSQGLPVGLQLIGPKYSEETIYQAAAAFEATTDYHKQPVIF  
GGDN

**SP030 nucleotide (SEQ ID NO:43)**

CTTTACAGGTAAACAACACTACAAGTCGGCGACAAGGCGCTTGATTTTTCTCTTACTACAACAGATCTTTC  
TAAAAAATCTCTGGCTGATTTTGATGGCAAGAAAAAGTCTTGAGTGTCTGTTCCCTTCTATCGATACAGG  
CATCTGCTCAACTCAAACACGTCGTTTTAATGAAGAATTGGCTGGACTGGACAACACGGTCGTATTGAC  
TGTTTCAATGGACCTACCTTTTGCTCAAAAACGTTGGTGGCTGCTGAAGGCCTTGACAATGCCATTAT  
GCTTTCAGACTACTTTGACCATTCTTTGGGCGCGATTATGCCCTCTTGATCAACGAATGGCACCTATT  
AGCACGCGCAGTCTTTGTCTCGATACTGACAATACGATTGCTACGTTGAATACGTGGATAATATCAA  
TTCTGAGCCAAACTTCGAA

**SP030 amino acid (SEQ ID NO:44)**

FTGKQLQVGDKALDFSLTTDLSSKSLADFDGKKKVLSVVPISIDTGICSTQTRRFNEELAGLDNTVVLT  
VSMDLPPAQKRWCGAEGLDNAIMLSDFDHSFGRDYALLINEWHLLARAVFVLDTDNTIRYVEYVDNIN  
SEPNE

**SP031 nucleotide (SEQ ID NO:45)**

CCAGGCTGATACAAGTATCGCAGACATTCAAAAAAGAGGCGAACTGGTTGTCTGGTGTCAAACAAGACGT  
TCCCAATTTTGGTTACAAGATCCCAAGACCGGTACTTATTCTGGTATCGAAACCGACTTGCCCAAGAT  
GGTAGCTGATGAACTCAAGGTCAAGATTCGCTATGTGCCGGTTACAGCACAAACCCGCGCCCCCTTCT  
AGACAAATGAACAGGTGATATGGATATCGCGACCTTTACCATCACGGACGAACGCAAAAACTCTACAA  
CTTTACCAGTCCCTACTACACAGACGCTTCTGGATTTTTGGTCAATAAATCTGCCAAAAATCAAAAAGAT  
TGAGGACCTAAACGGCAAAACCATCGGAGTCGCCCCAAGGTTCTATCACCCAACGCCTGATTACTGAACT  
GGGTAAAAAGAAAGGTCTGAAGTTTAAATTCGTCGAACTTGGTTCTTACCCAGAATTGATTACTTCCCT  
GCACGCTCATCGTATCGATACCTTTTCCGTTGACCGCTCTATTCTATCTGGCTACACTAGTAAACGGAC  
AGCACTACTAGATGATAGTTTCAAGCCATCTGACTACGGTATTGTTACCAAGAAATCAAATACAGAGCT  
CAACGACTATCTTGATAACTTGGTTACTAAATGGAGCAAGGATGGTAGTTTGCAGAACTTTATGACCG  
TTACAAGCTCAAACCATCTAGCCATACTGCAGAT

**SP031 amino acid (SEQ ID NO:46)**

QADTSIADIQKRGELVVGKQDVPNFGYXDPKGTGYSIGIETDLAKMVADELKVKIRYVPVTAQTRGPLL  
DNEQVMDIATFTITDERKKLYNFTSPYYTDASGFLVNKSAKIKKIEDLNGKTIGVAQGSITQRLITEL  
GKKKGLKFKFVELGSPYELITSLHAHRIDTFSVDRSILSGYTSKRTALLDDSFKPSDYGIVTKKSNTL  
NDYLDNLVTKWSKDGSLQKLYDRYKLPSSHTAD

**SP032 nucleotide (SEQ ID NO:47)**

GTCTGTATCATTGAAAACAAAGAAACAAACCGTGGTGTCTTgACTTTCATCTCTCAAGACCAAAT  
CAAACCAGAAATTGGACCGTGTCTTCAAGtCAGTGAAGAAATCTCTTAATGTTCCAGGTTTCCGTAAAGG  
TCACCTTCCACGCCCTATCTTCGACCAAAAATTTGGTGAAGAAGCTCTTTATCAAGATGCAATGAACGC  
ACTTTTGCCAAACGCTTATGAAGCAGCTGTAAAAGAAGCTGGTCTTGAAGTGGTTGCCCCAACCAAAAAT  
TGACGTAACCTCAATGGAAAAAGGTCAAGACTGGGTTATCACTGCTGAAGTCGTTACAAAACCTGAAGT  
AAAATTGGGTGACTACAAAACCTTGAAGTATCAGTTGATGTAGAAAAAGAAGTAACAGCCTGATGT  
CGAAGAGCGTATCGAACGCGAACGCAACAACCTGGCTGAATTGGTTATCAAGGAAGCTGCTGCTGAAAA  
CGGCGACACTGTTGTGATCGACTTCGTTGGTTCTATCGACGGTGTGTAATTTGACGGTGGAAAAGGTGA  
AACTTCTCACTTGGACTTGGTTTCAGGTCAATTCATCCCTGGTTTCGAAGACCAATTGGTAGGTCACTC  
AGCTGGCGAAACCGTTGATGTTATCGTAACATTCCCAGAAGACTACCAAGCAGAAGACCTTGCAGGTAA  
AGAAGCTAAATTCGTGACAACTATCCACGAAGTAAAAGCTAAAGAAGTTCCGGCTCTTGACGATGAACT  
TGCAAAAGACATTGATGAAGAAGTTGAAACACTTGCTGACTTGAAAGAAAAATACAGCAAAGAATTGGC  
TGCTGCTAAAGAAGAAGCTTACAAAGATGCAGTTGAAGGTGCAGCAATTGATACAGCTGTAGAAAATGC  
TGAAATCGTAGAACTTCCAGAAGAAATGATCCATGAAGAAGTTACCGTTTCAAGTAAATGAATTCCTTG  
GAATTTGCAACGTCAAGGGATCAACCCTGACATGTACTTCCAAATCACTGGAACACTCAAGAAGACCT  
TCACAACCAATACCAAGCAGAAGCTGAGTCACGTACTAAGACTAACCTTGTTATCGAAGCAGTTGCCAA  
AGCTGAAGGATTTGATGCTTCAGAAGAAGAAATCCAAAAAGAAGTTGAGCAATTGGCAGCAGACTACAA

Table 1

CATGGAAGTTGCACAAGTTCAAAACTTGCTTTTCAGCTGACATGTTGAAACATGATATCACTATCAAAAA  
AGCTGTTGAATTGATCACAAGCACAGCAACAGTAAAA

**SP032 amino acid (SEQ ID NO:48)**

SVSFENKETNRGVLFTTISQDQIKPELDRVFKSVKSLNVPGRKGLPRPIFDQKFGEEALYQDAMNA  
LLPNAYEAAVKEAGLEVVAQPKIDVTSMEKGQDWVITAEEVVKPEVKLGDKYKNLEVSVDVEKEVTDADV  
EERIERERNLAEELVIKEAAAENGDTVIDFVGSIDGVEFDGGKGENFSLGLSGQFIPGFEDQLVGH  
AGETVDVIVTFPEDYQAEDLAGKEAKFVTTIHEVKAKEVPALDDELAKDIDEEVETLADLKEKYSKELA  
AAKEEAYKDAVEGAAIDTAVENAEIVELPEEMIHEEVHRSVNEFLGNLQRQGINPDMYFQITGTTQEDL  
HNQYQAEAESRTKTNLVIEAVAKAEGFDASEEEIQKEVEQLAADYNMEVAQVQNLLSADMLKHDITIKK  
AVELITSTATVK

**SP033 nucleotide (SEQ ID NO:49)**

TGGTCAAAAGGAAAGTCAGACAGGAAAGGGGATGAAATTGTGACCAGTTTTATCCTATCTACGCTAT  
GGTTAAGGAAGTATCTGGTGACTTGAATGATGTTCCGATGATTCAGTCAAGTAGTGGTATTCACCTCTT  
TGAACCTTCGGCAAATGATATCGCAGCCATCTATGATGCAGATGTCTTGTGTACCATTCTCATACT  
CGAATCTTGGGCAGGAAGTCTGGATCCAAATCTAAAAAATCCAAAGTGAAGGTCTTAGAGGCTTCTGA  
GGGAATGACCTTGAACGTGTCCCTGGACTAGAGGATGTGGAAGCAGGGGATGGAGTTGATGAAAAAAC  
GCTCTATGACCTCACACATGGCTAGATCCTGAAAAGCTGGAGAAGAAGCCCAAATTATCGCTGATAA  
ACTTTCAGAGGTGGATAGTGAGCATAAAGAGACTTATCAAAAAATGCGCAACCTTTATCAAAAAAGCT  
CAGGAAT

**SP033 amino acid (SEQ ID NO:50)**

GQKESQTGKGMKIVTSFYPIYAMVKEVSGDLNDVRMIQSSSGIHSFEPSSANDIAAIYDADVYVHSHTL  
ESWAGSLDPNLKKSQVLEASEGMTLERVPGLEDVEAGDGVDEKTLYPDHTWLDPEKAGEEAQIIADK  
LSEVDSEHKETYQKNAQPLSKLRN

**SP034 nucleotide (SEQ ID NO:51)**

GAAGGATAGATATATTTTAGCATTTGAGACATCCTGTGATGAGACCAGTGTGCGCGTCTTGAAAAACGA  
CGATGAGCTCTTGTCCAATGTCATTGCTAGTCAAAATGAGAGTCACAAACGTTTTGGTGGCGTAGTGCC  
CGAAGTAGCCAGTCTGTCACCATGTCGAGGTCAATACAGCCTGTATCGAGGAGGCATTGGCAGAAGCAGG  
GATTACCGAAGAGGACGTGACAGCTGTTGCGGTTACCTACGGACCAGGCTTGGTTCGGAGCCTTGCTAGT  
TGGTTTTGTGCTGAGCTGCCAAGGCCTTTGCTTGGGCTCACGGACTTCCACTGATTCTGTTAATCATATGGC  
TGGGCACCTCATGGCAGCTCAGAGTGTGGAGCCTTTGGAGTTTCCCTTGCTAGCCCTCTTGGTCAGCGG  
CGGACACACAGAGTTGGTTTATGTTTCGGAGGCAGGAGATTATAAGATTGTTGGGGAAACCCGTGATGA  
TGCGGTTGGTGAGGCTTATGATAAGGTCGGCCGTGTCATGGGCTTGACCTATCCTGCAGGTCGTGAGAT  
TGACGAGCTGGCTCATCAGGGGCAGGATATTTATGATTTCCCCCGTGCCATGATTAAGGAAGATAATCT  
GGAGTTCTCCTTCTCAGGTTTGAAATCTGCCTTTATCAATCTTCATCACAAATGCCGAGCAAAAGGGAGA  
AAGCCTGTCTACAGAAGATTTGTGTGCTTCCTTCCAAGCAGCAGTTATGGACATTCTCATGGCAAAAAC  
CAAGAAGGCTTTGGAGAAATATCCTGTTAAATCCTAGTTGTGGCAGGTGGTGTGGCAGCCAATAAAGG  
TCTCAGAGAACGCCTAGCAGCCGAAATCACAGATGTCAAGGTTATCATCCCCCTCTGCGACTCTGCGG  
AGACAATGCAGGTATGATTGCCTATGCCAGCGTCAGCNAGTGGAACAAAGAAAACCTCGCAGGCTGGGA  
CCTCAATGCCAAACCAAGTCTTGCCTTGATACCATGGAA

**SP034 amino acid (SEQ ID NO:52)**

KDRYILAFETSCDETSVAVLKNDDLLSNVIASQIESHKRFGGVVPEVASRHHVEVITACIEEALAEAG  
ITEEDVTAVAVTYGPGVLGALLVGLSAAKFAWAHGLPLIPVNHMAGHLMAAQSVPELEFPLLALLVSG  
GHTELVVYSEAGDYKIVGETRDDAVGEAYDKVGRVMGLTYPAGREIDELAHQGGDIYDFPRAMIKEDNL  
EFSFSGLKSAFINLHHNAEQKGESLSTEDLCASFQAAMVMDILMAKTKKALEKYPVKILVVAGGVAANKG  
LRERLAAEITDVKVIIPLRLCGDNAGMIAYASVSXWNKENFAGWDLNAKPSLAFDTME

**SP035 nucleotide (SEQ ID NO:53)**

GGTAGTTAAAGTTGGTATTAACGGTTTCGGACGTATCGGTGCTCTTGCTTTCCGTCGTATCCAAAACGT  
AGAAGGTGTTGAAGTTACACGCATCAACGACCTTACAGATCCAGTTATGCTTGACACTTGTGAAATA  
CGACACAACCTCAAGGTCGTTTCGACGGTACTGTTGAAGTTAAAGAAGGTGGATTGTAAGTTAACGGTAA  
ATTTCATCAAAGTTTCTGCTGAACGTGATCCAGAACAAATCGACTGGGCTACTGACGGTGTAGAAATCGT  
TCTTGAAGCTACTGGTTTCTTTGCTAAGAAAGAAGCAGCTGAAAAACACCTTAAAGGTGGAGCTAAAAA

Table 1

AGTTGTTATCACTGCTCCTGGTGGAAACGACGTTAAACAGTTGTATTCAACACTAACCACGACGTTCT  
TGACGGTACTGAAACAGTTATCTCAGGTGCTTCATGTACTACAACTGCTTGGCTCCAATGGCTAAAGC  
TCTTCAAGACAACCTTTGGTGTGTTGAAGGATTGATGACTACTATCCACGCTTACACTGGTGACCAAAT  
GATCCTTGACGGACCACACCGTGGTGGTGACCTTCGCCGTGCTCGCGCTGGTGTGCAAAACATCGTTCC  
TAACTCAACTGGTGTGCAAAAGCTATCGGTCTTGTAATCCCAGAATTGAATGGTAAACTTGACGGATC  
TGCACAACCGCTTCCAACCTCCAACCTGGATCAGTTACTGAATTGGTAGCAGTTCTTGAAAAAGACGTTAC  
TGTTGATGAAGTGAACGCAGCTATGAAAGCAGCTTCAAACGAATCATACGGTTACACAGAAGATCCAAT  
CGTATCTTCAGATATCGTAGGTATGTCTTACGGTTCATTGTTTGACGCAACTCAAACCTAAAGTTCTTGA  
CGTTGACGGTAAACAATTGGTTAAAGTTGTATCATGGTACGACAACGAAATGTCATACACTGCACAACCT  
TGTTCTGACTCTTGGAATACTTCGCAAAAATTGC

**SP035 amino acid (SEQ ID NO:54)**

VVKVGINGFGRIGRLAFRRIQNVEGVETTRINDLTDPMVLAHLLKYDITQGRFDGTVEVKEGGFEVNGK  
FIKVSARDEPEQIDWATDGEIVLEATGFFAKKEAAEKHLKGGAKKVITAPGNDVKTVVFNTNHDVL  
DGTETVISGASCTTNCLAPMAKALQDNFVVEGLMTTIHAYTGDQMILDGPHRGDLRRRAGAANIVP  
NSTGAAKAIGLVIPELNGKLDGSAQVRPTPTGSVTELVAVLEKNVTVDENVNAAMKAASNESYGYTEDPI  
VSSDIVGMSYGSFLDATQTKVLDVDGKQLVKVSWYDNEMSYTAQLVRTLGLLRKNC

**SP036 nucleotide (SEQ ID NO:55)**

TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGTTAAGGAAAATAATCGTGTTCCTATATAGATGG  
AAAACAAGCGACGCAAAAAACGGAGAATTTGACTCCTGATGAGGTTAGCAAGCGTGAAAGGAATCAATGC  
TGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTCACTTCACATGGCGACCACTATCATTATTA  
CAATGGTAAGGTTTCCTTATGACGCTATCATCAGTGAAGAATTACTCATGAAAGATCCAAACTATAAGCT  
AAAAGATGAGGATATTGTTAATGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGGAAAAACTATGT  
TTACCTTAAGGATGCTGCCACGCGGATAACGTCCGTACAAAAGAGGAAATCAATCGACAAAAACAAGA  
GCATAGTCAACATCGTGAAGGTGGAACCTCAAGAAACGATGGTGTCTGTCCTTGGCAGCTTCGCAAGG  
ACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGGATACTGGTGATGCTTA  
TATCGTTTCCTCATGGAGATCATTACCATTACATTCCTAAGAATGAGTTATCAGCTAGCGAGTTGGCTGC  
TGCAGAAGCCTTCCTATCTGGTCGAGGAAATCTGTCAAATTCAGAACCTATCGCCGACAAAATAGCGA  
TAACACTCAAGAACAACACTGGGTACCTTCTGTAAGCAATCCAGGAACCTACAAATACTAACACAAGCAA  
CAACAGCTAACACTAACAGTCAAGCAAGTCAAAGTAATGACATTGATAGTCTCTTGAAACAGCTCTACAA  
ACTGCCCTTTGAGTCAACGACATGTAGAATCTGATGGCCTTGTCTTTGATCCAGCACAAATCACAAGTCG  
AACAGCTAGAGGTGTTGCAGTGCCACACGGAGATCATTACCACTTCATCCCTTACTCTCAAATGTCTGA  
ATTGGAAGAACGAATCGCTCGTATTATTCCTTCTCGTTATCGTTCAAACCATTTGGGTACCAGATTCAAG  
GCCAGAACAACCAAGTCCACAACCGACTCCGGAACCTAGTCCAGGCCCGCAACCTGCACCAAACTTTAA  
AATAGACTCAAATCTCTCTTTGGTTAGTCAGCTGGTACGAAAAGTTGGGGAAGGATATGTATTCGAAGA  
AAAGGGCATCTCTCGTTATGTCTTTGCGAAAGATTTACCATCTGAAACTGTTAAAAATCTTGAAAGCAA  
GTTATCAAAAACAAGAGAGTGTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCCTCGTGACCA  
AGAATTTTATGATAAAGCATATAATCTGTAACTGAGGCTCATAAAGCCTTGTGTTGNAAAATAAGGGTCG  
TAATCTGATTTCCAAGCCTTAGACAAATATTAGAAGCTTGAATGATGAATCGACTAATAAAGAAAA  
ATTGGTAGATGATTTATTGGCATTCTTAGCACCATTACCCATCCAGAGCGACTTGGCAAACCAAATTC  
TCAAATTGAGTATACTGAAGACGAAGTTCGTATTGCTCAATTAGCTGATAAGTATACAACGTCAGATGG  
TTACATTTTGTATGAACATGATATAATCAGTGATGAAGGAGATGCATATGTAACGCCCTCATATGGGCCA  
TAGTCACTGGATTGGAAAAGATAGCCTTTCTGATAAGGAAAAAGTTGCAGCTCAAGCCTATACTAAAGA  
AAAAGGTATCCTACCTCCATCTCCAGACGCAGATGTTAAAGCAAATCCAACCTGGAGATAGTGCAGCAGC  
TATTTACAATCGTGTGAAAGGGGAAAAACGAATTCCTACTCGTTCGACTTCCATATATGGTTGAGCATA  
AGTTGAGGTTAAAAACGGTAATTTGATTATTCCTCATAAGGATCATTACCATAATATTAAATTTGCTTG  
GTTTGATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAAGATTTGTTTGCGACGATTAAGTA  
CTACGTAGAACACCCCTGACGAACGTCCACATTCTAATGATGGATGGGGCAATGCCAGTGAGCATGTGTT  
AGGCAAGAAAGACCACAGTGAAGATCCAAATAAGAACCTCAAAGCGGATGAAGAGCCAGTAGAGGAAAC  
ACCTGCTGAGCCAGAAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAGAAGCAGAAGT  
TTTGCTTGCGAAAGTAACGGATTCTAGTCTGAAAGCCAATGCAACAGAACTCTAGCTGGTTTACGAAA  
TAATTTGACTCTTCAAATTATGGATAACAATAGTATCATGGCAGAAGCAGAAAAATTACTTGCGTTGTT  
AAAAGGAAGTAATCCTTCATCTGTAAGTAAGGAAAAAATAAAC

**SP036 amino acid (SEQ ID NO:56)**

SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGDHYHY  
NGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGKYVIKVDGKYVYVYLDAAHADNVRTKEEINRQKQE

Table 1

60

HSQHREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIIPKNELSASELAA  
 AEAFLSGRGNLSNSRTYRRQNSDNTSRTNWVPVSVNPGTTNTNTSNNSTNSQASQSNIDISLLKQLYK  
 LPLSQRHVESDGLVFDPAQITSRTARGVAVPHGDHYHFIPYSQMSELEERIARIIPLRYSNHWVPDSR  
 PEQSPQPTPEPSPGPQAPNLKIDSNSSLVSQLVRKVGEGYVFEEKGISRYVFAKDLPSETVKNLESK  
 LSKQESVSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFXNKGRNSDFQALDKLLERLNDESTNKEK  
 LVDDLLAFLAPITHPERLGKPNSEQIEYTEDEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGH  
 SHWIGKDSLSDEKVAQAAYTKEKGILPPSPDADVKNPTGDSAAAIYNRVKGEKRIPLVRLPYMVEHT  
 VEVKNGNLIIPHKDHYHNKFAWFDDHTYKAPNGYTTLEDLFATIKYYVEHPDERPHSNDGWNASEHVL  
 GKDKHSEDPNKNFKADEEPVEETPAEPEVPQVETEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRN  
 NLTLQIMDNNSIMAEAEKLLALLKGSNPSSVSKEKIN

**SP038 nucleotide (SEQ ID NO:57)**

TACTGAGATGCATCATAATCTAGGAGCTGAAAAGCGTTTCAGCAGTGGCTACTACTATCGATAGTTTAA  
 GGAGCGAAGTCAAAAAGTCAGAGCACTATCTGATCCAAATGTGCGTTTTGTTCCCTTCTTTGGCTCTAG  
 TGAATGGCTTCGTTTTGACGGTGCTCATTCTGCGGTATTAGCTGAGAAATACAATCGTTCCTACCGTCC  
 TTATCTTTTAGGACAGGGGGAGCTGCATCGCTTAACCAATATTTTGAATGCAACAGATGTTACCACA  
 GCTGGAGAATAACAAGTTGTGTATGTTATCTCACCTCAGTGGTTCAGTAAAAATGGCTATGATCCAGC  
 AGCCTTCCAGCAGTATTTTAATGGAGACCAGTTGACTAGTTTTCTGAAACATCAATCTGGGGATCAGGC  
 TAGTCAATATGCAGCGACTCGCTTACTGCAACAGTTCCCAAACGTAGCTATGAAGGACCTGGTTCAGAA  
 GTTGGCAAGTAAAGAAGAAATTGTCGACAGCAGACAATGAAATGATTGAATTATTGGCTCGTTTTAATGA  
 ACGCCAAGCTTCCTTTTTTGGTTCAGTTTTTCGGTTAGAGGCTATGTTAACTACGATAAGCATGTAGCTAA  
 GTATTTAAAAATCTTGCCAGACCAGTTTTCTTATCAGGCAATAGAAGATGTTGTCAAAGCAGATGCTGA  
 AAAAAATACTTCCAATAATGAGATGGGAATGGAAAATTATTTCTATAATGAGCAGATCAAGAAGGATTT  
 GAAGAAATTAAAGGATTCTCAGAAAAGCTTTACCTATCTCAAGTCGCCAGAGTATAATGNNTTGCAGTT  
 GGTTTTAACACAGTTTCTAAATCTAAGGTAAACCCGATTTTTATCATTCCACCTGTTAATAAAAAATG  
 GATGNACTATGCTGGTCTACGAGAGGATATGTACCAACAAACGGTGCAGAAGATTTCGCTACCAGTTAGA  
 AAGTCAAGGTTTTACCAATATAGCAGATTTTTCTAAGGACGGCGGGGAGCCTTTCTTTATGAAGGACAC  
 CATTACCTTGGTTGGTTGGGTGGGTGGCTTTTGACAAGGCAGTTGATCCTTTCCATCCAATCCAC  
 ACCAGCTCCGACTTACCATCTGAATGAGCGCTTTTTTCAGCAAAGATTGGGCGACTTATGATGGAGATGT  
 CAAAGAA

**SP038 amino acid (SEQ ID NO:58)**

TEMHNLGAEKRSVAVATTIDSFKERSQKVRALSDPNVRFVFPFGSSEWLRFDGAHSAVLAEKYNRSYRP  
 YLLGQGGAAASLNQYFGMQMLPQLENKQVYVVISPOWFSKNGYDPAAFQQYFNGDQLTSFLKHQSGDQA  
 SQYAATRLQLQFPNVAMKDLVQKLASKEELSTADNEMIELLARFNERQASFFGQFSVRGYVNYDKHVAK  
 YLKILPDQFSYQAIEDVVKADAENKTSNNEMGMENFYNEQIKKDLKKLKDQSQSFTYLYKSPEYNXLQL  
 VLTQFSKSKVNPFIIPVKNKWMXYAGLREDMYQQTQVKIRYQLESQGFNIADFSKDGGEPPFMKDT  
 IHLGWLGLAFDCAVDPLSNPTPAPTYHLNERFFSKDWATYDGDVKE

**SP039 nucleotide (SEQ ID NO:59)**

GGTTTTGAGAAAGTATTTGCAGGGGGCCCTGATTGAGTCGATTGAGCAAGTGAAAAATGACCGTATTGT  
 GGAAATTACAGTTTCCAATAAAAAACGAGATTGGAGACCATATCCAGGCTACCTTGATTATCGAAATTAT  
 GGGGAAACACAGTAATATTTCTACTGGTCGATAAAAGCAGTCATAAAATCCTCGAAGTTATCAAACACGT  
 CGGCTTTTCACAAAATAGCTACCGCACCTTACTTCCAGGATCGACCTATATCGCTCCGCCAAGTACAAA  
 ATCTCTCAATCCTTTTACTATCAAGGATGAAAAGCTCTTTGAAATCCTGCAAACCAAGAACTAACAGC  
 AAAAAATCTTCAAAGCCTCTTTCAAGGTCTGGGACGCGATACGGCAAATGAATTGGAAAGGATACTGGT  
 TAGTGAAAAACTTTCCGCTTTCCGAAATTTTTTCAATCAAGAAACCAAGCCATGCTTGACTGAGACTTC  
 CTTGAGTCCAGTTCCTTTTGCAAATCAGGTGGGAGAGCCTTTTGCAAATCTTTCTGATTTGTTGGACAC  
 CTACTATAAGGATAAGGCTGAGCGCGACCGCGTCAAACAGCAGGCCAGTGAAGTTCGTCGTGTTGA  
 AAATGAAGTTTCAGAAAAACCGACACAACTCAAAAAACAGGAAAAAGAGTTACTGGCGACAGACAACGC  
 TGAAGAATTTTCGTCAAAAAGGAGAATTGCTGACAACCTTCCTCCACCAAGTGCTAACGACCAAGACCA  
 GGTATCCTAGACAACACTACTATACCAACCAACCTATCATGATTGCGCTTGATAAGGCTCTGACTCCCAA  
 CCAGAATGCCCAACGCTATTTTAAACGGTATCAGAACTCAAAGAAGCTGTCAAATACTTGACTGATTT  
 GATTGAAGAAACCAAAGCCACTATTCTCTATCTGGAAGTGTAGAAACCGTCCTCAACCAAGCTGGACT  
 GGAAGAAATCGCTGAAATCCGTGAAGAATTGATTCAAACAGGTTTTATCCGCAGAAGACAACGGGAGAA  
 AATCCAGAAACGCAAAAACTAGAACAAATATCTAGCAAGCGATGGCAAAACCATCATCTATGTCGGACG  
 AAACAATCTTCAAATGAGGAATTGACCTTTAAATGGCCCGCAAGGAGGAACCTTGGTTCCATGCTAA  
 GGACATTCCTGGAAGCCATGTTGTCTCTCAGGAAATCTTGACCATCTGATGCAGTCAAGACAGACGC

WO 98/18930

PCT/US97/19422

Table 1

61

AGCAGAGTTAGCTGCCTACTTCTCTCAAGGGCGCCTGTGCAATCTGGTGCAGGTAGATATGATTGAAGT  
 CAAAAAACAATAAACCAACTGGTGGAAAACCCGGCTTTGTCACTTACACAGGACAAAAGACCCTCCG  
 CGTCACACCAGACTCCAAAAAATTGCATCCATGAAAAAATCC

**SP039 amino acid (SEQ ID NO:60)**

VLRKYLOGALIESIEQVENDRIVEITVSNKNEIGDHIQATLIIIMGKHSNILLVDKSSHKILEVIKHV  
 GFSQNSYRLLPGSTYIAPPSTKSLNPFTIKDEKLFEILQTQELTAKNLQSLFQGLGRDTANELERILV  
 SEKLSAFRNFFNQETKPCLTETSFSPVPFANQVGEPPANLSDLLDTYYKDKAERDRVKQQASELIRRVE  
 NELQKNRHKLKKQEKELLATDNAEEFRQKGELLTFLHQVPNDQDQVILDNYVTNQPIIMIALDKALTPN  
 QNAQRYFKRYQKLKEAVKYLTDLIEETKATILYLESVETVLNQAGLEEIAEIREELIQTGFIRRRQREK  
 IQKRKKLEQYLASDGKTIIVVGRNNLQNEELTFKMARKEELWFHAKDIPGSHVVISGNLDPDAVKTD  
 AELAAVFSQGRLSNLVQVDMIEVKKLNKPTGGKPGFVITYTGQKTLRVTPDSKKIASMKKS

**SP040 nucleotide (SEQ ID NO:61)**

GACAACATTTACTATCCATACAGTAGAGTCAGCACCAGCAGAAGTGAAAGAAATCTTGAACAGTAGA  
 AAAAGACAACAATGGCTATATTCCCAACCTAATCGGTCTCTTGGCCAAATGCCCCGACTGTTTTAGAAGC  
 CTACCAAATGTCTCATCTATCCACCGTCGCAACAGCCTGACACCCGTTGAGCGTGAAGTGGTGCAAT  
 CACGGCAGCCGTGACCAATGGTTGTGCCTTCTGTGTCGCAGGTCACACAGCCTTTTCCATCAAACAAAT  
 CCAGATGAATGATGACTTGATTCAAGCTCTTCGCAATCGTACTCCAATTGAAACAGATCCATAAATTGGA  
 TACCCTAGCTAAGTTTACCTTGGCAGTTATCAATACCAAGGGTCGTGTAGGAGATGAAGCCTTGTCTGA  
 GTTTTTAGAAGCTGGCTACACTCAACAAAATGCCTTGGATGTGGTTTTTGGTGTGAGCCTAGCAATCCT  
 CTGTAACATATGCCAACAACTTAGCTAATACCAATTAATCCAGAATTGCAACCTTATGCC

**SP040 amino acid (SEQ ID NO:62)**

TTFTIHTVESAPAEVKEILETVEKDNNGYIPNLIGLLANAPTIVLEAYQIVSSIHRNLSLTPVEREVVQI  
 TAAVTNGCAFCVAGHTAFSIKQIQMNDLLIQLARNRTP IETDPKLDTLAKFTLAVINTKGRVGDALSE  
 FLEAGYTQQNALDVVFGVSLAILCNYANNLANTPINPELQPYA

**SP041 nucleotide (SEQ ID NO:63)**

GGCTAAGGAAAGAGTGATGTACTAGCTTATAAACAGGGGTTGTTTGAACGAGAGAGCAGGCCAAGCG  
 AGGTGTGATGCGCTGGCCTAGTCGTAGCAGTCCTTAATGGAGAACGGTTTGACAAGCCAGGAGAGAAAAT  
 TCCAGATGACACCGAATTAAAACTCAAGGGGGAGAACTCAAGTATGTCAGCCGTGGTGGTTTGAAACT  
 GGAAAAGGCCTTGCAGGTCTTTGATTTGTGCGGTGGATGGCGCGACTACGATTGATATCGGGGCTCTAC  
 TGGAGGTTTTTACCGATGTATGCTACAGAATAGTGCCAAGTTGGTCTTTGCAGTCGATGTTGGTACCAA  
 TCAGTTGGCTTGGAAATTACGCCAAGACCCACGAGTTGTGAGCATGGAGCAGTTCAATTTCCGCTATGC  
 TGAAAAGACTGATTTGAGCAGGAGCCGAGCTTTGCCAGTATTGATGTGAGTTTCATTTCCCTTAGTCT  
 GATTTTGCCAGCCTTGCACCGTGTCTTGGCTGATCAAGGTCAGGTGGTAGCACTTGTCAAACCTCAGTT  
 TGAGGCAGGACGTGAGCAGATTGGGAAAAATGGAATTATTCGAGATGCTAAGGTTTCATCAGAATGTCCT  
 TGAATCTGTAACAGCTATGGCAGTAGAGGTAGGTTTTTTCAGTCCTTGGCTTGGACTTTTCTCCCATCCA  
 AGGTGGACATGGAAATATTGAATTTTTAGCGTATTTGAAAAAAGAAAAGTCAGCAAGCAATCAGATTCT  
 TGCTGAGATTAAAGAAGCAGTAGAGAGGGCGCATAGTCAATTTAAAAATGAA

**SP041 amino acid (SEQ ID NO:64)**

AKERVDVLAYKQGLFETREQAKRGVMAGLVVAVLNGERFDKPGKEIPDDTELKLGKELKYVSRGGLKL  
 EKALQVFDLSVDGATTIDIGASTGGFTDVMLQNSAKLVFAVDVGTNQLAWKLRQDPRVVSMEQFNFRYA  
 EKTDFEQEPSFASIDVSFISLSLILPALHRVLADQGVVALVKPQFEAGREQIGKNGIIRDAKVHQNVL  
 ESVTAMAVEVGFVGLGLDFSPIQGGHGNIEFLAYLKKEKSASNQILAEIKEAVERAHSQFKNE

**SP042 nucleotide (SEQ ID NO:65)**

TTGTTCCATGAACTTGGTCGTCACCAAGCTGGTCAGGTTAAGAAAGAGTCTAATCGAGTTTCTTATAT  
 AGATGGTGATCAGGCTGGTCAAAAGGCAGAAAACCTTGACACCAGATGAAGTCAGTAAGAGGGAGGGGAT  
 CAACGCCGAACAAATNGTNATCAAGATTACGGATCAAGGTTATGTGACCTCTCATGGAGACCATTATCA  
 TTAATAATGGCAAGGTTCCCTTATGATGCCATCATCAGTGAAGAGCTCCTCATGAAAGATCCGAATTA  
 TCAGTTGAAGGATTGAGACATTGTCAATGAAATCAAGGTTGGTTATGTCATTAAAGGTAACGGTAAATA  
 CTATGTNTACCTTAAGGATGCAGCTCATGCGGATAGATATTTCGGACAAAAGAAGAGATTAAACGTCAGAA  
 GCAGGAACCGAGTCATAATCATAACTCAAGAGCAGATAATGCTGTTGCTGCAGCCAGAGCCCAAGGACG  
 TTATACAACGGATGATGGGTATATCTTCAATGCATCTGATATCATTGAGGACACGGGTGATGCTTATAT  
 CGTTCCTCACGGCGACCATTACCATTACATTCCTAAGAATGAGTTATCAGCTAGCGAGTTAGCTGCTGC

WO 98/18930

PCT/US97/19422

Table 1

62

AGAAGCCTATTGGAATGGGAAGCAGGGATCTCGTCCTTCTTCAAGTTCTAGTTATAATGCAAATCCAGC  
 TCAACCAAGATTGTCTAGAGAACCACAATCTGACTGTCACTCCAACCTATCATCAAAATCAAGGGGAAAA  
 CATTTCAAGCCTTTTACGTGAATTGTATGCTAAACCTTATCAGAACGCCATGTGGAATCTGATGGCCT  
 TATTTTCGACCCAGCGCAAATCACAAGTCGAACCGCCAGAGGTGTAGCTGTCCCTCATGGTAACCATTA  
 CCACCTTTATCCCTTATGAACAAATGTCTGAATTGGAAAAACGAATTGCTCGTATTATCCCCTTCGTTA  
 TCGTTCAAACCATTTGGGTACCAGATTCAAGACCAGAACAACCAAGTCCACAATCGACTCCGGAACCTAG  
 TCCAAGTCCGCAACCTGCACCAAATCCTCAACCAGCTCCAAGCAATCCAATTGATGAGAAATTGGTCAA  
 AGAAGCTGTTTCGAAAAGTAGGCGATGGTTATGTCTTTGAGGAGAATGGAGTTTCTCGTTATATCCCAGC  
 CAAGGATCTTTTCAGCAGAAACAGCAGCAGGCATTGATAGCAAACTGGCCAAGCAGGAAAGTTTATCTCA  
 TAAGCTAGGAGCTAAGAAAACCTGACCTCCCATCTAGTGATCGAGAATTTTACAATAAGGCTTATGACTT  
 ACTAGCAAGAATTACCAAGATTTACTTGATAATAAAGGTGCGACAAGTTGATTTTGAGGCTTTGGATAA  
 CCTGTTGGAACGACTCAAGGATGTCNCAAGTGATAAAGTCAAGTTAGTGANGATATTCTTGCCTTCTT  
 AGCTCCGATTCGTATCCAGAACGTTTAGGAAAACCAAATGCGCAAATTACCTACACTGATGATGAGAT  
 TCAAGTAGCCAAGTTGGCAGGCAAGTACACAACAGAAGACGGTTATATCTTTGATCCTCGTGATATAAC  
 CAGTGATGAGGGGGATGCCTATGTAACCTCCACATATGACCCATAGCCACTGGATTAATAAAGATAGTTT  
 GTCTGAAGCTGAGAGAGCGGCAGGCCAGGCTTATGCTAAAGAGAAAGGTTTGACCCCTCCTTCGACAGA  
 CCATCAGGATTACAGAAATACTGAGGCAAAAGGAGCAGAAGCTATCTACAACCGCGTGAAAGCAGCTAA  
 GAAGGTGCCACTTGATCGTATGCCTTACAATCTTCAATATACTGTAGAAGTCAAAAACGGTAGTTTAAT  
 CATACCTCATTATGACCATTACCATAACATCAAATTTGAGTGGTTTGACGAAGGCCTTTATGAGGCACC  
 TAAGGGGTATACTCTTGAGGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAACGTCC  
 GCATTACAGATAATGGTTTTGGTAACGCTAGCGACCATGTTCAAAGAAACAAAAATGGTCAAGCTGATAC  
 CAATCAAACGGAAAAACCAAGCGAGGAGAAACCTCAGACAGAAAAACCTGAGGAAGAAACCCCTCGAGA  
 AGAGAAACCGCAAAGCGAGAAACCAGAGTCTCCAAAACCAACAGAGGAACCAAGAATCACCAGAGGA  
 ATCAGAAGAACCTCAGGTCGAGACTGAAAAGGTTGAAGAAAAAAGTGAAGAGAGGCTGAAGATTTACTTGG  
 AAAAATCCAGGAT

**SP042 amino acid (SEQ ID NO:66)**

CSYELGRHQAGQVKESNRVSYIDGDAQGQKAENLTPDEVSKREGINAEQXVIKITDQGYVTSHGDHYH  
 YYNGKVPYDAIISEELMKDPNYQLKDSIDVNEIKGGYVIKVNKYYVYLKDAHADNIRTKEEIKRQK  
 QERSHNHNSRADNAVAARAQGRYTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNELASASELAA  
 EAYWNGKQGSRPSSSSSYNANPAQPRLSNHNLTVPPTYHQNQGENISSLLRELYAKPLSERHVESDGL  
 IFDPAQITSRTARGVAVPHGNHYHFIPEYQMSLEKRIARIIPLYRSNHWVPDSRPEQSPQSTPEPS  
 PSPQPAPNPQPAPSNPIDEKLVKEAVRKVGDGTVFEENGVSRYIPAKDLAETAAGIDSKLAKQESLSH  
 KLGAKKTDLPSSDREFYNKAYDLLARIHQDLDNKGKQVDFEALDNLRLKDVXSDKVKLVXDILAF  
 APIRHPERLGPNAQITYTDDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMTSHWIKKDSL  
 SEAERAAAQAYAKEKGLTPPSTDHODSGNTEAKGAEAIYNRVKAAKKVPLDRMPYNLQYTVVEKNGSLI  
 IPHYDHYHNIKFEWFDEGLYEAPKGYTLEDLLATVKYVVEHPNERPHSDNGFGNASDHVQRNKGQADT  
 NQTEKPSEKPKQTEKPEEETPREKPKQSEKPEPKPTPEEPEESPEESEEPQVETEKVEEKLREAEDLLG  
 KIQD

**SP043 nucleotide (SEQ ID NO:67)**

TTATAAGGGTGAATTAGAAAAAGGATACCAATTTGATGGTTGGGAAATTTCTGGTTTTCGAAGGTAAAA  
 AGACGCTGGCTATGTTATTAATCTATCAAAAGATACCTTTATAAAACCTGTATTCAAGAAAATAGAGGA  
 GAAAAAGGAGGAAGAAAATAAACCTACTTTTGATGTATCGAAAAAGAAAGATAACCCACAAGTAAACCA  
 TAGTCAATTAAATGAAAGTCACAGAAAAGAGGATTTACAAAGAGAAGAGCATTACAAAAATCTGATTC  
 AACTAAGGATGTTACAGCTACAGTTCTTGATAAAAACAATATCAGTAGTAATCAACTACTAACAATCC  
 TAATAAG

**SP043 amino acid (SEQ ID NO:68)**

YKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKKEENKPTFDVSKKKDNPQVNH  
 SQLNESHKEDLQREEHSQKSDSTKDVTATVLDKNNISSKSTTNPNK

**SP044 nucleotide (SEQ ID NO:69)**

GAATGTTCAAGGCTCAAGAAAGTTCAGGAAATAAAATCCACTTTATCAATGTTCAAGAAGGTGGCAGTGA  
 TGCGATTATCTTGAAGCAATGGACATTTTGCCATGGTGGATACAGGAGAAGATTATGATTTCCAGGA  
 TGGAAGTGATTCTCGCTATCCATGGAGAGAAGGAATTGAAACGTCTTATAAGCATGTTCTAACAGACCG  
 TGTCTTTTCGTCGTTTGAAGGAATTGGGTGTCCAAAACTTGATTTTATTTTGGTGACCCATACCCACAG  
 TGATCATATTGGAATGTTGATGAATTACTGTCTACCTATCCAGTTGACCGAGTCTATCTTAAGAAATA

Table 1

63

TAGTGATAGTCGTATTACTAATTCTGAACGCTCTATGGGATAATCTGTATGGCTATGATAAGGTTTTTACAGACTGCTGCAGAAAAAGGTGTTTCAGTTATTCAAAAATATCACACAAGGGGATGCTCATTTTCAGTTTGGGGACATGGATATTTCAGCTCTATAATTATGAAAAATGAACTGATTTCATCGGGTGAATTAAAGAAAAATTTGGGATGACAATTCCAATTCCCTTGATTAGCGTGGTGAAAGTCAATGGCAAGAAAAATTTACCTTTGGGGGCGATTTAGATAATGTTTCATGGAGCAGAAGACAAGTATGGTCTCTCATTGGAAAAAGTTGATTTGATGAAGTTTAATCATCACCATGATACCAACAAATCAAATACCAAGGATTTTCATTAAAAATTTGAGTCCGAGTTTGATGTTTCAAACCTTCGGATAGTCTACCTTGGAAAAATGGTGGTGATAGTGAGTATGTTAATTGGCTCAAAGACGAGGAATTGAGAGAATCAACGCAGCCAGCAAAGACTATGATGCAACAGTTTTTGATATTCGAAAAGACGGTTTTTGTCATATTTCAACATCCTACAAGCCGATTCCAAGTTTTCAAGCTGGTTGGCATAAGAGTGCATATGGGAAGTGGTGGTATCAAGCGCCTGATTCTACAGGAGAGTATGCTGTGCGTTGGAATGAAATCGAAGGTGAATGGTATTACTTTAAACCAAACGGGTATCTTGTTACAGAATCAATGGAAAAATGGAACAATCATTTGGTTCTATTTGACAGACTCTGGTGCTTCTGCTAAAAATTTGGAAGAAAAATCGCTGGAATCTGGTATTAATTTAAACAAAGAAAACAGATGGAAATTGGTTGGATTCAAGATAAAGAGCAGTGGTATTATTTGGATGTGATGGTTCTATGAAGACAGGATGGCTTCAATATATGGGGCAATGGTATTACTTTGCTCCATCAGGGGA

**SP044 amino acid (SEQ ID NO:70)**

NVQAQESSGNKIHFINVQEGSDAIILESNGHFAMVDTGEDYDFPDGSDSRYPWREGIETSYKHVLTDRVFRRLKELGVQKLDLFILVTHTHSDHIGNVDELLSTYPVDRVYLKKYSDSRITNSERLWDNLYGYDKVLQTAAEKGVSVIQNITQGDHFQFGDMDIQLYNYENETDSSGELKKIWDDNSNLSISVVKVNGKKIYLGGLDINVHGAEDKYGPLIGKVDLMKFNHHHTNKSNTKDFIKNLSPLIVQTSDSL PWKNGVDSEYVNWLKE RGIERINAASKDYDATVFDIRKDG FVNISTSYKPIPSFQAGWHKSAYGNWWYQAPDSTGEYAVGWNEIEGEWYYFNQTGILLQNWKKWNHWFYLTDSGASAKNWKKIAGIWIYYFNKENQMEIGWIQDKEQWYYL DVGSMKTGWLQYMGQWYYFAPSGE

**SP045 nucleotide (SEQ ID NO:71)**

CTTGGGTGTAACCCATATCCAGCTCCTTCCAGTCTTGTCTTACTACTTTGTCAATGAATTGAAAAACCATGAACGCTTGTCTGACTACGCTTCAAGCAACAGCAACTACAACCTGGGGATATGACCCTCAAACTACTTCTCCTTGACTGGTATGTACTCAAGCGATCCTAAGAATCCAGAAAAACGAATCGCAGAATTTAAAAACCTCATCAACGAAATCCACAAACGTGGTATGGGAGCTATCCTAGATGTCGTTTATAACACACAGCCAAAGT CGATCTCTTTGAAGATTGGAACCAAACTACTACCCTTTATGGATGCCGATGGCACACCTCGAAGTAGCTTTGGTGGTGGAGCGCTTGGGACAAACCCACTATGACCAAAACGGCTCCTAATTGACTCTATCAAATA CCTAGTTGATACCTACAAAGTGGATGGCTTCCGTTTCGATATGATGGGAGACCATGACGCCGCTTCTATCGAAGAAGCTTACAAGGCTGCACGCGCCCTCAATCCAAACCTCATCATGCTTGGTGAAGGTTGGAGAACCTATGCCGGTGATGAAAACATGCCTACTAAAGCTGCTGACCAAGATTGGATGAAACATACCGATACTGT CGCTGTCTTTTTCAGATGACATCCGTAACAACCTCAAATCTGGTTATCCAAACGAAGGTCAACCTGCCTTTATCACAGGTGGCAAGCGTGATGTCAACACCATCTTTAAAAATCTCATTTGCTCAACCAACTAAGTTTGAAGCTGACAGCCCTGGAGATGTCAATACATCGCAGCCCATGATAACTTGACCCTCTTTGACATCATTGCCCCAGTCTATCAAAAAAGACCCAAGCAAGGCTGAGAACTATGCTGAAATCCACCGTCGTTTACGACTTGGAAATCTCATGGTCTTGACAGCTCAAGGAAGTCCATTTATCCACTCCGGTCAGGAATATGGACGTACTAAACAATTCCTGACCCAGCCTACAAGACTCCAGTAGCAGAGGATAAGGTTCCAAACAAATCTCACTTGTTGCGTGATAAGGACGGCAACCCATTTGACTATCCTTACTTCATCCATGACTCTTACGATTCTAGTGA TGCAGTCAACAAGTTTGACTGGACTAAGGCTACAGATGGTAAAGCTTATCCTGAAAATGTCAAGAGCCGTGACTATATGAAAGGTTTGATTGCCCTTCGTCAATCTACAGATGCCTTCCGACTTAAGAGTCTTCAAGATATCAAAGACCGTGTCACCTCATCACTGTCCCAGGCCAAAATGGTGTGGAAAAAGAGGATGTAGTGAT TGGCTACCAAATCACTGCTCCAAACGGCGATATCTACGCAGTCTTTGTCAATGCGGATGAAAAAGCTCGCGAATTTAATTTGGGAAGTGCCTTTGCACATCTAAGAAATGCGGAAGTTTGGCAGATGAAAACCAAGCAGGACCAGTCGGAATTGCCAACCCGAAAGGACTTGAATGGACTGAAAAAGGCTTGAATTTGAATGCCCTTACAGCTACTGTTCTTCGAGTCTCTCAAAAATGGAAGTACCCATGAGTCAACTGCAGAAAGAGAAACCAGACTCAACCCCTTCCAAGCCTGAACATCAAAAATGAAGCTTCTCACCCTGCACATCAAGACCCAGCTCCAGAGCTAGACCTGATTCTACTAAACCAGATGCCAAAGTAGCTGATGCGGAAAATAAACCTAGCCAAGCTACAGCTGATTCAAGCTGAACAACCAGCACAAAGACACAAGCATCATCTGTAAAAGAAGCGGTTTCGAAACGAATCGGTAGAAAACCTTAGCAAGGAAAATATACCTGCAACCCAGATAAACAAGCTGAA

**SP045 nucleotide (SEQ ID NO:72)**

LGVTHIQLLPVLSYFVNELKNHERLSDYASSNSNYNWGYDPQNYFSLTGMYSDDPKNPEKRIAEFKNL INEIHKRGMGAILDVVYNHTAKVDLFEDELPNYYHFMADGTPRTSFGGGRLGTHHMTKRLLIDS IKYLVDTYKVDGFRFDDMMGDHDAASIEEAYKAARALNPNIIMLGEGWRTYAGDENMPTKAADQDWMKHTD TV

WO 98/18930

PCT/US97/19422

Table 1

64

AVFSDDIRNNLKSYPNEGQPAFITGGKRDVNTIFKNLIAQPTNFEADSPGDVVIQYIAAHDNLTFLDII  
 AQSIIKKDPSKAENYAEIHRRLRLGNLMVLTAQGTPIHSGQEYGRTKQFRDPAYKTPVAEDKVPNKS  
 LRLDKDGNPFDPYFIHDSYDSSDAVNKFDTWKATDGKAYPENVKSRDYMKGILALRQSTDAFRLKSLQD  
 IKDRVHLITVPGQNGVEKEDVIGYQITAPNGDIYAVFVNADEKAREFNLGTAFALHRLNAEVLADENQA  
 GPVGIANPKGLEWTEKGLKLNALTATVLRVSQNGTSHESTAEKPDSTPSKPEHQNEASHPAHQDPAPE  
 ARPDSTKPDAKVADAENKPSQATADSQAEQPAQEAQASSVKEAVRNESVENSSKENIPATPDKQAE

**SP046 nucleotide (SEQ ID NO:73)**

TAGTGATGGTACTTGGCAAGGAAAACAGTATCTGAAAGAAGATGGCAGTCAAGCAGCAAATGAGTGGGT  
 TTTNGATACTCATTATCAATCTTGGTTCTATATAAAAGCAGATGCTAACTATGCTGAAAATGAATGGCT  
 AAAGCAAGGTGACGACTATTTTACCTCAAATCTGGTGGCTATATGGCCAAATCAGAATGGGTAGAAGA  
 CAAGGGAGCCTTTTATTATCTTGACCAAGATGGAAAGATGAAAAGAAATGCTTGGGTAGGAACTTCCTA  
 TGTTGGTGCAACAGGTGCCAAAGTAATAGAAGACTGGGTCTATGATTCTCAATACGATGCTTGGTTTTA  
 TATCAAAGCAGATGGACAGCACGCAGAGAAAGAATGGCTCCAAATTAAGGGAAGGACTATTATTTCAA  
 ATCCGGTGGTTATCTACTGACAAGTCAGTGGATTAATCAAGCTTATGTGAATGCTAGTGGTGCCAAAGT  
 ACAGCAAGGTTGGCTTTTTGACAAACAATACCAATCTTGGTTTTACATCAAAGAAAATGGAACTATGC  
 TGATAAAGAATGGATTTTCGAGAATGGTCACATATTATCTAAAATCCGGTGGCTACATGGCAGCCAA  
 TGAATGGATTTGGGATAAGGAATCTTGGTTTTATCTCAAATTTGATGGGAAAATGGCTGAAAAAGAATG  
 GGTCTACGATTCTCATAGTCAAGCTTGGTACTACTTCAAATCCGGTGGTTACATGACAGCCAATGAATG  
 GATTTGGGATAAGGAATCTTGGTTTTACCTCAAATCTGATGGGAAAATAGCTGAAAAAGAATGGGTCTA  
 CGATTCTCATAGTCAAGCTTGGTACTACTTCAAATCTGGTGGCTACATGGCGAAAAATGAGACAGTAGA  
 TGGTTATCAGCTTGGAAAGCGATGGTAAATGGCTTGGAGGAAAAACTACAAATGAAAATGCTGCTTACTA  
 TCAAGTAGTGCTGTACAGCCAATGTTTATGATTCAGATGGTGAAAAGCTTTCCTATATATCGCAAGG  
 TAGTGTCGTATGGCTAGATAAGGATAGAAAAAGTGATGACAAGCGCTTGGCTATTACTATTTCTGGTTT  
 GTCAGGCTATATGAAAACAGAAGATTTACAAGCGCTAGATGCTAGTAAGGACTTTATCCCTTATTATGA  
 GAGTGATGGCCACCGTTTTTATCACTATGTGGCTCAGAATGCTAGTATCCCAGTAGCTTCTCATCTTTC  
 TGATATGGAAGTAGGCAAGAAATATTATTCGGCAGATGGCCTGCATTTTGATGGTTTTAAGCTTGAGAA  
 TCCCTTCCTTTTCAAAGATTTAACAGAGGCTACAACTACAGTGCTGAAGAATTGGATAAGGTATTTAG  
 TTTGCTAAACATTAACAATAGCCTTTTGGAGAACAAGGGCGCTACTTTTAAGGAAGCCGAAGAACATTA  
 CCATATCAATGCTCTTTATCTCCTTGGCCATAGTGCCCTAGAAAAGTAAGTGGGGAAGAAGTAAATGTC  
 CAAAGATAAGAATAATTTCTTTGGCATTACAGCCTATGATACGACCCCTTACCTTTCTGCTAAGACATT  
 TGATGATGTGGATAAGGGAATTTTAGGTGCAACCAAGTGGATTAAGGAAAATATATCGATAGGGGAAG  
 AACTTTCTTGGAAACAAGGCTTCTGGTATGAATGTGGAATATGCTTCAGACCCCTTATTGGGGCGAAAA  
 AATTGCTAGTGTGATGATGAAAATCAATGAGAAGCTAGGTGGCAAAGAT

**SP046 amino acid (SEQ ID NO:74)**

SDGTWQKGQYLKEDGSQAANEVXDTHYQSWFYIKADANYAENEWLKQDDYFYLKSGGYMAKSEWVED  
 KGAFYYLDQDGKMKRNAWVGTSYVGATGAKVIEDWVYDSQYDAWFYIKADGQHAKEWELQIKGKDYFYK  
 SGGYLLTSQWINQAYVNASGAKVQQGWLFDKQYQSWFYIKENGNADKEWIFENGHYLYLKSGGYMAAN  
 EWIWDKESWFYLFKDGKMAKEWVYDSHSQAWYFYKSGGYMTANEWIWDKESWFYLFKSDGKIAKEWVY  
 DSHSQAWYFYKSGGYMAKNETVDGYQLGSDGKWLGGKTTNENAAYYQVVPVTANVYDSGDEKLSYISQ  
 SVVWLDKDRKSDDKRLAITISGLSGYMKTEDLQALDASKDFIPYYESDGHFRFYHYVAQNASIPVASHLS  
 DMEVGKKYYSADGLHFDGFKLENPFLFKDLTEATNYSAEELDKVFSLLNINNSLLENKGATFKEAEHY  
 HINALYLLAHSALSNWGRSKIAKDKNNFFGITAYDTPYLSAKTFDDVDKGILGATKWIKENYIDRGR  
 TFLGNKASGMNVEYASDPYWGEKIASVMMKINEKLGGKD

**SP048 nucleotide (SEQ ID NO:75)**

TGGGATTCAATATGTCAGAGATGATACTAGAGATAAAGAAGAGGGAATAGAGTATGATGACGCTGACAA  
 TGGGGATATTATTGTAAAAGTAGCGACTAAACCTAAGGTAGTAACCAAGAAAATTTCAAGTACGCGAAT  
 TCGTTATGAAAAGATGAAACAAAAGACCGTAGTGAAAATCCTGTTACAATTGATGGAGAGGATGGCTA  
 TGTAACTACGACAAGGACCTACGATGTTAATCCAGAGACTGGTTATGTTACCGAACAGGTTACTGTTGA  
 TAGAAAAGAAGCCACGGATACAGTTATCAAAGTTCAGCTAAAAGCAAGGTTGAAGAAGTTCTTGTTC  
 ATTTGCTACTAAATATGAAGCAGACAATGACCTTCTGCAGGACAGGAGCAAGAGATTACTCTAGGAAA  
 GAATGGGAAAACAGTTACAACGATAACTTATAATGTAGATGGAAAGAGTGGACAAGTAAGTGAAGTAC  
 TTTAAGTCAAAAAAAGACTCTCAAACAAGAGTTGTTAAAAAAGaACCArKCCCCAAGTTCTTGTCCA  
 AGAAATTCCAATCGAAACAGAATATCTCGATGGCCCAACTCTTGATAAAaGTCAAGAAGTAGAAGAAGT  
 AGGAGAAATTGGTAAATTACTCTTACTACAATCTATACTGGTAGATGAACGTGATGGAACAATTGAAGA  
 AACTACTTCTCGTCAAATTACTAAAGAGATGGTAAAAAGACGTATAAGGAGAGGGACGAGAGAACCCTGA

WO 98/18930

PCT/US97/19422

Table 1

65

AAAAGTTGTTGTTCCCTGAGCAATCATCTATTCCTTCGTATCCTGTATCTGTTACATCTAACCAAGGAAC  
 AGATGTAGCAGTAGAACCAGCTAAAGCAGTTGCTCCAACAACAGACTGGAAACAAGAAAATGGTATGTG  
 GTATTTTTATAATACTGATGGTTCCATGGCAACAGGTTGGGTACAAGTTAATAGTTCATGGTACTACCT  
 CAACAGCAACGGTTCTATGAAAGTCAATCAATGGTTCCAAGTTGGTGGTAAATGGTATTATGTAAATAC  
 ATCGGGTGAGTTAGCGGTCAATACAAGTATAGATGGCTATAGAGTCAATGATAATGGTGAATGGGTGCG  
 T

**SP048 amino acid (SEQ ID NO:76)**

GIQYVRDDTRDKEEGIEYDDADNGDIIVKVATKPKVVTKKISSSTRIRYEKDETKDRSENPVITIDGEDGY  
 VTTTRTYDVNPETGYVTEQVTVDRKEATDTVIKVPKSKVEEVLPFATKYEADNDLSAGQEIEITLKG  
 NGKTVTTITYNVDGKSGQVTESTLSQKKDSQTRVVKRTRXPQVLVQEIPETIEYLDGPTLDKSQEVEEV  
 GEIGKLLLLQSILVDERDGTIEETTSRQITKEMVKRRIRRGRTREPEKVVVPEQSSIPSYPSVTSNQG  
 DVAVEPAKAVAPTDDWKQENGMMWFYNTDGSMTGWVQVNSSWYYLNSNGSMKVNQWFQVGGKWWYVNT  
 SGELAVNTSIDGYRVNDNGEWR

**SP049 nucleotide (SEQ ID NO:77)**

GGATAATAGAGAAGCATTAAAAACCTTTATGACGGGTGAAAATTTTATCTCCAACATTATCTAGGAGC  
 ACATAGGGAAGAATAAATGGAGAGCATGGCTATACCTCCGTGTTGGGCACCTAATGCTCAGGCTGT  
 TCACTTGGTTGGTGATTTTACCAACTGGATTGAAAATCAGATTCCAATGGTAAGAAATGATTTTGGGGT  
 CTGGGAAGTCTTTACCAATATGGCTCAAGAAGGGCATATTTACAAATATCATGTACACGTCAAAATGG  
 TCATCAACTGATGAAGATTGACCTTTTGGCTGTCAGGTATGAGGCTCGTCCAGGAACAGGGGCAATCGT  
 AACAGAGCTTCCTGAGAAGAAATGGAAGGATGGACTTTGGCTGGCACGAAGAAAACGTTGGGGCTTTGA  
 AGAGCGTCCTGTCAATATTTATGAAGTTCACGCTGGATCATGGAAAAGAAATCTGATGGCAGTCCTTA  
 TAGTTTTGCCCAGCTCAAGGATGAACTCATTCCCTTATCTCGTTGAAATGAACTATACTCATATTGAGTT  
 TATGCCCTTGATGTCCCATCCTTTGGGCTTGAGTTGGGGGTATCAGCTTATGGGTACTTCGCTTTAGA  
 GCATGCTTATGGCCGACCAGAGGAGTTTCAAGATTTTGTG

**SP049 amino acid (SEQ ID NO:78)**

DNREALKTFMTGENFYLOHYLGAHREELNGEHGYTFRVWAPNAQAVHLVGDFTNWIENQIPMVRNDFGV  
 WEVFTNMAQEGHIYKYHVTRQNGHQLMKIDPFAVRYEARPGTGAIVTELPEKKWKDGLWLARRKRWGF  
 ERPVNIYEVHAGSWKRNSDGSFSAQLKDELPYLVEMNYTHIEFMPLMSHPLGLSWGYQLMGYFALE  
 HAYGRPEEFQDFV

**SP050 nucleotide (SEQ ID NO:79)**

AGATTTTGTGCGAGGAGTGTCATACCCATAATATTGGGGTTATTGTGGACTGGGTACCAGNTCACTTTAC  
 CATCAACGATGATGCCTTAGCCTATTATGATGGGACACCGACTTTTGAATACCAAGACCATAATAAGGC  
 TCATAACCATGGTTGGGGTGCCCTTAATTTTGACCTTGGAAGAAATGAAGTCCAGTCCTTCTTAATTTT  
 TTGCATTAAGCATTGGATTGATGTCTATCATTGGATGGTATTTCGTGTGGATGCTGTTAGCAACATGCT  
 CTATTTGGACTATGATGATGCTCCATGGACACCTAATAAGATGGCGGAAATCTCAACTATGAAGGTTA  
 TTATTTCTTTCAGCGCTTGAATGAGGTTATTAAGTTAGAATATCCAGATGTGATGATGATTGCAGAAGA  
 AAGTTCGTCTGCGATCAAGATTACGGGAATGAAAGAGATTGGTGGTCTAGGATTTGACTACAAATGGAA  
 CATGGGCTGGATGAATGATATCCTCCGTTTCTACGAAGAAGATCCGATCTATCGTAAATATGACTTTAA  
 CCTGGTGACTTTTCACTTTATGTATGTTTNCAGGAGAATTATCTCTTGCCATTCTCGCACGATGAAGT  
 GGTTTCATGGCAAGAAGAGTATGATGCATAAGATGTGGGGAGATCGTTACAATCAATTTCGACGGCTTGCG  
 CAATCTCTATACGTACCAAATTTGTCACCTGGTAAGAAATGCTCTTCATGGGTAGCGAATACGGTCA  
 ATTCTTAGAATGGAATCTGAAGAACAGTTGGAATGGTCTAACCTAGAAGACCCAATGAATGCTAAGAT  
 GAAGTATTTTCGCTTCTCAGCTAAACAGTTTACAAAGATCATCGCTGTCTGTGGGAAATTGATACCAG  
 CTATGATGGTATTGAAATCATTGATGCGGATAATCGAGACCAGAGTGTTCTTTCTTTTATTCGTAAGGG  
 TAAAAAGGGA

**SP050 amino acid (SEQ ID NO:80)**

DFVEECHTHNIGVIVDWVPXHFTINDDALAYDGTPTFEYQDHNKAHNHGWGALNFDLGKNEVQSFLIS  
 CIKHWIDVYHLDGIRVDAVSNNMLYLDYDDAPWTPNKGNNLYEGYYFLQRLNEVIKLEYPDVMMIAEE  
 SSSAIKITGMKEIGGLGFDYKWNMGWMMNDILRFYEEDPIYRKYDFNLVTF SFMYVXKENYLLPF SHDEV  
 VHGGKSMHMKMWGDRYNQFAGLRNLTYQICHPGKKLLFMGSEYQGFLWKSEEQLEWSNLEDPMNAKM  
 KYFASQLNQFYKDHRCLEWIDTSYDGIIEIDADNRDQSVLSFIRKGGKG

**SP051 nucleotide (SEQ ID NO:81)**

Table 1

ATCTGTAGTTTATGCGGATGAAACACTTATTACTCATCTGCTGAGAAACCTAAAGAGGAAAAAATGAT  
 AGTAGAAGAAAAGGCTGATAAAGCTTTGGAAACTAAAAATATAGTTGAAAGGACAGAACAAAGTGAACC  
 TAGTTCAACTGAGGCTATTGCATCTGAGNAGAAAGAAGATGAAGCCGTAACCTCCAAAAGAGGAAAAAGT  
 GTCTGCTAAACCGGAAGAAAAAGCTCCAAGGATAGAAATCACAAGCTTCAAATCAAGAAAAACCGCTCAA  
 GGAAGATGCTAAAGCTGTAACAAATGAAGAAGTGAATCAAATGATTGAAGACAGGAAAGTGGATTTTAA  
 TCAAAATTGGTACTTTAAACTCAATGCAAATTTCTAAGGAAGCCATTAAACCTGATGCAGACGTATCTAC  
 GTGGAAAAAATTAGATTTACCGTATGACTGGAGTATCTTTAACGATTTTCGATCATGAATCTCCTGCACA  
 AAATGAAGGTGGACAGCTCAACGGTGGGGAAGCTTGGTATCGCAAGACTTTCAAAGTATGAAAAAGA  
 CCTCAAGAAAAATGTTTCGCCTTACTTTTGATGGCGTCTACATGGATTCTCAAGTTTATGTCAATGGTCA  
 GTTAGTGGGGCATTATCCAAATGGTTATAACCAGTTCTCATATGATATCACCAAATACCTTCAAAAAGA  
 TGGTCGTGAGAATGTGATTGCTGTCCATGCAGTCAACAAACAGCCAAGTAGCCGTTGGTATTCAGGAAG  
 TGGTATCTATCGTGATGTGACTTTACAAGTGACAGATAAGGTGCATGTTGAGAAAAATGGGACAACTAT  
 TTTAACACCAAACTTGAAGAACAACAACATGGCAAGGTTGAAACTCATGTGACCAGCAAAATCGTCAA  
 TACGGACGACAAAGACCATGAACCTGTAGCCGAATATCAAATCGTTGAACGAGGTGGTTCATGCTGTAAC  
 AGGCTTAGTTCTGACAGCGAGTCTGACCTTAAAAGCACATGAATCAACAAGCCTAGATGCGATTTTGA  
 AGTTGAAAGACCAAACTCTGGACTGTTTTAAATGACAAACCTGCCTTGTACGAATTGATTACGCGTGT  
 TTACCGTGACGGTCAATTGGTTGATGCTAAGAAGGATTTGTTTGGTTACCGTTACTATCACTGGACTCC  
 AAATGAAGGTTTCTCTTTGAATGGTGAACGTATTAATTCATGGAGTATCCTTGCACCACGACCATGG  
 GCGCTTGGAGCAGAAGAAAACTATAAAGCAGAAATATCGCCGCTCTCAAACAAATGAAGGAGATGGGAGT  
 TAACTCCATCCGTACAACCCACAACCTGCTAGTGAGCAAACTTGCAAATCGCAGCAGAACTAGGTTT  
 ACTCGTTTCAAGGAAGAGGCCTTTGATACGTGGTATGGTGGCAAGAAACCTTATGACTATGGACGTTTCTT  
 TGAAAAAGATGCCACTCACCCAGAAGCTCGAAAAGGTGAAAAATGGTCTGATTTTGACCTACGTACCAT  
 GGTGCAAGAGGCAAAAACAACCTGCTATCTTCATGTGGTCAATTGGTAATGAAATAGGTGAAGCTAA  
 TGGTGATGCCCCTCTTTAGCAACTGTTAAACGTTTGGTTAAGGTTATCAAGGATGTTGATAAGACTCG  
 CTATGTTACCATGGGAGCAGATAAATTCGTTTCGGTAATGGTAGCGGAGGGCATGAGAAAAATGCTGA  
 TGAACCTCGATGCTGTTGGATTAACTATTCTGAAGATAATTACAAAGCCCTTAGAGCTAAGCATCCAAA  
 ATGGTTGATTTATGGATCAGAAACATCTTCAGCTACCCGTACACGTGGAAGTTACTATCGCCCTGAACG  
 TGAATTGAAACATAGCAATGGACCTGAGCGTAATTATGAACAGTCAGATTATGGAATGATCGTGTGGG  
 TTGGGGGAAAACAGCAACCGCTTCATGGACTTTTGACCGTGACAACGCTGGCTATGCTGGACAGTTTAT  
 CTGGACAGGTACGGACTATATTGGTGAACCTACACCATGGCACAACCAAAATCAAACCTCCTGTTAAGAG  
 CTCTTACTTTGGTATCGTAGATACAGCCGGCATTCCAAAACATGACTTCTATCTCTACCAAAGC

**SP051 amino acid (SEQ ID NO:82)**

SVVYADETLITHTAKEPKKEKMIVEEKADKALETKNIVERTEQSEPSSTEAIASEXKEDEAVTPKEEKV  
 SAKPEEKAPRIESQASNQEKPLKEDAKAVTNEEVNQMIEDRKVDNFQNWYFKLNANSKEAIKPDADVST  
 WKKLDLPYDWSIFNDFDHESPAQNEGGQLNGGEAWYRKTFKLDEKDLKKNVRLTFDGVYMDSQVYVNGQ  
 LVGHYPNGYNQFSYDITKYLQKDGRENVIHAVHNKQPSRWYSGSGIYRDVTLQVTDKVVHVEKNGTTI  
 LTPKLEEQQHKGKVEHVTSKIVNTDDKDELVAEYQIVERGGHAVTGLVRTASRTLKAHESTSLDAILE  
 VERPKLWTVLNDKPALYELITRVYRDGQLVDAKKDLFGYRYHWTNPEGFSLNGERIKFHGVSLLHHDHG  
 ALGAEENYKAEYRRLKQMKEMGVNSIRTHNPASEQTLQIAAELGLLVQEEAFDTWYGGKKPYDYGRFF  
 EKDATHPARKGEKWSDFDLRTMVERGKNNPAIFMWSIGNEIGEANGDAHSLATVKRLVKVIKDVDKTR  
 YVTMGADKFRFNGSGGHEKIADELDAVGFNYSEDNYKALRAKHPKWLIIYGSETSSATRTRGSYYRPER  
 ELKHSNGPERNYEQSDYGNDRVGWGTATASWTFDRDNAGYAGQFIWTGTDYIGEPTPWHNQNTPVKS  
 SYFGIVDTAGIPKHDFYLYQS

**SP052 nucleotide (SEQ ID NO:83)**

TTACTTTGGTATCGTAGATACAGCCGGCATTCCAAAACATGACTTCTATCTCTACCAAAGCCAATGGGT  
 TTCTGTTAAGAAGAAACCGATGGTACACCTTCTTCCTCACTGGAAGTGGGAAAAACAAAGAATTAGCATC  
 CAAAGTAGCTGACTCAGAAGGTAAGATTCAGTTTCGTGCTTATTTCGAATGCTTCTAGTGTAGAATTGTT  
 CTTGAATGGAAAATCTCTTGGTCTTAAGACTTTCAATAAAAAACAAACCAGCGATGGGCGGACTTACCA  
 AGAAGGTGCAATGCTAATGAACCTTATCTTGAATGGAAAGTTGCCTATCAACCAGGTACCTTGAAGC  
 AATTGCTCGTGATGAATCTGGCAAGGAAATGCTCGAGATAAGATTACGACTGCTGGTAAGCCAGCGGC  
 AGTTCTGCTTATTAAAGGAAGACCATGCGATTGCAGCAGATGGAAAAGACTTGACTTACATCTACTATGA  
 AATTGTTGACAGCCAGGGGAATGTGGTTCCAACCTGCTAATAATCTGGTTCGCTTCCAATTGCATGGCCA  
 AGGTCAACTGGTGGTGTAGATAACCGGAGAACAGCCAGCCGTGAACGCTATAAGGCGCAAGCAGATGG  
 TTCTTGGATTTCGTAAAGCATTAAATGGTAAAGGTGTTGCCATTGTCAAATCAACTGAACAAGCAGGGAA  
 ATTCACCTGACTGCCCCTCTGATCTCTTGAATCGAACCAAGTCACTGTCTTTACTGGTAAGAAAGA  
 AGGACAAGAGAAGACTGTTTTGGGGACAGAAGTGCCAAAAGTACAGACCATTATTGGAGAGGCACCTGA

Table 1

67

AATGCCTACCACTGTTCCGTTTGTATACAGTGATGGTAGCCGTGCAGAACGTCCTGTAACCTGGTCTTC  
 AGTAGATGTGAGCAAGCCTGGTATTGTAACGGTGAAAGGTATGGCTGACGGACGAGAAGTAGAAGCTCG  
 TGTAAGATGATTGCTCTTAAATCAGAGCTACCAGTTGTGAAACGTATTGCTCCAAATACTGACTTGAA  
 TTCTGTAGACAAATCTGTTTCCCTATGTTTTGATTGATGGAAGTGTTGAAGAGTATGAAGTGGACAAGTG  
 GGAGATTGCCGAAGAAGATAAAGCTAAGTTAGCAATTCCAGGTTCTCGTATTCAAGCGACCGGTTATTT  
 AGAAGGTCAACCAATTCATGCAACCCTTGTGGTAGAAGAAGGCAATCCTGCGGCACCTGCAGTACCAAC  
 TGTAAACGGTTGGTGGTGAGGCAGTAACAGGCTTACTAGTCAAAAAACCAATGCAATACCGCACTCTTGC  
 TTATGGAGCTAAGTTGCCAGAAGTCACAGCAAGTGCTAAAAATGCAGCTGTTACAGTTCTTCAAGCAAG  
 CGCAGCAAACGGCATGCGTGCGAGCATCTTTATTACGCTTAAAGATGGTGGCCCTCTTCAAACCTATGC  
 AATTCAATTCCTTGAAGAAGCGCCAAAAATTGCTCACTTGAGCTTGCAAGTGGAAAAAGCTGACAGTCT  
 CAAAGAAGACCAAATGTCAAATTGTCGGTTCGAGCTCACTATCAAGATGGAACGCAAGCTGTATTACC  
 AGCTGATAAAGTAACCTTCTCTACAAGTGGTGAAGGGGAAGTCGCAATTCGTAAAGGAATGCTTGAGTT  
 GCATAAGCCAGGAGCAGTCACTCTGAACGCTGAATATGAGGGAGCTAAAGACCAAGTTGAACTCACTAT  
 CCAAGCCAATACTGAGAAGAAGATTGCGCAATCCATCCGTCCTGTAAATGTAGTGACAGATTTGCATCA  
 GGAACCAAGTCTTCCAGCAACAGTAACAGTTGAGTATGACAAAGGTTCCCTAAAACCTATAAAGTCAC  
 TTGGCAAGCTATTCCGAAAGAAAACTAGACTCCTATCAAACATTTGAAGTACTAGGTAAAGTTGAAGG  
 AATTGACCTTGAAGCGCGTGCAAAAGTCTCTGTAGAAGGTATCGTTTCAGTTGAAGAAGTCAGTGTGAC  
 AACTCCAATCGCAGAAGCACCACAATTACCAGAAAGTGTTCGGACATATGATTCAAATGGTTCACGTTTC  
 ATCAGCTAAGGTTGCATGGGATGCGATTCTGTCAGAGCAATACGCTAAGGAAGGTGCTCTTTACAGTTAA  
 TGGTTCGCTTAGAAGGTACGCAATTAACA

**SP052 amino acid (SEQ ID NO:84)**

YFGIVDTAGIPKHFYLYQSQWVSVKKKPMVHLLPHWNWENKELASKVADSEGKIPVRAYSNAASSVELF  
 LNGKSLGLKFTFNKKQTS DGRTYQEGANANELYLEWKVAYQPGTLEAIARDESGKEIARDKITTAGKPAA  
 VRLIKEDHAIAADGKDLTYIYYEIVDSQGNVPTANNLVRFLHGGQQLVGVDNGEQASRERYKAQADG  
 SWIRKAFNGKGVAIVKSTEQAGKFTLTAHSDLLKSNQVTVFTGKKEGQEKTVLGTEVPKVQTIIGEAPE  
 MPTTVPFVYSDGSRAERPVTWSSVDVSKPGIVTVKGMADGREVEARVEVIALKSELPVVKRIAPNTDLN  
 SVDKSVSYVLIDGSVEEYEVDKWEIAEEDKAKLAIPGSRIQATGYLEGQPIHATLVVEGNPAAPAVPT  
 VTVGGEAVTGLTSQKPMQYRTLAYGAKLPEVTASAKNAAVTVLQASAANGMRASIFIQPKDGGPLQTYA  
 IQFLEEAPKIAHLSLQVEKADSLKEDQTVKLSVRAHYQDGTQAVLPADKVTFSTSGEGEVAIRKGMLEL  
 HKPGAVTLNAEYEGAKDQVELTIQANTEKKIAQSIRPVNVVTDLHQEPSLPATVTVVEYDKGFPKTHKVT  
 WQAIPEKLDYSYQTFEVLGKVEGIDLEARAKVSVEGIVSVEEVSVTTPIAEAPQLPESVRTYDSNGHVS  
 SAKVAWDAIRPEQYAKEGVFTVNGRLEGTQLT

**SP053 nucleotide (SEQ ID NO:85)**

AGCTAAGGTTGCATGGGATGCGATTTCGTCCAGAGCAATACGCTAAGGAAGGTGTCTTTACAGTTAATGG  
 TCGCTTAGAAGGTACGCAATTAACAATAAAGTTCATGTTTCGCGTATCTGCTCAAAGTACGCAAGGTGC  
 AAACATTTCTGACCAATGGACCGGTTTCAAGATTGCCACTTGCCTTTGCTTCAGACTCAAATCCAAGCGA  
 CCCAGTTTCAAATGTTAATGACAAGCTCATTCTCTACAATAACCAACCAGCCAATCGTTGGACAACTG  
 GAATCGTACTAATCCAGAAGCTTCAGTCCGTGTTCTGTTTGGAGATTGAGGTATCTTGAGCAACGCTC  
 CGTTGATAATCTAAGTGTGCGATTCCATGAAGACCATGGAGTTGGTGTACCGAAGTCTTATGTGATTGA  
 GTATTATGTTGGTAAGACTGTCCCAACAGCTCCTAAAAACCTAGTTTTGTTGGTAATGAGGACCATGT  
 CTTTAATGATTCTGCCAACTGGAAACAGTTACTAATCTAAAAGCCCCCTGCTCAACTCAAGGCTGGAGA  
 AATGAACCACTTTAGCTTTGATAAAGTTGAAACCTATGCTGTTTCGTATTTCGCATGGTTAAAGCAGATAA  
 CAAGCGTGAACGTCTATCACAGAGGTACAAATCTTTGCGAAACAAGTTGCGGCAGCCAAGCAAGGACA  
 AACAAGAAATCCAAGTTGACGGCAAAGACTTAGCAAACTTCAACCCTGATTTGACAGACTACTACCTTGA  
 GTCTGTAGATGGAAAAGTTCCGGCAGTCACAGCAAGTGTTAGCAACAATGGTCTCGCTACCGTCGTTCC  
 AAGCGTTTCGTGAAGGTGAGCCAGTTTCGTGTCTATCGCGAAAGCTGAAAATGGCGACATCTTAGGAGAATA  
 CCGTCTGCATTTCACTAAGGATAAGAGCTTACTTTCTCATAAACCAGTTGCTGCGGTTAAACAAGCTCG  
 CTTGCTACAAGTAGGTCAAGCACTTGAATTGCCGACTAAGGTTCCAGTTTACTTCACAGGTAAAGACGG  
 CTACGAAAACAAAAGACCTGACAGTTGAATGGGAAGAAGTTCCAGCGGAAAATCTGACAAAAGCAGGTCA  
 ATTTACTGTTTCGAGGCCGTGTCCTTGGTAGTAACCTTGTGCTGAGATCACTGTACGAGTGACAGACAA  
 ACTTGGTGAGACTCTTTCACATAACCCTAATATGATGAAAACAGTAACCAGGCCCTTTCCTTCAGCAAC  
 CAATGATATTGACAAAACTCTCATGACCGCGTTGACTATCTCAATGACGGAGATCATTAGAAAAATCG  
 TCGTTGGACAACTGGTCAACCAACACCATCTTCTAATCCAGAAGTATCAGCGGGTGTGATTTCCGTGA  
 AAATGGTAAGATTGTAGAACGACTGTTACACAAGGAAAAGTTAGTTCTTTGCAGATAGTGGTACCGGA  
 TGCACCATCTAACTCGTTTTAGAACGCTATGTCGGTCCAGAGTTTGAAGTGCCAACCTACTATTCAA  
 CTACCAAGCCTACGACGCAGACCATCCATTCAACAATCCAGAAAATTGGGAAGCTGTTCTTTATCGTGC

WO 98/18930

PCT/US97/19422

Table 1

68

GGATAAAGACATTGCAGCTGGTGTATGAAATCAACGTAACATTTAAAGCTATCAAAGCCAAAGCTATGAG  
 ATGGCGTATGGAGCGTAAAGCAGATAAGAGCGGTGTTGCGATGATTGAGATGACCTTCCTTGACCAAG  
 TGAATTGCCTCAAGAAAGCACTCAATCAAAGATTCTTGTAGATGGAAAAGAACTTGCTGATTTGCTGA  
 AAATCGTCAAGACTATCAAATTACCTATAAAGGTCAACGGCCAAAAGTCTCAGTTGAAGAAAACAATCA  
 AGTAGCTTCAACTGTGGTAGATAGTGGAGAAGATAGCTTCCAGTACTTGTTTCGCCTCGTTTCAGAAAG  
 TGGAAAACAAGTCAAGGAATACCGTATCCACTTGACTAAGGAAAAACCAGTTTCTGAGAAGACAGTTGTC  
 TGCTGTACAGAAGATCTTCCAAAAATCGAATTTGTTGAAAAAGATTTGGCATAACAAGACAGTTGAGAA  
 AAAAGATTCAACACTGTATCTAGGTGAACTCGTGTAGAACAAAGAAGGAAAAAGTTGGAAGAAGACGTAT  
 CTTTACAGCGATTAATCCTGATGGAAGTAAGGAAGAAAAACTCCGTGAAGTGGTAGAAGTTCCGACAGA  
 CCGCATCGTCTTGGTTGGAACCAACCAGTAGCTCAAGAAGCTAAAAAACCAAGTGTGAGAAAAGC  
 AGATACAAAACCAATTGATTCAAGTGAAGCTAGTCAAACTAATAAAGCCAG

**SP053 amino acid (SEQ ID NO:86)**

AKVAWDAIRPEQYAKEGVFTVNGRLEGTQLTTLKLVHVRVSAQTEQGANISDQWTGSELPLAFASDSNPSD  
 PVSNNVDKLISSNNQPANRWTNNWRNTPNASVGVLFSGILSKRVDNLVSGFHEDHGVGVPKSYVIE  
 YYVGKTVPTAPKNPSFVGNEDHVFNDSANWKPVTLNKAQKAGEMNHSFDKVVETAVRIRMVADN  
 KRGTSTIEVQIFAKQVAAKQGGQTRIQVDGKDLANFNPDLDYYLESDGKVPVAVTASVSNGLATVVP  
 SVREGEPVRVIAKAENGDI LGEYRLHFTKDKSLSHKPVAAVKQARLLQVGQALELPTKVPVYFTGKDG  
 YETKDLTVEWEEVPAENLTAKGQFTVRGRVLGSLNVAEITVRVTDKLGETLSDNPNYDENSNOAFASAT  
 NDIDKNSHDRVLDYLDGHSNRRWTNWSPTSSNPEVSAGVIFRENGKIVERTVTQGVQFFADSGTD  
 APSKLVRLERYVGPFEFVPTYYSNYQAYDADHPNPNENWEAVPYRADKDIAAGDEINVTFKAIKAKAMR  
 WRMERKADKSGVAMIEMTFLAPSELPQESTQSKILVDGKELADFAENRQDYQITYKGQRPKVSVEENNQ  
 VASTVVDSDGSDSFPVLVRLVSESGQVKEYRIHLTKEKPVSEKTVAAVQEDLPKIEFVEKDLAYKTVEK  
 KDSTLYLGETRVEQEGKVGKERIFTAINPDGSKEEKLREVVVPTDRIVLVGTPVQAQEAKKPQVSEKA  
 DTKPIDSSSEASQTNKAQ

**SP054 nucleotide (SEQ ID NO:87)**

CTATCACTATGTAAATAAAGAGATTATTTCAACAAGAGCTAAAGATTTAATTCAGACAGGAAAGCCTGA  
 CAGGAATGAAGTTGTATATGGTTTGGTGTATCAAAAAGATCAGTTGCCTCAAACAGGGACAGAA

**SP054 amino acid (SEQ ID NO:88)**

YHYVNKEIIISQEAKDLIQTGKPDRENVVYGLVYQKDQLPQTGTE

**SP055 nucleotide (SEQ ID NO:89)**

TGAGACTCCTCAATCAATAACAAATCAGGAGCAAGCTAGGACAGAAAAACCAAGTAGTAGAGACAGAGGA  
 AGCTCCAAAAGAAGAAGCACCTAAACAGAGAAGTCCAAAAGGAAGAACCAAAATCGGAGGTAAAACC  
 TACTGACGACACCCTTCTTAAAGTAGAAGAGGGGAAAAGAAGATTTCAGCAGAACCAGCTCCAGTTGAAGA  
 AGTAGGTGGAGAAGTTGAGTCAAAACCAGAGGAAAAAGTAGCAGTTAAGCCAGAAAAGTCAACCATCAGA  
 CAAACCAGCTGAGGAATCAAAAGTTGAACAAGCAGGTGAACCAGTCGCGCCAAGAGAAGACGAAAAGGC  
 ACCAGTCGAGCCAGAAAAGCAACCAGAAGCTCCTGAAGAAGAGAAGGCTGTAGAGGAAACACCGAAACA  
 AGAAGAGTCAACTCCAGATACCAAGGCTGAAGAACTGTAGAACCAAAAGAGGAGACTGTTAATCAATC  
 TATTGAACAACCAAAAGTTGAAACGCCTGCTGTAGAAAAACAAACAGAACCAACAGAGGAACCAAAAGT  
 TGAACAAGCAGGTGAACCAGTCGCGCCAAGAGAAGACGAACAGGCACCAACGGCACCAGTTGAGCCAGA  
 AAAGCAACCAGAAGTTCTGAAGAAGAGAAGGCTGTAGAGGAAACACCGAAACCAGAAGATAAAATAAA  
 GGGTATTGGTACTAAAGAACCAGTTGATAAAAGTGAGTTAAATAATCAAATTGATAAAGCTAGTTTCAGT  
 TTCTCCTACTGATTAT

**SP055 amino acid (SEQ ID NO:90)**

ETPQSITNQEQARTENQVVEETEEAPKEEAPKTEESPKKEPKSEVKPTDDTLPKVEEGKEDSAEPAPVEE  
 VGGEVESKPEEKVAVKPESQPSDKPAEESKVEQAGEPVAPREDEKAPVEPEKQPEAPEEEKAVEETPKQ  
 EESTPDTKAEETVEPKEETVNQSI EQPKVETPAVEKQTEPTEEPKVEQAGEPVAPREDEQAPTAPVEPE  
 KQPEVPEEEKAVEETPKPEDKIKGIGTKPEVDKSELNNQIDKASSVSPTDY

**SP056 nucleotide (SEQ ID NO:91)**

GGATGCTCAAGAACTGCGGGAGTTCACTATAAATATGTGGCAGATTCAGAGCTATCATCAGAAGAAAA  
 GAAGCAGCTTGTCTATGATATCCGACATACGTGGAGAATGATGATGAACTTATTATCTTGTTTATAA  
 GTTAAATTCTCAAAATCAACTGGCGGAATTGCCAAATACTGGAAGCAAGAATGAGAGGCCAA

Table 1

**SP056 amino acid (SEQ ID NO:92)**

DAQETAGVHYKYVADSELSSEEKKQLVYDIPTYVENDDETYVLVYKLNSQNQLAELPNTGSKNERQ

**SP057 nucleotide (SEQ ID NO:93)**

CGACAAAGGTGAGACTGAGGTTCAACCAGAGTCGCCAGATACTGTGGTAAGTGATAAAGGTGAACCAGA  
GCAGGTAGCACCCTTCCAGAATATAAGGGTAATATTGAGCAAGTAAAACCTGAAACTCCGGTTGAGAA  
GACCAAAGAACAAGGTCCAGAAAAAAGTGAAGAAGTTCCAGTAAAACCAACAGAAACACCAGTAAA  
TCCAAATGAAGGTACTACAGAAGGAACCTCAATTCAAGAAGCAGAAAAATCCAGTTCAACCTGCAGAAGA  
ATCAACAACGAATTCAGAGAAAGTATCACCAGATACATCTAGCAAAAAATACTGGGGAGGTGTCCAGTAA  
TCCTAGTGATTGACAAACCTCAGTTGGAGAATCAAATAAACAGAACATAATGACTCTAAAAATGAAAA  
TTCAGAAAAAAGTGTAGAAGAAGTTCCAGTAAATCCAAATGAAGGCACAGTAGAAGGTACCTCAAAATCA  
AGAAACAGAAAAACAGTTCAACCTGCAGAAGAAACACAAACAACTCTGGGAAAAATAGCTAACGAAAA  
TACTGGAGAAGTATCCAATAAACCTAGTGATTCAAACCACAGTTGAAGAATCAAATCAACCAGAAAA  
AAACGGAACTGCAACAAAAACAGAAATTCAGGTAATACAACATCAGAGAATGGACAAACAGAACCCAGA  
ACCATCAAACGGAAATTCAACTGAGGATGTTTCAACCGAATCAAACACATCCAATTCAAATGGAAACGA  
AGAAATTAAACAAGAAATGAAGTAGACCCTGATAAAAAGGTAGAGAACCAGAGAAAAACACTTGAATT  
AAGAAAT

**SP057 amino acid (SEQ ID NO:94)**

DKGETEVQPESPDTVVSDKGEPEQVAPLPEYKGNIEQVKPETPVEKTKEQGPEKTEEVVPKPTETPVN  
PNEGTTGETSIQEAENPVQPAEESTTNSEKVSPTSSKNTGEVSSNPSTSTSVGESNKPENHDSKNEN  
SEKTVVEVPVNPNEGTVEGTSNQETEKPVQPAEETQTN SGKIANENTGEVSNKP SDSKPPVEESNQPEK  
NGTATKPENSGNTTSENGQTEPEPSNGNSTEDVSTESNTSNSNGNEEIKQENELDPDKKVEEPEKTLLE  
RN

**SP058 nucleotide (SEQ ID NO:95)**

AAATCAATTGGTAGCACAAAGATCCAAAAGCACAAAGATAGCACTAAACTGACTGCTGAAAAATCAACTGT  
TAAAGCACCTGCTCAAAGAGTAGATGTAAAAGATATAACTCATTTAACAGATGAAGAAAAAGTTAAGGT  
TGCTATTTTACAAGCAAATGGTTTCAGCATTAGACGGAGCGACAATCAATGTAGCTGGAGATGGTACAGC  
AACATCACATTCCCAGATGGTTTCAGTAGTGACGATTCTAGGAAAAGATACAGTTCAACAATCTGCGAA  
AGGTGAATCTGTAACCTCAAGAAGCTACACCAGAGTATAAGCTAGAAAAATACACCAGGTGGAGATAAGGG  
AGGCAATACTGGAAGCTCAGATGCTAATGCCAATGAAGGCGGTGGTAGCCAGGCGGGTGGATCAGCTCA  
CACAGGTTCAAAAACCTCAGCTCAATCACAAGCTTCTAAGCAATTAGCTACTGAAAAAGAAATCAGCTAA  
AAATGCCATTGAAAAAGCAGCCAAGGACAAGCAGGATGAAATCAAAGGCGCACCGCTTTCTGTATAAGA  
AAAAGCAGAACTTTTAGCAAGAGTGGAAGCAGAAAAACAAGCAGCTCTCAAAGAGATTGAAAAATGCCAA  
AATATGGAAGATGTGAAGGAAGCAGAAACGATTGGAGTGCAAGCCATTGCCATGGTTACAGTTCCATAA  
GAGACCAGTGGCTCCTAAT

**SP058 amino acid (SEQ ID NO:96)**

NQLVAQDPKAQDSTKLTAEKSTVKAPAQRVDVKDITHLTDEEKVKVAILQANGSALDGATINVAGDGTA  
TITFPDGSVVTILGKDTVQSSAKGESVTQEATPEYKLENTPGGDKGGNTGSSDANANEGGGSQAGGSAH  
TGSQNSAQSQASKQLATEKESAKNAIEKAAKDKQDEIKGAPLSDKEKAELLARVEAEKQAALKEIENAK  
TMEDVKEAETIGVQAIAMVTVPKRPVAPN

**SP059 nucleotide (SEQ ID NO:97)**

CAAACAGTCAGCTTCAGGAACGATTGAGGTGATTTACAGAGAAAAATGGCTCTGGGACACGGGGTGCCTT  
CACAGAAATCACAGGGATTCTCAAAAAAGACGGTGATAAAAAAATTGACAACACTGCCAAAACAGCTGT  
GATTCAAAATAGTACAGAAGGTGTTCTCTCAGCAGTTCAAGGGAATGCTAATGCTATCGGCTACATCTC  
CTTGGGATCTTTAACGAAATCTGTCAAGGCTTTAGAGATTGATGGTGTCAAGGCTAGTCGAGACACAGT  
TTTAGATGGTGAATACCTCTTCAACGTCCCTTCAACATTGTTTGGTCTTCTAATCTTTCCAAGCTAGG  
TCAAGATTTTATCAGCTTTATCCACTCCAAACAAGGTCAACAAGTGGTCACAGATAATAAATTTATTGA  
AGCTAAAACCGGAAACACCGGAATATACAAGCCAACACTTATCAGGCAAGTTGTCTGTTGTAGGTTCCAC  
TTCAGTATCTTCTTTAATGGAAAAATTAGCAGAAGCTTATAAAAAAGAAAATCCAGAAGTTACGATTGA  
TATTACCTCTAATGGGTCTTCAGCAGGTATTACCGCTGTTAAGGAGAAAACCGCTGATATTGGTATGGT  
TTCTAGGGAATTAATCCTGAAGAAGGTAAGAGTCTCACCCATGATGCTATTGCTTTAGACGGTATTGC  
TGTTGTGGTCAATAATGACAATAAGGCAAGCCAAGTCAGTATGGCTGAACCTGCAGACGTTTTTAGTGG  
CAAATTAACCACCTGGGACAAGATTAAA

WO 98/18930

PCT/US97/19422

Table 1

70

**SP059 amino acid (SEQ ID NO:98)**

KQSASGTIEVISRENGSGTRGAFTEITGILKKDGDKKIDNTAKTAVIQNSTEGVLSAVQGNANAIGYIS  
 LGS�TKSVKALEIDGVKASRDVLDGEYPLQRPFNIVWSSNLSKLGQDFISFIHSKQGOQVVDNKFIE  
 AKTETTEYTSQHLSGKLSVVGSTSVSSLMELAEAYKKENPEVTIDITSNGSSAGITAVKEKTADIGMV  
 SRELTPEEGKSLTHDAIALDGIADVNNNDNKASQVSMELADVFSGKLTTWDKIK

**SP060 nucleotide (SEQ ID NO:99)**

ATTCGATGATGCGGATGAAAAGATGACCCGTGATGAAATTGCCTATATGCTGACAAATAGTGAAGAAAC  
 ATTGGATGCTGATGAGATTGAGATGCTACAAGGTGTCTTTTCGCTCGATGAACTGATGGCACGAGAGGT  
 TATGGTTCCCTCGAACGGATGCCTTTATGGTGGATATTCAGGATGATAGTCAAGCCATTATCCAAAGTAT  
 TTTAAACAAAATTATTCTCGTATCCCGGTTTATGATGGGGATAAGGACAATGTAATTGGAATCATTCA  
 CACCAAGAGTCTCCTTAAGGCAGGCTTTGTGGACGTTTGTGACAATATTGTTTGAAGAGAATTTTACA  
 AGATCCACTTTTTGTACCTGAACTATTTTGTGGATGACTTGCTAAAAGAACTGCGAAATACCCAAAG  
 ACAAATG

**SP060 amino acid (SEQ ID NO:100)**

FDDADEKMTRDEIAYMLTNSEETLDADEIEMLQGVFSLDELMAREVMVPRTD AFMVDIQDDSQAI IQSI  
 LKQNYSRIPVYDGDKNVIGIIHTKSLLKAGFVDGFDNIVWKRILQDPLFVPETIFVDDLLKELRNTQR  
 QM

**SP062 nucleotide (SEQ ID NO:101)**

GGAGAGTCGATCAAAAGTAGATGAAGCTGTGTCTAAGTTTGAAAAGGACTCATCTTCTTCGTCAAGTTC  
 AGACTCTTCCACTAAACCGGAAGCTTCAGATACAGCGAAGCCAAACAAGCCGACAGAACCAGGAGAAAA  
 GGTAGCAGAAGCTAAGAAGAAGGTTGAAGAAGCTGAGAAAAAAGCCAAGGATCAAAAAGAAGAAGATCG  
 TCGTAAC TACCAACCATTA CTACAAAACGCTTGAAC TTGAAATTGCTGAGTCCGATGTGGAAGTTAA  
 AAAAGCGGAGCTTGAAC TAGTAAAAGTGAAAGCTAACGAACCTCGAGACGAGCAA

**SP062 amino acid (SEQ ID NO:102)**

ESRSKVDEAVSKFEKDSSSSSSSDSSTKPEASDTAKPNKPTEPGEKVAEAKKKVEEA EKKAKDQKEEDR  
 RNYPTITYKTLELEIAESDVEVKAELELVKVKANEPRDEQ

**SP063 nucleotide (SEQ ID NO:103)**

ATGGACAACAGGAAACTGGGACGAGGTTATATCTGGTAAGATTGACAAGTACAAAGATCCAGATATTCC  
 AACAGTTGAATCACAAGAAGTTACGTCAGACTCTAGTGATAAAGAAATAACGGTAAGGTATGACCGTTT  
 ATCAACACCAGAAAAACCAATCCCAACCAAAATCCAGAGCATCCAAGTGTTCCGACACCAAACCCAGA  
 ACTACCAAATCAAGAGACTCCAACACCAGATAAACCAACTCCAGAACCAGGTACTCCAAAAACTGAAAC  
 TCCAGTGAATCCAGACCCAGAAGTTCGGACTTATGAGACAGGTAAGAGAGAGGAATTGCCAAACACAGG  
 TACAGAAGCTAAT

**SP063 amino acid (SEQ ID NO:104)**

WTTGNWDEVISGKIDKYKDPDIPTVESQEVTS DSSDKEITVRYDRLSTPEKPI PQPNPEHPSVPTPNPE  
 LPNQETPTPDKPTPEPGTPKTETPVNPDPEVPTYETGKREELPNTGTEAN

**SP064 nucleotide (SEQ ID NO:105)**

CGATGGGCTCAATCCAACCCAGGTCAAGTCTTACCTGAAGAGACATCGGGAACGAAAGAGGGTGACTT  
 ATCAGAAAAACCAGGAGACACCGTTCTCACTCAAGCGAAACCTGAGGGCGTTACTGGAAATACGAATTC  
 ACTTCCGACACCTACAGAAAGAACTGAAGTGAGCGAGGAAACAAGCCCTTCTAGTCTGGATACACTTTT  
 TGAAAAAGATGAAGAAGCTCAAAAAATCCAGAGCTAACAGATGTCTTAAAGAAACTGTAGATACAGC  
 TGATGTGGATGGGACACAAGCAAGTCCAGCAGAACTACTCCTGAACAAGTAAAAGGTGGAGTGAAAGA  
 AAATACAAAAGACAGCATCGATGTTCTGTGCTTATCTTGAAAAAGCTGAAGGGAAAGGTCTTTTCAC  
 TGCCGGTGTAACCAAGTAATTCCTTATGAAC TATTGCTGGTGATGGTATGTTAAC TCGTCTATTACT  
 AAAAGCTTCGGATAATGCTCCTTGGTCTGACAATGGTACTGCTAAAAATCCTGCTTTACCTCCTCTTGA  
 AGGATTAACAAAAGGGAAATACTTCTATGAAGTAGACTTAAATGGCAATACTGTTGGTAAACAAGGTCA  
 AGCTTTAATTGATCAACTTCGCGCTAATGGTACTCAAACCTTATAAAGCTACTGTTAAAGTTTACGGAA  
 TAAAGACGGTAAAGCTGACTTGACTAATCTAGTTGCTACTAAAAATGTAGACATCAACATCAATGGATT  
 AGTTGCTAAAGAAACAGTTCAAAAAGCCGTTGCAGACAACGTTAAAGACAGTATCGATGTTCCAGCAGC  
 CTACCTAGAAAAAGCCAAGGTGAAGGTCCATTACAGCAGGTGTCAACCATGTGATTCCATACGAAC  
 CTTGCGAGGTGATGGCATGTTGACTCGTCTCTTGCTCAAGGCATCTGACAAGGCACCATGGTCAGATAA

Table 1

CGGCGACGCTAAAAACCCAGCCCTATCTCCACTAGGCGAAAAACGTGAAGACCAAAGGTCAATACTTCTA  
TCAANTAGCCTTGGACGGAAATGTAGCTGGCAAAGAAAAACAAGCGCTCATTGACCAGTTCCGAGCAAA  
NGGTACTCAAACCTTACAGCGCTACAGTCAATGTCTATGGTAACAAAGACGGTAAACCAGACTTGGACAA  
CATCGTAGCAACTAAAAAAGTCACCTATTAACATAAACGGTTTAATTTCTAAAGAAAACAGTTCAAAAAGC  
CGTTGCAGACAACGTTAANGACAGTATCGATGTTCCAGCAGCCTACCTAGAAAAAGCCAAGGGTGAAGG  
TCCATTACAGCAGGTGTCAACCATGTGATTCCATACGAACTCTTCGAGGTGATGGTATGTTGACTCG  
TCTCTTGCTCAAGGCATCTGACAAGGCACCATGGTCAGATAACGGNGACGCTAAAAACCCAGCNCATC  
TCCACTAGGTGAAAACGTGAAGACCAAAGGTCAATACTTCTATCAANTAGCCTTGGACGGAAATGTAGC  
TGGCAAAGAAAAACAAGCGCTCATTGACCAGTTCCGAGCAAAACGGTACTCAAACCTTACAGCGCTACAGT  
CAATGTCTATGGTAACAAAGACGGTAAACCAGACTTGGACAACATCGTAGCAACTAAAAAAGTCACTAT  
TAAGATAAATGTTAAAGAAACATCAGACACAGCAAATGGTTCATTATCACCTTCTAACTCTGGTTCTGG  
CGTGACTCCGATGAATCACAATCATGCTACAGGTACTACAGATAGCATGCCTGCTGACACCATGACAAG  
TTCTACCAACACGATGGCAGGTGAAAACATGGCTGCTTCTGCTAACAAGATGTCTGATACGATGATGTC  
AGAGGATAAAGCTATG

**SP064 amino acid (SEQ ID NO:106)**

DGLNPTPGQVLPEETSGTKEGDLSEKPGDVLTVQAKPEGVTGNTNSLPTPTERTVSEETSPSSLDLTF  
EKDEEAQKNPELTDVLKETVDADVDGTQASPAETTPEQVKGGVKENTKDSIDVPAAYLEKAEGKGPFT  
AGVNQVIPYELFAGDGMLTRLLLKASDNAPWSDNGTAKNPALPPLEGLTKGKYFYEVDLNGNTVGKQGO  
ALIDQLRANGTQTYKATVKVYGNKDGKADLTNLVATKNVDININGLVAKETVQKAVADNVKDSIDVPA  
YLEKAKGEGPFTAGVNHVIPYELFAGDGMLTRLLLKASDKAPWSDNGDAKNPALSPLGENVKTKGQYFY  
QXALDGNVAGKEKQALIDQFRAXGTQTYSATVNVYGNKDGKPDLDNIVATKKVTININGLISKETVQKA  
VADNVXDSIDVPAAYLEKAKGEGPFTAGVNHVIPYELFAGDGMLTRLLLKASDKAPWSDNGDAKNPALS  
PLGENVKTKGQYFYQXALDGNVAGKEKQALIDQFRANGTQTYSATVNVYGNKDGKPDLDNIVATKKVTI  
KINVKETSDTANGSLSPSNSGSGVTPMNHNHATGTTDSMPADTMTSSTNTMAGENMAASANKMSDTMMS  
EDKAM

**SP065 nucleotide (SEQ ID NO:107)**

TTCCAATCAAAAAACAGGCAGATGGTAAACTCAATATCGTGACAACCTTTTACCCTGTCTATGAATTTAC  
CAAGCAAGTCGCAGGAGATACGGCTAATGTAGAACTCCTAATCGGTGCTGGGACAGAACCTCATGAATA  
CGAACCATCTGCCAAGGCAGTTGCCAAAATCCAAGATGCAGATACCTTCGTTTATGAAAATGAAAACAT  
GGAAACATGGGTACCTAAATTGCTAGATACCTTGATAAGAAAAAAGTGAAAACCATCAAGGCGACAGG  
CGATATGTTGCTCTTGCCAGGTGGCGAGGAAGAAGAGGGAGACCATGACCATGGAGAAGAAGGTCATCA  
CCATGAGTTTGACCCCCATGTTTGGTTATCACCAGTTCGTGCCATCAAACTAGTAGAGCACCATCCGCG  
ACACTTGTCAGCAGATTATCCTGATAAAAAAGAGACCTTTGAGAAGAATGCAGCTGCCTATATCGAAAA  
ATTGCAAGCCTTGATAAGGCTTACGCAGAAGGTTTGTCTCAAGCAAAACAAAAGAGCTTTGTGACTCA  
ACACGCAGCCTTTAACTATCTTGCCTTGGACTATGGGACTC

**SP065 amino acid (SEQ ID NO:108)**

SNQKQADGKLNIVTTFYPVYEFTKQVAGDTANVELLIGAGTEPHEYEPSAKAVAKIQDADTFVYENENM  
ETWVPKLLDLDLKKVKTIKATGDMLLLPGGEEEEGDHHDHGEEGHHHEFDPHVWLSVPVRAIKLVEHHP  
HLSADYPDKKETFEKNAAAYIEKLQALDKAYAEGLSQAKQKSFVTQHAAFNYLALDYGT

**SP067 nucleotide (SEQ ID NO:109)**

TATCACAGGATCGAACGGTAAGACAACCACAACGACTATGATTGGGGAAGTTTTGACTGCTGCTGGCCA  
ACATGGTCTTTTATCAGGGAATATCGGCTATCCAGCTAGTCAGGTGCTCAATAGCATCAGATAAGGA  
CACGCTTGTTATGGAACCTTCTTCTTTCCAACCTCATGGGTGTTCAAGAATTCCATCCAGAGATTGCGGT  
TATTACCAACCTCATGCCAACTCATATCGACTACCATGGGTCAATTTTCGGAATATGTAGCAGCCAAGTG  
GAATATCCAGAACAAGATGACAGCAGCTGATTTCTTGTCTTGAACCTTAATCAAGACTTGGCAAAAGA  
CTTGACTTCCAAGACAGAAGCCACTGTTGTACCATTTTCAACACTTGAAAAGGTTGATGGAGCTTATCT  
GGAAGATGGTCAACTCTACTTCCGTGGTGAAGTAGTCATGGCAGCGAATGAAATCGGTGTTCCAGGTAG  
CCACAATGTGGAATGCCCCTTGGGACTATTGCTGTAGCCAAGCTTCGTGATGTGGACAATCAAACCAT  
CAAGGAAACTCTTTCAGCCTTCCGTGGTGTCAAACACCGTCTCCAGTTTGTGGATGACATCAAGGGTGT  
TAAATCTATAACGACAGTAAATCAACTAATATCTTGGCTACTCAAAAAGCCTTGTGAGGATTTGACAA  
CAGCAAGGTTCGCTTGTGATTGCAGGTGGTTTGGACCGTGGCAATGAGTTTGACGAATTGGTGCCAGACAT  
TACTGGACTCAAGAAGATGGTCATCTGGGTCAATCTGCAGAACGTGTCAAACGGGCAGCAGACAAGGC  
TGGTGTGCTTATGTGGAGGCGACAGATATTGCAGATGCGACCCGCAAGGCCTATGAGCTTGGGACTCA

Table 1

AGGAGATGTGGTTCTTCTTAGTCCTGCCAATGCTAGCTGGGATATGTATGCTAACTTTGAAGTACGTGG  
CGACCTCTTTATCGACACAGTAGCGGAGTTAAAAGAA

**SP067 amino acid (SEQ ID NO:110)**

GITGSNGKTTTTMIGEVLTAAGQHLLSGNIGYPASQVAQIASDKDTLVMELSSFQLMGVQEFHPEIA  
VITNLMPTHIDYHGSFSEYVAAKWNIONKMTAADFLVLNFNQDLAKDLTSKTEATVVPFSTLEKVDGAY  
LEDGQLYFRGEVMAANEIGVPGSHNVENALATIAVAKLRDNDQTIKETLSAFGGVKHRLQFVDDIKG  
VKFYNDKSTNIIATQKALSGFDNSKVLIAGGLDRGNEFDELVPDITGLKKMVLGQSAERVKRAADK  
AGVAYVEATDIADATRKAYELATQGDVLLSPANASWDMYANFEVRGDLFIDTVAELKE

**SP068 nucleotide (SEQ ID NO:111)**

AAGTTCATCGAAGATGGTTGGGAAGTCCACTATATCGGGACAAGTGTGGTATCGAACACCAAGAAATC  
CTTAAGTCAGGTTTGGATGTCACCTTCCATTCTATTGCGACTGGAAAATTGCGTCGCTATTTCTCTTGG  
CAAAATATGCTGGACGTCTTCAAAGTTGGTTGGGAATTGTCCAATCGCTCTTTATCATGTTGCGACTG  
CGTCCACAGACCCCTTTTTTCAAAGGGGGGCTTTGTCTCAGTACCGCTGTTATCGCTGCGCGTGTGTCA  
GGAGTGCCTGTCTTTATTCACGAATCTGACCTGTCTATGGGCTTGGCCAATAAAATCGCCTATAAATTT  
GCGACTAAGATGTATTCAACCTTTGAACAAGCTTCGAGTTTGGCTAAGGTTGAGCATGTGGGAGCGG

**SP068 amino acid (SEQ ID NO:112)**

SSSKMVGKSTISGTSVVSNTKKSLSQVWMSPIILLRLENCVAISLGKICWTSSKLVGELSNRSLSCCDC  
VHRPFFQRGALSQYRLLSLRVCQECLSLFTNLCLWAWPIKSPINLRLRCIQPLNKLRLVWLRSLMWER

**SP069 nucleotide (SEQ ID NO:113)**

ATCGCTAGCTAGTGAAATGCAAGAAAGTACACGTAAATTCAGGTTACTGCTGACCTAACAGATGCCGG  
TGTTGGAACGATTGAAGTTCCTTTGAGCATTGAAGATTTACCCAATGGGCTGACCGCTGTGGCGACTCC  
GCAAAAAATTACAGTCAAGATTGGTAAGAAGGCTCAGAAGGATAAGGTAAAGATTGTACCAGAGATTGA  
CCCTAGTCAAATTGATAGTCGGGTACAAAATTGAAAATGTCTATGGTGTCAGATAAAGAAGTGTCTATTAC  
GAGTGACCAAGAGACATTGGATAGAATTGATAAGATTATCGCTGTTTTGCCAACTAGCGAACGTATAAC  
AGGTAATTACAGTGGTTTCAGTACCTTTGCAAGGCAATCGACCGCAATGGTGTTGTCTTACCGGCATTAT  
CACTCCGTTTGATACAATAATGAAGGTGACTACAAAACAGTAGCACCAAGTTCAAGCACATCAAATTC  
AAGTACAAGCAGTTCATCGGAGACATCTTCGTCAACGAAAGCAACTAGTTCAAAAACGAAT

**SP069 amino acid (SEQ ID NO:114)**

SLASEMQESTRKFKVTADLTDAVGVTIEVPLSIEDLPNGLTAVATPQKITVKIGKKAQKDKVKIVPEID  
PSQIDSRVQIENVMVSDKEVSITSDQETLDRIDKIIAVLPTSERITGNYSGSVPLQAI DRNGVVLPAVI  
TPFDTIMKVTTKPVAPSSSTSNSSTSSSSETSSSTKATSSKTN

**SP070 nucleotide (SEQ ID NO:115)**

GCACCAGATGGGGCACAAGGTTCAAGGATCAGATGTTGAAAAGTACTACTTTACCCAACGCGGTCTTGA  
GCAGGCAGGAATTACCATTTCTCTTTTGTATGAAAAAATCTAGACGGTGATATGGAAATTATCGCTGG  
AAATGCCTTTTCGTCCAGATAACAACGTCGAAATTGCCTATGCGGACCAAAATGGTATCAGCTACAAACG  
TTACCATGAGTTTCTAGGTAGCTTTATGCGTGACTTTGTTAGCATGGGAGTAGCAGGAGCACATGGAAA  
AATTCAACGACAGGTATGTTGTCTCATGTCTTGTCTCACATTACAGATACCAGCTTCTTGATTGGAGA  
TGGGACAGGTGCTGGTTTCGGCCAATGCCAAATATTTTGTCTTTGAATCTGACGAATATGAGCGTCACTT  
CATGCCTTACCACCCAGAATACTCTATTATCACCAACATTGACTTTGACCATCCAGATTATTTCAACAG  
TCTCGAGGATGTTTTTAATGCCTTTAACGACTATGCCAAACAAATCACCAAGGGTCTTTTTGTCTATGG  
TGAAGATGCTGAATTGCGTAAGATTACGTCTGATGCACCAATTTATTATTATGGTTTTGAAGCTGAAGG  
CAATGACTTTGTAGCTAGTGATCTTCTTCGTTCAATAACTGGTTCAACCTTCACCGTTCAATTTCCGTGG  
ACAAAATTGGGGCAATTCCACATTCACCACTTTGGTTCGTCACAATATCATGAATGCGACAGCCGTTAT  
TGGTCTTCTTTACACAGCAGGATTGATTTGAACTTGGTGCGTGAGCACTTGAAAACATTTGCCGGTGT  
TAAACGTCGTTTCACTGAGAAAAATTGTCAATGATACAGTGATTATCGATGACTTTGCCACCATCCAAAC  
AGAAATTATTGCGACCTTGGATGCGGCTCGTCAGAAATACCCAAGCAAGGAAATTGTAGCAGTCTTTCA  
ACCGCATACCTTTACAAGAACCATTGCCTTGTGGACGACTTTGCCCATGCTTTAAACCAAGCAGATGC  
TGTTTATCTAGCGCAAAATTTATGGCTCGGCTCGTGAAGTAGATCATGGTGACGTTAAGGTAGAAGACCT  
AGCCCAACAAAATCAACAAAAACCAAGTGATTACTGTTGAAAATGTTTCTCCACTCCTTAGACCATGA  
CAATGCTGTTTACGTCCTTTATGGGAGCAGGAGACATCCAAACCTATGAATACTCATTTGAGCGTCTCTT  
GTCTAACTTGACAAGCAATGTTCAA

Table 1

**SP070 amino acid (SEQ ID NO:116)**

HQMGHKVQGS DVEKYFTQ RGLEQAGITILPFDEKNLDG DMEIIAGNAFRPDNNVEIAYADQNGISYKR  
YHEFLGSFMRDFVSMGVAGAHGKTSTTGMLSHVLSHITDTSFLIGDGTGRGSANAKYFVFESDEYERHF  
MPYHPEYSIIITNIDFDHPDYFTSLEDVFNAFNDYAKQITKGLFVYGEDAELRKITS DAPIYYYGF EAEG  
NDFVASDLLRSITGSTFTVHFRGQNLGQFH IPTFGRHNIMNATAVIGLLYTAGFDLNLVREHLKTFAGV  
KRRFTEKIVNDTVIIDDFAHHPTEIIATLDAARQKYP SKEIVAVFQPHFTFTRTIALLDDFAHALNQADA  
VYLAQIYGSAREVDHGDVKVEDLANKINKKHQVITVENVSPLLDH DNAVYVFMGAGDIQTYEYSFERLL  
SNLTSNVQ

**SP071 nucleotide (SEQ ID NO:117)**

TTTTAACCCAAC TGTGGTACTTTCTTTT TACTGCAGGATTGAGCTTGT TAGTTTTATTGGTTTCTAA  
AAGGGAAAATGGAAAGAAACGACTTGTTCATTTTCTGCTGTTGACTAGCATGGGAGTTCAA TTGTTGCC  
GGCCAGTGCTTTTGGGTTGACCAGCCAGATTTTATCTGCCTATAATAGTCAGCTTTCTATCGGAGTCGG  
GGAACATTTACCAGAGCCTCTGAAAATCGAAGGTTATCAATATATTGGTTATATCAAAACTAAGAAACA  
GGATAATACAGAGCTTTCAAGGACAGTTGATGGGAAATACTCTGCTCAAAGAGATAGTCAACCAAAC TC  
TACAAAAACATCAGATGTAGTTCATTCAGCTGATTTAGAAATGGAACCAAGGACAGGGGAAGGTTAGTTT  
ACAAGGTGAAGCATCAGGGGATGATGGACTTTCAGAAAAATCTTCTATAGCAGCAGACAATCTATCTTC  
TAATGATTCAATTCGCAAGTCAAGTTGAGCAGAATCCGGATCACAAGGAGAATCTGTAGTTCGACCAAC  
AGTGCCAGAACAAGGAAATCCTGTGTCTGCTACAACGGTGCAGAGTGCAGGAAGAGGAAGTATTGGCGAC  
GACAAATGATCGACCAGAGTATAAACTTCCATTGGAAACCAAAGGCACGCAAGAACCCGGTCATGAGGG  
TGAAGCCGCAGTCCGTGAAGACTTACCAGTCTACACTAAGCCACTAGAAACCAAAGGTACACAAGGACC  
CGGACATGAAGGTGAAGCTGCAGTTCCGCGAGGAAGAACCAGCTTACACAGAACCGTTAGCAACGAAAGG  
CACGCAAGAGCCAGGTCATGAGGGCAAAGCTACAGTCCGCGAAGAGACTCTAGAGTACACGGAACCGGT  
AGCGACAAAAGGCACACAAGAACCCGAACATGAGGGCGAaCGGsCAGTAGAAGAAGAACTTCCGGCTTT  
AGAGGTCACTACACGAAATAGAACGGAAATCCAGAATATTCTTTATACAACAGAAGAAATTCAGGATCC  
AACACTTCTGAAAAATCGTCGTAAGATTGAACGACAAGGGCAAGCAGGGACACGTACAATTC AATATGA  
AGACTACATCGTAAATGGTAATGTCTGTAGAAACTAAAGAAAGTGTACGAACTGAAGTAGCTCCGGTCAA  
CGAAGTCGTTAAAGTAGGAACACTTGTGAAAGTTAAACCTACAGTAGAAATTACAACTTAAACAAAAGT  
TGAGAACAAAAATCTATAACTGTAAGTTATAACTTAATAGACACTACCTCAGCATATGTTTCTGCAAA  
AACGCAAGTTTTCCATGGAGACAAGCTAGTTAAAGAGGTGGATATAGAAAATCCTGCCAAAGAGCAAGT  
AATATCAGGTTTTAGATTACTACACACCGTATACAGTTAAAACACACCTAACTTATAATTTGGGTGAAAA  
TAATGAGGAAAATACTGAAACATCAACTCAAGATTTCCAATTAGAGTATAAGAAAATAGAGATTAAAGA  
TATTGATTCAGTAGAAATTATACGGTAAAGAAAATGATCGTTATCGTAGATATTTAAGTCTAAGTGAAGC  
GCCGACTGATACGGCTAAATACTTTGTAAAAGTGAAATCAGATCGCTTCAAAGAAATGTACCTACCTGT  
AAAATCTATTACAGAAAATACGGATGGAACGTATAAAGTGACGGTAGCCGTTGATCAACTTGTGCAAGA  
AGGTACAGACGGTTACAAAGATGATTACACATTTACTGTAGCTAAATCTAAAGCAGAGCAACCAGGAGT  
TTACACATCCTTTAAACAGCTGGTAACAGCCATGCAAAGCAATCTGTCTGGTGTCTATACATTGGCTTC  
AGATATGACCGCAGATGAGGTGAGCTTAGGCGATAAGCAGACAAGTTATCTCACAGGTGCATTTACAGG  
GAGCTTGATCGGTTCTGATGGAACAAAATCGTATGCCATTTATGATTTGAAGAAACCATTATTTGATAC  
ATTAAATGGTGTCTACAGTTAGAGATTTGGATATTTAAACTGTTTCTGCTGATAGTAAAGAAAATGTCCG  
AGCGCTGGCGAAGGCAGCGAATAGCGCGAATATTAATAATGTTGCAGTAGAAGGAAAAATCTCAGGTGC  
GAAATCTGTTGCGGGATTAGTAGCGGAGCGCAACAAATACAGTGATAGAAAACAGCTCGTTTACAGGGAA  
ACTTATCGCAAATCACCAGGACAGTAATAAAAAATGATACTGGAGGAATAGTAGGTAATATAACAGGAAA  
TAGTTCGAGAGTTAATAAAGTTAGGGTAGATGCCTTAATCTCTACTAATGCACGCAATAATAACCAAAC  
AGCTGGAGGGATAGTAGGTAGATTAGAAAATGGTGCATTGATATCTAATTCGGTTGCTACTGGAGAAAT  
ACGAAATGGTCAAGGATATTCTAGAGTCGGAGGAATAGTAGGATCTACGTGGCAAAACGGTCGAGTAAA  
TAATGTTGTGAGTAACGTAGATGTTGGAGATGGTTATGTTATCACCGGTGATCAATACGCAGCAGCAGA  
TGTGAAAAATGCAAGTACATCAGTTGATAATAGAAAAGCAGACAGATTCTGCTACAAAATTATCAAAAGA  
CCAAATAGACGCGAAAGTTGCTGATTATGGAATCACAGTAACTCTTGATGATACTGGGCAAGATTTAAA  
ACGTAATCTAAGAGAAGTTGATTATACAAGACTAAATAAAGCAGAAGCTGAAAGAAAAGTAGCTTATAG  
CAACATAGAAAACTGATGCCATTCTACAATAAAGACCTAGTAGTTCACTATGGTAACAAAGTAGCGAC  
AACAGATAAACTTTTACACTACAGAATTGTTAGATGTTGTGCGGATGAAAGATGATGAAGTAGTAACGGA  
TATTAATAATAAGAAAAATTCATAAATAAAGTTATGTTACATTTCAAAGATAATACAGTAGAATACCT  
AGATGTAACATTCAAAGAAAACCTTCATAAACAGTCAAGTAATCGAATACAATGTTACAGGAAAAGAATA  
TATATTACACCAGAAGCATTGTTTCAGACTATACAGCGATAACGAATAACGTACTAAGCGACTTGCA  
AAATGTAACACTTAAC

**SP071 amino acid (SEQ ID NO:118)**

Table 1

74

FNPTVGTFLFTAGLSLLVLLVSKRENGKKRLVHFLLLTSMGVQLLPASAFGLTSQILSAYNSQLSIGVG  
 EHLPEPLKIEGYQYIGYIKTKKQDNTELSRTVDGKYSAQRDSQPNSTKTSADVHSADLEWNQGGKVS  
 QGEASGDDGLSEKSSIAADNLSSNDSFASQVEQNPDKHGESVVRPTVPEQGNPVSATTVQSAEEVLAT  
 TNDREYKLPLETGKTQEPGHEGEAAVREDLPVYTKPLETKGTQGPHEGEAAVREEEPAYTEPLATKG  
 TQEPGHEGKATVREETLEYTEPVATKGTQEPHEGERXVEEELPALEVTTNRNTEIQNIPTTTEEIQDP  
 TLLKNRRKIERQGGAGTRTIQYEDYIVNGNVVETKEVSRTVAPVNEVVKVGTLVKVKPTVEITNLTKV  
 ENKKSITVSYNLIDTTSAYVSAKTQVFHGDKLVEVDIENPAKEQVISGLDYTPYTVKTHLTYNLGEN  
 NEENTETSTQDFQLEYKKIEIKDIDSVELYKGENDRYRRYLSLSEAPDTAKYFVKVKSDFKEMYL  
 KSTITENTDGTYKVTVAVDQLVEEGTDGYKDDYFTVAKSKAEQPGVYTSFKQLVTAMQSNLSGVYTLAS  
 DMTADEVSLGDKQTSYLTGAFTGSLIGSDGTSYAIYDLKKPLFDTLNGATVRDLDIKTVSADSKENVA  
 ALAKAANSANINNVAVEGKISGAKSVAGLVSATNTVIENSSFTGKLIANHQDSNKNNDTGGIVGNITGN  
 SSRVNVKVRVDALISTNARNNNQTAGGIVGRLENGALISNSVATGEIRNGQGYSRVGGIVGSTWQNGRVN  
 NVVSNVDVGDGYVITGDQYAAADVKNASTSVDNRKADRFATKLSKDQIDAKVADYGITVTLDDTGQDLK  
 RNLREVDTYRLNKAERKVAYSNIEKLMFPYKNDLVVHYGNKVATTDKLYTTELLDVVPMKDDEVTD  
 INNKKNSINKVMLHFKDNTVEYLDVTFKENFINSQVIEYNVTGKEYIFTPEAFVSDYTAITNNVLSDLQ  
 NVTNLN

**SP072 nucleotide (SEQ ID NO:119)**

TTTTAACCCAACTGTTGGTACTTTCTCTTTTACTGTCAGGATTGAGCTTGTTAGTTTTATTGGTTTCTAA  
 AAGGGAAAATGGAAAGAAACGACTTGTTCATTTTCTGCTGTTGACTAGCATGGGAGTTCAATTGTTGCC  
 GGCCAGTGCTTTTGGGTTGACCAGCCAGATTTTATCTGCCTATAATAGTCAGCTTTCTATCGGAGTCGG  
 GGAACATTTACCAGAGCCTCTGAAAAATCGAAGGTTATCAATATATTGGTTATATCAAACTAAGAAACA  
 GGATAATACAGAGCTTTCAAGGACAGTTGATGGGAAATACTCTGCTCAAAGAGATAGTCAACCAAATC  
 TACAAAAACATCAGATGTAGTTCATTCAGCTGATTTAGAATGGAACCAAGGACAGGGGAAGGTTAGTTT  
 ACAAGGTGAAGCATCAGGGGATGATGGACTTTCAGAAAAATCTTCTATAGCAGCAGACAATCTATCTTC  
 TAATGATTTCATTCGCAAGTCAAGTTGAGCAGAATCCGGATCACAAAGGAGAATCTGTAGTTCGACCAAC  
 AGTGCCAGAACAAAGGAAATCCTGTGTCTGCTACAACGGTGTCAGAGTGCGGAAGAGGATATTGGCGAC  
 GACAAATGATCGACCAGAGTATAAACTTCCATTGCAAAACCAAGGCACGCAAGAACCCGGTCATGAGGG  
 TGAAGCCGCAGTCCGTGAAGACTTACCAGTCTACAGTACAGCCACTAGAAACCAAGGTACACAAGGACC  
 CGGACATGAAGGTGAAGCTGCAGTTCGCGGAGGAAGAACCAGCTTACACAGAACCCTTAGCAACGAAAGG  
 CACGCAAGAGCCAGGTCTAGAGGGCAAAGCTACAGTCCGCGAAGAGACTCTAGAGTACACGGAACCCGT  
 AGCGACAAAAGGCACACAAGAACCCGAACATGAGGGCGAaCGGsCAGTAGAAGAAGAATTCGGGCTTT  
 AGAGGTCACTACACGAAATAGAACGGAATCCAGAATATTCCTTATACAACAGAAGAAATTCAGGATCC  
 AACACTTCTGAAAAATCGTCGTAAGATTGAACGACAAGGGCAAGCAGGGACACGTACAATTCATATGA  
 AGACTACATCGTAAATGGTAATGTCTGTAGAACTAAAGAAGTGTACGAACTGAAGTAGCTCCGGTCAA  
 CGAAGTCGTTAAAGTAGGAACACTTGTGAAAGTTAAACCTACAGTAGAAATTACAACTTAACAAAAGT  
 TGAGAACAAAAATCTATAACTGTAAGTTATAACTTAATAGACACTACCTCAGCATATGTTTCTGCAAA  
 AACGCAAGTTTTCCATGGAGACAAGCTAGTTAAAGAGGTGGATATAGAAAATCCTGCCAAAGAGCAAGT  
 AATATCAGGTTTAGATTACTACACACCGTATACAGTTAAAACACACCTAACTTATAATTTGGGTGAAAA  
 TAATGAGGAAAATACTGAAACATCAACTCAAGATTTCCAATTAGAGTATAAGAAAATAGAGATTAAAGA  
 TATTGATTTCAGTAGAATTATACGGTAAAGAAAATGATCGTTATCGTAGA

**SP072 amino acid (SEQ ID NO:120)**

FNPTVGTFLFTAGLSLLVLLVSKRENGKKRLVHFLLLTSMGVQLLPASAFGLTSQILSAYNSQLSIGVG  
 EHLPEPLKIEGYQYIGYIKTKKQDNTELSRTVDGKYSAQRDSQPNSTKTSADVHSADLEWNQGGKVS  
 QGEASGDDGLSEKSSIAADNLSSNDSFASQVEQNPDKHGESVVRPTVPEQGNPVSATTVQSAEEVLAT  
 TNDREYKLPLETGKTQEPGHEGEAAVREDLPVYTKPLETKGTQGPHEGEAAVREEEPAYTEPLATKG  
 TQEPGHEGKATVREETLEYTEPVATKGTQEPHEGERXVEEELPALEVTTNRNTEIQNIPTTTEEIQDP  
 TLLKNRRKIERQGGAGTRTIQYEDYIVNGNVVETKEVSRTVAPVNEVVKVGTLVKVKPTVEITNLTKV  
 ENKKSITVSYNLIDTTSAYVSAKTQVFHGDKLVEVDIENPAKEQVISGLDYTPYTVKTHLTYNLGEN  
 NEENTETSTQDFQLEYKKIEIKDIDSVELYKGENDRYRR

**SP073 nucleotide (SEQ ID NO:121)**

TCGTAGATATTTAAGTCTAAGTGAAGCGCCGACTGATACGGCTAAATACTTTGTAAAAGTGAAATCAGA  
 TCGCTTCAAAGAAATGTACCTACCTGTAAAATCTATTACAGAAAATACGGATGGAACGTATAAAGTGAC  
 GGTAGCCGTTGATCAACTTGTCTGAAGAAGGTACAGACGGTTACAAAGATGATTACACATTTACTGTAGC  
 TAAATCTAAAGCAGAGCAACCAAGGTTTACACATCCTTTAAACAGCTGGTAACAGCCATGCAAGCAA  
 TCTGTCTGGTGTCTATACATTGGCTTCAGATATGACCGCAGATGAGGTGAGCTTAGGCGATAAGCAGAC

Table 1

75

AAGTTATCTCACAGGTGCATTTACAGGGAGCTTGATCGGTTCTGATGGAACAAAATCGTATGCCATTTA  
 TGATTTGAAGAAACCATTATTTGATACATTAAATGGTGCTACAGTTAGAGATTTGGATATTTAAACTGT  
 TTCTGCTGATAGTAAAGAAAATGTCGCAGCGCTGGCGAAGGCAGCGAATAGCGCGAATATTAATAATGT  
 TGCAGTAGAAGGAAAAATCTCAGGTGCGAAATCTGTTGCGGGATTAGTAGCGAGCGCAACAAAATACAGT  
 GATAGAAAAACAGCTCGTTTTACAGGGAACTTATCGCAAATCACCAGGACAGTAATAAAAAATGATACTGG  
 AGGAATAGTAGGTAATATAACAGGAAATAGTTCGAGAGTTAATAAAAGTTAGGGTAGATGCCCTTAATCTC  
 TACTAATGCACGCAATAATAACCAAACAGCTGGAGGGATAGTAGGTAGATTAGAAAATGGTGCATTGAT  
 ATCTAATTCGGTTGCTACTGGAGAAATACGAAATGGTCAAGGATATTTCTAGAGTCCGAGGAATAGTAGG  
 ATCTACGTGGCAAAACGGTCGAGTAAATAATGTTGTGAGTAACGTAGATGTTGGAGATGGTTATGTTAT  
 CACCGGTGATCAATACGCAGCAGCAGATGTGAAAAATGCAAGTACATCAGTTGATAATAGAAAAGCAGA  
 CAGATTCGCTACAAAATTATCAAAAGACCAAATAGACGCGAAAGTTGCTGATTATGGAATCACAGTAAC  
 TCTTGATGATACTGGGCAAGATTTAAAACGTAATCTAAGAGAAGTTGATTATACAAGACTAAATAAAGC  
 AGAAGCTGAAAGAAAAGTAGCTTATAGCAACATAGAAAACTGATGCCATTCTACAATAAAGACCTAGT  
 AGTTCATATGGTAACAAAGTAGCGACAACAGATAAACTTTACACTACAGAATTGTTAGATGTTGTGCC  
 GATGAAAGATGATGAAGTAGTAACGGATATTAATAATAAGAAAAATTCATAAAATAAAGTTATGTTACA  
 TTTCAAAGATAATACAGTAGAATACCTAGATGTAACATTCAAAGAAAACCTTCATAAACAGTCAAGTAAT  
 CGAATACAATGTTACAGGAAAAGAATATATATTCACACCAGAAGCATTGTTTCAGACTATACAGCGAT  
 AACGAATAACGTACTAAGCGACTTGCAAAATGTAACACTTAAC

**SP073 amino acid (SEQ ID NO:122)**

RRYLSLSEAPDITAKYFVKVKSDFKEMYLPVKSITENTDGYKVTVAVDQLVEEGTDGYKDDYTFTVA  
 KSKAEQPGVYTSFKQLVTAMQSNLSGVYTLASDMTAEVSLGDKQTSYLTGAFTGSLIGSDGTSYAIY  
 DLKKPLFDLTNGATVRDLDIKTVSADSKENVAALAKAANSANINNVAVEGKISGAKSVAGLVASATNTV  
 IENSSFTGKLIANHQDSNKNDDGGIVGNITGNSSRVNKNVRVDALISTNARNNNQTAGGIVGRLENGALI  
 SNSVATGEIRNGQGSYRVGGIVGSTWQNGRVNNVSNVDVGDGYVITGDQYAAADVKNASTSVDNRKAD  
 RFATKLSKDQIDAKVADYGITVTLDDTGQDLKRNLRVDYTRLNKAERKVAISNIEKLMFPYFNKDLV  
 VHYGNKVATTDKLYTTELLDVVPMKDDEVVTDINNKKNSINKVMLHFKDNTVEYLDVTFKENFINSQVI  
 EYNVTGKEYIFTPEAFVSDYTAITNNVLSDLQNVTLN

**SP074 nucleotide (SEQ ID NO:123)**

CTTTGGTTTTGAAGGAAGTAAGCGTGGACAATTTGCTGTAGAAGGAATCAATCAACTTCGTGAGCATGT  
 AGACACTCTATTGATTATCTCAAAACAACAATTTGCTTGAAATTGTTGATAAGAAAACACCGCTTTTGGGA  
 GGCTCTTAGCGAAGCGGATAACGTTCTTCGTCAAGGTGTTCAAGGGATTACCGATTGATTACCAATCC  
 AGGATTGATTAACTTGGACTTTGCCGATGTGAAAACGGTAATGGCAAACAAAGGGAATGCTCTTATGGG  
 TATTGGTATCGGTAGTGGAGAAGAACGTGTGGTAGAAGCGGCACGTAAGGCAATCTATTACCACCTTCT  
 TGAAACAACATATTGACGGTGCTGAGGATGTTATCGTCAACGTTACTGGTGGTCTTGACTTAACCTTGAT  
 TGAGGCAGAAGAGGCTTCACAAATTGTGAACCAGGCAGCAGGTCAAGGAGTGAACATCTGGCTCGGTAC  
 TTCAATTGATGAAAGTATGCGTGATGAAATTCGTGTAACAGTTGTTGCAACGGGTGTTTCGTCAAGACCG  
 CGTAGAAAAGGTTGTGGCTCCACAAGCTAGATCTGCTACTAACTACCGTGAGACAGTGAAACCAGCTCA  
 TTCACATGGCTTTGATCGTCATTTTGATATGGCAGAAACAGTTGAATTGCCAAAACAAAATCCACGTCCG  
 TTTGGAACCAACTCAGGCATCTGCTTTTGGTGATTGGGATCTTCGCCGTGAATCGATTGTTTCGTACAAC  
 AGATTCAAGTCGTTTCTCAGTCGAGCGCTTTGAAGCCCCAATTTACAAAGATGAAGATGAATTGGATAC  
 ACCTCCATTTTCAAAAATCGT

**SP074 amino acid (SEQ ID NO:124)**

FGFEGSKRGQFAVEGINQLREHVDTLIIISNNNLEIVDKKTPLEALSEADNVLRQGVQGITDLITNP  
 GLINLDFADVKTVMANKGNALMGIGSGEERVVEAARKAIYSPLLETTIDGAEDVINVTGGDLTLI  
 EAEESQIVNQAAAGQVNIWLGTSIDESMRDEIRVTVVATGVRQDRVEKVVAPOARSATNYRETVKPAH  
 SHGFDRHFDMAETVELPKQNPRRLEPTQASAFGDWDLRRESIVRTTDSVSPVERFEAPISQDEDELDT  
 PPFKNR

**SP075 nucleotide (SEQ ID NO:125)**

CTACTACCTCTCGAGAGAACTGACCTAGAGGTGACCGTTTTTTGACCATGAGCAAGGTCAAGCCACCAA  
 GGCCGACGAGGAATATCAGTCCTTGGTTTTTCCAAACGCCGTAATAAAGCCTGGTACAAGATGGCGCG  
 TTGGGGGGTGATTTTTATGTGATTTATTAGCTAATTTAGAGAAATCAGGACAAGAAATCGACTTTTA  
 CCAGCGTTCGGGAGTCTTCTCTTGAAAAAGGATGAATCCAATTTGGAAGAATTTATCAACTGGCCCT  
 CCAGCGCAGAGAAGAATCTCCCTTGATAGGGCAATTAGCCATTCTGAACCAAGCCTCAGCTAATGAATT  
 ATTCCCTGGTTTGCAGGGATTGACCGCTGCTCTATGCTTCTGGTGGAGCGAGAGTAGATGGCCAAT

Table 1

TTTAGTGACTCGTTTGGCTGGAAGTCAGTCATGTCAAGCTGGTCAAAGAAAAAGTGACTCTGACACCGTT  
AGCATCAGGCTACCAGATTGGTGAAGAGGAGTTTGGAGCAGGTTATTTTGGCGACGGGAGCTTGGTTGGG  
GGACATGTTAGAGCCTTTAGGTTATGAAGTGGATGTCCGTCCTCAAAAAGGACAACACGAGATTATCA  
GCTTGCCCAAGACATGGAAGATTACCCTGTTGTTCATGCCAGAAGGGGAGTGGGATTTGATTCCCTTTGC  
AGGTGGGAAATTATCCTTAGGCGCTACCCACGAAAATGACATGGGATTTGATTTGACGGTAGATGAAAC  
CTTGCTCCAACAAATGGAGGAGGCCACCTTGACTCACTATCTGATTTTGGCTGAAGCTACTTCAAAATC  
TGAGCGTGTGGAAATCCGTGCCTACACCAGTGATTTCTCTCCTTTCTTTGGGCAGGTGCCTGACTTAAC  
TGGTGTCTATGCAGCCAGTGGACTAGGTTTCATCAGGCCTCACAACCTGGTCCTATCATTTGGTTACCATCT  
AGCCCAACTGATCCAAGACAAGGAGTTGACCTTGACCCCTCTAAATTACCCAATTGAAACTATGTCAA  
ACGAGTAAAAAGCGAA

**SP075 amino acid (SEQ ID NO:126)**

YYLSRESLDLEVTVFDHEQGQATKAAAGIIISPWFSKRNRKAWYKMARLGADFYVDLLADLEKSGQEIDFY  
QRSVFLKKDESNLEELYQLALQRREESPLIGQLAILNQASANELFPGLQGFDRLLYASGGARVDGQL  
LVTRLLEVSHVKLVKEKVTLTPLASGYQIGEEFEQVILATGAWLGDMLEPLGYEVDVRPQKQQLRDYQ  
LAQDMEDYPVVMPEGEWDLIPFAGGKLSLGATHENDMGFDLTVDETLLQQMEEATLTHYLILAEATSKS  
ERVGIRAYTSDFSFFGQVDPDLTGVIYAASGLGSSGLTTGPIIGYHLAQLIQDKELTDLPLNYPNIENYVK  
RVKSE

**SP076 nucleotide (SEQ ID NO:127)**

TAAGGTCAAAAGTCAGACCGCTAAGAAAGTGCTAGAAAAGATTGGAGCTGACTCGGTTATCTCGCCAGA  
GTATGAAATGGGGCAGTCTCTAGCACAGACCATTCTTTCCATAATAGTGTTGATGTCTTTTCAGTTGGA  
TAAAAATGTGTCTATCGTGGAGATGAAAATTCCTCAGTCTTGGGCAGGTCAAAGTCTGAGTAAATTAGA  
CCTCCGTGGCAAATACAATCTGAATATTTTGGGTTTCCGAGAGCAGGAAAATCCCCATTGGATGTTGA  
ATTTGGACCAGATGACCTCTTGAAAGCAGATACCTATATTTTGGCAGTCATCAACAACCAGTATTTGGA  
TACCCTA

**SP076 amino acid (SEQ ID NO:128)**

KVKSQTAKKVLEKIGADSVISPEYEMQSLAQILFHNSVDVFQLDKNVSIVEMKIPQSWAGQSLSKLD  
LRGKYNLNLGFRQENSPLDVEFGPDDLLKADTYILAVINNQYLDTL

**SP077 nucleotide (SEQ ID NO:129)**

TGACGGGTCTCAGGATCAGACTCAGGAAATCGCTGAGTGTTTAGCTAGCAAGTATCCTAATATCGTTAG  
AGCCATCTATCAGGAAAATAAATGCCATGGCGGTGCGGTCAATCGTGGCTTGGTAGAGGCTTCTGGGCG  
CTATTTTAAAGTAGTTGACAGTGATGACTGGGTGGATCCTCGTGCCTACTTGAAAATTCCTGAAACTTG  
CAGGAACTTGAGAGCAAAGGTCAAGAGGTGGATGTCTTTG

**SP077 amino acid (SEQ ID NO:130)**

DGSQDQTQEI AECLASKYPNIVRAIYQENKCHGGAVNRGLVEASGRYFKVVDSDDWVDPRAYLKILETC  
RNLRAKVKRWMSL

**SP078 nucleotide (SEQ ID NO:131)**

TAGAGGCTTTGCCAAATGGTGGGAAGGGCACGAGCGTCGAAAAGAGGAACGCTTTGTCAAACAAGAAGA  
AAAAGCTCGCCAAAAGGCTGAGAAAGAGGCTAGATTAGAACAAGAAGAGACTGAAAAAGCCTTACTCGA  
TTTGCTCCTGTTGATATGGAACGGGTGAAATTCAGACAGAGGAAGCTGTTCAAAATCTTCCACCTAT  
TCCAGAAGAAAAGTGGGTGGAACCAGAAATCATCCTGCCTCAAGCTGAACTTAAATTCCTGAACAGGA  
AGATGACTCAGATGACGAAGATGTTCAAGTCGATTTTTCAGCCAAAAGAAGCCCTTGAATACAACTTCC  
AAGCTTACAACCTTTTGCACCAGATAAACCAAAAGATCAGTCTAAAGAGAAGAAAATTGTCAGAGAAAA  
TATCAAAATCTTAGAAGCAACCTTTGCTAGCTTTGGTATTAAGGTAACAGTTGAACGGGCCGAAATTGG  
GCCATCAGTGACCAAGTATGAAGTCAAGCCGGCTGTTGGTGAAGGGTCAACCGCATTTCCAATCTATC  
AGATGACCTCGCTCTAGCCTTGGCTGCCAAAGATGTCCGGATTGAAGCACCATCCCTGGGAAATCCCT  
AATCGGAATTGAAGTGCCCAACTCCGATATTGCCACTGTATCTTTCCGAGAACTATGGGAACAATCGCA  
AACGAAAGCAGAAAATTTCTTGGAATTCCTTTAGGGAAGGCTGTTAATGGAACCGCAAGAGCTTTTGA  
CCTTTCTAAAATGCCCCACTTGCTAGTTGCAGGTTCAACGGGTTCAAGGAAGTCAGTACAGTTAACGG  
CATTATTGCTAGCATTCTCATGAAGGCGAGACCAGATCAAGTTAAATTTATGATGGTCGATCCCAAGAT  
GGTTGAGTTATCTGTTTACAATGATATTCACCTCTTTGATTCCAGTCGTGACCAATCCACGCAAAGC  
CAGCAAGGCTCTGCAAAGGTTGTGGATGAAATGGAAAACCGTTATGAACTCTTTGCCAAGGTGGGAGT  
TCGGAATATTGCAGGTTTTAATGCCAAGGTAGAAGAGTTCAATTCACAGTCTGAGTACAAGCAAATTC

WO 98/18930

PCT/US97/19422

Table 1

77

GCTACCATTTCATTGTCGTGATTGTGGATGAGTTGGCTGACCTCATGATGGTGGCCAGCAAGGAAGTGGGA  
 AGATGCTATCATCCGTCCTTGGGCAGAAGGCGCGTGCTGCAGGTATCCACATGATTCTTGCAACTCAGCG  
 TCCATCTGTTGATGTCATCTCTGGTTTGATTAAAGGCCAATGTTCCATCTCGTGTAGCATTTCGCGTTTC  
 ATCAGGAACAGACTCCCGTACGATTTTGGATGAAAAATGGAGCAGAAAAAATTCTTGGTCGAGGAGACAT  
 GCTCTTTTAAACCGATTGATGAAAAATCATCCAGTTCGTCTCCAAGGCTCCTTTATCTCGGATGACGATGT  
 TGAGCGCATTGTGAACCTTCATCAAGACTCAGGCAGATGCAGACTACGATGAGAGTTTTGATCCAGGTGA  
 GGTCTCTGAAAAATGAAGGAGAATTTTCGGATGGAGATGCTGGTGGTGAATCCGCTTTTTGAAGAAGCTAA  
 GTCTTTGGTTATCGAAACACAGAAAGCCAGTGCGTCTATGATTACGCGTCGTTTATCAGTTGGATTAA  
 CCGTGCGACCCGTCATGGAAGAACTGGAGATAGCAGGTGTCATCGGTCCAGCTGAAGGTACCAAAC  
 TCGAAAAGTGTACAACAA

**SP078 amino acid (SEQ ID NO:132)**

RGFAKWWEHERRKEERFVKQEEKARQKAEKEARLEQEETEKALLDLPPVDMETGEILTEEAVQNLPPPI  
 PEEKWVEPEIILPQAEKFPEQEDSDDEDVQVDFSAKEALEYKLPQLFAPDKPKDQSKEKKIVREN  
 IKILEATFASFGIKVTVRAEIGPSVTKYEVKPAVGVRVNRISNLSDDLALALAAKDVRIEAPIPGKSL  
 IGIEVPNSDIATVSFRELWEQSQTKAENFLEIPLGKAVNGTARAFDL SKMPHLLVAGSTGSGKSVAVNG  
 I IASILMKARPDQVKFMMVDPKMWELSVYNDI PHLLIPVVTNPRKASKALQKVVDENRYELFAKGV  
 RNIAGFNAKVEEFNSQSEYKQIPLPFIVIVDELADLMMVASKEVEDAIIRLGQKARAAGIHMILATQR  
 PSVDVISGLIKANVPSRVAFAVSSGTDSTILDENGAEKLLGRGDMLFKPIDENHPVRLQGSFI SDDDV  
 ERIVNFIKTQADADYDESFDPGEVSENEGEFSGDAGGDPLFEEAKSLVIETQKASASMIQRRLSVGFN  
 RATRLMEELEIAGVIGPAEGTKPRKVLQQ

**SP079 nucleotide (SEQ ID NO:133)**

TCAAAAAGAGAAGGAAAACCTTGGTTATTGCTGGGAAAATAGGTCCAGAACCCAGAAATTTTGGCCAATAT  
 GTATAAGTTGCTGATTGAAGAAAATACCAGCATGACTGCGACTGTTAAACCGAATTTTGGGAAGACAAG  
 CTTCTTTTATGAAGCTCTGAAAAAAGGCGATATTGACATCTATCCTGAATTTACTGGTACGGTGACTGA  
 AAGTTTGCTTCAACCATCACCAAGGTGAGTCATGAACCAGAACAGGTTTATCAGGTGGCGCGTGATGG  
 CATTGCTAAGCAGGATCATCTAGCCTATCTCAAACCCATGTCTTATCAAAACACCTATGCTGTAGTGT  
 TCCGAAAAGATTGCTCAAGAATATGGCTTGAAGACCATTTCAGACTTGAAAAAAGTGAAGGAGCAGTT  
 GAAGGCAGGTTTTACTCTGAGTTTAAACGACCGTGAAGATGGAATAAAGGCTTGCAATCAATGTATGG  
 TCTCAATCTCAATGAGCACCATTGAGCCAGCCCTTCGCTATCAGGCTATTTCAGTCAGGGGATATTCA  
 AATCACGGATGCCTATTTCGACTGATGCGGAATTGGAGCGTTATGATTTACAGGTCTTGGAAGATGACAA  
 GCAACTCTTCCCACCTTATCAAGGGGCTCCACTCATGAAAGAAAGCTCTTCTCAAGAAACACCCAGAGTT  
 GGAAAGAGTTCTTAATACATTGGCTGGTAAGATTACAGAAAGCCAGATGAGCCAGCTCAACTACCAAGT  
 CGGTGTTGAAGGCAAGTCAGCAAAGCAAGTAGCCAAGGAGTTTCTCCAAGAAACAAGGTTTGTGAAGAA  
 A

**SP079 amino acid (SEQ ID NO:134)**

QKEKENLVIAGKIGPEPEILANMYKLLIEENTSMTATVKPNFGKTSFLYEALKKGDIDIIYPEFTGTVTE  
 SLLQPSPKVSHEPEQVYQVARDGIAKQDHLAYLKPMYQNTYAVAVPKKIAQEYGLKTI SDLKKVEGQL  
 KAGFTLEFNDREDGNKGLQSMYGLNLNVATIEPALRYQAIQSGDIQITDAYSTDAELERYDLQVLEDDK  
 QLFPPYQGAAPLMKEALLKKHPELERVNLTLGKITESQMSQLNYQVGVGEGKSAKQVAKFLQEQGLLKK

**SP080 nucleotide (SEQ ID NO:135)**

ACGTTCTATTGAGGACCACTTTGATTCAAACCTTCGAATTGGAATATAACCTCAAAGAAAAAGGGAAAAC  
 AGATCTTTTGAAGCTAGTTGATAAAACAACCTGACATGCGTCTGCATTTTATCCGCCAAACTCATCCACG  
 CGGTCTCGGAGATGCTGTTTTGCAAGCCAAGGCTTTCGTCCGAAATGAACCTTTTGTCTGTTATGCTTGG  
 TGATGACTTGATGGATATCACAGACGAAAAGGCTGTTCCACTTACCAAACAACCTCATGGATGACTACGA  
 GCGTACCCACGCGTCTACTATCGCTGTATGCCAGTCCCTCATGACGAAGTATCTGCTTACGGGGTTAT  
 TGCTCCGCAAGGCGAAGGAAAAGATGGTCTTTACAGTGTGAAACCTTTGTTGAAAAACCAGCTCCAGA  
 GGACGCTCCTAGCGACCTTGCTATTATCGGACGCTACCTCCTCACGCCTGAAATTTTTGAGATTCTCGA  
 AAAGCAAGCTCCAGGTGCAGGAAATGAAATTCAGCTGACAGATGCAATCGACACCCCTCAATAAAACACA  
 ACGTGTATTTGCTCGTGAGTTCAAAGGGGCTCGTTACGATGTCCGAGACAAGTTTGGCTTCATGAAAAAC  
 ATCCATCGACTACGCCCTCAAACACCCACAAGTCAAAGATGATTGAAGAATTACCTCATCCAACCTTGG  
 AAAAGAATTGACTGAGAAGGAA

Table 1

**SP080 amino acid (SEQ ID NO:136)**

RSIEDHFDNSFELEYNLKEKGKTDLLKLVDKTTDMRLHFIQTHPRGLGDAVLQAKAFVGNPFVVMGLG  
DDLMDITDEKAVPLTKQLMDDYERTHASTIAVMPVPHDEVSAYGVIAPQGEGKDGLYSVETFVEKPAPE  
DAPSDLAIIGRYLLTPEIFEILEKQAPGAGNEIQLTDAIDTLNKTQRVFAREFKGARYDVGDKFGFMKT  
SIDYALKHPQVKDDLLKNYLIQLGKELTEKE

**SP081 nucleotide (SEQ ID NO:137)**

CGCTCAAAATACCAGAGGTGTTTCAGCTAATCGAGCACGTTTCTCCTCAAATGTTGAAAGCCCAATTGGA  
GAGTGTCTTTTCTGATATTCCACCTCAGGCTGTAAAACTGGAATGTTGGCTACTACTGAAATCATGGA  
AATCATCCAACCTTATCTTAAAAAAGTGGATTGTCCCTATGTCCTTGATCCTGTTATGGTTGCTACAAG  
TGGAGATGCCTTGATTGACTCAAATGCTAGAGACTATCTCAAAACAACTTACTACCTCTAGCAACTAT  
TATTACGCCAAATCTTCTCCTGAAGCAGAAGAGATTGTTGGTTTTTCAATCCATGACCCCGAAGACATGCA  
GCGTGCTGGTTCGCTGATTTTAAAAGAAATTTGGTCTCAGTCTGTGGTTATCAAAGGCGGACATCTCAA  
AGGTGGTGCTAAAGATTTCTCTTTACCAAGAATGAACAATTTGTCTGGGAAAGCCCACGAATTCAAAC  
CTGTACACCCCATGGTACT

**SP081 amino acid (SEQ ID NO:138)**

AQNTGRGVQLIEHVSPQMLKAQLESVFSIPPQAVKTMGLATTEIMEIIQPYLKKLDCPYVLDPVMVATS  
GDALIDSNARDYLKTNLLPLATIITPNLPEAEIEVGFSDPEDMQRAGRLILKEFGPQSVVIKGGHLK  
GGAKDFLFTKNEQFVWESPRIQTCHTHGT

**SP082 nucleotide (SEQ ID NO:139)**

AATTGTACAATTAGAAAAAGATAGCAAAATCAGACAAAGAACAAGTTGATAAACTATTTGAATCATTTGA  
TGCATCTTCAGATGAATCTATTTCTAAATTTAAAAGAATCTCTGAACTTCACTTAAAACCGATGCAGG  
TAAAGACTTATCTTAATAACAAAGTCAAAGAATCATCTAAAGCAATTGTAGATTTTTCATTTGCAAAAAGG  
TTTGGCTTATGATGTTAAAGATTGATGACAAATTTAAAGATAAAGCAACTCTTGAAACAAATGTAAA  
AGAAATTACAAAACAAATTGATTTTATCAAAAAAGTTGATGAACTTTTAAACAAGAGAATTTGGAAGA  
AACTCTTAAATCTCTAAATGATCTTGTGATAAATATCAAAAACAAATCGAACTTTTGAAGAAAGAAGA  
AGAAAAAGCTGCTGAAAAAGCTGCTGAAAAAGCAAAGGAATCTTCTAGTCAAAGTAATCTTCTGGTAG  
TGCTTCTAATGAGTCTTATAATGGATCTTCCAATTCAAATGTAGATTATAGTTTCATCTGAACAACTAA  
TGGATATTCAAATAATTATGGCGGTCAAGATTATTCTGGTTCAGGAGATAGTTCAACAAATGGTGGATC  
ATCAGAACAATATTCATCTAGCAATTCAAACAGCGGAGCAAATAATGTCTACAGATATAAAGGCACTGG  
TGCTGACGGCTATCAAAGATACTACTACAAAGATCATAATAATGGAGATGTGTATGATGACGATGGAAA  
TTACCTTGGGAACCTTTGGTGGCGGCATTGCAGAACCTAGTCAACGC

**SP082 amino acid (SEQ ID NO:140)**

IVQLEKDSKSDKEQVDKLFESFDASSDESISKLELSESLKTDAGKDYLNKVKESSKAIVDFHLQKG  
LAYDVKDSDDKFKDKATLETNVKEITKQIDFIKKVDETFKQENLEETLKSLNDLVDKYQKQIELLKKEE  
EKAAEKAAEKAKESSQSNSSGSASNESYNGSSNSNVYSSSEQTNGYSNNYGGQDYSGSGDSSTNGGS  
SEQYSSNSNSGANNVRYKGTGADGYQRYYYKDHNNGDVYDDGNYLGNFGGGIAEPSQR

**SP083 nucleotide (SEQ ID NO:141)**

TCTGACCAAGCAAAAAGAAGCAGTCAATGACAAAAGGAAAAGCAGCTGTTGTTAAGGTGGTGGAAAGCCA  
GGCAGAACTTTATAGCTTAGAAAAGAAATGAAGATGCTAGCCTAAGAAAGTTACAAGCAGATGGACGCAT  
CACGGAAGAACAGGCTAAAGCTTATAAAGAATACAATGATAAAAATGGAGGAGCAAATCGTAAAGTCAA  
TGAT

**SP083 amino acid (SEQ ID NO:142)**

LTKQKEAVNDKKGAAVVKVVESQAELYSLEKNEDASLRKLQADGRITEEQAKAYKEYNDKNGGANRKVN  
D

**SP084 nucleotide (SEQ ID NO:143)**

GTCCGGCTCTGTCCAGTCCACTTTTTCAGCGGTAGAGGAACAGATTTTCTTTATGGAGTTTGAAGAACT  
CTATCGGGAAACCCAAAAACGCAGTGTAGCCAGTCAGCAAAAGACTAGTCTGAACCTTAGATGGGCAGAC  
GCTTAGCAATGGCAGTCAAAGTTGCCAGTCCCTAAAGGAATTCAGGCCCATCAGGCCAAAGTATTAC  
ATTTGACCGAGCTGGGGGCAATTCGTCCCTGGCTAAGGTTGAATTTAGACCAGTAAAGGAGCGATTCCG  
CTATCAATTATATCTAGGAAATGGAAAAATTAAACGCATTAAGGAAACAAAAAT

Table 1

**SP084 amino acid (SEQ ID NO:144)**

SGSVQSTFSAVEEQIFFMEFEELYRETQKRSVASQKTSNLNDGQTLNNGSQKLPVPRKGIQAPSGQSIT  
FDRAGGNSSLAKVEFQTSKGAIRYQLYLGNKIKRIKETKN

**SP085 nucleotide (SEQ ID NO:145)**

GGGACAAATTCAAAAAATAGGCAAGAGGAAGCAAAATCTTGCAAAAGGAAGTCTTGAGGGTAGC  
TAAGATGGCCCTGCAGACGGGGCAAAATCAGGTAAGCATCAACGGAGTTGAGATTCAGGTATTTCTAG  
TGAAAAAGGATTGGAGGTCTACCATGGTTCAGAACAGTTGTTGGCAATCAAAGAGCCA

**SP085 amino acid (SEQ ID NO:146)**

GQIQKNRQEEAKILQKEEVLRVAKMALQTGQNQVSINGVEIQVFSSEKGLEVYHGSEQLLAIKEP

**SP086 nucleotide (SEQ ID NO:147)**

TGCTACCAGCAACAAAGCGAGCAAAAGGAGTGGCTCTTGTGTTGTGGACCAACTTGAGGTAGAATTAGA  
CCGTTCGCAGTTCGAAAAAGTAGAAGGCAATCGCCTATACATGAAGCAAGATGGCAAGGACATCGCCAT  
CGGTAAGTCAAAGTCAGATGATTTCCGTAAACGAATGCTCGTGGTCGAGGTTATCAGCCTATGGTTTA  
TGGACTCAAATCTGTACGGATTACAGAGGACAATCAACTGGTTCGCTTTCATTTCCAGTTCCAAAAAGG  
CTTAGAAAGGGAGTTTCATCTATCGTGTGGAAAAAGAAAAAGT

**SP086 amino acid (SEQ ID NO:148)**

RYQQQSEQKEWLLFVDQLEVELDRSQFEKVEGNRLYMKQDGKDIAIGKSKSDDFRKTNARGRGYQPMVY  
GLKSVRITEDNQLVRHFHFQFKGLEREFYIRVEKEKS

**SP087 nucleotide (SEQ ID NO:149)**

GAACCGACAAGTCGCGCCACTATCAAGACTATGCTTTGAATAAAGAAAAATTGGTTGCTTTTGCTATGGC  
TAAACGAACCAAAGATAAGGTTGAGCAAGAAAGTGGGGAACAGTTTTTTAATCTAGGTCAGGTAAGCTA  
TCAAAACAAGAAACTGGCTTAGTGACGAGGGTTCGTACGGATAAGAGCCAATATGAGTTTCTGTTTCC  
TTCAGTCAAAATCAAAGAAGAGAAAAAGAGATAAAAAAGGAAGGTTAGCGACCGATTCAAGCGAAAAAGT  
GGAGAAGAAAAAATCAGAAGAGAAGCCTGAAAAGAAAGAGAATTCA

**SP087 amino acid (SEQ ID NO:150)**

NRQVAHYQDYALNKEKLVAFAMAKRTKDKVEQESGEQFFNLGQVSYQNKKTGLVTRVRTDKSQYEFLLP  
SVKIKEEKRDKKEEVATDSSEKVEKKKSEEKPEKKENS

**SP088 nucleotide (SEQ ID NO:151)**

GGTTGTCGGCTGGCAATATATCCCGTTTCCATCTAAAGGTAGTACAATTGGTCCTTACCCAAATGGTAT  
CAGATTAGAAGGTTTTCCAAAGTCAGAGTGGTACTACTTCGATAAAAATGGAGTGCTACAAGAGTTTGT  
TGGTTGGAACATTAGAGATTAAACTAAAGACAGTGTGGAAGAAAGTACGGGGAAAAACGTGAAGA  
TTCAGAAGATAAAGAAGAGAAGCGTTATTATACGAACTATTACTTTAATCAAAATCATTCTTTAGAGAC  
AGGTTGGCTTTATGATCAGTCTAACTGGTATTATCTAGCTAAGACGGAAATTAATGGAGAAACTACCT  
TGGTGGTGAAAGACGTGCGGGGTGGATAAACGATGATTGACTTGGTACTACCTAGATCCAACAACTGG  
TATTATGCAAACAGGTTGGCAATATCTAGGTAATAAGTGGTACTACCTCCGTTCCCTCAGGAGCAATGGC  
CACTGGCTGGTATCAGGAAGGTACCACTTGGTATTATTTAGACCACCCAAATGGCGATATGAAAAACAGG  
TTGGCAAAACCTTGGGAACAAATGGTACTATCTCCGTTTCATCAGGAGCTATGGCAACTGGTTGGTATCA  
AGATGGTTCAACTTGGTACTACCTAAATGCAGGTAATGGAGACATGAAGACAGGTTGGTTCCAGGTCAA  
TGGCAACTGGTACTATGCTTATAGCTCAGGTGCTTTGGCAGTGAATACGACCGTAGATGGCTATTCTGT  
CAACTATAATGGCGAATGGGTTCGG

**SP088 amino acid (SEQ ID NO:152)**

VVGWQYIPFPSKGSTIGPYPNGIRLEGFPKSEWYYFDKNGVLQEFVGWKTLEIKTKDSVGRKYGEKRED  
SEDKEEKRYTNYFNQNSLETGWLYDQSNWYYLAKTEINGENYLGGERAGWINDDSTWYYLDPTTG  
IMQTGWQYLGKWKYYLRSSGAMATGWYQEGTTWYYLDHPNGDMKTGWQNLGNKWKYYLRSSGAMATGWYQ  
DGSTWYYLNAAGNDMKTGWFWQVNGNWYYAYSSGALAVNTTVDGYSVNYNGEWVR

**SP089 nucleotide (SEQ ID NO:153)**

GGCCAAATCAGAATGGGTAGAAGACAAGGGAGCCTTTTATTATCTTGACCAAGATGGAAAGATGAAAAG  
AAATGCTTGGGTAGGAACTTCCTATGTTGGTGCAACAGGTGCCAAAGTAATAGAAGACTGGGTCTATGA  
TTCTCAATACGATGCTTGGTTTTATATCAAAGCAGATGGACAGCACGCAGAGAAAGAATGGCTCCAAAT

Table 1

80

TAAAGGGAAGGACTATTATTTCAAATCCGGTGGTTATCTACTGACAAAGTCAGTGGATTAATCAAGCTTA  
TGTGAATGCTAGTGGTGCCAAAGTACAGCAAGGTTGGCTTTTGGACAAACAATACCAATCTTGGTTTTTA  
CATCAAAGAAAATGGAACTATGCTGATAAAGAATGGATTTTCGAGAATGGTCACTATTATTATCTAAA  
ATCCGGTGGCTACATGGCAGCCAATGAATGGATTTGGGATAAGGAATCTTGGTTTTATCTCAAATTTGA  
TGGGAAAATGGCTGAAAAAGAATGGGTCTACGATTCTCATAGTCAAGCTTGGTACTACTTCAAATCCGG  
TGGTTACATGACAGCCAATGAATGGATTTGGGATAAGGAATCTTGGTTTTATCTCAAATCTGATGGGAA  
AATAGCTGAAAAAGAATGGGTCTACGATTCTCATAGTCAAGCTTGGTACTACTTCAAATCCGGTGGTTA  
CATGACAGCCAATGAATGGATTTGGGATAAGGAATCTTGGTTTTACCTCAAATCTGATGGGAAAATAGC  
TGAAAAAGAATGGGTCTACGATTCTCATAGTCAAGCTTGGTACTACTTCAAATCTGGTGGCTACATGGC  
GAAAAATGAGACAGTAGATGGTTATCAGCTTGGAGCGATGGTAAATGGCTTGGAGGAAAAACTACAAA  
TGAAAAATGCTGCTTACTATCAAGTAGTGCCTGTTACAGCCAATGTTTATGATTGAGATGGTGAAAAGCT  
TTCCTATATATCGCAAGGTAGTGTCTGATGGCTAGATAAGGATAGAAAAAGTGATGACAAGCGCTTGGC  
TATTACTATTTCTGGTTTGTCTAGGCTATATGAAAACAGAAGATTTACAAGCGCTAGATGCTAGTAAGGA  
CTTTATCCCTTATTATGAGAGTGATGGCCACCGTTTTTATCACTATGTGGCTCAGAATGCTAGTATCCC  
AGTAGCTTCTCATCTTTCTGATATGGAAGTAGGCAAGAAAATATTATTCGGCAGATGGCCTGCATTTTGA  
TGGTTTTAAGCTTGAGAATCCCTTCTTTTCAAAGATTTAACAGAGGCTACAAACTACAGTGCTGAAGA  
ATTGGATAAGGTATTTAGTTTGGCTAAACATTAACAATAGCCTTTTGGAGAACAAGGGCGCTACTTTTAA  
GGAAGCCGAAGAACATTACCATATCAATGCTCTTTATCTCCTTGCCCATAGTGGCCTAGAAAAGTAACTG  
GGGAAGAAGTAAATTGCCAAAGATAAGAATAATTTCTTTGGCATTACAGCCTATGATACGACCCCTTA  
CCTTTCTGCTAAGACATTTGATGATGTGGATAAGGGAATTTTAGGTGCAACCAAGTGGATTAAGGAAAA  
TTATATCGATAGGGGAAGAACTTCTCTTGAAACAAGGCTTCTGGTATGAATGTGGAATATGCTTCAGA  
CCCTTATTGGGGCGAAAAATTGCTAGTGTGATGATGAAAATCAATGAGAAG

**SP089 amino acid (SEQ ID NO:154)**

AKSEWVEDKGAFFYLDQDGKMKRNAWVGTSYVGATGAKVIEDWVYDSQYDAWFYIKADGQHAKEKWLQI  
KGKDYYFKSGGYLLTSQWINQAYVNASGAKVQQGWLFQYQSWFYIKENGNADKEWIFENGHYLLK  
SGGYMAANEWIWDKESWFLKFDGKMAEKWVYDSHSQAWYFYSKSGGYMTANEWIWDKESWFLKSDGK  
IAEKWVYDSHSQAWYFYSKSGGYMTANEWIWDKESWFLKSDGKIAEKWVYDSHSQAWYFYSKSGGYMA  
KNETVDGYQLGSDGKWLGGKTTNENAAYYQVVPVTANVYDSGEEKLSYISQGSVWLDKDRKSDDKRLA  
ITISGLSGYMKTEDLQALDASKDFIPYYESDGHFRFYHYVAQNASIPVASHLSDMEVGKKYYSADGLHFD  
GFKLENPFLFKDLTEATNYSAEELDKVFSLLNINNSLLENKGATFKEAEEHYHINALYLLAHSALSNW  
GRSKIADKNNFFGITAYDTTPYLSAKTFDDVDKGILGATKWIKENYIDRGRTFLGNKASGMNVEYASD  
PYWGEKIASVMMKINEK

**SP090 nucleotide (SEQ ID NO:155)**

ATTTGCAGATGATTCTGAAGGATGGCAGTTTGTCCAAGAAAATGGTAGAACCTACTACAAAAGGGGGA  
TCTAAAAGAAACCTACTGGAGAGTATAGATGGGAAGTACTATTATTTTGATCCTTTATCCGGAGAGAT  
GGTTGTCTGGCTGGCAATATATACCTGCTCCACACAAGGGGGTTACGATTGGTCCTTCTCCAAGAATAGA  
GATTGCTCTTAGACCAGATTGGTTTTATTTTGGTCAAGATGGTGTATTACAAGAATTTGTTGGCAAGCA  
AGTTTTTAGAAGCAAAAACCTGCTACGAATACCAACAAACATCATGGGGAAGAATATGATAGCCAAGCAGA  
GAAACGAGTCTATTATTTTGAAGATCAGCGTAGTTATCATACTTTAAAAACTGGTTGGATTTATGAAGA  
GGGTCATTGGTATTATTTACAGAAGGATGGTGGCTTTGATTTCGCGCATCAACAGATTGACGGTTGGAGA  
GCTAGCACGTGGTTGGGTTAAGGATTACCCTCTTACGTATGATGAAGAGAAGCTAAAAGCAGCTCCATG  
GTACTATCTAAATCCAGCAACTGGCATTATGCAACAGGTTGGCAATATCTAGGTAATAGATGGTACTA  
CCTCCATTCTGTCAGGAGCTATGGCAACTGGCTGGTATAAGGAAGGCTCAACTTGGTACTATCTAGATGC  
TGAAAAATGGTGATATGAGAACTGGCTGGCAAAACCTTGGGAACAAATGGTACTATCTCCGTTTCATCAGG  
AGCTATGGCAACTGGTTGGTATCAGGAAAGTTTCGACTTGGTACTATCTAAATGCAAGTAATGGAGATAT  
GAAAACAGGCTGGTTCCAAGTCAATGGTAACTGGTACTATGCCTATGATTACAGGTGCTTTAGCTGTAA  
TACCACAGTAGGTGGTTACTACTTAACTATAATGGTGAATGGGTTAAG

**SP090 amino acid (SEQ ID NO:156)**

VFADDSEGWFVQENGRTYYKGDLEKTYWRVIDGKYIYFDPLSGEMVVGWQYIPAPHKGVITIGSPRI  
EIALRPDWFYFGQDGVLEFVGKQVLEAKTATNTNKHHEEYDSQAEKRVYFYFEDQRSYHTLKTGWIYE  
EGHWYYLQKDGDFDSRINRLTVGELARGWVKDYPLTYDEEKLKAAPWYYLNPATGIMQTGWQYLGNRWY  
YLHSSGAMATGWYKEGSTWYYLDAENGDMRTGWQNLGNKWYYLRSSGAMATGWYQESSTWYYLNASNGD  
MKTGWFQVNGNWYYAYDSGALAVNTTVGGYYLNYNGEWWK

Table 1

**SP091 nucleotide (SEQ ID NO:157)**

TGTCGCTGCAAAATGAACTGAAGTAGCAAAAACCTTCGCAGGATACAACGACAGCTTCAAGTAGTTCAGA  
 GCAAAATCAGTCTTCTAATAAAACGCAACGAGCGCAGAAGTACAGACTAATGCTGCTGCCACTGGGA  
 TGGGGATTATTATGTAAAGGATGATGGTTCTAAAGCTCAAAGTGAATGGATTTTTTGACAACTACTATAA  
 GGCTTGGTTTTATATTAATTCAGATGGTCGTTACTCGCAGAATGAATGGCATGGAAATTACTACCTGAA  
 ATCAGGTGGATATATGGCCCAAAACGAGTGGATCTATGACAGTAATTACAAGAGTTGGTTTTATCTCAA  
 GTCAGATGGGGCTTATGCTCATCAAGAATGGCAATTGATTGGAAAATAAGTGGTACTACTTCAAGAAGTG  
 GGGTTACATGGCTAAAAGCCAATGGCAAGGAAGTTATTTCTTGAATGGTCAAGGAGCTATGATGCAAAA  
 TGAATGGCTSCTATGATCCAGCCTATTCTGCTTATTTTTATCTAAAATCCGATGGAACTTATGCTAACC  
 AAGAGTGGCAAAAAGTGGGCGGCAATGGTACTATTTCAAGAAGTGGGGCTATATGGCTCGGAATGAGT  
 GGCAAGGCAACTACTATTTGACTGGAAGTGGTGCCATGGCGACTGACGAAGTGATTATGGATGGTACTC  
 GCTATATCTTTGCGGCCTCTGGTGAGCTCAAAGAAAAAAAAGATTGAATGTCGGCTGGGTTACACAGAG  
 ATGGTAAGCGCTATTTCTTTAATAATAGAGAAGAACAAGTGGGAACCGAACATGCTAAGAAAGTCATTG  
 ATATTAGTGAGCACAATGGTCGTATCAATGATTGGAAAAAGGTTATTGATGAGAACGAAGTGGATGGTG  
 TCATTGTTCTGCTAGGTTATAGCGGTAAAGAAGACAAGGAATTGGCGCATAACATTAAGGAGTTAAACC  
 GTCTGGGAATTCCTTATGGTGTCTATCTCTATACCTATGCTGAAAATGAGACCGATGCTGAGAGTGACG  
 CTAAACAGACCATTGAACTTATAAAGAAATACAATATGAACCTGTCTTACCCTATCTATTATGATGTTG  
 AGAATTGGGAATATGTAAATAAGAGCAAGAGAGCTCCAAGTGATACAGGCACTTGGGTTAAAATCATCA  
 ACAAGTACATGGACACGATGAAGCAGGCGGGTTATCAAAATGTGTATGTCTATAGCTATCGTAGTTTAT  
 TACAGACGCGTTTAAAACACCCAGATATTTTAAAACATGTAAACTGGGTAGCGGCCTATACGAATGCTT  
 TAGAATGGGAAAACCTCATTATTCAGGAAAAAAGGTTGGCAATATACCTCTTCTGAATACATGAAAG  
 GAATCCAAGGGCGCGTAGATGTACGCGTTTGGTAT

**SP091 amino acid (SEQ ID NO:158)**

VAANETEVAKTSQDTTASSSSSEQNQSSNKTQTSAEVQTNAAAHWDGDYVVKDDGSKAQSEWIFDNYYK  
 AWFYINSDGRYSQNEWHGNYLKSQGYMAQNEWIYDSNYKSWFYLKSDGAYAHQEWQLIGNKWWYFCKW  
 GYMAKSQWQGSYFLNGQGAMMQNEWLYDPAYSAYFYLKSDGTYANQEWQKVGKWWYFCKWGYMARNEW  
 QGNYLLTSGAMATDEVIDMDGTRYIFAASGELKEKKDLNVGWVHRDGRYFFNNREEQVQTEHAKKVID  
 ISEHNGRINDWKKVIDENEVDGVIVRLGYSGKEDKELAHNIKELNRLGIPYGVYLYTYAENETDAESDA  
 KQTIELIKKYNMNLSPYIYDVENWEYVNKSKRAPSDTGTWVKIINKYMDTMKQAGYQNVVYSYRSL  
 QTRLKHPDILKHVNWVAAYTNALEWENPHYSKKGWQYTSSEYMKGIQGRVDVSVWY

**SP092 nucleotide (SEQ ID NO:159)**

TACGCTCTCAGCCTACTTTTGTAAGAGCAGAAGAATCTCCACAAGTTGTCGAAAAATCTTCATTAGAGAA  
 GAAATATGAGGAAGCAAAAGCAAAAGCTGATACTGCCAAGAAAGATTACGAAACGGCTAAAAAGAAAGC  
 AGAAGACGCTCAGAAAAAGTATGAAGATGATCAGAAGAGAACTGAGGAGAAAGCTCGAAAAAGAACGAGA  
 AGCATCTCAAAAATTGAATGATGTGGCGCTTGTGTTCAAAATGCATATAAAGAGTACCGAGAAGTTCA  
 AAATCAACGTAGTAAATATAAATCTGACGCTGAATATCAGAAAAAATTAACAGAGGTCGACTCTAAAAT  
 AGAGAAGGCTAGGAAAGAGCAACAGGACTTGCAAAATAAATTTAATGAAGTAAGAGCAGTTGTAGTTCC  
 TGAACCAAATGCGTTGGCTGAGACTAAGAAAAAGCAGAAGAAGCTAAAGCAGAAGAAAAAGTAGCTAA  
 GAGAAAATATGATTATGCAACTCTAAAGGTAGCACTAGCGAAGAAAGAAAGTAGAGGCTAAGGAACTTGA  
 AATTGAAAAAATTCAATATGAAATTTCTACTTTGGAACAAGAAGTTGCTACTGCTCAACATCAAGTAGA  
 TAATTTGAAAAAACTTCTTGCTGGTGCGGATCCTGATGATGGCACAGAAGTTATAGAAGCTAAATTTAAA  
 AAAAGGAGAAGCTGAGCTAAACGCTAAACAAGCTGAGTTAGCAAAAAACAAACAGAACTTGAAAAACT  
 TCTTGACAGCCTTGATCCTGAAGGTAAGACTCAGGATGAATTAGATAAAGAAGCAGAAGAAGCTGAGTT  
 GGATAAAAAAGCTGATGAACTTCAAAATAAAGTTGCTGATTTAGAAAAAGAAATTAGTAACCTTGAAAT  
 ATTACTTGAGGGGGCTGATNCTGAAGATGATACTGCTGCTCTTCAAAATAAATTAGCTACTAAAAAAGC  
 TGAATTGGAAAAAATCAAAAAGAATTAGATGCAGCTCTTAATGAGTTAGGCCCTGATGGAGATGAAGA  
 AGAAACTCCAGCGCCGGCTCCTCAACCAGAGCAACCAGCTCCTGCACCAAAACCAGAGCAACCAGCTCC  
 AGCTCCAAAACCAGAGCAACCAGCTCCTGCACCAAAACCAGAGCAACCAGCTCCAGCTCCAAAACCAGA  
 GCAACCAGCTCCAGCTCCAAAACCAGAGCAACCAGCTAAGCCGGAGAAACCAGCTGAAGAGCCTACTCA  
 ACCAGAAAAACCAGCCACTCCAAAACAGGCTGGAACAAGAAAAACGGTATGTGGTATTTCTACAATAC  
 TGATGGTTCAATGGCAATAGGTTGGCTCCAAAACAACGGTTCATGGTACTACCTAAACGCTAACGGCGC  
 TATGGCAACAGGTTGGGTGAAAGATGGAGATACCTGGTACTATCTTGAAGCATCAGGTGCTATGAAAGC  
 AAGCCAATGGTTCAAAGTATCAGATAAATGGTACTATGTCAACAGCAATGGCGCTATGGCGACAGGCTG  
 GCTCCAATACAATGGCTCATGGTACTACCTCAACGCTAATGGTGATATGGCGACAGGATGGGCTCAATA  
 CAACGGTTTCATGGTATTACCTCAACGCTAATGGTGATATGGCGACAGGATGGGCTAAAGTCAACGGTTTC  
 ATGGTACTACCTAAACGCTAACGGTGCTATGGCTACAGGTTGGGCTAAAGTCAACGGTTTCATGGTACTA

Table 1

CCTAAACGCTAACGGTTCAATGGCAACAGGTTGGGTGAAAGATGGAGATACCTGGTACTATCTTGAAGC  
ATCAGGTGCTATGAAAGCAAGCCAATGGTTCAAAGTATCAGATAAATGGTACTATGTCAATGGCTTAGG  
TGCCCTTGCACTCAACACAACCTGTAGATGGCTATAAAGTCAATGCCAATGGTGAATGGGT

**SP092 amino acid (SEQ ID NO:160)**

TSQPTFVRAEESPVVEKSSLEKKYEEAKAKADTAKKDYETAKKKAEDAQKKYEDDQKRTEEKARKEAE  
ASQKLNNDVALVVQNAYKEYREVQNORSKYKSDAEYQKKLTEVDSKIEKARKEQQDLQNKFNVRVAVVP  
EPNALAETKKKAEEAKAEKVAKRKYDYATLKVALLAKKEVEAKELEIEKLQYEISTLEQEVATAQHQVD  
NLKKLLAGADPDDGTEVIEAKLKKGEAELNAKQAEALAKKQTELEKLLDSDPEGKTQDELDEKEAEEAEL  
DKKADELQNKVADLEKEISNLEILLGGADXEDDTAALQNKLATKKAELEKTQKELDAALNELGPDGDEE  
ETPAPAPQPEQPAPAPKPEQPAPAPKPEQPAPAPKPEQPAPAPKPEQPAPAPKPEQPAKPEKPAEEPTQ  
PEKPATPKTGWKQENGWYFYNTDGSMAIGWLQNGSWYYLNANGAMATGWVKDGDWYYLEASGAMKA  
SQWFKVSDKWYYVNSNGAMATGWLQYNGSWYYLNANGDMATGWLQYNGSWYYLNANGDMATGWAKVNGS  
WYYLNANGAMATGWAKVNGSWYYLNANGSMATGWVKDGDWYYLEASGAMKASQWFKVSDKWYYVNLG  
ALAVNTTVDGYKVNANGEWV

**P093 nucleotide (SEQ ID NO:161)**

TGGACAGGTGAAAGGTCATGCTACATTTGTGAAATCCATGACAACTGAAATGTACCAAGAACAACAGAA  
CCATTCTCTCGCCTACAATCAACGCTTGGNTTCGCAAAATCGCATTGTAGATCCTTTTTTGGCGGAGGG  
ATATGAGGTCAATTACCAAGTGCTGACGACCGCTGATGCAGTCTATGGTTACTTGTCTATTCCAAGTTT  
GGAAATCATGGAGCCGGTTTATTTGGGAGCAGATTATCATCATTTAGGGATGGGCTTGGCTCATGTGGA  
TGGTACACCGCTGCCTCTGGATGGTACAGGGATTTCGCTCAGTGATTGCTGGGCACCGTGCAGAGCCAAAG  
CCATGTCCTTTTTCCGCCATTTGGATCAGCTAAAAGTTGGAGATGCTCTTTATTATGATAATGGCCAGGA  
AATTGTAGAATATCAGATGATGGACACAGAGATTATTTTACCGTCGGAATGGGAAAAATTAGAATCGGT  
TAGCTCTAAAAATATCATGACCTTGATAACCTGCGATCCGATTCCTACCTTTAATAAACGCTTATTAGT  
GAATTTTGAACGAGTCGCTGTTTATCAAAAATCAGATCCACAAACAGCTGCAGTTGCGAGGGTTGCTTT  
TACGAAAGAAGGACAATCTGTATCGCGTGTGCAACCTCTCAATGGTTG

**SP093 amino acid (SEQ ID NO:162)**

GQVKGHATFVKSMTTTEMYQEQQNHSLAYNQRLXSQNRIVDPLAEGYEVNYQVSDPDPAVYGYLSIPSL  
EIMEPVYLGADYHHLGMGLAHVDGTPLPLDGTGIRSVIAGHRAEP SHVFFRHLQDKVGDALYYDNGQE  
IVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAVYQKSDPQTAAVARVAF  
TKEGQSVSRVATSQWL

**SP094 nucleotide (SEQ ID NO:163)**

GATTGCTCCTTTGAAGGATTTGAGAGAAAACCATGTTGGAAATTGCTTCTGGTGCTCAAAATCTTCGTGC  
CAAGGAAGTTGGTGCCTATGAACTGAGAGAAGTAACCTCGCCAATTTAATGCTATGTTGGATCAGATTGA  
TCAGTTGATGGTAGCTATTTCGTAGCCAGGAAGAAACGACCCGTCAGTACCAACTTCAAGCCCTTTCGAG  
CCAGATTAATCCACATTTCTCTATAACACTTTGGACACCATCATCTGGATGGCTGAATTTTCATGATAG  
TCAGCGAGTGGTGCAGGTGACCAAGTCCTTGGCAACCTATTTCCGCTTGGCGCTCAATCAAGGCAAGGA  
CTTGATTGTCTCTCTGACGAAATCAATCATGTCCGCCAGTATCTCTTTATCCAGAAACAACGCTATGG  
AGATAAGCTGGAATACGAAATTAATGAAAATGTTGCCTTTGATAATTTAGTCTTACCCAAGCTGGTCTCT  
ACAACCCCTTGTAGAAAATGCTCTTTACCATGGCATTAAAGGAAAAGGAAGGTGAGGGCCATATTAAGT  
TTCTGTCCAGAAACAGGATTCCGGGATTGGTCATCCGATTGAGGATGATGGCGTTGGCTTCCAAGATGC  
TGGTGATAGTAGTCAAAGTCAACTCAAACGTGGGGGAGTTGGTCTTCAAAATGTCGATCAACGGCTCAA  
ACTTCATTTTGGAGCCAATTACCATATGAAGATTGATTCTAGACCCCAAAAAGGACGAAAGTTGAAAT  
ATATATAAATAGAATAGAACTAGC

**SP094 amino acid (SEQ ID NO:164)**

IAPLKDILRETMLEIASGAQNLRAKEVGAYELREVTRQFNAMLDQIDQLMVAIRSQEETTRQYQLQALSS  
QINPHFLYNTLDIT IIMAEFHDSQRVVQVTKSLATYFRLALNQKDLICLSDEINHVRQYLF IQKQRYG  
DKLEYEINENVAFDNLVLPKLVLPVENALYHGIKEKEGQGHIKLSVQKQDSGLVIRIEDDGVGFQDA  
GDSSQSOLKRGVGLQNVDRQLKLHFGANYHMKIDSRPQKGTKEIYINRIETS

**SP095 nucleotide (SEQ ID NO:165)**

TAGGTCATATGGGACTTTTTTCTACAACAAAATAGGCTCCATAATATCTATAAGGGATTACCCACTA  
CAAATATTATAGAGCCGAAAATTCACATCTAATATATGCAGACTACTTTGAAATGAAATTAATAAATTA  
ATTAAAGGATGACACAAAAGTTTTTGAAAATCTACATTCAAATTTGTAGAAGGATATAAATATACCT

Table 1

GACAGAATCTAAAGAATCTGGAATTAAACAAATGGACAATGTCATAAAATATTTTGAGTTTATTGAATC  
TAAAAGTATTGCTTTATATTTTCAAAAACGATTAAATGAGCTGATAGAT

**SP095 amino acid (SEQ ID NO:166)**

RSYGTFFLQQNRLHNIYKGFTHYKYRAENSHLIYADYFEMKLKLLKDDTKVFEKSTFKFVEGYKIYL  
TESKESGIKQMDNVIKYFEFIESKSIALYFQKRLNELID

**SP096 nucleotide (SEQ ID NO:167)**

CAACGTTGAGAATTATTTGCGAATGTGTTTGGATAGCATTTCAGAATCAGACGTATCAAAATTTTGAGTG  
TTTATTAATCAATGATGGCTCTCCAGATCATTTCATCCAAAATATGTGAAGAATTTGTAGAGAAAGATTCTC  
GTTTTCAAATATTTTGAGAAAGCAAACGGCGGTCTTTTCATCAGCTCGTAACCTAGGTATTGAATGTTT  
GGGGGGGGCGGTACATTACTTTTGTAGACTC

**SP096 amino acid (SEQ ID NO:168)**

NVENYLRMC LDSIQNTYQNFEC LLINDGSPDHSSKICEEFVEKDSRFKYFEKANGGLSSARNLGIECS  
GGGVHYFCRL

**SP097 nucleotide (SEQ ID NO:169)**

CTACTATCAATCAAGTTCTTCAGCCATTGAGGGCCACCATTGAGGGCAACAGCCAAACGACCATCAGCCA  
GACTAGCCACTTTTATTCAGTCTTATATCAAAAACTAGAAACCACCTCGACTGGTTTGACCCAGCAGAC  
GGATGTTCTGGCCTATGCTGAGAATCCCAAGTCAAGACAAGGTGAGGGAATCCGAGATTTGTTTTGAC  
CATCTTGAAGTCAGATAAGGACTTGAAAACTGTTGTGCTGGTGACCAAATCTGGTCAGGTCAATTTCTAC  
AGATGACAGTGTGCAGATGAAAACTTCCTCTGATATGATGGCTGAGGATTGGTACCAAAGGCCATTCA  
TCAGGGAGCTATGCCTGTTTTGACTCCAGCTCGTAAATCAGATAGTCAGTGGGTCAATTTCTGTCACTCA  
AGAACTTGTGATGCAAAGGGAGCCAATCTTGGTGTGCTTCGTTTGGATATTTCTTATGAACTCTGGA  
AGCCTATCTCAATCAACTCCAGTTGGGGCAGCAGGGCTTTGCCTTCATTATCAATGAAAACCATGAATT  
TGTCTACCATCCTCAACACACAGTTTATAGTTTCGTCTAGCAAAATGGAGGCTATGAAACCCTACATCGA  
TACAGGTCAAGGTTATACTCCTGGTCACAAATCCTACGTCAGTCAAGAGAAGATTGCAGGAAGTATTG  
GACGGTGCTTGGCGTGTATCATTTGAAAAAGTTAGACCAGGTTTCGGAGTCAG

**SP097 amino acid (SEQ ID NO:170)**

YYQSSSSAIEATIEGNSQTTISQTSHFQSYIKKLETTSTGLTQQTDVLAYAENPSQDKVEGIRD LFLT  
ILKSDKDLKTVVLVTKSGQVISTDDSVQMKTSDDMAEDWYQKAIHQGAMPVLT PARKSDSQWVISVTQ  
ELVDAKGANLGVRLDISYETLEAYLNQLQLGQQGF AFIINENHEFVYHPQHTVYSSSSKMEAMPYID  
TGQGYTPGHKSYVSQEKIAGTDWTVLGVSSLEKLDQVRSQ

**SP098 nucleotide (SEQ ID NO:171)**

GACAAAACATTAAACGTCCTGAGGTTTATCACCTGCAGGGACTTTAGAGAAGCTAAAGGTAGCTGT  
TCAGTATGGAGCAGATGCTGTCTTTATCGGTGGTCAGGCCTATGGTCTTCGTAGCCGTGCGGGAACTT  
TACTTTTCAACAGATGGAAGAAGGCGTGCAGTTTTCGGCCAAGTATGGTGCCAAGGTCTATGTAGCGGC  
TAATATGGTTATGCACGAAGGAAATGAAGCTGGTGCTGGTGAGTGGTTCCGTAAACTGCGTGATATCGG  
GATTGCAGCAGTTATCGTATCTGACCCAGCCTTGATTATGATTGCAGTGACTGAAGCACCAGGCCTTGA  
AATCCACCTTTCTACCCAAGCCAGTGCCACTAATGAAACCCTTGAGTTCTGGAAAGAGCTAGGCTT  
GACTCGTGTGCTTTTAGCGCGTGAGGTTTCAATGGAAGAATTAGCTGAGATCCGCAAACGTACAGATGT  
TGAAATTGAAGCCTTTGTCCATGGAGCTATGTGTATTTTCACTCTGGACGTTGTACTCTTTCAAACCA  
CATGAGTATGCGTGATGCCAACCCTGGTGGATGTTCTCAGTCATGCCGTTGGAAATACGACCTTTACGA  
TATGCCATTTGGGAAAGAACGTAAGAGTTTGCAGGGTGAGATTCCAGAAGAATTTTCAATGTCAGCCGT  
TGACATGTCTATGATTGACCANATTCAGATATGATTGAAAATGGTGTGGACAGTCTAAAAATCGAAGG  
ACGTATGNAGTCTATTCAC TANGTATCAACAGTAACCAACTGCTACAAGGCGGCTGTGGATGCCTATCT  
TGAAAGTCTTGAAGTTTGAAGCTATCAAACAAGACTTGGTGGACGAGATGTGGAAGGTTGCCCAACG  
TGAAGTGGCTACAGGATTTTACTATGGTACACCATCTGAAAATGAGCAGTTGTTTGGTGTCTCGTGTAA  
AATCCCTGAGTACAAGTTTGTGCTGAAGTGGTTTCTTATGATGATGCGGCACAAACAGCAACTATTTCG  
TCAACGAAACGTCATTAACGAAGGGGACCAAGTTGAGTTTATGGTCCAGGTTTCCGTCATTTTGAAC  
CTATATTGAAGATTTGCATGATGCTAAAGGCAATAAAATCGACCGCGCTCCAAATCCAATGGAAGTATT  
GACTATTAAAGTCCCAACCTGTTCAATCAGGAGACATGGTTTCGAGCTCTTAAAGAGGGGCTTATCAA  
TCTTTATAAGGAAGATGGAACCAGCGTCACAGTTCGTGCT

WO 98/18930

PCT/US97/19422

Table 1

84

**SP098 amino acid (SEQ ID NO:172)**

TKTLKRPEVLSPAGTLEKLKVAVQYGADAVFIGGQAYGLRSRAGNFTFEQMEEGVQFAAKYGAKVYVAA  
 NMVMHEGNEAGAGEWFRKLRDIGIAAVIVSDPALIMIAVTEAPGLEIHLSTQASATNYETLEFWKELGL  
 TRVVLAREVSMEEAEIRKRTDVEIEAFVHGAMCISYSGRCTLSNHMSMRDANRGGCSQSCRWKYDLYD  
 MPFGKERKSLQGEIPEEFMSAVDMSMIDXIPDMIENGVDLSKIEGRMXSIHXVSTVTNICYKAAVDAYL  
 ESPEKFEAIKQDLVDEMWKVAQRELATGFYYGTPSENEQLFGARRKIPEYKFVAEVVSYDDAAQTATIR  
 QRNVINEGDQVEFYGPGRHFETYIEDLHDAKGNKIDRAPNPMELLTIKVPQPVQSGDMVRALKEGLIN  
 LYKEDGTSVTVRA

**SP099 nucleotide (SEQ ID NO:173)**

TTCTCAGGAGACCTTTAAAAATATCACCAATAGCTTCTCCATGCAAATCAATCGTCGCGTCAACCAAGG  
 AACGCCCTCGTGGTGTCTGGGAATATCAAGGGTGAAGACATCAAAAAATCACCGAAAACAAGGCCATTGA  
 GTCTTATGTCAAACGTATCAACGCTATCGGAGATTTGACTGGATATGACCTGATTGAAACGCCAGAAAC  
 CAAGAAGAATCTCACTGCTGATCGTGCCAAGCGTTTTGGAAGTAGCTTGATGATTACAGGTGTCAATGA  
 CTCCTCTAAAGAAGACAAGTTTGTCTCTGGTTCTTATAAACTAGTCGAAGGAGAGCACTTAACCAACGA  
 CGACAAGGATAAAATCCTCTTGCACAAGGACTTGGCAGCCAAACACGGCTGGAAAGTAGGGGACAAGGT  
 TAACTGGACTCTAATATCTACGATGCAGATAATGAAAAAGGAGCCAAGGAAACAGTTGAAGTGACAAT  
 CAAGGGACTCTTTGATGGTCATAATAAGTCAGCAGTAACCTACTCACAAGAACTTTACGAAAACACAGC  
 TATTACAGACATTCACACTGCTGCAAACTTTATGGATACACAGAAGACACAGCCATTTATGGGGACGC  
 AACCTTCTTTGTAACAGCAGACAAGAACTTGGATGATGTTATGAAAGAGTTGAATGGCATCAGTGGTAT  
 CAACTGGAAGAGCTACACACTCGTCAAGAGCTCCTCTAACTACCCAGCTCTTGAGCAATCTATCTCTCG  
 TATGTACAAGATGGCCAAC

**SP099 amino acid (SEQ ID NO:174)**

SQETFKNITNSFSMQINRRVNQGTPRGAGNIKGEDIKKITENKAIESYVKRINAIGDLTGYPDLIETPET  
 KKNLTADRAKFRGSSLMITGVNDSSKEDKFVSGSYKLVEGEHLTNDKDKILLHKDLAAKHGKVGDKV  
 KLDNSNIYDADNEKGAKETVEVTIKGLFDGHNKSAVTYSQELYENTAITDIHTAAKLYGYTEDTAIYGDA  
 TFFVTADKNLDDVMKELNGISGINWKSITLVKSSSNYPALQSIISGMYKMAN

**SP100 nucleotide (SEQ ID NO:175)**

AGTAAATGCGCAATCAAATTCATTAATATTAATAGATGAACCTGAAATCTCACTTCATCCGAGTGCAAT  
 CTATAAAATTTAAAGAGTTTTTACTTCAAGAGTGTTTAAATAAAAAACATCAAATTATTATCACTACACA  
 TTCTACACAACTTATAAAAGATTTTCTAGAGAAGCCGTGAACTTTTAGTGAAAAACGGAGAAAAGGT  
 AGATGTTATTGAAAAATATTGATTATCAGGATGCATTTTTTGAATTAGGTGATGTGTATCATTCTAGGAA  
 GATGATTTATGTTGAAGATAGACTAGCTAAATATATTCTAGAGTTTGTATCACTCATTCAAGGTAGTGA  
 GAATCTTAAACAGAAATTTAGTAGTGAGATATATTCTGGTGGAGCAAATCAAATAATTTGTAATAATAT  
 TTTAACTCATCGTATTTAGATTCCGATAACCATTATTTTTGGCTTGATGGAGATCAAAACACTAATGT  
 TAGTGAATCAAATAATTTAATGAACATCTTGAAAATGGTGTGTTATATCAGATAAAATTCCTGAATC  
 AGATAATAAAAAATCTTGATGATATTATAAAATTGATAANGGGATGTCCAATTAAATTTAATGTTTCAGG  
 TAATAAAGGGCAAAAAATAATATTGAATTAATTGCGAAACAAAGAAGCTTTATAGATTATTGGGCTAA  
 ATAC

**SP100 amino acid (SEQ ID NO:176)**

VNAQSNLSLILIDEPEISLHPSAIYKFKEFLLQECLNKKHQIIITTHSTQLIKDFPREAVKLLVKNGEKV  
 DVIENIDYQDAFFELGDVYHSRKMIYVEDRLAKYILEFVITHSGSENKQNLVVRYIPGGANQIIICNNI  
 LNSSYLDSDNHFWLDGDQNTNVSESNNLMNYLENGVVISDKIPESDNKNLDDIIKLIXGCPKIFNVSG  
 NKGQKNNIELIAKQRSFIDYWAKY

**SP101 nucleotide (SEQ ID NO:177)**

TTACCGCGTTCATCAAGATGTCAAACAAGTCATGACCTATCAACCCATGGTGCGAGAAATATTGAGTGA  
 ACAAGACACCCAGCAAACGAAGAGCTTGTGCTTGCTATGATTTATACTGAAACAAAAGGAAAAGAAGG  
 CGATGTTATGCAGTCTAGTGAGTCTGCAAGTGGTCCACCAACACCATCAATGATAATGCCTCTAGCAT  
 TCGGCAAGGCATTCAAACCTCTGACAGGCAATCTCTATCTGGCGCAGAAGAAGGGGGTAGATATCTGGAC  
 AGCTGTTCAAGCCTATAATTTTGGACCTGCCTATATCGATTTTATCGCCCAAAATGGCAAGGAAAATAC  
 CCTGGCTCTAGCCAAACAGTACTCTCGTGAGACTGTTGCCCCCTTGCTTGGTAATAGGACTGGAAAGAC  
 TTATAGTTATATTACCCCATTTCCATTTTTCACGGTGCTGAACCTATGTAAATGGAGGAAACTATTA  
 TTATTCTAGACAGGTACGACTTAACCTTTACATCATCAAATGTTTCACTCTCTTTCAACATCTGGC

Table 1

**SP101 amino acid (SEQ ID NO:178)**

YRVHQDVKQVMYQPMVREILSEQDTPANEELVLAMIYTETKGKEGDMQSSSESASGSTNTINDNASSI  
RQGIQTLTGNLYLAQKKGVDIWTAQAYNFGPAYIDFIAQNGKENTLALAKQYSRETVAPLLGNRTGKT  
YSYIHPIISIFHGAELYVNGGNYYSRQVRLNLYIIKCFTLFSTSG

**SP102 nucleotide (SEQ ID NO:179)**

GTGGATGGGCTTTAACTATCTTCGTATTGCGCGTGCGGCTAAAATTGTGGACAATGAGGAGTTTGAAGC  
CTTGATTTCGTACGGGTCAATTGATTGATTGCGCGACCCAGCAGAATTCCACAGAAAACATATCCTTGG  
TGCACGCAATATTCCTTCAAGTCAGTTGAAAAC TAGTCTTG CAGCCCTTCGTAAAGATAAACCTGTCCT  
TCTCTACGAAAACCAACGTGCGCAACGAGTTACAAATGCAGCTCTTTACTTGAAAAACAAGGTTTTTC  
TGAGATTTATATCCTTTCTTATGGCTTGGATTCTTGGAAGGGAAAGTGAAGACTAGC

**SP102 amino acid (SEQ ID NO:180)**

WMGFNYLRIRRAAKIVDNEEFELIRTGQLIDLDPAEFHRKHILGARNIPSSQLKTSLAALRKDKPVL  
LYENQRAQRVTNAALYLKKQGFSEIYILSYGLDSWKGVKTS

**SP103 nucleotide (SEQ ID NO:181)**

ACTAAACCAGCATCGTTTCGCAGGAAAATAAGGACAATAATCGTGTCTCTTATGTGGATGGCAGCCAGTC  
AAGTCAGAAAAGTGAAAAC TTGACACCAGACCAGGTTAGCCAGAAAAGAAGGAATTCAGGCTGAGCAAAAT  
TGTAATCAAAATTACAGATCAGGGCTATGTAACGTCACACGGTGACCACTATCATTACTATAATGGGAA  
AGTTCCTTATGATGCCCTCTTTAGTGAAGAACTCTTGATGAAGGATCCAAACTATCAACTTAAAGACGC  
TGATATTGTCAATGAAGTCAAGGGTGGTTATATCATCAAGGTCGATGGAAAATATTATGTCTACCTGAA  
AGATGCAGCTCATGCTGATAATGTTTCGAACTAAAGATGAAATCAATCGTCAAAAACAAGAACATGTCAA  
AGATAATGAGAAGGTTAACTCTAATGTTGCTGTAGCAAGGTCTCAGGGACGATATACGACAAATGATGG  
TTATGTCTTTAATCCAGCTGATATATCGAAGATACGGGTAATGCTTATATCGTTCCTCATGGAGGTCA  
CTATCACTACATTCCCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCTGGCTGG  
AAAAATATGCAACCGAGTCAGTTAAGCTATTCTTCAACAGCTAGTGACAATAACACGCAATCTGTAGC  
AAAAGGATCAACTAGCAAGCCAGCAAATAAATCTGAAAATCTCCAGAGTCTTTTGAAGGAACTCTATGA  
TTCACCTAGCGCCCAACGTTACAGTGAATCAGATGGCCTGGTCTTTGACCCTGCTAAGATTATCAGTCG  
TACACCAAATGGAGTTGCGATTCCGCAATGGCGACCATTACCACTTTATTCTTACAGCAAGCTTTCTGC  
CTTAGAAGAAAAGATTGCCAGAAATGGTGCCTATCAGTGGAAGTGGTCTACAGTTTCTACAAATGCAAA  
ACCTAATGAAGTAGTGTCTAGTCTAGGCAGTCTTTCAAGCAATCCTTCTTTTAAACGACAAGTAAGGA  
GCTCTCTTCAGCATCTGATGGTTATATTTTAAATCCAAAAGATATCGTTGAAGAAACGGCTACAGCTTA  
TATTGTAAGACATGGTGATCATTTCATTACATTCCAAAATCAAATCAAATGGGCAACCGACTCTTCC  
AAACAATAGTCTAGCAACACCTTCTCCATCTCTTCCAATCAATCCAGGAACCTCACATGAGAAACATGA  
AGAAGATGGATACGGATTGATGCTAATCGTATTATCGCTGAAGATGAATCAGGTTTTGTGATGAGTCA  
CGGAGACCACAATCATTATTTCTTCAAGAAG

**SP103 amino acid (SEQ ID NO:182)**

LNQHRSEQENKDNRRVSYVDGSQSSQKSENLPDQVSQKEGIQAEQIVIKITDQGYVTSBGDHYHYNGK  
VPYDALFSEELLMKDPNYQLKDADIVNEVKGGYLIKVDGKYVYVKDAAHADNVRTKDEINRQKQEHVK  
DNEKVNNSNAVARSQGRYTTNDGYVFNPAIIEDTGNAYIVPHGGHYHYIPKSDLSASELAAAKAHLA  
KNMQPSQLSYSSSTASDNNTQSVAKGSTSKPANKSENLSLLKELYDPSAQRYSES DGLVDFPAKII SR  
TPNGVAIPHGDHYHFIPYSKLSALEEKIARMVPISGTGSTVSTNAKPNEVVSSLSLSSNPSSLTTSKE  
LSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPNNSLATPSPSLPINPGTSHEKHE  
EDGYGFDANRIIAEDES GFVMSHGDHNYFFKK

**SP105 nucleotide (SEQ ID NO:183)**

TGACTACCTTGAAATCCCACCTTTACAGCTATCTTGGTGGATTCAACACTAAAGTTCTTCCAACCTCAAT  
GATGAACATCATCAACGGTGGTTCTCACTCTGACGCTCCAATCGCTTTCCAAGAGTTTCATGATCTTGCC  
AGTTGGTGCGCCAACATTTAAAGAAGCCCTTCGTTACGGTGCTGAAATCTTCCACGCTCTTAAGAAAAT  
CCTTAAATCACGTGGTTTGGAACTGCCGTAGGTGACGAAGGTGGATTGCTCCTGCTTTTGAAGGAAC  
TGAAGATGGTGTGAACTATCCTTGCTGCGATTGAAGCTGCTGGATATGTACCAGTTAAAGACGTATT  
TATCGGATTTGACTGTGCTTCATCAGAATTCACGATAAAGAACGTAAAGTTTACGACTACACTAAATT  
TGAAGGTGAAGGTGCTGCTTCGTACATCTGCAGAACAAATCGACTACCTTGAAGAATTGGTTAACAA  
ATACCCAATCATCACTATTGAAGATGGTATGGATGAAACGACTGGGATGGTTGGAAAGCTCTTACTGA  
ACGCTCTTGGTAAGAAAGTACAACCTGTTGGTGACGACTTCTTCGTAACAAACACTGACTACCTTGCACG

WO 98/18930

PCT/US97/19422

Table 1

86

TGGTATCCAAGAAGGTGCTGCTAACTCAATCCTTATCAAAGTTAACCAAATCGGTACTCTTACTGAAAC  
TTTTGAAGCTATCGAAATGGCTAAAGAAGCTGGTTACACTGCTGTTGTATCACACCGTTTACGGTGAAAC  
TGAAGATTCAACAATCGCTGATATTGCAGTTGCAACTAACGCAGGACAAATCAAGACTGGTTCCTTTT  
ACGTACAGACCGCATCGCTAAATACAACCAATTGCTTCGTATCGAAGACCAACTTGGTGAAGTAGCTGA  
ATATCGTGGATTGAAATCATTCTACAACCTTAAAAAA

**SP105 amino acid (SEQ ID NO:184)**

DYLEIPLYSYLGGFNTKVLPTPMNIINGGSHSDAPIAFQEFMILPVGAPTFKEALRYGAEIFHALKKI  
LKSRGLETAVGDEGGFAPRFEGTEDGVETILAAIEAAGYVPGKDVFIGFDCASSEFYDKERKVYDYTKF  
EGEGAAVRTSAEQIDYLEELVNKYPIITIEDGMDENDWDGWKALTERLGKKVQLVGDDFFVTNTDYLAR  
GIQEGAANSILIKVNQIGTLTETFEAIEMAKEAGYTAVVSHRSGETEDSTIADIAVATNAGQIKTGSLS  
RTDRIAKYNQLLRIEDQLGEVAEYRGLKSFYNLKK

**SP106 nucleotide (SEQ ID NO:185)**

TCGTATCTTTTTTGGAGCAATGTTGCGGTAGAAGGACATTCCATGGATCCGACCCTAGCGGATGGCGA  
AATTCCTCTCGTTGTAAACACCTTCTTATTGACCGTTTTTGATATCGTGGTGGCCCATGAGGAAGATGG  
CAATAAGGACATCGTCAAGCGCGTGATTGGAATGCCTGGCGACACCATTTCGTTACGAAAATGATAAACT  
CTACATCAATGACAAAGAAACGGACGAGCCTTATCTAGCAGACTATATCAAACGCTTCAAGGATGACAA  
ACTCCAAAGCACTTACTCAGGCAAGGGCTTTGAAGGAAATAAAGGAACCTTCTTTAGAAAGTATCGCTCA  
AAAAGCTCAAGCCTTACAGTTGATGTCAACTACAACACCAACTTTAGCTTTACTGTTCCAGAAGGAGA  
ATACCTTCTCCTCGGAGATGACCGCTTGGTTTCGAGCGACAGCCGCCACGTAGGTACCTTCAAAGCAAA  
AGATATCACAGGGGAAGCTAAATTCGCTTATGGCCAATCACCCGTATCGGAACATTT

**SP106 amino acid (SEQ ID NO:186)**

RIFFWSNVRVEGHSMDPTLADGEILFVVKHLPIDRFDIVVAHEEDGNKDIVKRVIGMPGDTIRYENDKL  
YINDKETDEPYLADYIKRFKDDKLQSTYSGKGFEGNKGTFFRSIAQKAQAFITVDVNYNTNFSFTVPEGE  
YLLLGDDRLVSSDSRHVGTFFKAKDITGEAKFRLWPITRIGTF

**SP107 nucleotide (SEQ ID NO:187)**

GGACTCTCTCAAAGATGTGAAAGCAAATGCTAGCGACAGCAAGCCTGCACAGGACAAGAAGGATGCAAA  
ACAAGGAACGGAAGATAGTAAGGATTCAGATAAGATGACTGAAACAAACTCAGTTCCGGCAGGAGTGAT  
TGTGGTCAGTCTACTTGCCCTCCTAGGCGTGATTGCCTTCTGGCTGATTCCGCGTAAGAAAGAGTCAGA  
AATCCAGCAATTAAAGCACGGAATTGATCAAGGTTCTAGGACAGCTAGATGCAGAAAAAGCGGATAAAAA  
AGTCCTTGCCAAAGCCCAAAACCTTCTCCAAGAAACCTTGATTTCGTGAAGAAGAAAATGGCTCAGC  
AGAGACAGAACTAACTAGTAGAGGAGCTTAAAGCAATCCTTGACAACTCAAG

**SP107 amino acid (SEQ ID NO:188)**

DSLKDVKANASDSKPAQDKKDAKQGTEDSKSDKMTETNSVPAGVIVVSLALLGVIAFWLIRRKKESE  
IQQLSTELIKVLGQLDAEKADKKVLAKAQNLLQETLDFVKEENGSAETETKLVEELKAILDKLK

**SP108 nucleotide (SEQ ID NO:189)**

CAAGAAATCCTATCATCTCTTCCAGAAGCAAACAGAGACGAGGGGAATTCAGACTCAGTTGATTGAAGA  
ATCGCTTAGTCAGCAGACTATAATCCAGTCCTTCAATGCTCAAACAGAATTTATCCAAAGATTGCGTGA  
GGCTCATGACAACTACTCAGGCTATTCTCAGTCAGCCATCTTTTATTCTTCAACGGTCAATCCTTCGAC  
TCGCTTTGTAAATGCACTCATTTATGCCCTTTTAGCTGGAGTAGGAGCTTATCGTATCATGATGGGTTT  
AGCCTTGACCGTCGGTCGTTTAGTGACTTTTTTGAAGTATGTTTCAAGCAATACACCAAGCCCTTTAACGA  
TATTTCTTCAAGTGTAGCTGAGTTGCAAGTGTCTGGCTTGCGTAGAGCGTATCTATGGAGTCTTAGA  
TAGCCCTGAAGTGGCTGAAACAGGTAAGGAAGTCTTGACGACAGTGACCAAGTTAAGGGAGCTATTTT  
CTTTAAACATGTCTCTTTTGGCTACCATCCTGAAAAAATTTTGATTAAAGGACTTGTCTATCGATATTT  
AGCTGGTAGTAAGGTAGCCATCGTTGGTCCGACAGGTGCTGGAAAATCAACTCTTATCAATCTCCTTAT  
GCGTTTTTATCCCATTAGCTCGGGAGATATCTTGCTGGATGGGCAATCCATTTATGATTATACACGAGT  
ATCATTTGAGACAGCAGTTTGGTATGGTGCTTCAAGAAACCTGGCTCACACAAGGGACCATTCATGATAA  
TATTGCCTTTGGCAATCCTGAAGCCAGTCGAGAGCAAGTAATTGCTGCTGCCAAAGCAGCTAATGACAGA  
CTTTTTCATCCAACAGTTGCCACAGGGATACGATACCAAGTTGGAAAATGCTGGAGAATCTCTCTGT  
CGGCAAGCTCAGCTCTTGACCATAGCCCGAGTCTTTCTGGCTATTCCAAAGATTCTTATCTTAGACGA  
GGCAACTTCTTCCATTGATACACGGACAGAAGTGCTGGTACAGGATGCCCTTTGCAAACTCATGAAGGG  
CCGCACAAGTTTCATCATTGCTCACCGTTTGTCAACCATTACAGGATGCGGATTTAATTCTTGTCTTAGT

Table 1

AGATGGTGATATTGTTGAATATGGTAACCATCAAGAACTCATGGATAGAAAGGGTAAGTATTACCAAAT  
GCAAAAAGCTGCGGCTTTTAGTTCTGA

A

**SP108 amino acid (SEQ ID NO:190)**

KKSYHLFQKQTETRGITQQLIEESLSQQTIIQSFNAQTEFIQRLREAHDNYSQYSQSAIFYSSVTNPST  
RFVNALIYALLAGVGAYRIMMGSAITVGRVLVTLNLYVQQYTKPFNDISSVLAELQSALACVERIYGVLD  
SPEVAETGKEVLTTSDQVKGAISFKHVSFGYHPEKILIKDLSIDIPAGSKVAIVGPTGAGKSTLINLLM  
RFYPISSGDILLDQSIYDYTRVSLRQQFGMVLQETWLTQGTIHDNIAFGNPEASREQVIAAKAANAD  
FFIQQLPQGYDTKLENAGESLSVGQAQLLTARVFLAIPKILILDEATSSIDTRTEVLVQDAFAKLMKG  
RTSFIIAHLSTIQDADLILVLVDGDIVYGNHQELMDRKGKYYQMOKAAAFSSE

**SP109 nucleotide (SEQ ID NO:191)**

ACGAAATGCAGGGCAGACAGATGCCTCGCAAATTGAAAAGGCGGCAGTTAGCCAAGGAGGAAAAGCAGT  
GAAAAAACAGAAATTAGTAAAGACGCAGACTTGCACGAAATTTATCTAGCTGGAGGTTGTTTCTGGGG  
AGTGGAGGAATATTTCTCACGTGTTCCCGGGGTGACGGATGCCGTTTCAGGCTATGCAAATGGTAGAGG  
AGAAACAACCAAGTACGAATTGATTAACCAAACAGGTCATGCAGAAACCGTCCATGTCACCTATGATGC  
CAAGCAAATTTCTCTCAAGGAAATCCTGCTTCACTATTTCCGCATTATCAATCCAACCAGCAAAAATAA  
ACAAGGAAATGATGTGGGGACCCAGTACCGTACTGGTGTATTATACACAGATGACAAGGATTTGGAAGT  
GATTAACCAAGTCTTTGATGAGGTGGCTAAGAAATACGATCAACCTCTAGCAGTTGAAAAGGAAAACCTT  
GAAGAATTTTGTGGTGGCTGAGGATTACCATCAAGACTATCTCAAGAAAAATCCAAATGGCTACTGCCA  
TATCAATGTTAATCAGGCGGCCTATCCTGTCTATGATGCCAGCAAATATCCAAAACCAAGTATGAGGA  
ATTGAAAAGACCCCTGTACCTGAGGAGTATGCAGTTACCCAGGAAAATCAAACAGAACGAGCTTTCTC  
AAACCGTTACTGGGATAAATTTGAATCCGGTATCTATGTGGATATAGCAACTGGGGAACCTCTCTTTTC  
ATCAAAAGACAAATTTGAGTCTGGTGTGGCTGGCCTAGTTTACCCAACCCATCAGTCCAGATGTTGT  
CACCTACAAGGAAGATAAGTCTACAATATGACCGTATGGAAGTGCGGAGCCGAGTAGGAGATTCTCA  
CCTTGGGCATGTCTTTACGGATGGTCCACAGGACAAGGGCGGCTTACGTTACTGTATCAATAGCCTCTC  
TATCCGCTTTATTCCCAAAGACCAAATGGAAGAAAAAGGCTACGCTTATTTACTAGATTATGTTGAT

**SP109 amino acid (SEQ ID NO:192)**

RNAGQTDASQIEKAASVQGGKAVKKTEISKDADLHEIYLAGGCFWGVVEEYFSRVPGVTDVSGYANGRG  
ETTKYELINQTHAETVHVYDAKQISLKEILLHYFRIINPTSKNKQGNVGTQYRTGVYYTDDKDLEV  
INQVFDEVAKKYDQPLAVEKENLKNFVVAEDYHQDYLLKNPNGYCHINVNQAAYPVIDASKYPKPSDEE  
LKKTLSPPEYAVTQENQTERAFSNRYWDFESGIYVDIATGEPLFSSKDKFESGCGWPSFTQPI SPDVV  
TYKEDKSYNMTRMEVRSRVGDSHLGHVFTDGPQDKGLRYCINSLSIRFIPKQDMEEKGYAYLLDYVD

**SP110 nucleotide (SEQ ID NO:193)**

TGTATAGTTTTTTAGCGCTTGTCTCTTAATCTGNTAAAAATGAAGAAAATACTTCTAAAGAGCATGCG  
CCTGATAAAATAGTTTTAGATCATGCTTTCCGTCAAACCTATATTAGATAAAAAACCTGAAAGAGTTGCA  
ACTATTGCTTGGGGAAATCATGATGTAGCATTAGCTTTAGGAATAGTTCTTGTGGATTTTCAAAAGCA  
AATTACGGTGTAAGTGCTGATAAAGGAGTTTTACCATGGACAGAAGAAAAAATCAAAGAACTAAATGGT  
AAAGCTAACCTATTTGACGATTTGGATGGACTTAACCTTTGAAGCAATATCAAATTTCTAAACCAGATGTT  
ATCTTAGCAGGTTATTCTGGTATAACTAAAGAAGATTATGACACTCTATCA

**SP110 amino acid (SEQ ID NO:194)**

CIVFSACSSNSXKNEENTSKEHAPDKIVLDHAFGQTILDKKPERVATIAWGNHDVALALGIVPVGFPSKA  
NYGVSADKGVLPWTEEKIKELNGKANLFDLDGLNFEAISNSKPDVILAGYSGITKEDYDTLS

**SP111 nucleotide (SEQ ID NO:195)**

GTGTGTCGAGCATATTCTGAAGCAAACCTATCAAAATATAGAAATTATTTTAGTTGATGACGGTTCTAC  
GGATAATTCTGGGGAAATTTGTGATGCTTTTATGATGCAAGATAATCGTGTGCGAGTATTGCATCAAGA  
AAATAAGGGGGGGGCGAGCACAAAGCTAAAAATATGGGGATTAGTGTAGCTAAGGGAGAGTACATCACGAT  
TGTTGATTCAGATGATATCCTAAAAGAAAATATGATTGAAACTCTTTATCAGCAAGTCCAAGAAAAGGA  
TGCAGATGTTGTTATAGGGAATTACTATAATTATGACGAAAGTGACGGGAATTTTATTTTATGTAAC  
AGGGCAAGATTTTTCGCTCGAAGAATTAGCTATACAAGAAATTATGAACCGTCAAGCAGGAGATTGGAA  
ATTCAATAGCTCGGCCTTTATATTGCGGACATTTAAGTTGATTAAAAAAGAATTATTCAATGAAGTTCA  
CTTTTCAAATGGTCGCCGCTTTGATGATGAAGCAACTATGCATCGCTTTTATCTTTTAGCCTCTAAAAT  
CGTCTTTATAAACGATAATCTCTATCTGTATAGAAGACGTTTACGGAAGCATCATGAGAACGGGAATTTGA

Table 1

TCTTTCCTGGGCAAGAGATATTGTTGAAGTGTTTTCTAAGAAAATATCGGATTGTGTCTTGGCTGGTTTT  
GGATGTCTCCGTTCTGCGTATTCGATTTGTCAATCTTTTAAAAAGATTATAAGCAAACCTTTAGAATACCA  
TCAATTAACAGATACTGAGGAATATAAAGATATTTGTTTCAGATTAAAGTTGTTTTTTGATGCAGAACA  
AAGAAATGGTAAAAGT

**SP111 amino acid (SEQ ID NO:196)**

CVEHILKQTYQNIIEILVDDGSTDNSGEICDAFMMQDNVRVRLHQENKGGAAQAKNMGISVAKGEYITI  
VDSDDIVKENMIETLYQQVQEKDADVIGNYYNDESDGNFYFYVTGQDFCVEELAIQEIMNRQAGDWK  
FNSSAFILPTFKLIKKELFNEVHFSNGRRFDDEATMHRFYLLASKIVFINDNLYLYRRRSGSIMRTEFD  
LSWARDIVEVFSKKISDCVLGLDVSVLRIKRVNLLKDYKQTFLEYHQLTDTEEYKDICFRLKLFFDAEQ  
RNGKS

**SP0112 nucleotide (SEQ ID NO:197)**

GTGTTTGGATAGCATTGAGAATCAGACGTATCAAAATTTTGAGTGTTTATTAATCAATGATGGCTCTCC  
AGATCATTTCATCCAAAATATGTGAAGAATTTGTAGAGAAAGATTCTCGTTTCAAATATTTTGAGAAAGC  
AAACGGCGGTCTTTTCATCAGCTCGTAACCTAGGTATTGAATGTTTCGGGGGGGGCGTACATTACTTTTGT  
AGACTCTGATGATTGGTTGGAACATGATGCTTTAGACCGATTATATGGTGCTTTGAAAAAGGAAAACGC  
AGATATTAGTATCGGGCGTTATAATCTTATGATGAAAACACGCTATGTGTATATGACTTATGTTACGGA  
TCCAGATGATTCTCTAGAAGTGATAGAAGGTAAAGCAATTATGGATAGGGAAGGTGTGGAAGAAGTCAG  
AAATGGGAAGTGGACTGTAGCTGTCTTGAAGTTATTCAAGAGAGAGTTACTACAAGATTTACCATTTCC  
TATAGGAAAAATTGCAGAGGATACTTACTGGACATGGAAGGTACTTCTAAGAGCTTCGAGGATAGTCTA  
TTTGAATCGTTGTGTTTACTGGTACCGTGTGTTGTTTATCTGATACTTTATCGAATACATGGAGTAAAA  
GCGTATGTATGATGAAATTGGGGCTAGGGAAGAAAAGATAGCTATTTTAGCAAGTTCAGACTATGACTT  
GACCAATCATATTTTGATTATAAAAATAGATTACAAAGAGTGATAGCAAAATTAGAAGAACAAAATAT  
GCAGTTCACAGAGATTACAGAAGAATGATGGAAGAAATGTCTTTACTTCCG

**SP0112 amino acid (SEQ ID NO:198)**

CLDSIQNQTYQNFECILLINDGSPDHSSKICEEFVEKDSRFKYFEKANGGLSSARNLGI ECSGGAYITFV  
DSDDWLEHDALDRLYGALKKENADISIGRYNSYDETRYVYMTYVTDPDSSLEVIEGKAIMDREGVEEVR  
NGNWTAVLKLKFKRELLQDLFPPIGKIAEDTYWTKVLLRASRIVLYLNRVYWRVGLSDTLSNTWSEK  
RMYDEIGAREEKIAILASSDYDLTNHILYKNRLQRVIAKLEEQNMQFTEIYRRMMEKLSLLP

**SP113 nucleotide (SEQ ID NO:199)**

GTGCCTAGATAGTATTATTACTCAAACATATAAAAAATATTGAGATTGTTGTCGTTAATGATGGTTCTAC  
GGATGCTTCAGGTGAAATTTGTAAAGAATTTTCAGAAATGGATCACCGAATTCTCTATATAGAACAAGA  
AAATGCTGGTCTTTCTGCCGCACGAAACACCGGTCTGAATAATATGTCCGGAATTTATGTGACCTTTGT  
GGACTCGGATGATTGGATTGAGCAAGATTATGTAGAACTCTATATAAAAAAATAGTAGAGTATCAGGC  
TGATATTGCAGTTGGTAATTATTATTCTTTCAACGAAAGTGAAGGAATGTTCTACTTTTCATATATTGGG  
AGACTCCTATTATGAGAAAGTATATGATAATGTTTCTATCTTTGAGAACTTGATGAACTCAAGAAAT  
GAAGAGTTTGTCTTTGATATCTGCTTGGGGTAAACTCTATAAGGCAAGATTGTTTGAGCAGTTGCGCTT  
TGACATAGGTAAATTAGGAGAAGATGGTTACCTCAATCAAAAGGTATATTTATTATCAGAAAAGGTAAT  
TTATTTAAATAAAAAGTCTTTATGCTTATCGGATTAGAAAAGGTAGTTTATCAAGAGTTTGGACAGAAAA  
GTGGATGCACGCTTTAGTTGATGCTATGTCTGAACGTATTACGCTACTAGCTAATATGGGTTATCCTCT  
AGAGAAACACTTGGCAGTTTATCGTCAGATGTTGGAAGTCAGTCTCGCCAACGGTCAAGCTAGTGGTTT  
ATCTGACACAGCAACGTATAAAGAGTTTGAAATGAAACAAAAGGCTTTTAAATCAGCTATCGAGACAAGA  
GGAAAGTGAAAAGAAAGCCATTGTCCTCGCAGCAAACTATGGCTATGTAGACCAAGTTTTAACGACAAT  
CAAGTCTATTTGTTATCATAATCGTTTCGATTTCGTTTTTATCTGATTATAGCGATTTTCCAAATGAATG  
GATTAAGCAATTAAATAAGCGCTTAGAGAAAGTTGACTCAGAAATTATTAATTGTGCGGTAACCTCTGA  
GCAAATTTTCATGTTATAAATCGGATATTAGTTACACAGTCTTTTACGCTATTTTCATAGCTGATTTCTG  
GCAAGAAGACAAGGCCCTCTACTTGGACTGTGATCTAGTTGTAACGAAAAATCTGGATGACTTGTTCG  
TACAGACTTACAAGATTATCCTTTGGCTGCTGTTAGAGATTTTGGGGGCAGAGCTTATTTTGGTCAAGA  
AATCTTTAATGCCGTGTTCTCTTGGTAAACAATGCTTTTGGAAAAAGAGAATATGACCCAAAAAT  
AATTGATGTAACCAATGAATGGCATGATAAGGTGGATCAGGCAGATCAGAGCATCTTGAATATGCTTTT  
TGAACATAAATGGTTGGAATTGGACTTTGATTATAATCATATTGTCATTCATAAAGAACTTTGCTGATTA  
TCAATTGCCCTGAGGGTCAGGATTATCCTGCTATTATTCACTATCTTCTCATCGGAAGCCGTGGAAAGA  
TTTGGCGGCCCAACCTATCGTGAAGTTTGGTGGTACTATCATGGGCTTGAATGGACAGAATTGGGACA  
AAACCATCATTTACATCCATTACAAAGATCTCACATCTATCCAATAAAGGAACCTTTCACTTGTCTAAT  
CTATACTGCCTCAGACCATATTGAACAAATTGAGACATTGGTTCAATCCTTGCCTGATATTCAGTTTAA

Table 1

GATAGCAGCTAGAGTAATAGTTAGTGATCGATTGGCTCAGATGACAATTTATCCAAACGTGACTATATT  
 TAACGGAATTCACATATTTGGTAGATGTCGATAATGAATTGGTAGAAACCAGTCAAGTACTTTTAGATAT  
 TAATCATGGCGAAAAGACAGAAGAAATTTCTCGATCAATTTGCTAATCTTGGCAAGCCTATCTTATCCTT  
 TGAAAATACTAAACCTATGAAGTAGGTGAGGAGCATATGCTGTTGACCAAGTCAAGCAATGATTGA  
 AAAATTGAGAGAAATAAGCAAA

**SP113 amino acid (SEQ ID NO:200)**

CLDSIITQTYKNIEIVVNDGSTDASGEICKEFSEMDHRILYIEQENAGLSAARNTGLNMSGNYVTFV  
 DSDDWIEQDYVETLYKKIVEYQADIAVGNYYSFNESEGMFYFHLGDSYIEKVYDNVSIFENLYETQEM  
 KSFALISAWGKLYKARLFEQLRFDIGKLGEDGYLNQKVYLLSEKVIYLNKSLYAYRIRKGSLSRVWTEK  
 WMHALVDAMSERITLLANMGYPLEKHLAVYRQMLEVSLANGQASGLSDTATYKEFEMQRLNQLSRQE  
 ESEKKAIVLAANYGYVDQVLTTIKSICYHNRSIRFYLIHSDFPNEWIKQLNKRLEKFDSEIINCRVTSE  
 QISCYKSDISYTVFLRYFIADVFQEDKALYLDCLVVTKNLDDLFDLQDYPLAAVRDFGGRAYFGQE  
 IFNAGVLLVNNAFWKKENMTQKLIDVTNEWHDKVDQADQSI LNMLFEHKWLELDFDYNHIVIHKQFADY  
 QLPEGQDYPALIIHYLSHRKPWKDLAAQTYREVWVYHGLEWTELQGNHHLHPLQRSHIY?IKEPFTCLI  
 YTASDHIEQIETLVQSLPDIQFKIAARVIVSDRLAQMTIYPNVTIFNGIHYLVVDVDELVETSQVLLDI  
 NHGEKTEEILDQFANLGKPILSFENTKTYEVGQEAYAVDQVQAMIEKLREISK

**SP114 nucleotide (SEQ ID NO:201)**

CATTGAGAAGCAGACCTATCAAAATCTGGAAATTATTCTTGTGATGATGGTGCAACAGATGAAAAGTGG  
 TCGCTTGTGTGATTCAATCGCTGAACAAGATGACAGGGTGTGAGTGCCTCATAAAAAAGAAAGGATT  
 GTCGCAAGCACGAAATGATGGGATGAAGCAGGCTCACGGGGATTATCTGATTTTATTGACTCAGATGA  
 TTATATCCATCCAGAAATGATTCAGAGCTTATATGAGCAATTAGTTCAAGAAGATGCGGATGTTTTGAG  
 CTGTGGTGTGATGAATGTCTATGCTAATGATGAAAGCCACAGTCAGCCAATCAGGATGACTATTTTGT  
 CTGTGATTCTCAAACATTTCTAAAGGAATACCTCATAGGTGAAAAAATACCTGGGACGATTTGCAATAA  
 GCTAATCAAGAGACAGATTGCAACTGCCCTATCCTTTCTAAGGGGTTGATTTACGAAGATGCCTATTA  
 CCATTTTGATTTAATCAAGTTGGCCAAGAAGTATGTGGTTAATACTAAACCCTATTATTACTATTTCCA  
 TAGAGGGGATAGTATTACGACCAAACCCTATGCAGAGAAGGATTTAGCCTATATTGATATCTACCAAAA  
 GTTTTATAATGAAGTTGTGAAAAACTATCTGACTTGAAAGAGGTCGCTTTTTCAGATTTGGCCTATGC  
 CCACCTCTTTTATTCTGGATAAGATGTTGCTAGATGATCAGTATAAACAGTTTGAAGCCTATTCTCAGAT  
 TCATCGTTTTTTTAAAGGCCATGCCCTTGTCTATTTCTAGGAATCCAATTTTCCGTAAGGGGAGAAGAAT  
 TAGTGCCTTTGGCCCTATTTCATAAATATTTCTTATATCGATTCTTATTACTGAAAAATATTGAAAAATC  
 TAAAAAATTACAT

**SP114 amino acid (SEQ ID NO:202)**

IQKQTYQNLEIILVDDGATDESGRLCDSIAEQDDRVSVLHKKNEGLSQARNDMKQAHGQYLIFIDSD  
 YIHPEMIQSLYEQLVQEDADVSSCGVMNVYANDESPQSANQDDYFVCDSTFLKEYLIGEKIPGTICNK  
 LIKRQIATALSFPKGLIYEDAYYHFDLIKLAKEYVNTKPYYYFHRGDSITTKPYAEKDLAYIDIYQK  
 FYNEVVKNPDLKEVAFFRLAYAHFFILDKMLDDQYKQFEAYSQIHRFLKGHAFASRNPIFRKGRRI  
 SALALFINISLYRFLLLKNIKSKKLH

**SP115 nucleotide (SEQ ID NO:203)**

TAAGGCTGATAATCGTGTTCAAATGAGAACGACGATTAATAATGAATCGCCATTGTTGCTTTCTCCGTT  
 GTATGGCAATGATAATGGTAACGGATTATGGTGGGGGAACACATTGAAGGGAGCATGGGAAGCTATTCC  
 TGAAGATGTAAAGCCATATGCAGCGATTGAACCTCATCTGCAAAAAGTCTGTAAACCAACAAGTTGTAT  
 TCCACGAGATACGAAAGAATTGAGAGAATGGTATGTCAAGATGTTGGAGGAAGCTCAAAGTCTAAACAT  
 TCCAGTTTTCTTGGTTATTATGTCGGCTGGAGAGCGTAATACAGTTCCCTCCAGAGTGGTTAGATGAACA  
 ATTCCAAAAGTATAGTGTGTTAAAAGGTGTTTTAAATATTGAGAATTATTGGATTTACAATAACCAGTT  
 AGCTCCGCATAGTGCTAAATATTTGGAAGTTTGTGCCAAATATGGAGCGCATTTTATCTGGCATGATCA  
 TGAAAAATGGTTCTGGGAAACTATTATGAATGATCCGACATTCTTTGAAGCGAGTCAAAAATATCATAA  
 AAATTTGGTGTGGCAACTAAAAATACGCCAATAAGAGATGATGCGGGTACAGATTCTATCGTTAGTGG  
 ATTTTGGTTGAGTGGCTTATGTGATAACTGGGGCTCATCAACAGATACATGGAAATGGTGGGAAAAACA  
 TTATACAAACACATTTGAAACTGGAAGAGCTAGGGATATGAGATCCTATGCATCGGAACCAGAATCAAT  
 GATTGCTATGGAAATGATGAATGTATATACTGGGGGAGGCACAGTTTATAATTTGAATGTGCCGCGTA  
 TACATTTATGACAAATGATGTACCAACTCCAGCATTTACTAAAGGTATTATTCCTTTCTTTAGACATGC  
 TATACAAAATCCAGCTCCAAGTAAGGAAGAAGTTGTAATAAGAAAAGCTGTATTTTGAATGGAGA  
 AGGTAGGATTAGTTCATTAAACGGATTTTATCAAGGACTTTATTCGAATGATGAAACCAATGCCTTTATA  
 TAATAATGGGAGATATCATATCTTCTGTAATACATGAGAAAATGATAAGGAAAAGATTTTCATCTAT

Table 1

90

ATTCCTTAATGCAAAAATTTTGGACTAAAAATAGTGAGGAATTGTCTAGTAAAGTCAACTATTTAAACTC  
 GCTTTTATCCAAAACCTTTATGAAGGAGATGGGTATGCTCAGCGTGTAGGTAATTCCTGGTATATTTATAA  
 TAGTAATGCTAATATCAATAAAAAATCAGCAAGTAATGTTGCCTATGTATACTAATAATACAAAAGTCGTT  
 ATCGTTAGATTTGACGCCACATACTTACGCTGTTGTTAAAGAAAATCCAAATAATTTACATATTTTATT  
 GAATAATTACAGGACAGATAAGACAGCTATGTGGGCATTATCAGGAAATTTTGATGCATCAAAAAGTTG  
 GAAGAAAGAAGAATTAGAGTTAGCGAACTGGATAAGCAAAAATTTATCCATCAATCCTGTAGATAATGA  
 CTTTAGGACAACAACACTTACATTAAAAGGGCATACTGGTCATAAACCTCAGATAAATAAGTGGCGA  
 TAAAAATCATTATACTTATACAGAAAATTTGGGATGAGAATACCCATGTTTATACCATTACGGTTAATCA  
 TAATGGAATGGTAGAGATGTCTATAAATACTGAGGGGACAGGTCCAGTCTCTTTCCCAACACCAGATAA  
 ATTTAATGATGGTAATTTGAATATAGCATATGCAAAACCAACAACACAAAGTTCTGTAGATTACAATGG  
 AGACCTTAATAGAGCTGTGGATGGTAACAGAAATGGTAATTTTAACTCTGGTTCGGTAACACACACTAG  
 GGCAGATAATCCCTCTTGGTGGGAAGTCGATTTGAAAAAATGGATAAAGTTGGGCTTGTTAAATTTA  
 TAATCGCACAGATGCTGAGACTCAACGTCTATCTAATTTT

**SP115 amino acid (SEQ ID NO:204)**

KADNRVQMRTTINNESPLLLSPLYGNDNGNGLWWGNTLKGAWAEPEDVKPYAAIELHPAKVCKPTSCI  
 PRDTKELREWVVKMLEEAQSLNIPVFLVIMSAGERNTVPPEWLDEQFQKYSVLKGVNLNIENYWIYNQL  
 APHSKYLEVCAKYGAHFIWHDHEKWFWEITMNDPTFFEASQKYHKNLVLATKNTPIRDAGTDSIVSG  
 FWLSGLCDNWGSSDTWKWWEKHYTNTFETGRARDMRSYASEPESMIAMEMMNVTYGGGTVYNFECAAY  
 TFMNDVPTPAFTKGIIPFFRHAIQNPAPSKKEEVNRTKAVFWNGEGRISSLNGFYQGLYSNDETMPY  
 NNGRYHILPVIHEKIDKEKISSIFPNAKILTKNSEELSSKVNYLNSLYPKLYEGDGYAQRVGNWSWYIYN  
 SNANINKNQVMLPMYTNNTKSLSLDLTPHTYAVVKENPNNLHILLNNYRTDKTAMWALSGNFDASKSW  
 KKEELELANWISKYNSINPVDNDFRTTTLTKLGHTGHKPKQINISGDKNHYTYTENWDENTHVYITIVNH  
 NGMVEMSINTEGTGPVSFPTPKDFNDGNLNIAYAKPTTQSSVDYNGDPNRAVDGNRNGNFNSGSVTHTR  
 ADNPSWWEVDLKKMDKVGLVKIYNRTDAETQRLSNF

**SP117 nucleotide (SEQ ID NO:205)**

CTGTGGCAATCAGTCAGCTGCTTCCAAACAGTCAGCTTCAGGAACGATTGAGGTGATTTACAGAGAAAA  
 TGGCTCTGGGACACGGGGTGCTTTCACAGAAATCACAGGGATTCTCAAAAAAGACGGTGATAAAAAAAT  
 TGACAACACTGCCAAAACAGCTGTGATTCAAAATAGTACAGAAGGTGTTCTCTCAGCAGTTCAAGGGAA  
 TGCTAATGCTATCGGCTACATCTCCTTGGGATCTTTAACGAAATCTGTCAAGGCTTTAGAGATTGATGG  
 TGCAAGGCTAGTCGAGACACAGTTTTAGATGGTGAATACCTCTTCAACGTCCCTTCAACATTGTTTG  
 GTCTTCTAATCTTTCCAAGCTAGGTCAAGATTTTATCAGCTTTATCCACTCCAAACAAGGTCAACAAGT  
 GGTCACAGATAATAAATTTATTGAAGCTAAAACCGAAACCACGGAATATACAAGCCAACACTTATCAGG  
 CAAGTTGTCTGTTGTAGGTTCCACTTCAGTATCTTCTTTAATGGAAAAATTAGCAGAAGCTTATAAAAA  
 AGAAAAATCCAGAAGTTACGATTGATATTACCTCTAATGGGTCTTCAGCAGGTATTACCGCTGTTAAGGA  
 GAAAACCGCTGATATTGGTATGGTTTCTAGGGAATTAACCTCTGAAGAAGGTAAGAGTCTACCCCATGA  
 TGCTATTGCTTTAGACGCTATTGCTGTTGTGGTCAATAATGACAATAAGGCAAGCCAAGTCAGTATGGC  
 TGAAGTTGCAGACGTTTTTAGTGGCAAATTAACCACCTGGGACAAGATTAAA

**SP117 amino acid (SEQ ID NO:206)**

CGNQSAASKQSASGTIEVISRENGSGTRGAFTETIGILKKDGDKIDNTAKTAVIQNSTEGLVLSAVQGN  
 ANAIGYISLGLTKSVKALEIDGVKASRDTVLGDGEYPLQRPFNIVWSSNLSKLGQDFISFIHKSQGGQV  
 VTDNKFIEAKTETETETYSQHLGSKLSVVGSTSVSSLMKLAAYKKENPEVTIDITSNGSSAGITAVKE  
 KTADIGMVSRELTPPEGKSLTHDAIALDGIIVVVNNDNKASQVSMALADVFSGLTWTWDKIK

**SP118 nucleotide (SEQ ID NO:207)**

TTGTCAACAACAACATGCTACTTCTGAGGGGACGAATCAAAGGCAAAGCAGTTCAGCGAAAGTTCCATG  
 GAAAGCTTCATACACCAACCTAAACAACCAGGTAAGTACAGAAGAGGTCAAATCTCTCTTATCAGCTCA  
 CTTGGATCCAAATAGTGTGATGCATTTTTTAATCTCGTTAATGACTATAATACCATTGTCCGCTCAAC  
 TGGCTTATCAGGAGATTTCACTTCTTTACTCACACCGAATACGATGTTGAGAAAAATCAGTCATCTCTG  
 GAATCAAAAAGAAGGGCGATTTTGTGTTGGGACCAACTGCCGTATCAATAGTTATTGTCTTTGAAAAAATC  
 AGTCACCATTCCAAAGCTTGAAAAGAATGACCAGTTGCTTTTCTAGATAATGATGCCATTGATAAAGG  
 AAAGGTCTTTGATTCAAGATAAGGAAGAGTTGATATTCTATTTTCGAGAGTTCCAAGTCAAGTCAAC  
 TACAGATGTCAAGGTTACGCTGAAAAGATGGAAGCATTCTTCTCACAATTTCAATTCAATGAAAAAGC  
 TCGAATGCTGTCTGTAGTCTTGCACGACAATTTGGATGGCGAGTATCTGTTTGTAGGCCACGTTGGGGT  
 CTTAGTACCTGCTGATGACGGTTTCTTATTTGTAGAGAAATTGACTTTCGAAGAGCCCTACCAAGCGAT

WO 98/18930

PCT/US97/19422

Table 1

91

TAAATTTGCTAGTAAGGAAGATTGCTACAAGTATTTGGGCACCAAGTATGCGGATTATACAGGCGAGGG  
ACTGGCTAAGCCTTTTATCATGGATAATGATAAGTGGGTAAACTT

**SP118 amino acid (SEQ ID NO:208)**

CQQQHATSEGTNQRQSSSAKVPWKASYTNLNNQVSTEEVKSLLSAHLDPNSVDAFFNLVNDYNTIVGST  
GLSGDFTSFTHTEYDVEKISHLWNQKKGDFVGTNCRINSYCLLKNSVTIPKLEKNDQLFLDNDIDKG  
KVFDSQDKEEFDILFSRVPTESTTDDVKVHAEKMEAFFSQFQFNEKARMLSVVLHDNLDGEYLFVGHVGV  
LVPADDGFLFVEKLTFEOPYQAIFASKEDCYKYLGTKYADYTGEGLAKPFI MDNDKWVKL

**SP119 nucleotide (SEQ ID NO:209)**

TTGTTTCAGGCAAGTCCGTGACTAGTGAACACCAAACGAAAGATGAAATGAAGACGGAGCAGACAGCTAG  
TAAAACAAGCGCAGCTAAAGGGAAAGAGGTGGCTGATTTTGAATTGATGGGAGTAGATGGCAAGACCTA  
CCGTTTATCTGATTACAAGGGCAAGAAAGTCTATCTCAAATTCTGGGCTTCTTGGTGTTCATCTGTCT  
GGCTAGTCTTCCAGATACGGATGAGATTGCTAAAGAAGCTGGTGATGACTATGTGGTCTTGACAGTAGT  
GTCACCAGGACATAAGGGAGAGCAATCTGAAGCGGACTTTAAGAATTGGTATAAGGGATTGGATTATAA  
AAATCTCCCAGTCTAGTTGACCCATCAGGCAAACTTTTGGAACTTATGGTGTCCGTTCTTACCCAAC  
CCAAGCCTTTATAGACAAAGAAGGCAAGCTGGTCAAAACACATCCAGGATTCATGGAAAAAGATGCAAT  
TTTGCAAACTTTGAAGGAATTAGCC

**SP119 amino acid (SEQ ID NO:210)**

CSGKSVTSEHQTKDEMKTETASKTSAAKGKEVADFELMGVDGKTYRLSDYKGGKVVYLKFWASWCSICL  
ASLPDTDEIAKEAGDDYVVLTVVSPGHKEQSEADFKNWKGLDYKNLPVLVDPGKLLLEYTVGRSYPT  
QAFIDKEGLVKTHPGFMEKDAILQTLKELA

**SP120 nucleotide (SEQ ID NO:211)**

CTCGCAAATTGAAAAGGCGGCAGTTAGCCAAGGAGGAAAAGCAGTGAAAAAACAGAAATTAGTAAAGA  
CGCAGACTTGCACGAAATTTATCTAGCTGGAGGTTGTTTCTGGGGAGTGGAGGAATATTTCTCACGTGT  
TCCCGGGGTGACGGATGCCGTTTTCAGGCTATGCAAATGGTAGAGGAGAAACAACCAAGTACGAATTGAT  
TAACCAAACAGGTGATGCAGAAACCGTCCATGTCACTATGATGCCAAGCAAAATTTCTCTCAAGGAAAT  
CCTGCTTCACTATTTCCGCATTATCAATCCAACCAGCAAAAAATAACAAGGAAATGATGTGGGGACCCA  
GTACCGTACTGGTGTATTATTACACAGATGACAAGGATTTGGAAGTGATTAACCAAGTCTTTGATGAGGT  
GGCTAAGAAATACGATCAACCTCTAGCAGTTGAAAAGGAAAACCTTGAAGAATTTTGTGGTGGCTGAGGA  
TTACCATCAAGACTATCTCAAGAAAAATCCAAATGGCTACTGCCATATCAATGTTAATCAGGCGGCCTA  
TCCTGTCTATGATGCCAGCAAAATATCCAAAACCAAGTGATGAGGAATTGAAAAGACCCTGTCACCTGA  
GGAGTATGCAGTTACCCAGGAAAATCAAACAGAACGAGCTTTCTCAAACCGTTACTGGGATAAATTTGA  
ATCCGGTATCTATGTGGATATAGCAACTGGGGAACCTCTCTTTTCATCAAAGACAAATTTGAGTCTGG  
TTGTGGCTGGCCTAGTTTACCCAAACCATCAGTCCAGATGTTGTCACCTACAAGGAAGATAAGTCCTA  
CAATATGACGCGTATGGAAGTGGGAGCCGAGTAGGAGATTCTCACCTTGGGCATGTCTTTACGGATGG  
TCCACAGGACAAGGGCGGCTTACGTTACTGTATCAATAGCCTCTCTATCCGCTTTATTCCCAAAGACCA  
AATGGAAGAAAAAGGTACGCTTATTTAC

**SP120 amino acid (SEQ ID NO:212)**

SQIEKA AVSQQGKAVKKTEISKDADLHEIYLAGGCFWGVVEEYFSRVPGVTDVSGYANGRGETTKYELI  
NQTGHAETVHVITYDAKQISLKEILLHYFRIINPTSKNKQGNVDVGTQYRTGVYYTDDKDLEVINQVFDEV  
AKKYDQPLAVEKENLKNFVVAEDYHQDYLLKNPNNGYCHINVNQAAYPVIDASKYPKPSDEELKKTLSPE  
EYAVTQENQTERAFSNRYWDFESGIYVDIATGEPLFSSKDKFESGCGWPSFTQPI SPDVVITYKEDKSY  
NMTRMEVRSRVGDSHLGHVFTDGPQDKGLRYCINSLRSIRFIPKQMEEEKGTLIY

**SP121 nucleotide (SEQ ID NO:213)**

TTGTCAGTCAGGTTCTAATGGTTCTCAGTCTGCTGTGGATGCTATCAAACAAAAAGGGAAATTAGTTGT  
GGCAACACAGTCCCTGACTATGCACCTTTGAATTTCAATCATTGGTTGATGGAAAGAACAGGTAGTCGG  
TGCAGACATCGACATGGCTCAGGCTATCGCTGATGAACCTGGGGTTAAGTTGGAAATCTCAAGCATGAG  
TTTTTGACAATGTTTTGACCAGTCTTCAAACCTGGTAAGGCTGACCTAGCAGTTGCAGGAATTAGTGCTAC  
TGACGAGAGAAAAAGAAGTCTTTGATTTTTCAATCCCATACTATGAAAACAAGATTAGTTTCTTGGTTCG  
TAAGGCTGATGTGGAATAACAAGGATTTAACTAGCCTAGAAAGTGCTAATATTGCAGCCCAAAAGG  
GACTGTTCCAGAATCAATGGTCAAGGAACAATTCGCAAAAGTTCAATTAACCTCCCTAACTAATATGGG  
TGAAGCAGTCAATGAATTGCAGGCTGGAAAAATAGATGCTGTTTATATGGATGAGCCTGTTGCACTTAG

Table 1

TTATGCTGCTAAAAACGCTGGCTTAGCTGTGCAACTGTCAGCTTGAAGATGAAGGACGGCGACGCCAA  
TGCC

**SP121 amino acid (SEQ ID NO:214)**

CQSGSNGSQSAVDAIKQKGKLVVATSPDYAPFEFQSLVDGKNQVVGADIDMAQAIADDELGVKLEISSMS  
FDNVLTSLOTGKADLAVAGISATDERKEVDFDSIPYYENKISFLVRKADVEKYKDLSLESANIAAQKG  
TVPESMVKEQLPKVQLTSLTNMGEAVNELQAGKIDAVHMDPEVALSYAAKNAGLAVATVSLKMKDGDAN  
A

**SP122 nucleotide (SEQ ID NO:215)**

GGAAACTTCACAGGATTTTAAAGAGAAGAAAACAGCAGTCATTAAGGAAAAAGAAGTTGTTAGTAAAAA  
TCCTGTGATAGACAATAACACTAGCAATGAAGAAGCAAAAATCAAAGAAGAAAATTCCAATAAATCCCA  
AGGAGATTATACGGACTCATTTGTGAATAAAAAACACAGAAAATCCCAAAAAAGAAGATAAAGTTGTCTA  
TATTGCTGAATTTAAAGATAAAGAATCTGGAGAAAAAGCAATCAAGGAACCTATCCAGTCTTAAGAATAC  
AAAAGTTTTATATACTTATGATAGAATTTTAAACGGTAGTGCCATAGAAACAACCTCCAGATAACTTGGGA  
CAAAATTAAACAAATAGAAGGTATTTTCATCGGTTGAAAGGGCACAAAAGTCCAACCCATGATGAATCA  
TGCCAGAAAGGAAATTGGAGTTGAGGAAGCTATTGATTACCTAAAGTCTATCAATGCTCCGTTTGGGAA  
AAATTTTGATGGTAGAGGTATGGTCATTTCAAATATCGATACTGGAACAGATTATAGACATAAGGCTAT  
GAGAATCGATGATGATGCCAAAGCCTCAATGAGATTTAAAAAAGAAGACTTAAAAGGCACTGATAAAAA  
TTATTGGTTGAGTGATAAAATCCCTCATGCGTTCAATTATTATAATGGTGGCAAAATCACTGTAGAAAA  
ATATGATGATGGAAGGGATTATTTTGACCCACATGGGATGCATATTGCAGGGATTCTTGCTGGAAATGA  
TACTGAACAAGACATCAAAAACTTTAAACGGCATAGATGGAATTGCACCTAATGCACAAATTTTCTCTTA  
CAAAATGTATTCTGACGCAGGATCTGGGTTTGCAGGTTGATGAAACAATGTTTCATGCTATTGAAGATTC  
TATCAAACACAACGTTGATGTTGTTTCGGTATCATCTGGTTTTACAGGAACAGGTCTTGTAGGTGAGAA  
ATATTGGCAAGCTATTCGGGCATTAAGAAAAGCAGGCATTCCAATGGTTGTGCGCTACGGGTAACCTATGC  
GACTTCTGCTTCAAGTTCTTCATGGGATTTAGTAGCAATAATCATCTGAAAATGACCGACACTGGAAA  
TGTAACACGAACTGCAGCACATGAAGATGCGATAGCGGTCGCTTCTGCTAAAAATCAAACAGTTGAGTT  
TGATAAAGTTAACATAGGTGGAGAAAGTTTTAAATACAGAAATATAGGGGCCTTTTTTCGATAAGAGTAA  
AATCACAACAAATGAAGATGGAACAAAAGCTCCTAGTAAATTAATAATTTGTATATATAGGCAAGGGGCA  
AGACAAAGATTGTTGATGTTTGGATCTTAGGGGCAAAATTCAGTAATGGATAGAATTTATACAAAGGA  
TTTTAAAAATGCTTTTTAAAAAAGCTATGGATAAGGGTGCACGCGCCATTATGGTTGTAAATACTGTAAA  
TTACTACAATAGAGATAATTGGACAGAGCTTCCAGCTATGGGATATGAAGCGGATGAAGGTACTAAAAG  
TCAAGTGTTTTCAATTTTCAAGGAGATGATGGTGTAAGCTATGGAACATGATTAATCCTGATAAAAAAAC  
TGAAGTCAAAGAAATAATAAAGAAGATTTTAAAGATAAATTGGAGCAATACTATCCAATTGATATGGA  
AAGTTTTTAATTCCAACAAACCGAATGTAGGTGACGAAAAAGAGATTGACTTTAAGTTTGCACCTGACAC  
AGACAAAGAACTCTATAAAGAAGATATCATCGTTCCAGCAGGATCTACATCTTGGGGGCCAAGAATAGA  
TTTACTTTTTAAACCCGATGTTTCAGCACCTGGTAAAAATATTAATCCACGCTTAATGTTATTAATGG  
CAAATCAACTTATGGCTATATGTCAGGAAGTAGTATGGCGACTCCAATCGTGGCAGCTTCTACTGTTTT  
GATTAGACCGAAATTAAAGGAAATGCTTGAAAGACCTGTATTGAAAAATCTTAAGGGAGATGACAAAAT  
AGATCTTACAAGTCTTACAAAATTTGCCCTACAAAATACTGCGCGACCTATGATGGATGCAACTTCTTG  
GAAAGAAAAAAGTCAATACTTTGCATCACCTAGACAACAGGGAGCAGGCCTAATTAATGTGGCCAATGC  
TTTGAGAAATGAAGTTGTAGCAACTTTCAAAAACACTGATTCTAAAGGTTTGGTAAACTCATATGGTTC  
CATTTCTCTTAAAGAAATAAAAGGTGATAAAAAATACTTTACAATCAAGCTTCACAATACATCAAACAG  
ACCTTTGACTTTTTAAAGTTTTCAGCATCAGCGATAACTACAGATTCTCTAACTGACAGATTAAAACCTTGA  
TGAAACATATAAAGATGAAAAATCTCCAGATGGTAAGCAAAATTGTTCCAGAAATTCACCCAGAAAAAGT  
CAAAGGAGCAAATATCACATTTGAGCATGATACTTTCACTATAGGCGCAAAATCTAGCTTTGATTTGAA  
TGCGGTTATAAATGTTGGAGAGGCCAAAAACAAAAATAAATTTGTAGAATCATTTATTCAATTTTGAGTC  
AGTGGAAGCGATGGAAGCTCTAAACTCCAGCGGGAAGAAAAATAAACTTCCAACCTTCTTTGTGCGATGCC  
TCTAATGGGATTTGCTGGGAATTGGAACCACGAACCAATCCTTGATAAATGGGCTTGGGAAGAAGGGTC  
AAGATCAAAAACACTGGGAGGTTATGATGATGATGGTAAACCGAAAAATTCAGGAACCTTAAATAAGGG  
AATTGGTGGAGAACATGGTATAGATAAATTTAATCCAGCAGGAGTTATACAAAATAGAAAAGATAAAAA  
TACAACATCCCTGGATCAAAATCCAGAATTATTTGCTTTCAATAACGAAGGGATCAACGCTCCATCATC  
AAGTGGTTCTAAGATTGCTAACATTTATCCTTTAGATTCAAATGGAATCCTCAAGATGCTCAACTTGA  
AAGAGGATTAACACCTTCTCCACTTGTATTAAGAAGTGCAGAAGAAGGATTGATT

**SP122 amino acid (SEQ ID NO:216)**

ETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDTSFVNKNTENPKKEDKVYV  
IAEFKDKESGEKAIKELSSLKNTKVLTYDRIFNGSAIETTPDNLDKIKQIEGISSVERAQKVQPMNH

Table 1

ARKEIGVEEAIDYLSINAPFGKNFDGRGMVISNIDTGTDIRHKAMRIDDDAKASMRFKKEDLKGTDKN  
YWLSDKIPHAFNYNNGKITVEKYDDGRDYFDPHGMHIAILAGNDTEQDIKNFNGIDGIAPNAQIFSY  
KMYSDAGSGFAGDETMFHAIEDSIKHNVDVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYA  
TSASSSSWDLVANNHLKMTDTGNVTRTAAHEDIAIVASAKNQTFEFDKVNIGGESFKYRNIGAFFDKSK  
ITTNEDGTKAPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVN  
YYNRDNWTEL PAMGYEADEGTSQVFSISGDDGVKLWNMINPDKKTEVKNRNKEDFKDKLEQYYPIDME  
SFNSNKNPVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPVDVSAPGKNIKSTLNVING  
KSTYGYMSGTSMATPIVAASTVLIRPKLKEMLERPVLKNLKGDDKIDLTSLTKIALQNTARPMMDATSW  
KEKSQYFASPRQQAGLINVANALRNEVVATFKNTDSKGLVNSYGSISLKEIKGDKKYFTIKLHNTSNR  
PLTFKVSASAITTDSLDRKLDETYKDEKSPDGKQIVPEIHPEKVKGANITFEHDTFTIGANSSFDLN  
AVINVGEAKNKNKFVESFIHFESVEAMEALNSSGKKINFQPSLSMPLMGFAGNWNHEPILDKWAVEEGS  
RSKTLGGYDDDGPPIPGTLNKGIGGEHGDIFKNPAGVIQNRKDKNTTSLDQNPFLFAFNNEGINAPSS  
SGSKIANIYPLDSNGNPQDAQLERGLTPSPLVLRSAEEGLI

**SP123 nucleotide (SEQ ID NO:217)**

TGTGGTCAAGTTGAGACTCCTCAATCAATAACAAATCAGGAGCAAGCTAGGACAGAAAACCAAGTAGT  
AGAGACAGAGGAAGCTCCAAAAGAAGAAGCACCTAAAACAGAAGAAAGTCCAAAGGAAGAACCAAAATC  
GGAGGTAAAACCTACTGACGACACCCCTTCTAAAGTAGAAGAGGGGAAAGAAGATTACAGCAGAACCCAGC  
TCCAGTTGAAGAAGTAGGTGGAGAAGTTGAGTCAAAACAGAGGAAAAAGTAGCAGTTAAGCCAGAAAG  
TCAACCATCAGACAAACCAGCTGAGGAATCAAAAGTTGAACAAGCAGGTGAACCCAGTCGCGCCAAGAGA  
AGACGAAAAGGCACCAGTCGAGCCAGAAAAGCAACCAGAAGCTCCTGAAGAAGAGAAGGCTGTAGAGGA  
AACACCGAAACAAGAAGAGTCACTCCAGATACCAAGGCTGAAGAACTGTAGAACCAAAAGAGGAGAC  
TGTTAATCAATCTATTGAACAACCAAAAGTTGAAACGCCTGCTGTAGAAAAACAAACAGAACCAACAGA  
GGAACCAAAAGTTGAACAAGCAGGTGAACCCAGTCGCGCCAAGAGAAGACGAACAGGCACCAACGGCACC  
AGTTGAGCCAGAAAAGCAACCAGAAGTTCTGAAGAAGAGAAGGCTGTAGAGGAAACACCGAAACCGAGA  
AGATAAAATAAAGGTATTGGTACTAAAGAACCAGTTGATAAAAGTGAGTTAAATAATCAAATTTGATAA  
AGCTAGTTCAAGTTTCTCTACTGATTATTCTACAGCAAGTTACAATGCTCTTGGACCTGTTTTAGAAAC  
TGCAAAAGGTGTCTATGCTTCAGAGCCTGTAAAACAGCCTGAGGTAAATAGCGAGACAAATAAACTTAA  
AACGGCTATTGACGCTCTAAACGTTGATAAAACTGAATTAACAATACGATTGCAGATGCAAAAACAAA  
AAGGAAAGAACATTACAGTGATAGAAGTTGGCAAAACCTCCAAACTGAAGTTACAAAGGCTGAAAAAGT  
TGCAGCTAATACAGATGCTAAACAAGTGAAGTTAACGAAGCTGTTGAAAAATTAAGTCAACTATTGA  
AAAATTGGTTGAATTATCTGAAAAGCCAATATTAAACATTGACTAGTACCGATAAGAAAATATTGGAACG  
TGAAGCTGTTGCTAAGTATACTCTAGAAAAACAAAACAAAACAAAATCAAATCAATCACAGCTGAATT  
GAAAAAAGGAGAAGAAGTTATTAATACTGTAGTCCTTACAGATGACAAGGTAACAACAGAACTATAAG  
CGCTGCATTTAAGAACCTAGAGTACTACAAAGAATACACCCTATCTACAACATGATTTACGACAGAGG  
TAACGGTGAAGAACTGAACTCTAGAAAATCAAAATATTCAATTAGATCTTAAAAAAGTTGAGCTTAA  
AAATATTAAACGTACAGATTTAATCAAATACGAAAATGAAAAAGAACTAATGAATCACTGATAACAAC  
TATTCCTGATGATAAGAGCAATTATTATTTAAAAATAACTTCAAATAATCAGAAAACCTACATTACTAGC  
TGTTAAAAATATAGAAGAACTACGGTTAACGGAACACCTGTATATAAAGTTACAGCAATCGCAGACAA  
TTTAGTCTCTAGAAGTCTGATAATAAATTTGAAGAAGAA

**SP123 amino acid (SEQ ID NO:218)**

VVEVETPQSITNQEQARTENQVVETEEAPKEEAPKTEESPKEEPKSEVKPTDDTLPKVEEGKEDSAEPA  
PVEEVGGEVESKPEEKVAVKPESQPSDKPAEESKVEQAGEPVAPREDEKAPVEPEKQPEAPEEEKAVEE  
TPKQEESTPDTKAEETVEPKEETVNQSIQPKVETPAVEKQTEPTEEPKVEQAGEPVAPREDEQAPTAP  
VEPEKQPEVPPEEEKAVEETPKPEDKIKIGITKEPVDKSELNNQIDKASSVSPTDYSTASYNALGPVLET  
AKGVYASEPVKQPEVNSETNKLKTAIDALNVDKTELNNTIADAKTKVKEHYSRWSWQNLQTEVTKAEKV  
AANTDAKQSEVNEAVEKLATIEKLVELSEKPIILTLSTDKKILEREAVAKYTLNQNKTIKSITAEL  
KKGEEVINTVVLTDKVTETETISAAFNLEYKEYTLSTMIYDRNGEETETLENQNIQLDLKKVELK  
NIKRTDLIKYENGKETNESLITTIIPDDKSNNYLLKITSNNQKTLLAVKNIEETTNGTPVYKVTAIADN  
LVSRTADNKFEEE

**SP124 amino acid (SEQ ID NO:219)**

AACACCTGTATATAAAGTTACAGCAATCGCAGACAATTTAGTCTCTAGAAGTCTGATAATAAATTTGA  
AGAAGAATACGTTCACTATATTGAAAAACCTAAAGTCCACGAAGATAATGTATATTATAATTTCAAAGA  
ATTAGTGGAAGCTATTCAAACGATCCTTCAAAGAATATCGTCTGGGACAAATCAATGAGCGCTAGAAA  
TGTTGTTCCCTAATGGAAAAATCATATATCACTAAAGAATTCACAGGAAAACTTTTAAAGTTCTGAAGGAAA  
ACAATTTGCTATTACTGAATTGGAACATCCATTATTTAATGTGATAACAAACGCAACGATAAATAATGT

Table 1

94

GAATTTTGAAAATGTAGAGATAGAACGTTCTGGTCAAGATAATATTGCATCATTAGCCAATACTATGAA  
 AGGTTCTTCAGTTATTACAAATGTCAAAATTACAGGCACACTTTTCAGGTCGTAATAATGTTGCTGGATT  
 TGTAAATAATATGAATGATGGAACCTCGTATTGAAAATGTTGCTTTCTTTGGCAAACCTACACTCTACAAG  
 TGGAAATGGCTCTCATACAGGGGGAATTGCAGGTACAACTATAGAGGAATTGTTAGAAAAGCATATGT  
 TGATGCTACTATTACAGGAAACAAAACACGCGCCAGCTTGTAGTTCCATAAGTAGATTATGGATTAAAC  
 TCTAGACCATCTTATTGGTACAAAAGCTCTCCTAAGTGAAGTCGGTTGTAAAAGGTAAAAATAGATGTTTC  
 AAATCCAGTAGAAGTTGGAGCAATAGCAAGTAAGACTTGGCCTGTAGGTACGGTAAGTAATTCTGTCAG  
 CTATGCTAAGATTATCCGTGGAGAGGAGTTATTCGGCTCTAACGACGTTGATGATTCTGATTATGCTAG  
 TGCTCATATAAAAAGATTTATATGCGGTAGAGGGATATTCGTCAGGTAATAGATCATTTAGGAAATCTAA  
 AACATTTACTAAATTAATAAGAAACAAGCTGATGCTAAAGTTACTACTTTCAATATTACTGCTGATAA  
 ATTAGAAAGTGATCTATCTCTCTTGCAAACTTAATGAAGAAAAAGCCTATTCTAGTATTCAAGATTA  
 TAACGCTGAATATAACCAAGCCTATAAAAATCTTGAAAAATTAATACCATTCTACAATAAAGATTATAT  
 TGTATATCAAGGTAATAAATTAATAAAGAACACCATCTAAATACTAAAGAAAGTTCTTCTGTTACCGC  
 GATGAACAACAATGAGTTTATCACAACCTAGATGAAGCTAATAAAATTAATTGTTCACTATGCGGACGG  
 TACAAAAGATTACTTTAACTTGTCTTCTAGCAGTGAAGGTTTAAGTAATGTAAAAGAATATACTATAAC  
 TGACTTAGGAATTAATATACACCTAATATCGTTCAAAAAGATAACACTACTCTTGTTAATGATATAAA  
 ATCTATTTTGAATCAGTAGAGCTTCAGTCTCAAACGATGTATCAGCATCTAAATCGATTAGGTGACTA  
 TAGAGTTAATGCAATCAAAGATTTATATTTAGAAGAAAGCTTCACAGATGTTAAAGAAAACTTAACAAA  
 CCTAATCACAAAATTAGTTCAAAACGAAGAACATCAACTAAATGATTCTCCAGCTGCTCGTCAAAATGAT  
 TCGTGATAAAGTCGAGAAAAACAAAGCAGCTTTATTACTAGGTTTAACTTACCTAAATCGTTACTATGG  
 AGTTAAATTTGGTGATGTTAATATTAAAGAATTAATGCTATTCAAACCAGATTTCTATGGTGAAAAAGT  
 TAGCGTATTAGACAGATTAATTGAAATCGGTTCTAAAGAGAACAAACATTAAGGTTTACGTACATTCTGA  
 CGCATTCGGTCAAGTA

**SP124 amino acid (SEQ ID NO:220)**

TPVYKVTAIADNLVSRTADNKFEEHYVHYIEKPKVHEDNVYYNFKELVEAIQNDPSKEYRLGQSMSARN  
 VVPNGKSYITKEFTGKLLSSEKQFAITELEHPLFNVTINATINNVPNFENVEIERSGQDNISLANMTK  
 GSSVITNVKITGTLGRNNVAGFVNMNDGTRIENVAFFGKLHSTSGNGSHTGGIAGTNYRGIVRKAYV  
 DATITGNKTRASLLVPKVDYGLTLDHLIGTKALLTESVVKGLDVSNNPVEVGAIASKTWVPVGTVSNVSV  
 YAKIIRGEELFGSNDVDDSDYASAHIKDLYAVEGYSSGNRSFRKSKTFTKLTKEQADAKVTTFNITADK  
 LESDLSPLAKLNEEKAYSSIQDYNAEYNQAYKNLEKLI PFYNKDYIVYQGNKLNKEHHLNTEKVL SVTA  
 MNNNEFITNLDEANKIIVHYADGTKDYFNLSSSSEGLSNVKEYTITDLGIKYTPNIVQKDNNTLVNDIK  
 SILESVELQSQTMYQHLNRLGDYRVNAIKDLYLEESFTDVKENLTNLITKLVQNEEHQLNDSPAARQMI  
 RDKVEKNKAALLLGLTYLNRYYGVKFGDVNIKELMLFKPDFYGEKVSVDRLIEIGSKENNIKGSRTFD  
 AFGQV

**SP125 nucleotide (SEQ ID NO:221)**

ATTAGACAGATTAATTGAAATCGGTTCTAAAGAGAACAACTTAAAGGTTACGTACATTTCGACGCATT  
 CGGTCAAGTATTGGCTAAATATACTAAATCAGGTAATTTAGATGCATTTTAAATTATAATAGACAATT  
 GTTCACAAATATAGACAATATGAACGATTGGTTTATTGATGCTACAGAAGACCATGTCTACATCGCAGA  
 ACGCGCTTCTGAGGTGCAAGAAATTAATAAATCTAAACATCGTGCATTTCGATAATTTAAAACGAAGTCA  
 CCTTAGAAATACTATACTCCCACTACTGAATATTGATAAAGCACATCTTTATTTAATTTCAAATTATAA  
 TGCAATTGCCTTTGGTAGTGACAGCGATTAGGTAAAAATCATTAGAAGATATTAAAGATATCGTTAA  
 CAAAGCTGCAGATGGTTATAGAACTATTATGATTTCTGGTATCGTCTAGCGTCTGATAACGTTAAACA  
 ACGACTACTAAGAGATGCTGTTATTCCTATTTGGGAAGGTTATAACGCTCCTGGTGATGGGTTGAAAA  
 ATATGGCCGCTATAATACCGACAAAGTATATACTCCTCTTAGAGAATTCTTTGGTCCTATGGATAAGTA  
 TTATAATTATAATGGAACAGGAGCTTATGCTGCTATATATCCTAACTCTGATGATATTAGAACTGATGT  
 AAAATATGTTTCATTTAGAAATGGTTGGTGAATACGGTATTTTACGTTTACACACATGAAACAACACACGT  
 CAACGACCGTGCGATTTACTTAGGTGGCTTTGGACACCGTGAAGGTACTGATGCTGAAGCATATGCTCA  
 GGGTATGCTACAACTCCTGTTACTGGTAGTGGATTGATGAGTTTGGTCTTTAGGTATTAATATGGT  
 ATTTAAACGCAAAAATGATGGGAATCAGTGGTATATTACAGATCCAAAACTCTAAAAACACGAGAAGA  
 TATTAATAGATATATGAAGGGTTATAATGACACTTAACTCTTCTTGATGAAATTGAGGCTGAATCTGT  
 GATTTCTCAACAAAATAAAGATTTAAATAGTGCATGGTTCAAAAAAATAGATAGAGAATACCGTGATAA  
 CAATAAATTAATCAATGGGATAAAAATTCGAAATCTAAGTCAAGAAGAGAAAAATGAATTAATTAATCA  
 ATCTGTTAATGATTTAGTTGATCAACAATTAATGACTAATCGCAATCCAGGTAATGGTATCTATAAACC  
 CGAAGCAATTAGCTATAACGATCAATCACCTTATGTAGGTGTTAGAATGATGACCGGTATCTACGGAGG  
 TAATACTAGTAAAGGTGCTCCTGGAGCTGTTTCATTCAAACATAATGCTTTTATAGATTATGGGGTTACTA  
 CGGATACGAAAATGGGTTCTTAGGTTATGCTTCAAATAAATATAAACAACAATCTAAAAACAGATGGTGA

WO 98/18930

PCT/US97/19422

Table 1

95

GTCTGTTCTAAGTGATGAATATATTATCAAGAAAAATATCTAACAATACATTTAATACTATTGAAGAATT  
 TAAAAAGCTTACTTCAAAGAAGTTAAAGATAAAGCAACGAAAGGATTAACAACATTCGAAGTAAATGG  
 TTCTTCCGTTTCATCATACGATGATTTACTGACATTGTTTAAAGAAGCTGTTAAAAAGATGCCGAAAC  
 TCTTAAACAAGAAGCAAACGGTAATAAACAGTATCTATGAATAATACAGTTAAATTTAAAGAAGCTGT  
 TTATAAGAACTTCTTCAACAAACAAATAGCTTTAAAACTTCAATCTTTAAA

**SP125 amino acid (SEQ ID NO:222)**

LDRLIEIGSKENNIKGSRTFDFGQVLAKYTKSGNLDAFLNYNRQLFTNIDNMNDWFI DATEDHVYIAE  
 RASEVEEIKNSKHRAFDNLKRSHLRNTILPLLNDKAHLYLISYNAIAFGSAERLGKKSLEDIKDIVN  
 KAADGYRNYDFWYRLASDNVQRLRLDAVIP I WEGYNAPGGWVEKYGRYNTDKVYTPLEFFGPMCKY  
 YNYNGTGAYAAIYPNSDDIRTDVKYVHLEMVGEYGISVYTHETTHVNDRAIYLGGFGHREGTDAEYAAQ  
 GMLQTPVTGSGFDEFGLGINMVFKRKNDGNQWYITDPKTLKTREDINRYMKGYNLTLTLLDEIEAESV  
 ISQQNKDLNSAWFKKIDREYRDNNKLNQWDKIRNLSQEEKNELNIQSVNDLVDQQLMTNRNPGNGIYKP  
 EASYNQSPYVGVMMTGIIYGGNTSKGAPGAVSFKHNAFRLWGYGYENGFLGYASNKYQQSKTDGE  
 SVLSDEYIIKKISNNTFNTIEEFKKAYFKEVKDKATKGLTTFEVNGSSVSSYDDLTLFKEAVKKDAET  
 LKQEANGNKTVSMNNTVKLKEAVYKKLLQQTNSFKTSIFK

**SP126 nucleotide (SEQ ID NO:223)**

TAAGACAGATGAACGGAGCAAGGTGTTTGACTTTTCCATTCCCTACTATACTGCAAAAAATAAACTCAT  
 TGTCAAAAAATCTGACTTGACTACTTATCAGTCTGTAAACGACTTGGCGCAGAAAAAGGTTGGAGCGCA  
 GAAAGGTTTCGATTCAAGAGACGATGGCGGAAAGATTTGCTACAAAATCTTCCCTCGTATCTCTGCTTAA  
 AAATGGGAATTTAATCACAGATTTAAAATCAGGACAAGTGGATGCCGTTATCTTTGAAGAACCCTGTTTC  
 CAAGGGATTTGTGGAAAAATAATCCTGATTTAGCAATCGCAGACCTCAATTTTGAAGAGCAAGATGA  
 TTCTTACGCGGTAGCCATgAAAAAGATAGCAAGAAATTGAAGAGGCAGTTCGATAAAACCATTCAAAA  
 GTTGAAGGAGTCTGGGGAATTAGACAAACTCATTGAGGAAGCCTTA

**SP126 amino acid (SEQ ID NO:224)**

KTDERSKVDFDSIPYYTAKNKLIVKKSDLTTYQSVNDLAQKKVGAQKGSIQETMAKDLLQNSSLVSLPK  
 NGNLITDLKSGQVDAVIFEEPVSCKGFVENNPDLAIADLNFEKEQDDSYAVAMKKDSKKLKRQFDKTIQK  
 LKESGELDKLISEAL

**SP127 nucleotide (SEQ ID NO:225)**

CTGTGAGAATCAAGCTACACCCAAAGAGACTAGCGCTCAAAAGACAATCGTCCTTGCTACAGCTGGCGA  
 CGTGCCACCATTGTGACTACGAAGACAAGGGCAATCTGACAGGCTTTGATATCGAAGTTTTAAAGGCAGT  
 AGATGAAAAACTCAGCGACTACGAGATTCAATTCCAAAGAACCGCCTGGGAGAGCATCTTCCCAGGACT  
 TGATTCTGGTCACTATCAGGCTGCGGCCAATAACTTGAGTTACACAAAAGAGCGTGCTGAAAAATACCT  
 TTACTCGCTTCCAATTTCCAACAATCCCCCTCGTCCTTGTCAGCAACAAGAAAAATCCTTTGACTTCTCT  
 TGACCAGATCGCTGGTAAAACAACACAAGAGGATACCGGAACCTCTAACGCTCAATTCATCAATAACTG  
 GAATCAGAAAACACACTGATAATCCCGCTACAATTAATTTTTCTGGTGAGGATATTGGTAAACGAATCCT  
 AGACCTTGCTAACGGAGAGTTTGATTTCTAGTTTTTGTACAAGGTATCCGTTCAAAGATTATCAAGGA  
 CCGTGGTTTTAGACCTCTCAGTCGTTGATTTACCTTCTGCAGATAGCCCCAGCAATTATATCATTTTCTC  
 AAGCGACCAAAAAGAGTTTTAAAGAGCAATTTGATAAAGCGCTCAAAGAACTCTATCAAGACGGAACCTT  
 TGAAAAACTCAGCAATACCTATCTAGGTGGTTCTTACCTCCCAGATCAATCTCAGTTACAA

**SP127 amino acid (SEQ ID NO:226)**

CENQATPKETSQAQKTIVLATAGDVPPFDYEDKGNLTGFDIEVLKAVDEKLSDYEIQFORTAWESIFPGL  
 DSGHYQAAANNLSYTKERAKEYLYSLPISNNPLVLVSNKKNPLTSLDQIAGKTTQEDTGTSNAQFINNW  
 NQKHTDNPATINFSGEDIGKRILDLANGEFDLFLVFDKVSQKIIKDRGLDLSVVDLPSADSPSNYIIFS  
 SDQKEFKEQFDKALKELYQDGTLEKLSNTYLGGSYLPDQSQLQ

Table 2  
*S. pneumoniae* Antigenic Epitopes

**SP001**

Lys-1 to Ile-10; Leu-13 to Lys-32; Arg-41 to Ile-51; Ser-85 to Glu-97; Ala-159 to His-168; Val-309 to Thr-318; Val-341 to Asn-352; Asn-415 to Met-430; Phe-454 to Asn-464; Ser-573 to Gly-591; Asn-597 to Thr-641; and Asn-644 to Ala-664.

**SP004**

Thr-9 to Thr-24; Ile-29 to Ala-48; Thr-49 to Val-56; Val-286 to Val-312; Pro-316 to Glu-344; Val-345 to Ile-367; Gln-368 to Val-399; Ser-400 to Glu-431; Asn-436 to Ala-457; Ile-467 to Ala-498; and Thr-499 to Glu-540.

**SP006**

Glu-1 to Lys-13; Pro-24 to Gly-36; Val-104 to Thr-112; Ala-118 to Asn-130; Trp-137 to Ala-146; Ser-151 to Ile-159; Ile-181 to Leu-188; and Pro-194 to Tyr-202.

**SP007**

Gly-1 to Asn-7; Tyr-24 to Gln-34; His-47 to Phe-55; Ser-60 to Ala-67; Ala-122 to Leu-129; Leu-221 to Lys-230; Val-236 to Phe-256; and Asp-271 to Gly-283; and Leu-291 to Asp-297.

**SP008**

Leu-4 to Lys-17; Gln-24 to Leu-32; Asp-60 to Ser-66; Ser-70 to Asp-76; Ala-276 to Lys-283; Asn-304 to Lys-311; and Thr-429 to Pro-437.

**SP009**

Thr-4 to Glu-11; Leu-50 to Asp-60; Ile-102 to Trp-123; and Ser-138 to Ile-157.

**SP010**

Phe-34 to Gly-41; Asp-44 to Lys-50; Leu-172 to Val-186; Leu-191 to Val-198; Ser-202 to Ile-209; and Val-213 to Leu-221.

**SP011**

Asn-2 to Thr-10; Asp-87 to Ala-102; Tyr-125 to Glu-132; Thr-181 to Tyr-189; Arg-217 to Thr-232; Asn-257 to Lys-264; Pro-271 to Ser-278; Tyr-317 to Ala-325; Glu-327 to Pro-337; and Thr-374 to Val-381.

**SP012**

Gly-1 to Lys-19; Phe-34 to Tyr-41; Leu-109 to Lys-126; and Leu-231 to Glu-247.

**SP013**

Ala-1 to Lys-12; Ile-42 to Pro-53; Leu-138 to Lys-146; Ile-205 to Lys-217; Ser-235 to Ile-251; and Ser-261 to Tyr-272.

**SP014**

Gly-1 to Val-16; Leu-35 to Leu-44; Asp-73 to Asp-81; Ile-83 to Asp-92; Glu-145 to Ile-153; Phe-188 to Asn-196; Ser-208 to Phe-215; Ile-224 to Leu-231; and Asn-235 to Ala-243.

**SP015**

Ser-1 to Pro-16; Asn-78 to Glu-88; Ala-100 to Val-108; Ala-122 to Thr-129; Thr-131 to Ser-137; Leu-201 to Ser-220; and Gly-242 to Val-251.

**Table 2**  
***S. pneumoniae* Antigenic Epitopes**

**SP016**

Gly-1 to Glu-20; Thr-30 to Val-38; Gln-94 to Asn-105; Lys-173 to Pro-182; Gly-189 to Arg-197; Ser-207 to Val-224; Pro-288 to Leu-298; Ala-327 to Ala-342; and Ser-391 to Ala-402.

**SP017**

Ser-1 to Thr-12; Ala-36 to Tyr-45; Gln-48 to Ile-54; Lys-59 to Lys-76; Tyr-113 to Leu-138; and Phe-212 to Asp-219.

**SP019**

Val-97 to Glu-117; Asp-163 to Leu-169; Thr-182 to Thr-191; and Lys-241 to Ser-250.

**SP020**

Asn-18 to Lys-25; Thr-47 to Glu-60; Trp-75 to Val-84; Gly-102 to Val-110; Pro-122 to Ala-131; and Glu-250 to Pro-258.

**SP021**

Ser-1 to Asp-8; Val-44 to Asp-54; Ala-117 to Val-125; Thr-165 to Thr-173; and Glu-180 to Pro-189.

**SP022**

Phe-5 to Lys-13; Thr-20 to Ser-36; Glu-59 to Lys-81; Tyr-85 to Gly-93; Trp-94 to Trp-101; and Thr-195 to Trp-208.

**SP023**

Gln-45 to Glu-59; Asp-69 to Pro-85; Lys-111 to Asn-121; Pro-218 to Ala-228; and Glu-250 to Asn-281.

**SP025**

Gln-14 to Thr-20; Gly-27 to Phe-33; Gly-63 to Glu-71; and Ile-93 to Phe-102.

**SP028**

Asp-171 to Pro-179; Tyr-340 to Glu-350; Pro-455 to Tyr-463; and Asp-474 to Pro-480.

**SP030**

Leu-22 to Leu-37; Trp-81 to Ala-90; Phe-101 to Ala-106; Thr-124 to Tyr-130; and Asn-138 to Glu-144.

**SP031**

Asp-8 to Val-16; Gly-27 to Thr-35; Gly-178 to Asp-195; Thr-200 to Asp-209; Trp-218 to Leu-224; and Lys-226 to Asp-241.

**SP032**

Ser-9 to Asp-28; Phe-31 to Val-40; Gly-42 to Arg-50; Ile-52 to Leu-60; Asp-174 to Phe-186; Leu-324 to Met-333; and Thr-340 to Asn-347.

**SP033**

Gln-2 to Ile-13; Phe-46 to Ile-53; and Asp-104 to Thr-121.

**SP034**

Glu-36 to Gly-43; Ala-188 to Asp-196; Trp-313 to Gly-320; and Leu-323 to Leu-329.

**Table 2**  
***S. pneumoniae* Antigenic Epitopes**

**SP035**

Arg-19 to Asp-36; Asp-47 to Val-57; Asn-134 to Thr-143; Asp-187 to Arg-196; and Glu-222 to Ser-230.

**SP036**

Arg-10 to Arg-17; Lys-29 to Ser-39; Ser-140 to Ala-153; Arg-158 to Tyr-169; Asp-175 to Ala-183; Gly-216 to Asn-236; Ala-261 to Leu-270; Arg-282 to Phe-291; and Thr-297 to Ala-305; Pro-342 to Gln-362; Phe-455 to Asp-463; His-497 to Thr-511; Ala-521 to Gly-529; Ile-537 to Val-546; Ile-556 to Ala-568; Pro-581 to Ser-595; Glu-670 to Ala-685; Ser-696 to Ala-705 and Leu-782 to Ser-791.

**SP038**

Glu-61 to Pro-69; Phe-107 to Ala-115; Leu-130 to Tyr-141; Ala-229 to Glu-237; Ser-282 to Asn-287; Ala-330 to Glu-338; and Tyr-387 to Glu-393.

**SP039**

Ser-28 to Asp-35; Pro-88 to Pro-96; Leu-125 to Arg-135; Phe-149 to Leu-157; Gln-246 to Val-254; Ala-357 to Thr-362; Gly-402 to Lys-411; and Leu-440 to Pro-448.

**SP040**

Thr-21 to Ile-30; His-54 to Gln-68; Arg-103 to Leu-117; and Thr-127 to Leu-136.

**SP041**

Gly-36 to Asp-49; Leu-121 to Val-128; and Ala-186 to Ile-196.

**SP042**

Gly-11 to Arg-19; Ile-23 to Lys-31; His-145 to Asn-151; Gln-159 to Asp-166; Ile-175 to Asp-181; Gly-213 to Tyr-225; Ile-283 to Val-291; Pro-329 to Glu-364; Arg-372 to Ser-386; Thr-421 to Phe-430; Leu-445 to Val-453; Ile-486 to Ala-497; Asp-524 to Ala-535; His-662 to Gly-674; and His-679 to Gln-702.

**SP043**

Lys-2 to Asp-12; Val-58 to Asn-68; Ser-87 to Asp-95; and Asp-102 to Lys-117.

**SP044**

Gln-3 to Lys-11; Asp-37 to Tyr-52; Glu-171 to Leu-191; His-234 to Asn-247; and Asn-283 to Ala-291.

**SP045**

Tyr-52 to Ile-63; Asp-212 to Gln-227; Ser-315 to Thr-332; Leu-345 to Phe-354; Asp-362 to Val-370; Thr-518 to Asn-539; Ala-545 to Lys-559; and Val-601 to Pro-610.

**SP046**

Gln-9 to Ala-18; Glu-179 to Lys-186; Lys-264 to Glu-271; Gly-304 to Glu-17; Ser-503 to Asn-511; Asn-546 to Thr-553; and Asn-584 to Asp-591.

**SP048**

**Table 2**  
***S. pneumoniae* Antigenic Epitopes**

Tyr-4 to Asp-25; Lys-33 to Val-70; Asp-151 to Thr-170; Asp-222 to Val-257; Thr-290 to Phe-301; and Gly-357 to Val-367.

**SP049**

Ala-23 to Arg-37; Tyr-85 to Gln-95; Glu-106 to Ile-118; Arg-131 to Ile-144; Gly-150 to Ser-162; and Ala-209 to Asp-218.

**SP050**

Asp-95 to Glu-113; Gly-220 to Gly-228; Asn-284 to Glu-295; Thr-298 to Val-315.

**SP051**

Lys-16 to Glu-50; Lys-57 to Asn-104; Ser-158 to Trp-173; Asp-265 to Pro-279; Val-368 to Tyr-386; Glu-420 to Ile-454; Pro-476 to Ile-516; Phe-561 to Gly-581; Thr-606 to Gly-664; and Glu-676 to Val-696.

**SP052**

Asn-41 to Tyr-60; Phe-80 to Glu-103; Ala-117 to Val-139; Ile-142 to Leu-155; Val-190 to Lys-212; Glu-276 to Phe-283; Arg-290 to Ser-299; Leu-328 to Val-351; Gly-358 to Thr-388; Glu-472 to Ala-483; Val-533 to Asn-561; Asp-595 to Val-606; Glu-609 to Val-620; Glu-672 to Ser-691.

**SP053**

Ala-62 to Val-101; Thr-147 to Leu-174; Lys-204 to Val-216; Gln-228 to Val-262; Ser-277 to Gly-297; Thr-341 to Gly-368; Thr-385 to Ala-409; Thr-414 to Ser-453; Asn-461 to Leu-490; Glu-576 to Thr-625; Gly-630 to Arg-639; and Asp-720 to Leu-740.

**SP054**

Glu-7 to Val-28; and Tyr-33 to Glu-44.

**SP055**

Pro-3 to Val-18; Thr-21 to Lys-53; Val-84 to Lys-99; Ile-162 to Val-172; and Val-204 to Ser-241.

**SP056**

Val-34 to Tyr-41; Leu-47 to Glu-55; and Pro-57 to Gln-66.

**SP057**

Asp-1 to Val-25; Pro-29 to Ile-80; Asn-96 to Val-145; and Pro-150 to Glu-172.

**SP058**

Ala-64 to Thr-70; Leu-82 to His-138; and Val-228 to Asn-236.

**SP059**

Val-10 to Thr-24; Ser-76 to Pro-102; Ser-109 to Ile-119; Ser-124 to Val-130; Thr-186 to Ile-194; and Asn-234 to Ser-243.

**SP060**

Leu-70 to Arg-76; and Val-79 to Ile-88.

**SP062**

Glu-14 to Lys-28; Ser-32 to Lys-46; and Glu-66 to Thr-74.

**Table 2**  
***S. pneumoniae* Antigenic Epitopes**

**SP063**

Ile-10 to Val-25; Val-30 to Thr-40; Asp-44 to Pro-54; Asn-57 to Val-63; Pro-71 to Val-100; and Thr-105 to Thr-116.

**SP064**

Pro-12 to Leu-32; Val-40 to Leu-68; Asp-95 to Ala-125; Ser-164 to Glu-184; Ser-314 to Glu-346; Asn-382 to Val-393; Leu-463 to Gln-498; Asn-534 to Lys-548; and Lys-557 to Gly-605.

**SP065**

Asn-2 to Ile-12; Ala-39 to Thr-61; and His-135 to Ala-155.

**SP067**

Gly-1 to Thr-13; Asp-203 to Asn-218; and Gly-240 to Asp-253.

**SP068**

Ser-2 to Ser-12; Val-17 to Gln-26; and Lys-54 to Cys-67.

**SP069**

Ser-32 to Thr-41; Pro-66 to Glu-80; Thr-110 to Val-122; and Val-147 to Thr-180.

**SP070**

Lys-6 to Tyr-16; Gln-19 to Ile-27; Arg-50 to Ala-58; Leu-112 to Val-128; Ile-151 to Asn-167; Leu-305 to Phe-321.

**SP071**

Gln-92 to Asn-158; Gln-171 to Gln-188; Val-204 to Val-240; Thr-247 to Ala-273; Glu-279 to Thr-338; Pro-345 to Glu-368; Asn-483 to Lys-539; Val-552 to Ala-568; Glu-575 to Ser-591; Ser-621 to Gly-640; Gln-742 to Gly-758.

**SP072**

Val-68 to Tyr-81; Tyr-86 to Val-121; Leu-127 to Gly-140; Gly-144 to Ala-155; Gln-168 to Val-185; Asp-210 to Thr-241; Glu-246 to Thr-269; Lys-275 to Tyr-295; Gly-303 to Pro-320; Arg-327 to Ile-335; Thr-338 to Thr-364; Tyr-478 to Phe-495; and Tyr-499 to Arg-521.

**SP073**

Glu-37 to Val-45; Glu-55 to Val-68; Thr-104 to Thr-119; Ile-127 to Tyr-135; Asn-220 to Ile-232; Thr-237 to Ala-250; Ser-253 to Ala-263; Glu-284 to Ile-297; and Met-438 to Asn-455.

**SP074**

Gly-2 to Ala-12; Gly-96 to Ile-110; and Thr-220 to Phe-239.

**SP075**

Phe-33 to Tyr-42; Gln-93 to Gly-102; and Val-196 to Asp-211.

**SP076**

Ser-64 to Leu-76; and Phe-81 to Ala-101.

**SP077**

Asp-1 to Glu-12; Tyr-26 to Val-36; and Val-51 to Thr-62.

**Table 2**  
***S. pneumoniae* Antigenic Epitopes**

**SP078**

Ala-193 to Ile-208; Tyr-266 to Asn-275; Glu-356 to Leu-369; Ala-411 to Gly-422; Ser-437 to Pro-464; Thr-492 to Glu-534; and Glu-571 to Gln-508.

**SP079**

Gly-11 to Leu-20; Lys-39 to Leu-48; Leu-72 to Val-85; Asn-147 to Ser-158; Ile-178 to Asp-187; Tyr-189 to Gln-201; and Leu-203 to Ala-216

**SP080**

Ser-2 to Glu-12; Gln-42 to Ala-51; Ala-116 to Ser-127; Phe-131 to Asp-143; and Ile-159 to Ile-171.

**SP081**

Gln-2 to Leu-9; Gln-49 to Cys-57; Ile-108 to Val-131; Gly-134 to Leu-145; and Trp-154 to Cys-162.

**SP082**

Ile-101 to Ser-187; Gly-191 to Asn-221; Arg-225 to Arg-236; Tyr-239 to Leu-255; and Gly-259 to Arg-268.

**SP083**

Ser-28 to Asp-70.

**SP084**

Leu-42 to Gln-66; Thr-69 to Lys-81; Glu-83 to Arg-92; and Gly-98 to Asn-110.

**SP085**

Gln-2 to Val-22; and Ser-45 to Glu-51.

**SP086**

Leu-18 to Gln-65; and Lys-72 to Val-83.

**SP087**

Ser-45 to Leu-53; and Thr-55 to Gln-63

**SP088**

Pro-8 to Ile-16; Leu-25 to Trp-33; Tyr-35 to Gln-43; Leu-51 to Val-59; Val-59 to Arg-67; Thr-55 to Tyr-63; Asn-85 to Gly-93; Thr-107 to Leu-115;  
Leu-115 to Trp-123; Ala-121 to Thr-129; Tyr-153 to Ala-161; His-176 to Gly-184; Tyr-194 to Ala-202; Ala-217 to Gly-225; and Asn-85 to Gly-93.

**SP089**

Trp-43 to Ala-51; Gln-68 to Phe-76; Val-93 to Gln-101; Phe-106 to Phe-114; Lys-117 to Lys-125; Trp-148 to Phe-156; Glu-168 to Gln-176; Ile-193 to Tyr-201; Lys-203 to Lys-211; Glu-212 to Gln-220; Ile-237 to Tyr-245; Lys-247 to Lys-255; Glu-256 to Gln-264; Met-275 to Gly-283; Lys-286 to Gly-294; Trp-292 to Glu-300; Asp-289 to Thr-297; Tyr-315 to Ser-323; Asp-334 to Lys-342; Pro-371 to Arg-379; Arg-485 to Asn-493; Lys-527 to Arg-535; Phe-537 to Met-545; and Tyr-549 to Glu-557.

**SP090**

**Table 2**  
***S. pneumoniae* Antigenic Epitopes**

Phe-2 to Gln-10; Gln-13 to Lys-21; Tyr-19 to Glu-27; Tyr-39 to Met-47; Pro-65 to Leu-73; Tyr-121 to His-129; Lys-147 to Ile-155; Gly-161 to Lys-169; Gly-218 to Trp-226; Asp-230 to Thr-238; Tyr-249 to Ala-257; and Ala-272 to Gly-280.

**SP091**

Ser-19 to Ser-27; Asn-25 to Thr-33; Val-51 to Gln-59; Asn-75 to Asn-83; Ile-103 to Trp-111; Tyr-113 to Ala-121; Leu-175 to Asn-183; Glu-185 to Trp-193; Ala-203 to Tyr-211; Val-250 to Phe-258; Asn-260 to Thr-268; Ser-278 to Asp-286; Tyr-305 to Leu-313; Asn-316 to Gly-324; Asn-374 to Asp-382; Asn-441 to Gly-449; and Ser-454 to Gln-462.

**SP092**

Arg-95 to Glu-103; Ala-216 to Val-224; Leu-338 to Glu-346; Pro-350 to Ala-358; Pro-359 to Ala-367; Pro-368 to Ala-376; Pro-377 to Ala-385; Pro-386 to Ala-394; Pro-395 to Ala-403; Pro-350 to Ala-358; Gln-414 to Lys-422; Pro-421 to Asn-429; Trp-465 to Tyr-473; Phe-487 to Tyr-495; Asn-517 to Gly-525; Trp-586 to Tyr-594; Phe-608 to Tyr-616; and Asp-630 to Gly-638.

**SP093**

Gln-30 to Ile-38; Gln-52 to Val-60; Ala-108 to His-116; Tyr-133 to Glu-141; Tyr-192 to Ala-200; and Phe-207 to Ser-215.

**SP094**

Ala-87 to Val-95; Leu-110 to Cys-118; Gln-133 to Leu-141; Ser-185 to Leu-193; Ile-195 to Gly-203; Asp-206 to Gln-214; Ser-211 to Gly-219; Ile-241 to Thr-249.

**SP095**

Arg-1 to Gln-9; Phe-7 to Asn-15; Thr-21 to Asn-30; Leu-46 to Phe-54; and Ser-72 to Met-80.

**SP096**

Gly-29 to Ile-37; Glu-52 to Ser-60; and Leu-64 to Gly-72.

**SP097**

Ala-11 to Thr-19; Glu-53 to Glu-61; Ser-91 to Lys-99; Thr-123 to Gln-131; and Gly-209 to Lys-217.

**SP098**

Thr-3 to Ser-11; Gly-38 to Phe-46; Tyr-175 to Asn-183; Met-187 to Cys-195; Gln-197 to Leu-205; Tyr-307 to Gln-315; Gly-318 to Tyr-326; Asn-348 to Val-356; Lys-377 to Pro-385; and Leu-415 to Val-423.

**SP099**

Arg-19 to Gly-27; Asp-76 to Ser-84; Val-90 to Lys-98; Phe-165 to Val-173; Leu-237 to Pro-245.

**SP100**

His-111 to Gln-119; Ser-141 to His-149; Asp-154 to Ser-162; Gln-158 to Gln-166; Asp-154 to Gln-166; Lys-180 to Gln-188; and Ser-206 to Gln-214.

**SP101**

**Table 2**  
***S. pneumoniae* Antigenic Epitopes**

Glu-23 to Glu-31; Glu-40 to Val-48; Gln-50 to Ser-58; Thr-61 to Ile-69; Leu-82 to Ile-90; Ala-108 to Leu-116; Gln-121 to Pro-129; and Leu-130 to Thr-138.

**SP102**

Asp-32 to His-40; Arg-48 to Lys-56; and Asp-102 to Thr-110.

**SP103**

Arg-5 to Gln-13; Gln-22 to Leu-30; Arg-151 to Gln-159; Arg-167 to Gln-175; Pro-189 to Glu-197; Gly-207 to Leu-215; Ser-219 to Gln-227; Ser-233 to Ser-241; Pro-255 to Asp-264; Lys-272 to Gly-280; Ser-318 to Val-326; Thr-341 to Asp-351; Asn-356 to Thr-364; Val-370 to Tyr-378; Ile-379 to Gln-387; and Met-435 to Tyr-443.

**SP105**

Asn-28 to Pro-36; Thr-77 to Phe-85; Arg-88 to Val-96; Gly-107 to Phe-115; Asp-169 to Asp-177; His-248 to Ser-256; and Ser-274 to Ala-282.

**SP106**

Val-10 to Thr-18; Ile-62 to Tyr-70; Ile-71 to Pro-79; Lys-86 to Gln-94; Lys-100 to Thr-108; Phe-132 to Leu-140; and Asp-145 to Arg-153.

**SP107**

Asp-33 to Val-41; and Arg-63 to Gln-71.

**SP108**

Lys-9 to Gln-17; Leu-44 to Ser-52; Ser-63 to Phe-71; Tyr-109 to Ser-117; Ile-183 to Ile-191; Pro-194 to Leu-202; Gly-257 to Gln-265; Ala-323 to Thr-331; and Leu-381 to Tyr-389.

**SP109**

Asn-2 to Gln-10; Ala-65 to Lys-73; Leu-76 to Glu-84; Thr-111 to Asp-119; Gln-116 to Tyr-124; Tyr-130 to Val-138; Asp-173 to Gly-181; Asp-196 to Ser-204; Asn-231 to Ser-239; Phe-252 to Ser-260; Phe-270 to Tyr-278; Val-291 to His-299; Asp-306 to Leu-314; and Pro-327 to Gly-335.

**SP110**

Ser-8 to Glu-16; Ile-37 to Val-45; Ala-107 to Val-115; and Gly-122 to Thr-130.

**SP111**

Asp-19 to Glu-28; Leu-43 to Ala-51; Asn-102 to Phe-110; Gln-133 to Ser-141; Phe-162 to Asp-170; Tyr-194 to Met-202; and Asp-273 to Ser-281.

Table 2  
*S. pneumoniae* Antigenic Epitopes

**SP112**

Asp-3 to Gln-11; Gly-21 to Ile-29; Ala-46 to Arg-54; Arg-98 to Arg-106; Thr-114 to Val-122; Gln-133 to Asn-141; and Leu-223 to Thr-231.

**SP113**

Asn-19 to Gly-27; Arg-54 to Ser-62; Val-69 to Gln-77; Ser-117 to Asn-125; Gly-164 to Leu-172; Tyr-193 to Ser-201; Cys-303 to Phe-311; His-315 to Ile-323; Arg-341 to Cys-349; Ile-347 to Ser-355; Arg-403 to Phe-411; Gln-484 to Pro-492; Ser-499 to Leu-507; Ile-541 to Thr-549  
Asn-622 to Ile-630; and Glu-645 to Gly-653.

**SP114**

Gly-17 to Leu-25; His-40 to Gln-48; Arg-49 to Arg-57; Ile-65 to Pro-73;  
Asn-101 to Asp-111; Gly-128 to Cys-136; Phe-183 to Thr-191; and Pro-268 to Ile-276.

**SP115**

Met-8 to Ser-16; Tyr-24 to Leu-32; Cys-68 to Leu-76; Ser-100 to Pro-108; Thr-193 to Thr-201; Gly-238 to Pro-250; Thr-280 to Phe-288; Pro-303 to Asn-312; Trp-319 to Leu-328; Leu-335 to Leu-344; Lys-395 to Ala-403; Asn-416 to Gln-424; Tyr-430 to Ser-438; Val-448 to Leu-456; Leu-460 to Thr-468; Pro-502 to Thr-510; Lys-515 to Ile-524; Gln-523 to His-532; Tyr-535 to Thr-543; Ser-559 to Pro-567; Thr-572 to Asn-580;  
Val-594 to Arg-602; Arg-603 to Asn-611; Thr-620 to Trp-628; and Tyr-644 to Arg-653.

**SP117**

Ala-6 to Gly-14; Ile-19 to Thr-27; Thr-99 to Leu-107; Ser-117 to Asp-125; His-131 to Val-139; Ile-193 to Gly-201; and Val-241 to Gln-249.

**SP118**

Ser-8 to Trp-23; His-46 to Ala-54; Asn-93 to Gly-101; Val-100 to Ser-108; Arg-155 to Asp-163; and His-192 to Leu-200.

**SP119**

Tyr-46 to Lys-54; Ser-93 to Ser-101; Trp-108 to Asn-116; Val-121 to Glu-129; and Tyr-131 to Gln-139.

**SP120**

Ala-57 to Lys-65; Leu-68 to Glu-76; Thr-103 to Tyr-116; Tyr-122 to Val-130; His-163 to Gly-173; Asp-188 to Ser-196; Ser-222 to Ser-231; Phe-244 to Ser-252; Pro-262 to Tyr-270; Val-283 to His-291; and Asp-298 to Leu-306.

**SP121**

Ser-3 to Ala-11; Asp-13 to Leu-21; Ser-36 to Val-44; and Gln-136 to Met-144.

**SP122**

Asn-28 to Lys-36; Glu-39 to Thr-50; Val-54 to Lys-62; Asn-106 to Leu-114; Phe-159 to Gly-167; Asn-172 to Arg-180; Glu-199 to Asn-207;

**Table 2**  
***S. pneumoniae* Antigenic Epitopes**

Lys-230 to His-241; Asn-252 to Gly-263; Met-278 to Ala-287; Thr-346 to Asp-354; Lys-362 to Thr-370; Asp-392 to Asn-405; Asp-411 to Ala-424; Gly-434 to Gly-443; Tyr-484 to Glu-492; Ile-511 to Leu-519; Asn-524 to Asp-538; Glu-552 to Ile-567; Val-605 to Lys-613; Phe-697 to Ala-705; Phe-722 to Leu-730; Leu-753 to Leu-761; Asp-787 to Gln-795; Leu-858 to Asn-866; Ala-892 to Thr-901; Gly-903 to Ile-913; Ile-921 to Asn-931; Asn-938 to Pro-951; Gly-960 to Lys-970; Leu-977 to Asp-985; and Leu-988 to Pro-996.

**SP123**

Val-4 to Asn-12; Glu-47 to Leu-55; Lys-89 to Glu-100; Ser-165 to Thr-173; Lys-234 to Val-242; Ser-258 to Ser-266; Glu-284 to Asn-292; Tyr-327 to Leu-335; Tyr-457 to Thr-465; Tyr-493 to Glu-501; Thr-506 to Tyr-514; Lys-517 to Thr-525; Asn-532 to Gly-540; and Arg-556 to Glu-564.

**SP124**

Arg-16 to Glu-24; Gln-52 to Arg-60; Asn-69 to Tyr-77; Glu-121 to Asn-129; Ala-134 to Val-142; Thr-151 to Ala-159; Asn-164 to Glu-172; His-181 to His-189; Thr-210 to Ala-218; Ser-244 to Val-252; Phe-287 to Tyr-297; Ser-312 to Thr-323; His-433 to Tyr-441; Ser-445 to Asn-453; Asn-469 to Thr-477; Asn-501 to Asn-509; Gln-536 to Ala-547; and Gln-608 to Asp-621.

**SP125**

Ser-9 to Asp-21; Ala-28 to Leu-36; Asn-49 to Phe-57; Val-137 to Arg-145; Asn-155 to Leu-163; Glu-183 to Asp-191; Gly-202 to Tyr-210; Pro-221 to Asp-229; Phe-263 to Ala-271; Phe-300 to Gln-308; Asp-313 to Glu-321; Asn-324 to Asp-332; Ile-346 to Asn-354; Asp-362 to Lys-370; Met-402 to Gly-410; Gly-437 to Gly-445; Ser-471 to Glu-483; Gly-529 to Asp-537; Gln-555 to Val-563; and Leu-579 to Lys-587.

**SP126**

Leu-22 to Thr-30; Val-65 to Leu-73; and Thr-75 to Asp-83.

**SP127**

Glu-2 to Ala-12; Asp-28 to Thr-36; Val-105 to Thr-113; Lys-121 to Thr-129; Trp-138 to Pro-146; Ser-152 to Ile-160; Lys-180 to Asp-188; Leu-194 to Asn-202; and Gly-228 to Thr-236.

Table 3  
*S. pneumoniae* ORF Cloning Primers

| Primer |        |                                               |          |
|--------|--------|-----------------------------------------------|----------|
| Name   | SEQ ID | Sequence                                      | RE       |
| SP001A | NO:227 | GACTGGATCCTAAAATCTACGACAATAAAAATC             | Bam HI   |
| SP001B | NO:228 | CTGAGTCGACTGGTTGTGCTGGTTGAG                   | Sal I    |
| SP004A | NO:229 | GTCAGGATCCAAATTACAATACGGACTATG                | Bam HI   |
| SP004B | NO:230 | CAGTGTGCGACTAACTCTAGGTCGGAAAC                 | Sal I    |
| SP006A | NO:231 | GACTGGATCCTGAGAATCAAGCTACACCCAAAGAG           | Bam HI   |
| SP006B | NO:232 | AGTCAAGCTTTTGTAACTGAGATTGATCTGG               | Hind III |
| SP007A | NO:233 | GACTGGATCCTGGTAACCGCTCTTCTCGTAACGCAGC         | Bam HI   |
| SP007B | NO:234 | AGTCAAGCTTTTTCAGGAAGCTTTTACGCTTCC             | Hind III |
| SP008A | NO:235 | AGTCAGATCTTGTGGAAATTTGACAGGTAACAGCAAAAAAGCTGC | Bgl II   |
| SP008B | NO:236 | ACTGAAGCTTTTTGTTTTTCAAGAATTCATCG              | Hind III |
| SP009A | NO:237 | GACTGGATCCTGGTCAAGGAAGCTTCTAAAGAC             | Bam HI   |
| SP009B | NO:238 | AGTCAAGCTTTCACAAATTCGTTGGTGAAGCC              | Hind III |
| SP010A | NO:239 | GACTGGATCCTAGCTCAGGTGGAAACGCTGGTTCATCC        | Bam HI   |
| SP010B | NO:240 | AGTCAAGCTTATCAACTTTTCCACCTTCAACAACC           | Hind III |
| SP011A | NO:241 | GTCAAGATCTCTCCAAGTATGGTAAATCTGCGGATGG         | Bgl II   |
| SP011B | NO:242 | AGTCCTGCAGATCCACATCCGCTTTCATCGGGTTAAAGAAGG    | Pst I    |
| SP012A | NO:243 | GACTGGATCCTGGGAAAAATTTCTAGCGAACTAGTGG         | Bam HI   |
| SP012B | NO:244 | GTCAGTGCAGCTGTCTCTTTTACTTCTTTGGTTGC           | Pst I    |
| SP013A | NO:245 | GACTGGATCCTGCTAGCGGAAAAAAGATACAACCTTCTGG      | Bam HI   |
| SP013B | NO:246 | CTGAAAGCTTTTTTGCCAATCCTTCAGCAATCTTGTC         | Hind III |
| SP014A | NO:247 | GACTAGATCTTGGCTCAAAAAATACAGCTTCAAGTCC         | Bgl II   |
| SP014B | NO:248 | AGTCCTGCAGGTTTTTGTGTTGCTTGGTATTGGTTCG         | Pst I    |
| SP015A | NO:249 | GACTGGATCCTAGTACAACTCAAGCACTAGTCAGACAGAG      | Bam HI   |
| SP015B | NO:250 | CAGTCTGCAGTTTCAAAGCTTTTTGTATGTCTTC            | Pst I    |
| SP016A | NO:251 | GACTGGATCCTGGCAATTCTGGCGGAAGTAAAGATGC         | Bam HI   |
| SP016B | NO:252 | AGTCAAGCTTGTTTCATAGCTTTTTTGATTGTTTCG          | Hind III |
| SP017A | NO:253 | GACTGGATCCTTCACAAGAAAAACAAAAATGAAGATGG        | Bam HI   |
| SP017B | NO:254 | AGTCAAGCTTATCGACGTAGTCTCCGCCTTC               | Hind III |
| SP019A | NO:255 | GACTGGATCCGAAAGGTCTGTGGTCAAATAATCTTACC        | Bam HI   |
| SP019B | NO:256 | AGTCAAGCTTAGAGTTAATCATGGTGTCTTGCCAATAGG       | Hind III |
| SP020A | NO:257 | GACTGGATCCAACTCAGAAAAGAAAGCAGACAATGC          | Bam HI   |
| SP020B | NO:258 | AGTCAAGCTTCCAACTGGTTGATCCAAACCATCTG           | Hind III |
| SP021A | NO:259 | GACTGGATCCTTCGAAAGGGTCAGAAGGTGCAGACC          | Bam HI   |
| SP021B | NO:260 | AGTCAAGCTTCTGTAGGCTTGGTGTGCCCCAGTTGC          | Hind III |
| SP022A | NO:261 | CTGAGGATCCGGGGATGGCAGCTTTTAAAAATC             | Bam HI   |
| SP022B | NO:262 | CAGTAAGCTTGTTTACCCATTACCATTTACC               | Hind III |
| SP023A | NO:263 | CAGTGGATCCAGACGAGCAAAAAATTAAG                 | Bam HI   |
| SP023B | NO:264 | TCAGAAGCTTGTTTACCCATTACCATTT                  | Hind III |
| SP025A | NO:265 | GACTGGATCCCTGTGGTGAGGAAGAACTAAAAAG            | Bam HI   |
| SP025B | NO:266 | CTGAGTCGACAATATTCTGTAGGAATGCTTCGAATTTG        | Sal I    |
| SP028A | NO:267 | CTGAGGATCCGACTTTTAACAATAAACTATTGAAGAG         | Bam HI   |
| SP028B | NO:268 | GTCAGTGCAGGTTGTACCTCCAAAAATCACGG              | Pst I    |
| SP030A | NO:269 | GACTGGATCCCTTTACAGGTAAACAACCTACAAGTCGG        | Bam HI   |
| SP030B | NO:270 | CAGTAAGCTTTTTCGAAGTTTGGCTCAGAATTG             | Hind III |
| SP031A | NO:271 | GACTGGATCCCCAGGCTGATACAAGTATCGCA              | Bam HI   |
| SP031B | NO:272 | CAGTAAGCTTATCTGCAGTATGGCTAGATGG               | Hind III |
| SP032A | NO:273 | GACTGGATCCGTCTGTATCATTTGAAAACAAAGAAAC         | Bam HI   |
| SP032B | NO:274 | CAGTCTGCAGTTTTACTGTTGCTGTGCTTGTG              | Pst I    |
| SP033A | NO:275 | ACTGAGATCTTGGTCAAAGGAAAGTCAGACAGGAAAGG        | Bgl II   |
| SP033B | NO:276 | CAGTAAGCTTATTCCTGAGCTTTTTTGATAAAGGTTGCGCA     | Hind III |
| SP034A | NO:277 | ACTGGGATCCGAAGGATAGATATATTTTAGCATTTGAGAC      | Bam HI   |
| SP034B | NO:278 | AGTCAAGCTTCCATGGTATCAAAGGCAAGACTTGG           | Hind III |
| SP035A | NO:279 | GTCAGGATCCGGTAGTTAAAGTTGGTATTAAACGG           | Bam HI   |
| SP035B | NO:280 | AGTCAAGCTTGCAATTTTTGCGAAGTATTCCAAGAG          | Hind III |
| SP036A | NO:281 | AGTCGGATCCTTCTTACGAGTTGGGACTGTATCAAGC         | Bam HI   |

**Table 3**  
***S. pneumoniae* ORF Cloning Primers**

| <b>Primer</b> |               |                                             |           |
|---------------|---------------|---------------------------------------------|-----------|
| <b>Name</b>   | <b>SEQ ID</b> | <b>Sequence</b>                             | <b>RE</b> |
| SP036B        | NO:282        | AGTCAAGCTTGTTTATTTTTCCTTACTTACAGATGAAGG     | Hind III  |
| SP038A        | NO:283        | AGTCGGATCCTACTGAGATGCATCATAATCTAGGAGC       | Bam HI    |
| SP038B        | NO:284        | TCAGCTCGAGTTCTTTGACATCTCCATCATAAGTCGC       | Xho I     |
| SP039A        | NO:285        | GACTGGATCCGGTTTTTGAGAAAGTATTTGCAGGGG        | Bam HI    |
| SP039B        | NO:286        | CAGTAAGCTTGGATTTTTTCATGGATGCAATTTTTTTTGG    | Hind III  |
| SP040A        | NO:287        | GACTGGATCCGACAACATTTACTATCCATACAGTAGAGTCAGC | Bam HI    |
| SP040B        | NO:288        | GACTAAGCTTGGCATAAGGTTGCAATTCTGGATTAATTGG    | Hind III  |
| SP041A        | NO:289        | GACTGGATCCGGCTAAGGAAAGAGTGGATG              | Bam HI    |
| SP041B        | NO:290        | GACTAAGCTTTTCATTTTTTAAATTGACTATGCGCCCG      | Hind III  |
| SP042A        | NO:291        | GACTGGATCCTTGTTCCCTATGAACTTGGTCGTCACC       | Bam HI    |
| SP042B        | NO:292        | CATGAAGCTTATCCTGGATTTTTTCCAAGTAAATCT        | Hind III  |
| SP043A        | NO:293        | GACTGGATCCTTATAAGGGTGAATTAGAAAAAGG          | Bam HI    |
| SP043B        | NO:294        | GACTAAGCTTCTTATTAGGATTGTTAGTAGTTG           | Hind III  |
| SP044A        | NO:295        | GACTGGATCCGAATGTTCAAGGCTCAAGAAAGTTCAGG      | Bam HI    |
| SP044B        | NO:296        | GACTAAGCTTTTCCCCTGATGGAGCAAAGTAATACC        | Hind III  |
| SP045A        | NO:297        | GACTGGATCCCTTGGGTGTAACCCATATCCAGCTCCTTCC    | Bam HI    |
| SP045B        | NO:298        | GACTGTCGACTTCAGCTTGTATCTGGGGTTGC            | Sal I     |
| SP046A        | NO:299        | GACTGGATCCTAGTGATGCTACTTGGCAAGGAAAACAG      | Bam HI    |
| SP046B        | NO:300        | ACTGCTGCAGATCTTTGCCACCTAGCTTCTCATTG         | Pst I     |
| SP048A        | NO:301        | GTCAGGATCCTGGGATTCAATATGTCAGAGATGATACTAG    | Bam HI    |
| SP048B        | NO:302        | CTAGAAGCTTACGCACCCATTCACCATTATCATTG         | Hind III  |
| SP049A        | NO:303        | GTCAGGATCCGGATAATAGAGAAGCATTAAAAACC         | Bam HI    |
| SP049B        | NO:304        | AGTCAAGCTTGACAAAATCTTGAAACTCCTCTGGTC        | Hind III  |
| SP050A        | NO:305        | GTCAGGATCCAGATTTTGTGCGAGGAGTGTCATACC        | Bam HI    |
| SP050B        | NO:306        | AGTCAAGCTTTCCCTTTTACCCTTACGAATCCAGG         | Hind III  |
| SP051A        | NO:307        | GACTGGATCCATCTGTAGTTTATGCGGATGAAACACTTATTAC | Bam HI    |
| SP051B        | NO:308        | GACTGTCGACGCTTTGGTAGAGATAGAAGTCATG          | Sal I     |
| SP052A        | NO:309        | GACTGGATCCTTACTTTGGTATCGTAGATACAGCCGGC      | Bam HI    |
| SP052B        | NO:310        | AGTCAAGCTTTGTAAATTGCGTACCTTCTAAGCGACC       | Hind III  |
| SP053A        | NO:311        | GACTGGATCCAGCTAAGGTTGCATGGGATGCGATTCCG      | Bam HI    |
| SP053B        | NO:312        | GACTGTCGACCTGGGCTTTATTAGTTTGACTAGC          | Sal I     |
| SP054A        | NO:313        | CAGTGGATCCCTATCAGTATGTAAATAAAAGAGA          | Bam HI    |
| SP054B        | NO:314        | ACTGAAGCTTTTCTGTCCCTGTTTGAGGCA              | Hind III  |
| SP055A        | NO:315        | CAGTGGATCCTGAGACTCCTCAATCAATAACAAA          | Bam HI    |
| SP055B        | NO:316        | ACGTAAGCTTATAATCAGTAGGAGAAACTGAACT          | Hind III  |
| SP056A        | NO:317        | CAGTGGATCCGGATGCTCAAGAAACTGCGG              | Bam HI    |
| SP056B        | NO:318        | GACTAAGCTTTTGCCTCTCATTCTTGCTTCC             | Hind III  |
| SP057A        | NO:319        | CAGTGGATCCCGACAAAGGTGAGACTGAG               | Bam HI    |
| SP057B        | NO:320        | ACGTAAGCTTATTTCTTAATTCAAGTGTTTCTCTG         | Hind III  |
| SP058A        | NO:321        | GACTGGATCCAAATCAATTGGTAGCACAAGATCC          | Bam HI    |
| SP058B        | NO:322        | CAGTGTGACATTAGGAGCCACTGGTCTC                | Sal I     |
| SP059A        | NO:323        | CAGTGGATCCCAAACAGTCAGCTTCAGGAAC             | Bam HI    |
| SP059B        | NO:324        | GACTCTGCAGTTTAATCTTGTCCTCCAGGTGG            | Pst I     |
| SP060A        | NO:325        | GACTGGATCCATTTCGATGATGCGGATGAAAAG           | Bam HI    |
| SP060B        | NO:326        | GACTAAGCTTCATTTGTCTTTGGGTATTTTCGCA          | Hind III  |
| SP062A        | NO:327        | CAGTGGATCCGGAGAGTCCGATCAAAAAGTAG            | Bam HI    |
| SP062B        | NO:328        | GTCAGTGCAGTTGCTCGTCTCGAGGTTG                | Pst I     |
| SP063A        | NO:329        | CAGTGGATCCATGGACAACAGGAAACTGGGAC            | Bam HI    |
| SP063B        | NO:330        | CAGTAAGCTTATTAGCTTCTGTACCTGTGTTTG           | Hind III  |
| SP064A        | NO:331        | GACTGGATCCCGATGGGCTCAATCCAACCCCAAGTCAAGTC   | Bam HI    |
| SP064B        | NO:332        | GACTCTGCAGCATAGCTTTATCCTCTGACATCATCGTATC    | Pst I     |
| SP065A        | NO:333        | GACTGGATCCTTCCAATCAAAAACAGGCAGATGG          | Bam HI    |
| SP065B        | NO:334        | GACTAAGCTTGAGTCCCATAGTCCAAGGCA              | Hind III  |
| SP067A        | NO:335        | AGTCGGATCCTATCACAGGATCGAACGGTAAGACAACC      | Bam HI    |
| SP067B        | NO:336        | ACTGGTCGACTTCTTTAACTCCGCTACTGTGTC           | Sal I     |

Table 3  
*S. pneumoniae* ORF Cloning Primers

| Primer |        |                                                 |          |
|--------|--------|-------------------------------------------------|----------|
| Name   | SEQ ID | Sequence                                        | RE       |
| SP068A | NO:337 | CAGTGGATCCAAGTTCATCGAAGATGGTTGGGAAGTCC          | Bam HI   |
| SP068B | NO:338 | GATCGTCGACCCGCTCCCATGCTCAACCTT                  | Sal I    |
| SP069A | NO:339 | TGACGGATCCATCGCTAGCTAGTGAAATGCAAGAAAG           | Bam HI   |
| SP069B | NO:340 | TGACAAGCTTATTCTGTTTTTGAAGTAGTTGCTTTTCGT         | Hind III |
| SP070A | NO:341 | GACTGGATCCGCACCAGATGGGGCACAAGGTTCAAGG           | Bam HI   |
| SP070B | NO:342 | TGACAAGCTTAACTTGTAAACGAACAGTTCAATCTG            | Hind III |
| SP071A | NO:343 | GACTAGATCTTTTTTAACCCAACTGTTGGTACTTTTCC          | Bgl II   |
| SP071B | NO:344 | TGACAAGCTTGTTAGGTGTTACATTTTGACCGTC              | Hind III |
| SP072A | NO:345 | ACTGAGATCTTTTTTAACCCAACTGTTGGTACTTTTCC          | Bgl II   |
| SP072B | NO:346 | GACTAAGCTTTCTACGATAACGATCATTTTCTTTACC           | Hind III |
| SP073A | NO:347 | GACTGTCGACTCGTAGATATTTAAGTCTAAGTGAAGCG          | Sal I    |
| SP073B | NO:348 | AGTCAAGCTTGTTAGGTGTTACATTTTGCAAGTC              | Hind III |
| SP074A | NO:349 | GACTGGATCCCTTTGGTTTTGAAGGAAGTAAG                | Bam HI   |
| SP074B | NO:350 | TGACCTGCAGACGATTTTGAAGAAATGGAGGTGTATC           | Pst I    |
| SP075A | NO:351 | CAGTGGATCCCTACTACCTCTCGAGAGAAAG                 | Bam HI   |
| SP075B | NO:352 | ACTGAAGCTTTTCGCTTTTACTCGTTTGACA                 | Hind III |
| SP076A | NO:353 | CAGTGGATCCCTAAGGTCAAAGTCAAGCCGCTAAGAAAGTGC      | Bam HI   |
| SP076B | NO:354 | CAGTAAGCTTTAGGGTATCCAAATACTGGTTGTTGATG          | Hind III |
| SP077A | NO:355 | TGACAGATCTTGACGGGTCTCAGGATCAGACTCAGG            | Bgl II   |
| SP077B | NO:356 | TGACAAGCTTCAAAGACATCCACCTCTTGACCTTTG            | Hind III |
| SP078A | NO:357 | GACTGGATCCCTAGAGGCTTTGCCAAATGGTGGGAAGGG         | Bam HI   |
| SP078B | NO:358 | GTCAGTCGACTTGTGTAAACACTTTTCGAGGTTTGGTACC        | Sal I    |
| SP079A | NO:359 | CAGTGGATCCCTCAAAAAGAGAAGGAAAACCTTG              | Bam HI   |
| SP079B | NO:360 | CAGTCTGCAGTTTCTTCAACAAACCTTGTTCCTG              | Pst I    |
| SP080A | NO:361 | CAGTGGATCCACGTTCTATTGAGGACCACTT                 | Bam HI   |
| SP080B | NO:362 | CAGTAAGCTTTTCCTTCTCAGTCAATTCTTTTCC              | Hind III |
| SP081A | NO:363 | GACTGGATCCCGCTCAAAATACCAAGGTTGTTTCCAG           | Bam HI   |
| SP081B | NO:364 | GACTAAGCTTAGTACCATGGGTGTGACAGGTTTGAA            | Hind III |
| SP082A | NO:365 | CTGAGGATCCAAATTGTACAATTGAAAAAGATAGC             | Bam HI   |
| SP082B | NO:366 | TGACAAGCTTGCCTTGCTAGGTTCTGCAATGCC               | Hind III |
| SP083A | NO:367 | GACTGGATCCCTCTGACCAAGCAAAAAGAAGCAGTCAATGA       | Bam HI   |
| SP083B | NO:368 | TCAGCAGCTGATCATTTGACTTTACGATTTGCTCC             | Bgl II   |
| SP084A | NO:369 | GACTGGATCCGTCGGCTCTGTCCAGTCCACTTTTTCAGCG        | Bam HI   |
| SP084B | NO:370 | TCAGAAGCTTATTTTTTGTTCCTTAATGCGTT                | Hind III |
| SP085A | NO:371 | GACTGGATCCGGGACAAATTCAAAAAAATAGGCAAGAGG         | Bam HI   |
| SP085B | NO:372 | GTCAAAGCTTTGGCTCTTTGATTGCCAACAACCTG             | Hind III |
| SP086A | NO:373 | GACTGGATCCCTCGCTACCAGCAACAAGCGAGCAAAAGG         | Bam HI   |
| SP086B | NO:374 | GACTAAGCTTACTTTTTTCTTTTCCACACGA                 | Hind III |
| SP087A | NO:375 | CAGTGGATCCGAACCGACAAGTCGCCCCACTATCAAGACT        | Bam HI   |
| SP087B | NO:376 | CTGAAAGCTTTGAATTCTCTTTCTTTTCAGGCT               | Hind III |
| SP088A | NO:377 | TCGAGGATCCGGTTGTGCGCTGGCAATATATCCCGT            | Bam HI   |
| SP088B | NO:378 | CAGTAAGCTTCCGAACCCATTGCGCCATTATAGTTGAC          | Hind III |
| SP089A | NO:379 | AGTCGGATCCGGCCAAATCAGAATGGGTAGAAGAC             | Bam HI   |
| SP089B | NO:380 | TGACCTGCAGCTTCTCATTTGATTTTCATCATCAC             | Pst I    |
| SP090A | NO:381 | GACTGGATCCATTTGCGAGATGATTCTGAAGGATGG            | Bam HI   |
| SP090B | NO:382 | TCAGCTGCAGCTTAAACCATTCACCATTTCTAGTTTAAAG        | Pst I    |
| SP091A | NO:383 | GACTGGATCCCTGTCGCTGCAAAATGAAACTGAAGTAGC         | Bam HI   |
| SP091B | NO:384 | GACTAAGCTTATACCAAACGCTGACATCTACGCG              | Hind III |
| SP092A | NO:385 | AGTCAGATCTTACGTCTCAGCCTACTTTTGTAAGAGC           | Bgl II   |
| SP092B | NO:386 | GACTAAGCTTAAACCATTCACCATTTGGCATTGAC             | Hind III |
| SP093A | NO:387 | CAGTGGATCCCTGGACAGGTGAAAGGTCATGCTACATTTGTG      | Bam HI   |
| SP093B | NO:388 | GACTAAGCTTCAACCATTTGAGACCTTGCAACAC              | Hind III |
| SP094A | NO:389 | GTCAGGATCCGATTGCTCCTTTGAAGGATTTGAGAGAAACC       | Bam HI   |
| SP094B | NO:390 | GACTAAGCTTCGATCAAAGATAAGATAAATATATATAAAGT       | Hind III |
| SP095A | NO:391 | GACTGGATCCCTAGGTCATATGGGACTTTTTTTCTACAACAAATAGG | Bam HI   |

**Table 3**  
***S. pneumoniae* ORF Cloning Primers**

| <b>Primer</b> |               |                                            |           |
|---------------|---------------|--------------------------------------------|-----------|
| <b>Name</b>   | <b>SEQ ID</b> | <b>Sequence</b>                            | <b>RE</b> |
| SP095B        | NO:392        | TGACAAGCTTATCTATCAGCTCATTTAATCGTTTTTG      | Hind III  |
| SP096A        | NO:393        | CTGAGGATCCCAACGTTGAGAATTATTTGCGAATG        | Bam HI    |
| SP096B        | NO:394        | TGACAAGCTTGAGTCTACAAAAGTAATGTAC            | Hind III  |
| SP097A        | NO:395        | GTCAGGATCCCTACTATCAATCAAGTTCTTCAGCC        | Bam HI    |
| SP097B        | NO:396        | TGACAAGCTTGACTGAGGCTTGGACCAGATTGAAAAG      | Hind III  |
| SP098A        | NO:397        | GACTGGATCCGACAAAAACATTAAACGTCCTGAGG        | Bam HI    |
| SP098B        | NO:398        | GACTAAGCTTAGCACGAACTGTGACGCTGGTTCC         | Hind III  |
| SP099A        | NO:399        | GACTGGATCCTTCTCAGGAGACCTTTAAAAATATC        | Bam HI    |
| SP099B        | NO:400        | GACTAAGCTTGTTGGCCATCTTGTACATACC            | Hind III  |
| SP100A        | NO:401        | GACTGGATCCAGTAAATGCGCAATCAAATTC            | Bam HI    |
| SP100B        | NO:402        | AGTCCTGCAGGTATTTAGCCCAATAATCTATAAAGCT      | Pst I     |
| SP101A        | NO:403        | CAGTGGATCCTTACCGCGTTCATCAAGATGTC           | Bam HI    |
| SP101B        | NO:404        | GACTAAGCTTGCCAGATGTTGAAAAGAGAGTG           | Hind III  |
| SP102A        | NO:405        | GACTGGATCCGTGGATGGGCTTTAACTATCTTCGTATTCG   | Bam HI    |
| SP102B        | NO:406        | AGTCAAGCTTGCTAGTCTTCACTTTCCCTTTCC          | Hind III  |
| SP103A        | NO:407        | GACTGTCGACACTAAACCAGCATCGTTTCGCAGGA        | Sal I     |
| SP103B        | NO:408        | CTGACTGCAGCTTCTTGAAGAAATAATGATTGTGG        | Pst I     |
| SP105A        | NO:409        | CAGTGGATCCTGACTACCTTGAAATCCCCTT            | Bam HI    |
| SP105B        | NO:410        | CAGTAAGCTTTTTTTTAAAGTTGTAGAATGATTTCAATC    | Hind III  |
| SP106A        | NO:411        | CAGTGTGACTCGTATCTTTTTTTGGAGCAATGTT         | Sal I     |
| SP106B        | NO:412        | GACTAAGCTTAAATGTTCCGATACGGGTGATTG          | Hind III  |
| SP107A        | NO:413        | CAGTGGATCCGGACTCTCTCAAAGATGTGAAAG          | Bam HI    |
| SP107B        | NO:414        | GACTAAGCTTCTTGAGTTTGTCAAGGATTGCTTT         | Hind III  |
| SP108A        | NO:415        | CAGTGGATCCCAAGAAATCCTATCATCTCTCCAGAAG      | Bam HI    |
| SP108B        | NO:416        | GACTAAGCTTTTCAGAACTAAAAGCCGCAGCTT          | Hind III  |
| SP109A        | NO:417        | GACTGGATCCACGAAATGCAGGGCAGACAG             | Bam HI    |
| SP109B        | NO:418        | CAGTAAGCTTATCAACATAATCTAGTAAATAAGCGT       | Hind III  |
| SP110A        | NO:419        | CAGTGGATCCTGTATAGTTTTTAGCGCTTGTTCTTC       | Bam HI    |
| SP110B        | NO:420        | GTCAAAGCTTTGATAGAGTGTACATAATCTTCTTTAG      | Hind III  |
| SP111A        | NO:421        | GACTGGATCCGTGTGTCGAGCATATTCTGAAG           | Bam HI    |
| SP111B        | NO:422        | CAGTAAGCTTACTTTTACCATTCTTTGTTCTGCATC       | Hind III  |
| SP112A        | NO:423        | GACTGTCGACGTGTTTGGATAGCATTGAGAATCAGACG     | Sal I     |
| SP112B        | NO:424        | CAGTAAGCTTCGGAAGTAAAGACAATTTTTTCC          | Hind III  |
| SP113A        | NO:425        | CAGTGGATCCGTGCCTAGATAGTATTATTACTCAAAC      | Bam HI    |
| SP113B        | NO:426        | GACTAAGCTTTTTTGCTTATTTCTCTCAATTTTTTC       | Hind III  |
| SP114A        | NO:427        | CAGTGGATCCCATTGAGAAGCAGACCTATCAAATC        | Bam HI    |
| SP114B        | NO:428        | ACTGAAGCTTATGTAATTTTTTAGATTTTTCAATATTTTCAG | Hind III  |
| SP115A        | NO:429        | AGTCGGATCCTAAGGCTGATAATCGTGTTCAAATG        | Bam HI    |
| SP115B        | NO:430        | GACTAAGCTTAAAATTAGATAGACGTTGAGT            | Hind III  |
| SP117A        | NO:431        | AGTCGGATCCCTGTGGCAATCAGTCAGCTGCTTCC        | Bam HI    |
| SP117B        | NO:432        | GACTGTCGACTTTAATCTTGTCCCAGGTGGTTAATTTGCC   | Sal I     |
| SP118A        | NO:433        | ACTGGTCGACTTGTCAACAACAACATGCTACTTCTGAG     | Sal I     |
| SP118B        | NO:434        | GACTCTGCAGAAGTTTAAACCACTTATCATTATCC        | Pst I     |
| SP119A        | NO:435        | ACTGGGATCCTTGTTTCAGGCAAGTCCGTGACTAGTGAAC   | Bam HI    |
| SP119B        | NO:436        | GACTAAGCTTGGCTAATTCCTTCAAAGTTTGCA          | Hind III  |
| SP120A        | NO:437        | AGTCGGATCCCTCGCAAATTGAAAAGGCGGCAGTTAGCC    | Bam HI    |
| SP120B        | NO:438        | GACTAAGCTTGTAATAAGCGTACCTTTTTCTTCC         | Hind III  |
| SP121A        | NO:439        | TCAGGGATCCTTGTCAGTCAGGTTCTAATGGTTCTCAG     | Bam HI    |
| SP121B        | NO:440        | AGTCAAGCTTGGCATTGGCGTCGCCGTCCTTC           | Hind III  |
| SP122A        | NO:441        | GACTGGATCCGGAACCTTCACAGGATTTTAAAGAGAAG     | Bam HI    |
| SP122B        | NO:442        | GACTGTCGACAAATCAATCCTTCTTCTGCACTTCT        | Sal I     |
| SP123A        | NO:443        | CAGTGGATCCTGTGGTCAAGTTGAGACTCCTCAATC       | Bam HI    |
| SP123B        | NO:444        | GACTAAGCTTTTCTTCAAATTTATTATCAGC            | Hind III  |
| SP124A        | NO:445        | AGTCGGATCCAACACCTGTATATAAAGTTACAGCAATCG    | Bam HI    |
| SP124B        | NO:446        | GACTGTCGACTACTTGACCGAATGCGTCAATGTACG       | Sal I     |

WO 98/18930

PCT/US97/19422

110

Table 3  
*S. pneumoniae* ORF Cloning Primers

| Primer      |               |                                     |           |
|-------------|---------------|-------------------------------------|-----------|
| <u>Name</u> | <u>SEQ ID</u> | <u>Sequence</u>                     | <u>RE</u> |
| SP125A      | NO: 447       | CTGAGGATCCATTAGACAGATTAATTGAAATCGG  | Bam HI    |
| SP125B      | NO: 448       | GACTGTCGACTTTAAAGATTGAAGTTTTAAAGCT  | Sal I     |
| SP126A      | NO: 449       | TGACGGATCCTAAGACAGATGAACGGAGCAAGGTG | Bam HI    |
| SP126B      | NO: 450       | CTGAAAGCTTTAAGGCTTCCTCAATGAGTTTGTCT | Hind III  |
| SP127A      | NO: 451       | GACTGGATCCCTGTGAGAATCAAGCTACACCCA   | Bam HI    |
| SP127B      | NO: 452       | CTGAAAGCTTTTGTAAGTGAAGATTGATCTGGGAG | Hind III  |

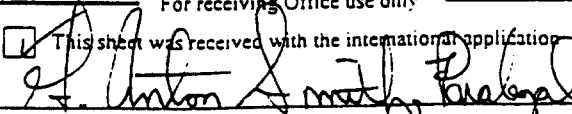
WO 98/18936

PCT/US97/19422

110/1

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

|                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A. The indications made below relate to the microorganism referred to in the description<br>on page <u>9</u> , line <u>12</u>                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                     |
| B. IDENTIFICATION OF DEPOSIT <span style="float: right;">Further deposits are identified on an additional sheet <input type="checkbox"/></span>                                                                                                                                                                                                                                                                               |                                                                                                                                                                                     |
| Name of depositary institution<br><b>American Type Culture Collection</b>                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                     |
| Address of depositary institution (including postal code and country)<br><b>12301 Parklawn Drive<br/>Rockville, Maryland 20852<br/>United States of America</b>                                                                                                                                                                                                                                                               |                                                                                                                                                                                     |
| Date of deposit<br><b>October 10, 1996</b>                                                                                                                                                                                                                                                                                                                                                                                    | Accession Number<br><b>55840</b>                                                                                                                                                    |
| C. ADDITIONAL INDICATIONS (leave blank if not applicable) <span style="float: right;">This information is continued on an additional sheet <input checked="" type="checkbox"/></span>                                                                                                                                                                                                                                         |                                                                                                                                                                                     |
| In respect of those designations in which a European Patent is sought a sample of the deposited microorganism will be made available until the publication the mention of the grant of the European patent or until the date on which application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4) EPC). |                                                                                                                                                                                     |
| D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                     |
|                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                     |
| E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                     |
| The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g. "Accession Number of Deposit")                                                                                                                                                                                                                                                           |                                                                                                                                                                                     |
|                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                     |
| <div>For receiving Office use only</div> <div><input checked="" type="checkbox"/> This sheet was received with the international application</div> <div></div> <div>Authorized officer</div> <div><b>12 DECEMBER 1997</b></div>                                                                                                            | <div>For International Bureau use only</div> <div><input type="checkbox"/> This sheet was received by the International Bureau on:</div> <div> </div> <div>Authorized officer</div> |

**SINGAPORE**

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert. The request to this effect must be filed by the applicant with the International Bureau before the completion of the technical preparations for international publication of the application.

**NORWAY**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Norwegian Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Norwegians Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Norwegian Patent Office or any person approved by the applicant in the individual case.

**AUSTRALIA**

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

**FINLAND**

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Registration), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art.

**ICELAND**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Icelandic Patent Office), or has been finally decided upon by the Icelandic Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected in the art.

Page 2

**DENMARK**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent Office or any person approved by the applicant in the individual case.

**SWEDEN**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PUT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant, any request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent Office or any person approved by the applicant in the individual case.

**UNITED KINGDOM**

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert. The request to this effect must be filed by the applicant with the International Bureau before the completion of the technical preparations for the International publication of the application.

**NETHERLANDS**

The applicant hereby requests that until the date of a grant of a Netherlands patent or until the date on which the application is refused or withdrawn or lapse, the microorganism shall be made available as provided in Rule 31F(1) of the Patent Rules only by the issue of a sample to an expert. The request to this effect must be furnished by the applicant with the Netherlands Industrial Property Office before the date on which the application is made available to the public under Section 22C or Section 25 of the Patents Act of the Kingdom of the Netherlands, whichever two dates occurs earlier.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Human Genome Sciences, Inc.

(ii) TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

(iii) NUMBER OF SEQUENCES: 452

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Gowling, Strathy & Henderson

(B) STREET: 2600-160 Elgin Street

(C) CITY: Ottawa

(D) STATE: Ontario

(E) COUNTRY: Canada

(F) ZIP: K1P 1C3

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

(B) COMPUTER: HP Vectra 486/33

(C) OPERATING SYSTEM: MSDOS version 6.2

(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 2,269,663

(B) FILING DATE: 1997/10/30

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/029,960

(B) FILING DATE: 1996/10/31

## (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Gowling, Strathy & Henderson
- (B) REGISTRATION NUMBER:
- (C) REFERENCE/DOCKET NUMBER: 08-882915CA

## (vi) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (613) 233-1781
- (B) TELEFAX: (613) 563-9869

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1999 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| TAAAATCTAC GACAATAAAA ATCAACTCAT TGCTGACTTG GGTCTGAAC GCCGCGTCAA   | 60  |
| TGCCCCAAGCT AATGATATTC CCACAGATTT GGTTAAGGCA ATCGTTTCTA TCGAAGACCA | 120 |
| TCGCTTCTTC GACCACAGGG GGATTGATAC CATCCGTATC CTGGGAGCTT TCTTGCGCAA  | 180 |
| TCTGCAAAGC AATTCCCTCC AAGGTGGATC AACTCTCACC CAACAGTTGA TTAAGTTGAC  | 240 |
| TTACTTTTCA ACTTCGACTT CCGACCAGAC TATTTCTCGT AAGGCTCAGG AAGCTTGGTT  | 300 |
| AGCGATTGAG TTAGAACAAA AAGCAACCAA GCAAGAAATC TTGACCTACT ATATAAATAA  | 360 |
| GGTCTACATG TCTAATGGGA ACTATGGAAT GCAGACAGCA GCTCAAAACT ACTATGGTAA  | 420 |
| AGACCTCAAT AATTTAAGTT TACCTCAGTT AGCCTTGCTG GCTGGAATGC CTCAGGCACC  | 480 |
| AAACCAATAT GACCCCTATT CACATCCAGA AGCAGCCCAA GACCGCCGAA ACTTGGTCTT  | 540 |
| ATCTGAAATG AAAAATCAAG GCTACATCTC TGCTGAACAG TATGAGAAAG CAGTCAATAC  | 600 |
| ACCAATTACT GATGGACTAC AAAGTCTCAA ATCAGCAAGT AATTACCCTG CTTACATGGA  | 660 |
| TAATTACCTC AAGGAAGTCA TCAATCAAGT TGAAGAAGAA ACAGGCTATA ACCTACTCAC  | 720 |
| AACTGGGATG GATGTCTACA CAAATGTAGA CCAAGAAGCT CAAAAACATC TGTGGGATAT  | 780 |
| TTACAATACA GACGAATACG TTGCCTATCC AGACGATGAA TTGCAAGTCG CTTCTACCAT  | 840 |
| TGTTGATGTT TCTAACGGTA AAGTCATTGC CCAGCTAGGA GCACGCCATC AGTCAAGTAA  | 900 |

113

```

TGTTCCTTC GGAATTAACC AAGCAGTAGA AACAAACCGC GACTGGGGAT CAACTATGAA      960
ACCGATCACA GACTATGCTC CTGCCTTGGA GTACGGTGTC TACGATTCAA CTGCTACTAT      1020
CGTTCACGAT GAGCCCTATA ACTACCCTGG GACAAATACT CCTGTTTATA ACTGGGATAG      1080
GGGCTACTTT GGCAACATCA CCTTGCAATA CGCCCTGCAA CAATCGCGAA ACGTCCCAGC      1140
CGTGGAAACT CTAAACAAGG TCGGACTCAA CCGCGCCAAG ACTTTCCTAA ATGGTCTAGG      1200
AATCGACTAC CCAAGTATTC ACTACTCAA TGCCATTTCA AGTAACACAA CCGAATCAGA      1260
CAAAAAATAT GGAGCAAGTA GTGAAAAGAT GGCTGCTGCT TACGCTGCCT TTGCAAATGG      1320
TGGAACCTAC TATAAACCAA TGTATATCCA TAAAGTCGTC TTAGTGATG GGAGTGAAAA      1380
AGAGTTCTCT AATGTCGGAA CTCGTGCCAT GAAGGAAACG ACAGCCTATA TGATGACCGA      1440
CATGATGAAA ACAGTCTTGA CTTATGGAAC TGGACGAAAT GCCTATCTTG CTTGGCTCCC      1500
TCAGGCTGGT AAAACAGGAA CCTCTAACTA TACAGACGAG GAAATTGAAA ACCACATCAA      1560
GACCTCTCAA TTTGTAGCAC CTGATGAACT ATTTGTGGC TATACGCGTA AATATTCAAT      1620
GGCTGTATGG ACAGGCTATT CTAACCGTCT GACACCACTT GTAGGCAATG GCCTTACGGT      1680
CGTGCCAAA GTTTACCGCT CTATGATGAC CTACCTGTCT GAAGGAAGCA ATCCAGAAGA      1740
TTGGAATATA CCAGAGGGGC TCTACAGAAA TGGAGAATTC GTATTTAAAA ATGGTGCTCG      1800
TTCTACGTGG AACTCACCTG CTCCACAACA ACCCCATCA ACTGAAAGTT CAAGCTCATC      1860
ATCAGATAGT TCAACTTCAC AGTCTAGCTC AACCCTCCA AGCACAATA ATAGTACGAC      1920
TACCAATCCT AACAATAATA CGCAACAATC AAATACAACC CCTGATCAAC AAAATCAGAA      1980
TCCTCAACCA GCACAACCA                                     1999

```

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 666 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

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

```

114

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

115

| 405 |     |     |     |     |     |     |     |     |     | 410 |     |     |     |     | 415 |  |  |  |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|--|--|
| Thr | Glu | Ser | Asp | Lys | Lys | Tyr | Gly | Ala | Ser | Ser | Glu | Lys | Met | Ala | Ala |  |  |  |  |
|     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |  |  |  |  |
| Ala | Tyr | Ala | Ala | Phe | Ala | Asn | Gly | Gly | Thr | Tyr | Tyr | Lys | Pro | Met | Tyr |  |  |  |  |
|     |     | 435 |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |  |  |  |  |
| Ile | His | Lys | Val | Val | Phe | Ser | Asp | Gly | Ser | Glu | Lys | Glu | Phe | Ser | Asn |  |  |  |  |
|     |     | 450 |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |  |  |  |  |
| Val | Gly | Thr | Arg | Ala | Met | Lys | Glu | Thr | Thr | Ala | Tyr | Met | Met | Thr | Asp |  |  |  |  |
| 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |  |  |  |  |
| Met | Met | Lys | Thr | Val | Leu | Thr | Tyr | Gly | Thr | Gly | Arg | Asn | Ala | Tyr | Leu |  |  |  |  |
|     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |  |  |  |  |
| Ala | Trp | Leu | Pro | Gln | Ala | Gly | Lys | Thr | Gly | Thr | Ser | Asn | Tyr | Thr | Asp |  |  |  |  |
|     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |  |  |  |  |
| Glu | Glu | Ile | Glu | Asn | His | Ile | Lys | Thr | Ser | Gln | Phe | Val | Ala | Pro | Asp |  |  |  |  |
|     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |  |  |  |  |
| Glu | Leu | Phe | Ala | Gly | Tyr | Thr | Arg | Lys | Tyr | Ser | Met | Ala | Val | Trp | Thr |  |  |  |  |
|     |     | 530 |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |  |  |  |  |
| Gly | Tyr | Ser | Asn | Arg | Leu | Thr | Pro | Leu | Val | Gly | Asn | Gly | Leu | Thr | Val |  |  |  |  |
| 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |  |  |  |  |
| Ala | Ala | Lys | Val | Tyr | Arg | Ser | Met | Met | Thr | Tyr | Leu | Ser | Glu | Gly | Ser |  |  |  |  |
|     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |  |  |  |  |
| Asn | Pro | Glu | Asp | Trp | Asn | Ile | Pro | Glu | Gly | Leu | Tyr | Arg | Asn | Gly | Glu |  |  |  |  |
|     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |  |  |  |  |
| Phe | Val | Phe | Lys | Asn | Gly | Ala | Arg | Ser | Thr | Trp | Asn | Ser | Pro | Ala | Pro |  |  |  |  |
|     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |  |  |  |  |
| Gln | Gln | Pro | Pro | Ser | Thr | Glu | Ser | Ser | Ser | Ser | Ser | Asp | Ser | Ser |     |  |  |  |  |
|     |     | 610 |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |  |  |  |  |
| Thr | Ser | Gln | Ser | Ser | Ser | Thr | Thr | Pro | Ser | Thr | Asn | Asn | Ser | Thr | Thr |  |  |  |  |
| 625 |     |     |     |     | 630 |     |     |     |     |     | 635 |     |     |     | 640 |  |  |  |  |
| Thr | Asn | Pro | Asn | Asn | Asn | Thr | Gln | Gln | Ser | Asn | Thr | Thr | Pro | Asp | Gln |  |  |  |  |
|     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     | 655 |     |  |  |  |  |
| Gln | Asn | Gln | Asn | Pro | Gln | Pro | Ala | Gln | Pro |     |     |     |     |     |     |  |  |  |  |
|     |     |     | 660 |     |     |     |     | 665 |     |     |     |     |     |     |     |  |  |  |  |

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1714 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

AAATTACAAT ACGGACTATG AATTGACCTC TGGAGAAAAA TTACCTCTTC CTAAAGAGAT

60

|                                                                    |      |
|--------------------------------------------------------------------|------|
| TTCAGGTTAC ACTTATATTG GATATATCAA AGAGGGAAAA ACGACTTCTG AGTCTGAAGT  | 120  |
| AAGTAATCAA AAGAGTTCAG TTGCCACTCC TACAAAACAA CAAAAGGTGG ATTATAATGT  | 180  |
| TACACCGAAT TTTGTAGACC ATCCATCAAC AGTACAAGCT ATTCAAGAAC AAACACCTGT  | 240  |
| TTCTTCAACT AAGCCGACAG AAGTTCAAGT AGTTGAAAAA CCTTTCTCTA CTGAATTAAT  | 300  |
| CAATCCAAGA AAAGAAGAGA AACAACTCTC AGATTCTCAA GAACAATTAG CCGAACATAA  | 360  |
| GAATCTAGAA ACGAAGAAAG AGGAGAAGAT TTCTCCAAAA GAAAAGACTG GGGTAAATAC  | 420  |
| ATTAAATCCA CAGGATGAAG TTTTATCAGG TCAATTGAAC AAACCTGAAC TCTTATATCG  | 480  |
| TGAGGAAACT ATGAGAGACAA AAATAGATTT TCAAGAAGAA ATTCAAGAAA ATCCTGATTT | 540  |
| AGCTGAAGGA ACTGTAAGAG TAAAACAAGA AGGTAAATTA GGTAAGAAAG TTGAAATCGT  | 600  |
| CAGAAATATTC TCTGTAAACA AGGAAGAAGT TTCGCGAGAA ATTGTTTCAA CTCAACGAC  | 660  |
| TGCGCCTAGT CCAAGAATAG TCGAAAAAGG TACTAAAAAA ACTCAAGTTA TAAAGGAACA  | 720  |
| ACCTGAGACT GGTGTAGAAC ATAAGGACGT ACAGTCTGGA GCTATTGTTG AACCCGCAAT  | 780  |
| TCAGCCTGAG TTGCCCGAAG CTGTAGTAAG TGACAAAGGC GAACCAGAAG TTCAACCTAC  | 840  |
| ATTACCCGAA GCAGTTGTGA CCGACAAAGG TGAGACTGAG GTTCAACCAG AGTCGCCAGA  | 900  |
| TACTGTGGTA AGTGATAAAG GTGAACCAGA GCAGGTAGCA CCGCTTCCAG AATATAAGGG  | 960  |
| TAATATTGAG CAAGTAAAC CTGAAACTCC GGTTGAGAAG ACCAAAGAAC AAGGTCCAGA   | 1020 |
| AAAACTGAA GAAGTTCCAG TAAAACCAAC AGAAGAAACA CCAGTAAATC CAAATGAAGG   | 1080 |
| TACTACAGAA GGAACCTCAA TTCAAGAAGC AGAAAATCCA GTTCAACCTG CAGAAGAATC  | 1140 |
| AACAACGAAT TCAGAGAAAG TATCACCAGA TACATCTAGC AAAAATACTG GGGAAGTGTC  | 1200 |
| CAGTAATCCT AGTGATTCTGA CAACCTCAGT TGGAGAATCA AATAAACCAG AACATAATGA | 1260 |
| CTCTAAAAAT GAAAATTCAG AAAAACTGT AGAAGAAGTT CCAGTAAATC CAAATGAAGG   | 1320 |
| CACAGTAGAA GGTACCTCAA ATCAAGAAAC AGAAAAACCA GTTCAACCTG CAGAAGAAAC  | 1380 |
| ACAAACAAAC TCTGGGAAAA TAGCTAACGA AAATACTGGA GAAGTATCCA ATAAACCTAG  | 1440 |
| TGATTCAAAA CCACCAGTTG AAGAATCAAA TCAACCAGAA AAAAACGGAA CTGCAACAAA  | 1500 |
| ACCAGAAAAT TCAGGTAATA CAACATCAGA GAATGGACAA ACAGAACCAG AACCATCAAA  | 1560 |
| CGGAAATTCA ACTGAGGATG TTTCAACCGA ATCAAACACA TCCAATTCAA ATGGAAACGA  | 1620 |
| AGAAATTAAA CAAGAAAATG AACTAGACCC TGATAAAAAG GTAGAAGAAC CAGAGAAAAC  | 1680 |
| ACTTGAATTA AGAAATGTTT CCGACCTAGA GTTA                              | 1714 |

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 571 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

117

(ii) MOLECULE TYPE: protein

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

```

Asn Tyr Asn Thr Asp Tyr Glu Leu Thr Ser Gly Glu Lys Leu Pro Leu
1          5          10          15

Pro Lys Glu Ile Ser Gly Tyr Thr Tyr Ile Gly Tyr Ile Lys Glu Gly
          20          25          30

Lys Thr Thr Ser Glu Ser Glu Val Ser Asn Gln Lys Ser Ser Val Ala
          35          40          45

Thr Pro Thr Lys Gln Gln Lys Val Asp Tyr Asn Val Thr Pro Asn Phe
50          55          60

Val Asp His Pro Ser Thr Val Gln Ala Ile Gln Glu Gln Thr Pro Val
65          70          75          80

Ser Ser Thr Lys Pro Thr Glu Val Gln Val Val Glu Lys Pro Phe Ser
          85          90          95

Thr Glu Leu Ile Asn Pro Arg Lys Glu Glu Lys Gln Ser Ser Asp Ser
100          105          110

Gln Glu Gln Leu Ala Glu His Lys Asn Leu Glu Thr Lys Lys Glu Glu
115          120          125

Lys Ile Ser Pro Lys Glu Lys Thr Gly Val Asn Thr Leu Asn Pro Gln
130          135          140

Asp Glu Val Leu Ser Gly Gln Leu Asn Lys Pro Glu Leu Leu Tyr Arg
145          150          155          160

Glu Glu Thr Met Glu Thr Lys Ile Asp Phe Gln Glu Glu Ile Gln Glu
165          170          175

Asn Pro Asp Leu Ala Glu Gly Thr Val Arg Val Lys Gln Glu Gly Lys
180          185          190

Leu Gly Lys Lys Val Glu Ile Val Arg Ile Phe Ser Val Asn Lys Glu
195          200          205

Glu Val Ser Arg Glu Ile Val Ser Thr Ser Thr Thr Ala Pro Ser Pro
210          215          220

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

Pro Glu Thr Gly Val Glu His Lys Asp Val Gln Ser Gly Ala Ile Val
245          250          255

Glu Pro Ala Ile Gln Pro Glu Leu Pro Glu Ala Val Val Ser Asp Lys
260          265          270

Gly Glu Pro Glu Val Gln Pro Thr Leu Pro Glu Ala Val Val Thr Asp
275          280          285

Lys Gly Glu Thr Glu Val Gln Pro Glu Ser Pro Asp Thr Val Val Ser
290          295          300

```

118

```

Asp Lys Gly Glu Pro Glu Gln Val Ala Pro Leu Pro Glu Tyr Lys Gly
305                      310                      315                      320

Asn Ile Glu Gln Val Lys Pro Glu Thr Pro Val Glu Lys Thr Lys Glu
                      325                      330                      335

Gln Gly Pro Glu Lys Thr Glu Glu Val Pro Val Lys Pro Thr Glu Glu
                      340                      345                      350

Thr Pro Val Asn Pro Asn Glu Gly Thr Thr Glu Gly Thr Ser Ile Gln
                      355                      360                      365

Glu Ala Glu Asn Pro Val Gln Pro Ala Glu Glu Ser Thr Thr Asn Ser
                      370                      375                      380

Glu Lys Val Ser Pro Asp Thr Ser Ser Lys Asn Thr Gly Glu Val Ser
385                      390                      395                      400

Ser Asn Pro Ser Asp Ser Thr Thr Ser Val Gly Glu Ser Asn Lys Pro
                      405                      410                      415

Glu His Asn Asp Ser Lys Asn Glu Asn Ser Glu Lys Thr Val Glu Glu
                      420                      425                      430

Val Pro Val Asn Pro Asn Glu Gly Thr Val Glu Gly Thr Ser Asn Gln
                      435                      440                      445

Glu Thr Glu Lys Pro Val Gln Pro Ala Glu Glu Thr Gln Thr Asn Ser
                      450                      455                      460

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

Asp Ser Lys Pro Pro Val Glu Glu Ser Asn Gln Pro Glu Lys Asn Gly
                      485                      490                      495

Thr Ala Thr Lys Pro Glu Asn Ser Gly Asn Thr Thr Ser Glu Asn Gly
                      500                      505                      510

Gln Thr Glu Pro Glu Pro Ser Asn Gly Asn Ser Thr Glu Asp Val Ser
                      515                      520                      525

Thr Glu Ser Asn Thr Ser Asn Ser Asn Gly Asn Glu Glu Ile Lys Gln
                      530                      535                      540

Glu Asn Glu Leu Asp Pro Asp Lys Lys Val Glu Glu Pro Glu Lys Thr
545                      550                      555                      560

Leu Glu Leu Arg Asn Val Ser Asp Leu Glu Leu
                      565                      570

```

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 748 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

TGAGAATCAA GCTACACCCA AAGAGACTAG CGCTCAAAG ACAATCGTCC TTGCTACAGC

60

119

```

TGGCGACGTG CCACCATTG ACTACGAAGA CAAGGGCAAT CTGACAGGCT TTGATATCGA      120
AGTTTTAAAG GCAGTAGATG AAAAAGTCAG CGACTACGAG ATTCAATTCC AAAGAACCGC      180
CTGGGAGAGC ATCTTCCCAG GACTTGATTC TGGTCACTAT CAGGCTGCGG CCAATAACTT      240
GAGTTACACA AAAGAGCGTG CTGAAAAATA CCTTTACTCG CTTCCAATTT CCAACAATCC      300
CCTCGTCCTT GTCAGCAACA AGAAAAATCC TTTGACTTCT CTTGACCAGA TCGCTGGTAA      360
AACAAACACAA GAGGATACCG GAACTTCTAA CGCTCAATTC ATCAATAACT GGAATCAGAA      420
ACACACTGAT AATCCCGCTA CAATTAATTT TTCTGGTGAG GATATTGGTA AACGAATCCT      480
AGACCTTGCT AACGGAGAGT TTGATTTCTT AGTTTTTGAC AAGGTATCCG TTCAAAGAT      540
TATCAAGGAC CGTGGTTTAG ACCTCTCAGT CGTTGATTTA CCTTCTGCAG ATAGCCCCAG      600
CAATTATATC ATTTTCTCAA GCGACCAAAA AGAGTTTAAA GAGCAATTTG ATAAAGCGCT      660
CAAAGAACTC TATCAAGACG GAACCCTTGA AAAAGTCAGC AATACCTATC TAGGTGGTTT      720
TTACCTCCCA GATCAATCTC AGTTACAA                                         748

```

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 249 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

```

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

```

120

|                                                                 |     |             |
|-----------------------------------------------------------------|-----|-------------|
| 130                                                             | 135 | 140         |
| Pro Ala Thr Ile Asn Phe Ser Gly Glu Asp Ile Gly Lys Arg Ile Leu |     |             |
| 145                                                             | 150 | 155 160     |
| Asp Leu Ala Asn Gly Glu Phe Asp Phe Leu Val Phe Asp Lys Val Ser |     |             |
|                                                                 | 165 | 170 175     |
| Val Gln Lys Ile Ile Lys Asp Arg Gly Leu Asp Leu Ser Val Val Asp |     |             |
|                                                                 | 180 | 185 190     |
| Leu Pro Ser Ala Asp Ser Pro Ser Asn Tyr Ile Ile Phe Ser Ser Asp |     |             |
|                                                                 | 195 | 200 205     |
| Gln Lys Glu Phe Lys Glu Gln Phe Asp Lys Ala Leu Lys Glu Leu Tyr |     |             |
|                                                                 | 210 | 215 220     |
| Gln Asp Gly Thr Leu Glu Lys Leu Ser Asn Thr Tyr Leu Gly Gly Ser |     |             |
|                                                                 | 225 | 230 235 240 |
| Tyr Leu Pro Asp Gln Ser Gln Leu Gln                             |     |             |
|                                                                 | 245 |             |

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 985 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| TGGTAACCGC TCTTCTCGTA ACGCAGCTTC ATCTTCTGAT GTGAAGACAA AAGCAGCAAT  | 60  |
| CGTCACTGAT ACTGGTGGTG TTGATGACAA ATCATTCAAC CAATCAGCTT GGAAGGTTT   | 120 |
| GCAGGCTTGG GGTAAAGAAC ACAATCTTTC AAAAGATAAC GGTTCCTACTT ACTTCCAATC | 180 |
| AACAAGTGAA GCTGACTACG CTAACAACCTT GCAACAAGCG GCTGGAAGTT ACAACCTAAT | 240 |
| CTTCGGTGTT GGTTTTGCCC TTAATAATGC AGTTAAAGAT GCAGCAAAAG AACACACTGA  | 300 |
| CTTGAACATAT GTCTTGATTG ATGATGTGAT TAAAGACCAA AAGAATGTTG CGAGCGTAAC | 360 |
| TTTCGCTGAT AATGAGTCAG GTTACCTTGC AGGTGTGGCT GCAGCAAAAA CAACTAAGAC  | 420 |
| AAAACAAGTT GGTTTGTAG GTGGTATCGA ATCTGAAGTT ATCTCTCGTT TTGAAGCAGG   | 480 |
| ATTCAAGGCT GGTGTGCGT CAGTAGACCC ATCTATCAAA GTCCAAGTTG ACTACGCTGG   | 540 |
| TTCATTTGGT GATGCCGCTA AAGGTAAAC AATTGCAGCC GCACAATACG CAGCCGGTGC   | 600 |
| AGATATTGTT TACCAAGTAG CTGGTGGTAC AGGTGCAGGT GTCTTTGCAG AGGCAAAATC  | 660 |
| TCTCAACGAA AGCCGTCCTG AAAATGAAAA AGTTTGGGTT ATCGGTGTTG ATCGTGACCA  | 720 |
| AGAAGCAGAA GGTAAATACA CTTCTAAAGA TGGCAAAGAA TCAAACCTTG TTCTTGATC   | 780 |
| TACTTTGAAA CAAGTTGGTA CAACTGTAAA AGATATTTCT AACAAGGCAG AAAGAGGAGA  | 840 |
| ATTCCCTGGC GGTCAAGTGA TCGTTTACTC ATTGAAGGAT AAAGGGGTTG ACTTGGCAGT  | 900 |

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 328 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

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

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1404 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

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

|            |            |            |            |            |            |      |
|------------|------------|------------|------------|------------|------------|------|
| TGTGGAAATT | TGACAGGTAA | CAGCAAAAAA | GCTGCTGATT | CAGGTGACAA | ACCTGTTATC | 60   |
| AAAATGTACC | AAATCGGTGA | CAAACCAGAC | AACTTGGATG | AATTGTTAGC | AAATGCCAAC | 120  |
| AAAATCATTG | AAGAAAAAGT | TGGTGCCAAA | TTGGATATCC | AATACCTTGG | CTGGGGTGAC | 180  |
| TATGGTAAGA | AAATGTCAGT | TATCACATCA | TCTGGTGAAA | ACTATGATAT | TGCTTTTGCA | 240  |
| GATAACTATA | TTGTAAATGC | TCAAAAAGGT | GCTTACGCTG | ACTTGACAGA | ATTGTACAAA | 300  |
| AAAGAAGGTA | AAGACCTTTA | CAAAGCACTT | GACCCAGCTT | ACATCAAGGG | TAATACTGTA | 360  |
| AATGGTAAGA | TTTACGCTGT | TCCAGTTGCA | GCCAACTTGG | CATCATCTCA | AACTTTTGCC | 420  |
| TTCAACGGAA | CTCTCCTTGC | TAAATATGGT | ATCGATATTT | CAGGTGTTAC | TTCTTACGAA | 480  |
| ACTCTTGAGC | CAGTCTTGAA | ACAAATCAAA | GAAAAAGCTC | CAGACGTAGT | ACCATTTGCT | 540  |
| ATTGGTAAAG | TTTTCATCCC | ATCTGATAAT | TTGACTACC  | CAGTAGCAAA | CGGTCTTCCA | 600  |
| TTCGTTATCG | ACCTTGAAGG | CGATACTACT | AAAGTTGTAA | ACCGTTACGA | AGTGCCCTGT | 660  |
| TTCAAAGAAC | ACTTGAAGAC | TCTTCACAAA | TTCTATGAAG | CTGGCTACAT | TCCAAAAGAC | 720  |
| GTCGCAACAA | GCGATACTTC | CTTTGACCTT | CAACAAGATA | CTTGGTTCGT | TCGTGAAGAA | 780  |
| ACAGTAGGAC | CAGCTGACTA | CGGTAACAGC | TTGCTTTTAC | GTGTTGCCAA | CAAAGATATC | 840  |
| CAAATCAAAC | CAATTACTAA | CTTCATCAAG | NAAAACCAAA | CAACACAAGT | TGCTAACTTT | 900  |
| GTCTCTCTAA | ACAACCTCTA | GAACAAAGAA | AAATCAATGG | AAATCTTGAA | CCTCTTGAAT | 960  |
| ACGAACCCAG | AACTCTTGAA | CGGTCTTGTT | TACGGTCCAG | AAGGCAAGAA | CTGGGAAAAA | 1020 |

123

```

ATTGAAGGTA AAGAAAACCG TGTTGCGGTT CTTGATGGCT ACAAAGGAAA CACTCACATG      1080
GGTGGATGGA ACACTGGTAA CAACTGGATC CTTTACATCA ACGAAAACGT TACAGACCAA      1140
CAAATCGAAA ATTCTAAGAA AGAATTGGCA GAAGCTAAAG AATCTCCAGC GCTTGGATTT      1200
ATCTTCAATA CTGACAATGT GAAATCTGAA ATCTCAGCTA TTGCTAACAC AATGCAACAA      1260
TTTGATACAG CTATCAACAC TGGTACTGTA GACCCAGATA AAGCGATTCC AGAATTGATG      1320
GAAAAATTGA AATCTGAAGG TGCCTACGAA AAAGTATTGA ACGAAATGCA AAAACAATAC      1380
GATGAATTCT TGAAAAACAA AAAA                                           1404

```

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 468 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

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

```

124

Tyr Pro Val Ala Asn Gly Leu Pro Phe Val Ile Asp Leu Glu Gly Asp  
 195 200 205  
 Thr Thr Lys Val Val Asn Arg Tyr Glu Val Pro Arg Phe Lys Glu His  
 210 215 220  
 Leu Lys Thr Leu His Lys Phe Tyr Glu Ala Gly Tyr Ile Pro Lys Asp  
 225 230 235 240  
 Val Ala Thr Ser Asp Thr Ser Phe Asp Leu Gln Gln Asp Thr Trp Phe  
 245 250 255  
 Val Arg Glu Glu Thr Val Gly Pro Ala Asp Tyr Gly Asn Ser Leu Leu  
 260 265 270  
 Ser Arg Val Ala Asn Lys Asp Ile Gln Ile Lys Pro Ile Thr Asn Phe  
 275 280 285  
 Ile Lys Xaa Asn Gln Thr Thr Gln Val Ala Asn Phe Val Ile Ser Asn  
 290 295 300  
 Asn Ser Lys Asn Lys Glu Lys Ser Met Glu Ile Leu Asn Leu Leu Asn  
 305 310 315 320  
 Thr Asn Pro Glu Leu Leu Asn Gly Leu Val Tyr Gly Pro Glu Gly Lys  
 325 330 335  
 Asn Trp Glu Lys Ile Glu Gly Lys Glu Asn Arg Val Arg Val Leu Asp  
 340 345 350  
 Gly Tyr Lys Gly Asn Thr His Met Gly Gly Trp Asn Thr Gly Asn Asn  
 355 360 365  
 Trp Ile Leu Tyr Ile Asn Glu Asn Val Thr Asp Gln Gln Ile Glu Asn  
 370 375 380  
 Ser Lys Lys Glu Leu Ala Glu Ala Lys Glu Ser Pro Ala Leu Gly Phe  
 385 390 395 400  
 Ile Phe Asn Thr Asp Asn Val Lys Ser Glu Ile Ser Ala Ile Ala Asn  
 405 410 415  
 Thr Met Gln Gln Phe Asp Thr Ala Ile Asn Thr Gly Thr Val Asp Pro  
 420 425 430  
 Asp Lys Ala Ile Pro Glu Leu Met Glu Lys Leu Lys Ser Glu Gly Ala  
 435 440 445  
 Tyr Glu Lys Val Leu Asn Glu Met Gln Lys Gln Tyr Asp Glu Phe Leu  
 450 455 460  
 Lys Asn Lys Lys  
 465

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 937 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

125

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| TGGTCAAGGA ACTGCTTCTA AAGACAACAA AGAGGCAGAA CTTAAGAAGG TTGACTTTAT  | 60  |
| CCTAGACTGG ACACCAAATA CCAACCACAC AGGGCTTTAT GTTGCCAAGG AAAAAGGTTA  | 120 |
| TTTCAAAGAA GCTGGAGTGG ATGTTGATTT GAAATTGCCA CCAGAAGAAA GTTCTTCTGA  | 180 |
| CTTGTTATC AACGGAAAGG CACCATTTCG AGTGTATTTT CAAGACTACA TGGCTAAGAA   | 240 |
| ATTGGAAAAA GGAGCAGGAA TCACTGCCGT TGCAGCTATT GTTGAACACA ATACATCAGG  | 300 |
| AATCATCTCT CGTAAATCTG ATAATGTAAG CAGTCCAAAA GACTTGGTTG GTAAGAAATA  | 360 |
| TGGGACATGG AATGACCCAA CTGAACTTGC TATGTTGAAA ACCTTGGTAG AATCTCAAGG  | 420 |
| TGGAGACTTT GAGAAGGTTG AAAAAGTACC AAATAACGAC TCAAACCTCA TCACACCGAT  | 480 |
| TGCCAATGGC GTCTTTGATA CTGCTTGGAT TTACTACGGT TGGGATGGTA TCCTTGCTAA  | 540 |
| ATCTCAAGGT GTAGATGCTA ACTTCATGTA CTTGAAAGAC TATGTCAAGG AGTTTACTA   | 600 |
| CTATTCACCA GTTATCATCG CAAACAACGA CTATCTGAAA GATAACAAAG AAGAAGCTCG  | 660 |
| CAAAGTCATC CAAGCCATCA AAAAAGGCTA CCAATATGCC ATGGAACATC CAGAAGAAGC  | 720 |
| TGCAGATATT CTCATCAAGA ATGCACCTGA ACTCAAGGAA AAACGTGACT TTGTCATCGA  | 780 |
| ATCTCAAAAA TACTTGTCAA AAGAATACGC AAGCGACAAG GAAAAATGGG GTCAATTTGA  | 840 |
| CGCAGCTCGC TGGAAATGCTT TCTACAAATG GGATAAAGAA AATGGTATCC TTAAAGAAGA | 900 |
| CTTGACAGAC AAAGGCTTCA CCAACGAATT TGTGAAA                           | 937 |

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 312 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

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

(2) INFORMATION FOR SEQ ID NO: 13:

(A) LENGTH: 799 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

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

|            |            |             |            |            |            |     |
|------------|------------|-------------|------------|------------|------------|-----|
| TAGCTCAGGT | GGAAACGCTG | GTTTCATCCTC | TGGAAAAACA | ACTGCCAAAG | CTCGCACTAT | 60  |
| CGATGAAATC | AAAAAAGCG  | GTGAAC TGCG | AATCGCCGTG | TTTGGAGATA | AAAAACCGTT | 120 |
| TGGCTACGTT | GACAATGATG | GTTCTACCAA  | GGTACGCTAC | GATATTGAAC | TAGGGAACCA | 180 |
| ACTAGCTCAA | GACCTTGGTG | TCAAGGTTAA  | ATACATTTC  | GTCGATGCTG | CCAACCGTGC | 240 |

127

```

GGAATACTTG ATTTCAAACA AGGTAGATAT TACTCTTGCT AACTTTACAG TAACTGACGA      300
ACGTAAGAAA CAAGTTGATT TTGCCCTTCC ATATATGAAA GTTCTCTGCG GTGTCGTATC      360
ACCTAAGACT GGTCTCATT CAGACGTCAA ACAACTTGAA GGTAACCT TAATTGTCAC      420
AAAAGGAACG ACTGCTGAGA CTTATTTTGA AAAGAATCAT CCAGAAATCA AACTCCAAAA      480
ATACGACCAA TACAGTGACT CTTACCAAGC TCTTCTTGAC GGACGTGGAG ATGCCTTTTC      540
AACTGACAAT ACGGAAGTTC TAGCTTGGGC GCTTGAAAAT AAAGGATTTG AAGTAGGAAT      600
TACTTCCCTC GGTGATCCCC ATACCATTGC GGCAGCAGTT CAAAAAGGCA ACCAAGAATT      660
GCTAGACTTC ATCAATAAAG ATATTGAAAA ATTAGGCAAG GAAACTTCT TCCACAAGGC      720
CTATGAAAAG ACACTTCACC CAACCTACGG TGACGCTGCT AAAGCAGATG ACCTGGTTGT      780
TGAAGGTGGA AAAGTTGAT                                          799

```

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 266 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

Ser Ser Gly Gly Asn Ala Gly Ser Ser Ser Gly Lys Thr Thr Ala Lys
1              5              10              15

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

Val Phe Gly Asp Lys Lys Pro Phe Gly Tyr Val Asp Asn Asp Gly Ser
35            40            45

Thr Lys Val Arg Tyr Asp Ile Glu Leu Gly Asn Gln Leu Ala Gln Asp
50            55            60

Leu Gly Val Lys Val Lys Tyr Ile Ser Val Asp Ala Ala Asn Arg Ala
65            70            75            80

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

Val Thr Asp Glu Arg Lys Lys Gln Val Asp Phe Ala Leu Pro Tyr Met
100           105           110

Lys Val Ser Leu Gly Val Val Ser Pro Lys Thr Gly Leu Ile Thr Asp
115          120          125

Val Lys Gln Leu Glu Gly Lys Thr Leu Ile Val Thr Lys Gly Thr Thr
130          135          140

Ala Glu Thr Tyr Phe Glu Lys Asn His Pro Glu Ile Lys Leu Gln Lys
145          150          155          160

```

128

Tyr Asp Gln Tyr Ser Asp Ser Tyr Gln Ala Leu Leu Asp Gly Arg Gly  
 165 170 175  
 Asp Ala Phe Ser Thr Asp Asn Thr Glu Val Leu Ala Trp Ala Leu Glu  
 180 185 190  
 Asn Lys Gly Phe Glu Val Gly Ile Thr Ser Leu Gly Asp Pro Asp Thr  
 195 200 205  
 Ile Ala Ala Ala Val Gln Lys Gly Asn Gln Glu Leu Leu Asp Phe Ile  
 210 215 220  
 Asn Lys Asp Ile Glu Lys Leu Gly Lys Glu Asn Phe Phe His Lys Ala  
 225 230 235 240  
 Tyr Glu Lys Thr Leu His Pro Thr Tyr Gly Asp Ala Ala Lys Ala Asp  
 245 250 255  
 Asp Leu Val Val Glu Gly Gly Lys Val Asp  
 260 265

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1189 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| CTCCAACATAT GGTAAATCTG CGGATGGCAC AGTGACCATC GAGTATTTCA ACCAGAAAAA | 60  |
| AGAAATGACC AAAACCTTGG AAGAAATCAC TCGTGATTTT GAGAAGGAAA ACCCTAAGAT  | 120 |
| CAAGGTCAAA GTCGTCAATG TACCAAATGC TGGTGAAGTA TTGAAGACAC GCGTTCTCGC  | 180 |
| AGGAGATGTG CCTGATGTGG TCAATATTTA CCCACAGTCC ATCGAACTGC AAGAATGGGC  | 240 |
| AAAAGCAGGT GTTTTTGAAG ATTTGAGCAA CAAAGACTAC CTGAAACGCG TGAAAAATGG  | 300 |
| CTACGCTGAA AAATATGCTG TAAACGAAAA AGTTTACAAC GTTCCTTTTA CAGCTAATGC  | 360 |
| TTATGGAATT TACTACAACA AAGATAAATT CGAAGAACTG GGCTTGAAGG TTCCTGAAAC  | 420 |
| CTGGGATGAA TTTGAACAGT TAGTCAAAGA TATCGTTGCT AAAGGACAAA CACCATTGG   | 480 |
| AATGTCAGGT GCAGATGCTT GGACACTCAA TGGTTACAAT CAATTAGCCT TTGCGACAGC  | 540 |
| AACAGGTGGA GGAAAAGAAG CAAATCAATA CCTTCGTTAT TCTCAACCAA ATGCCATTAA  | 600 |
| ATTGTCGGAT CCGATTATGA AAGATGATAT CAAGGTCATG GACATCCTTC GCATCAATGG  | 660 |
| ATCTAAGCAA AAGAACTGGG AAGGTGCTGG CTATACCGAT GTTATCGGAG CCTTCGCACG  | 720 |
| TGGGGATGTC CTCATGACAC CAAATGGGTC TTGGGCGATC ACAGCGATTA ATGAACAAAA  | 780 |
| ACCGAACTTT AAGATTGGGA CCTTCATGAT TCCAGGAAAA GAAAAAGGAC AAAGCTTAAC  | 840 |
| CGTTGGTGCG GGAGACTTGG CATGGTCTAT CTCAGCCACC ACCAAACATC CAAAAGAAGC  | 900 |
| CAATGCCTTT GTGGAATATA TGACCCGTCC AGAAGTCATG CAAAAATACT ACGATGTGGA  | 960 |

129

CGGATCTCCA ACAGCGATCG AAGGGGTCAA ACAAGCAGGA GAAGATTCAC CGCTTGCTGG 1020  
 TATGACCGAA TATGCCTTTA CGGATCGTCA CTTGGTCTGG TTGCAACAAT ACTGGACCAG 1080  
 TGAAGCAGAC TTCCATACCT TGACCATGAA CTATGTCTTG ACCGGTGATA AACAAGGCAT 1140  
 GGTCAATGAT TTGAATGCCT TCTTTAACCC GATGAAAGCG GATGTGGAT 1189

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 396 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

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

130

Asn Trp Glu Gly Ala Gly Tyr Thr Asp Val Ile Gly Ala Phe Ala Arg  
 225 230 235 240  
 Gly Asp Val Leu Met Thr Pro Asn Gly Ser Trp Ala Ile Thr Ala Ile  
 245 250 255  
 Asn Glu Gln Lys Pro Asn Phe Lys Ile Gly Thr Phe Met Ile Pro Gly  
 260 265 270  
 Lys Glu Lys Gly Gln Ser Leu Thr Val Gly Ala Gly Asp Leu Ala Trp  
 275 280 285  
 Ser Ile Ser Ala Thr Thr Lys His Pro Lys Glu Ala Asn Ala Phe Val  
 290 295 300  
 Glu Tyr Met Thr Arg Pro Glu Val Met Gln Lys Tyr Tyr Asp Val Asp  
 305 310 315 320  
 Gly Ser Pro Thr Ala Ile Glu Gly Val Lys Gln Ala Gly Glu Asp Ser  
 325 330 335  
 Pro Leu Ala Gly Met Thr Glu Tyr Ala Phe Thr Asp Arg His Leu Val  
 340 345 350  
 Trp Leu Gln Gln Tyr Trp Thr Ser Glu Ala Asp Phe His Thr Leu Thr  
 355 360 365  
 Met Asn Tyr Val Leu Thr Gly Asp Lys Gln Gly Met Val Asn Asp Leu  
 370 375 380  
 Asn Ala Phe Phe Asn Pro Met Lys Ala Asp Val Asp  
 385 390 395

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 775 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| TGGGAAAAAT TCTAGCGAAA CTAGTGGAGA TAATTGGTCA AAGTACCACT CTAACAAGTC  | 60  |
| TATTACTATT GGATTTGATA GTACTTTTGT TCCAATGGGA TTTGCTCAGA AAGATGGTTC  | 120 |
| TTATGCAGGA TTTGATATTG ATTTAGCTAC AGCTGTTTTT GAAAAATACG GAATCACGGT  | 180 |
| AAATTGGCAA CCGATTGATT GGGATTTGAA AGAAGCTGAA TTGACAAAAG GAACGATTGA  | 240 |
| TCTGATTTGG AATGGCTATT CCGCTACAGA CGAACGCCGT GAAAAGGTGG CTTTCAGTAA  | 300 |
| CTCATATATG AAGAATGAGC AGGTATTGGT TACGAAGAAA TCATCTGGTA TCACGACTGC  | 360 |
| AAAGGATATG ACTGGAAAGA CATTAGGAGC TCAAGCTGGT TCATCTGGTT ATGCGGACTT  | 420 |
| TGAAGCAAAT CCAGAAATTT TGAAGAATAT TGTCGCTAAT AAGGAAGCGA ATCAATACCA  | 480 |
| AACCTTTAAT GAAGCCTTGA TTGATTTGAA AAACGATCGA ATTGATGGTC TATTGATTGA  | 540 |
| CCGTGTCTAT GCAAACCTATT ATTTAGAAGC AGAAGGTGTT TTAAACGATT ATAATGTCTT | 600 |

131

TACAGTTGGA CTAGAAACAG AAGCTTTTGC GGTGGAGCC CGTAAGGAAG ATACAAACTT 660  
 GGTTAAGAAG ATAAATGAAG CTTTTTCTAG TCTTTACAAG GACGGCAAGT TCCAAGAAAT 720  
 CAGCCAAAAA TGGTTTGGAG AAGATGTAGC AACCAAAGAA GTAAAAGAAG GACAG 775

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 258 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

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

132

|     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|
| 225 |     | 230 |     | 235 |     | 240 |
| Ser | Gln | Lys | Trp | Phe | Gly | Glu |
|     |     |     |     | Asp | Val | Ala |
|     |     |     |     | Thr | Lys | Glu |
|     |     |     |     | Val | Lys | Glu |
|     |     |     |     | 245 | 250 | 255 |

Gly Gln

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 868 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|            |             |            |            |            |             |     |
|------------|-------------|------------|------------|------------|-------------|-----|
| TGCTAGCGGA | AAAAAAGATA  | CAACTTCTGG | TCAAAACTA  | AAAGTTGTTG | CTACAAACTC  | 60  |
| AATCATCGCT | GATATTACTA  | AAAATATTGC | TGGTGACAAA | ATTGACCTTC | ATAGTATCGT  | 120 |
| TCCGATTGGG | CAAGACCCAC  | ACGAATACGA | ACCACTTCCT | GAAGACGTTA | AGAAAACCTC  | 180 |
| TGAGGCTAAT | TTGATTTTCT  | ATAACGGTAT | CAACCTTGAA | ACAGGTGGCA | ATGCTTG GTT | 240 |
| TACAAAATTG | G TAGAAAATG | CCAAGAAAAC | TGAAAACAAA | GA TACTTCG | CAGTCAGCGA  | 300 |
| CGGCGTTGAT | GTTATCTACC  | TTGAAGGTCA | AAATGAAAAA | GGAAAAGAAG | ACCCACACGC  | 360 |
| TTGGCTTAAC | CTTGAAAACG  | GTATTATTTT | TGCTAAAAAT | ATCGCCAAAC | AATTGAGCGC  | 420 |
| CAAAGACCCT | AACAATAAAG  | AATTCTATGA | AAAAAATCTC | AAAGAATATA | CTGATAAGTT  | 480 |
| AGACAAACTT | GATAAAGAAA  | GTAAGGATAA | ATTTAATAAG | ATCCCTGCTG | AAAAGAAACT  | 540 |
| CATTGTAACC | AGCGAAGGAG  | CATTCAAATA | CTTCTCTAAA | GCCTATGGTG | TCCCAAGTGC  | 600 |
| TTACATCTGG | GAAATCAATA  | CTGAAGAAGA | AGGAACTCCT | GAACAAATCA | AGACCTTGGT  | 660 |
| TGAAAAACTT | CGCCAAACAA  | AAGTTCCATC | ACTCTTTGTA | GAATCAAGTG | TGGATGACCG  | 720 |
| TCCAATGAAA | ACTGTTTCTC  | AAGACACAAA | CATCCCAATC | TACGCTCAAA | TCTTTACTGA  | 780 |
| CTCTATCGCA | GAACAAGGTA  | AAGAAGGCGA | CAGCTACTAC | AGCATGATGA | AATACAACCT  | 840 |
| TGACAAGATT | GCTGAAGGAT  | TGGCAAAA   |            |            |             | 868 |

## (2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 289 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

133

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

## (2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1546 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                    |      |
|--------------------------------------------------------------------|------|
| TGGCTCAAAA AATACAGCTT CAAGTCCAGA TTATAAGTTG GAAGGTGTAA CATTCCCGCT  | 60   |
| TCAAGAAAAG AAAACATTGA AGTTTATGAC AGCCAGTTCA CCGTTATCTC CTAAAGACCC  | 120  |
| AAATGAAAAG TTAATTTTGC AACGTTTGGG GAAGGAAACT GGC GTTCATA TTGACTGGAC | 180  |
| CAACTACCAA TCCGACTTTG CAGAAAAACG TAAC TTGGAT ATTTCTAGTG GTGATTTACC | 240  |
| AGATGCTATC CACAACGACG GAGCTTCAGA TGTGGACTTG ATGAACTGGG CTAAAAAAGG  | 300  |
| TGTTATTATT CCAGTTGAAG ATTTGATTGA TAAATACATG CCAAATCTTA AGAAAATTTT  | 360  |
| GGATGAGAAA CCAGAGTACA AGGCCTTGAT GACAGCACCT GATGGGCACA TTTACTCATT  | 420  |
| TCCATGGATT GAAGAGCTTG GAGATGGTAA AGAGTCTATT CACAGTGTCA ACGATATGGC  | 480  |
| TTGGATTAAC AAAGATTGGC TTAAGAAACT TGGTCTTGAA ATGCCAAAAA CTACTGATGA  | 540  |
| TTTGATTAAA GTCCTAGAAG CTTTCAAAAA CGGGGATCCA AATGGAAATG GAGAGGCTGA  | 600  |
| TGAAATTCCA TTTTCATTTA TTAGTGGTAA CGGAAACGAA GATTTTAAAT TCCTATTTGC  | 660  |
| TGCATTTGGT ATAGGGGATA ACGATGATCA TTTAGTAGTA GGAAATGATG GCAAAGTTGA  | 720  |
| CTTCACAGCA GATAACGATA ACTATAAAGA AGGTGTCAAA TTTATCCGTC AATTGCAAGA  | 780  |
| AAAAGGCCTG ATTGATAAAG AAGCTTTCGA ACATGATTGG AATAGTTACA TTGCTAAAGG  | 840  |
| TCATGATCAG AAATTTGGTG TTTACTTTAC ATGGGATAAG AATAATGTTA CTGGAAGTAA  | 900  |
| CGAAAGTTAT GATGTTTTAC CAGTACTTGC TGGACCAAGT GGTCAAAAAC ACGTAGCTCG  | 960  |
| TACAAACGGT ATGGGATTTG CACGTGACAA GATGGTTATT ACCAGTGTA AAAAAACCT    | 1020 |
| AGAATTGACA GCTAAATGGA TTGATGCACA ATACGCTCCA CTCCAATCTG TGCAAAATAA  | 1080 |
| CTGGGGAAC TACGGAGATG ACAAACAACA AAACATCTTT GAATTGGATC AAGCGTCAAA   | 1140 |
| TAGTCTAAAA CACTTACCAC TAAACGGAAC TGCACCAGCA GAACTTCGTC AAAAGACTGA  | 1200 |
| AGTAGGAGGA CCACTAGCTA TCCTAGATTC ATACTATGGT AAAGTAACAA CCATGCCTGA  | 1260 |
| TGATGCCAAA TGGCGTTTGG ATCTTATCAA AGAATATTAT GTTCCTTACA TGAGCAATGT  | 1320 |
| CAATAACTAT CCAAGAGTCT TTATGACACA GGAAGATTG GACAAGATTG CCCATATCGA   | 1380 |
| AGCAGATATG AATGACTATA TCTACCGTAA ACGTGCTGAA TGGATTGTAA ATGGCAATAT  | 1440 |
| TGATACTGAG TGGGATGATT ACAAGAAAGA ACTTGAAAAA TACGGACTTT CTGATTACCT  | 1500 |
| CGCTATTAAA CAAAAATACT ACGACCAATA CCAAGCAAAC AAAAAC                 | 1546 |

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 515 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

135

(ii) MOLECULE TYPE: protein

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

```

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

```

136

Val Leu Pro Val Leu Ala Gly Pro Ser Gly Gln Lys His Val Ala Arg  
 305 310 315 320  
 Thr Asn Gly Met Gly Phe Ala Arg Asp Lys Met Val Ile Thr Ser Val  
 325 330 335  
 Asn Lys Asn Leu Glu Leu Thr Ala Lys Trp Ile Asp Ala Gln Tyr Ala  
 340 345 350  
 Pro Leu Gln Ser Val Gln Asn Asn Trp Gly Thr Tyr Gly Asp Asp Lys  
 355 360 365  
 Gln Gln Asn Ile Phe Glu Leu Asp Gln Ala Ser Asn Ser Leu Lys His  
 370 375 380  
 Leu Pro Leu Asn Gly Thr Ala Pro Ala Glu Leu Arg Gln Lys Thr Glu  
 385 390 395 400  
 Val Gly Gly Pro Leu Ala Ile Leu Asp Ser Tyr Tyr Gly Lys Val Thr  
 405 410 415  
 Thr Met Pro Asp Asp Ala Lys Trp Arg Leu Asp Leu Ile Lys Glu Tyr  
 420 425 430  
 Tyr Val Pro Tyr Met Ser Asn Val Asn Asn Tyr Pro Arg Val Phe Met  
 435 440 445  
 Thr Gln Glu Asp Leu Asp Lys Ile Ala His Ile Glu Ala Asp Met Asn  
 450 455 460  
 Asp Tyr Ile Tyr Arg Lys Arg Ala Glu Trp Ile Val Asn Gly Asn Ile  
 465 470 475 480  
 Asp Thr Glu Trp Asp Asp Tyr Lys Lys Glu Leu Glu Lys Tyr Gly Leu  
 485 490 495  
 Ser Asp Tyr Leu Ala Ile Lys Gln Lys Tyr Tyr Asp Gln Tyr Gln Ala  
 500 505 510  
 Asn Lys Asn  
 515

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 895 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| TAGTACAAAC TCAAGCACTA GTCAGACAGA GACCACTAGC TCTGCTCCAA CAGAGGTAAC | 60  |
| CATTAAAAGT TCACTGGACG AGGTCAAAGT TTCCAAAGTT CCTGAAAAGA TTGTGACCTT | 120 |
| TGACCTCGGC GCTGCGGATA CTATTCGCGC TTTAGGATTT GAAAAAATA TCGTCGGAAT  | 180 |
| GCCTACAAAA ACTGTTCCGA CTTATCTAAA AGACCTAGTG GGAAGTGTCA AAAATGTTGG | 240 |
| TTCTATGAAA GAACCTGATT TAGAAGCTAT CGCCGCCCTT GAGCCTGATT TGATTATCGC | 300 |

137

```

TTCGCCACGT ACACAAAAAT TCGTAGACAA ATTCAAAGAA ATCGCCCCAA CCGTTCTCTT      360
CCAAGCAAGC AAGGACGACT ACTGGACTTC TACCAAGGCT AATATCGAAT CCTTAGCAAG      420
TGCCTTCGGC GAAACTGGTA CACAGAAAGC CAAGGAAGAA TTGACCAAGC TAGACAAGAG      480
CATCCAAGAA GTCGCTACTA AAAATGAAAG CTCTGACAAA AAAGCCCTTG CGATCCTCCT      540
TAATGAAGGA AAAATGGCAG CCTTTGGTGC CAAATCTCGT TTCTCTTTCT TGTACCAAAC      600
CTTGAAATTC AAACCAACTG ATACAAAATT TGAAGACTCA CGCCACGGAC AAGAAGTCAG      660
CTTGAAAGT GTCAAAGAAA TCAACCCTGA CATCCTCTTT GTCATCAACC GTACCCTTGC      720
CATCGGTGGG GACAACTCTA GCAACGACGG TGTCCTAGAA AATGCCCTTA TCGCTGAAAC      780
ACCTGCTGCT AAAAATGGTA AGATTATCCA ACTAACACCA GACCTCTGGT ATCTAAGCGG      840
AGGCGGACTT GAATCAACAA AACTCATGAT TGAAGACATA CAAAAAGCTT TGAAA          895

```

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 298 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

Ser Thr Asn Ser Ser Thr Ser Gln Thr Glu Thr Ser Ser Ser Ala Pro
1              5              10              15

Thr Glu Val Thr Ile Lys Ser Ser Leu Asp Glu Val Lys Leu Ser Lys
20              25              30

Val Pro Glu Lys Ile Val Thr Phe Asp Leu Gly Ala Ala Asp Thr Ile
35              40              45

Arg Ala Leu Gly Phe Glu Lys Asn Ile Val Gly Met Pro Thr Lys Thr
50              55              60

Val Pro Thr Tyr Leu Lys Asp Leu Val Gly Thr Val Lys Asn Val Gly
65              70              75              80

Ser Met Lys Glu Pro Asp Leu Glu Ala Ile Ala Ala Leu Glu Pro Asp
85              90              95

Leu Ile Ile Ala Ser Pro Arg Thr Gln Lys Phe Val Asp Lys Phe Lys
100             105             110

Glu Ile Ala Pro Thr Val Leu Phe Gln Ala Ser Lys Asp Asp Tyr Trp
115             120             125

Thr Ser Thr Lys Ala Asn Ile Glu Ser Leu Ala Ser Ala Phe Gly Glu
130             135             140

Thr Gly Thr Gln Lys Ala Lys Glu Glu Leu Thr Lys Leu Asp Lys Ser
145             150             155             160

```

138

Ile Gln Glu Val Ala Thr Lys Asn Glu Ser Ser Asp Lys Lys Ala Leu  
 165 170 175  
 Ala Ile Leu Leu Asn Glu Gly Lys Met Ala Ala Phe Gly Ala Lys Ser  
 180 185 190  
 Arg Phe Ser Phe Leu Tyr Gln Thr Leu Lys Phe Lys Pro Thr Asp Thr  
 195 200 205  
 Lys Phe Glu Asp Ser Arg His Gly Gln Glu Val Ser Phe Glu Ser Val  
 210 215 220  
 Lys Glu Ile Asn Pro Asp Ile Leu Phe Val Ile Asn Arg Thr Leu Ala  
 225 230 235 240  
 Ile Gly Gly Asp Asn Ser Ser Asn Asp Gly Val Leu Glu Asn Ala Leu  
 245 250 255  
 Ile Ala Glu Thr Pro Ala Ala Lys Asn Gly Lys Ile Ile Gln Leu Thr  
 260 265 270  
 Pro Asp Leu Trp Tyr Leu Ser Gly Gly Gly Leu Glu Ser Thr Lys Leu  
 275 280 285  
 Met Ile Glu Asp Ile Gln Lys Ala Leu Lys  
 290 295

## (2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1261 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| TGGCAATTCT GGC GGAAGTA AAGATGCTGC CAAATCAGGT GGTGACGGTG CAAAAACAGA | 60  |
| AATCACTTGG TGGGCATTCC CAGTATTTAC CCAAGAAAAA ACTGGTGACG GTGTTGGAAC  | 120 |
| TTATGAAAAA TCAATCATCG AAGCGTTTGA AAAAGCAAAC CCAGATATAA AAGTGAAATT  | 180 |
| GGAAACCATC GACTTCAAGT CAGGTCCTGA AAAAATCACA ACAGCCATCG AAGCAGGAAC  | 240 |
| AGCTCCAGAC GTACTCTTTG ATGCACCAGG ACGTATCATC CAATACGGTA AAAACGGTAA  | 300 |
| ATTGGCTGAG TTGAATGACC TCTTCACAGA TGAATTTGTT AAAGATGTCA ACAATGAAAA  | 360 |
| CATCGTACAA GCAAGTAAAG CTGGAGACAA GGCTTATATG TATCCGATTA GTTCTGCCCC  | 420 |
| ATTCTACATG GCAATGAACA AGAAAATGTT AGAAGATGCT GGAGTAGCAA ACCTTGTAAG  | 480 |
| AGAAGGTTGG ACAACTGATG ATTTTGAAAA AGTATTGAAA GCACTTAAAG ACAAGGGTTA  | 540 |
| CACACCAGGT TCATTGTTCA GTTCTGGTCA AGGGGGAGAC CAAGGAACAC GTGCCTTTAT  | 600 |
| CTCTAACCTT TATAGCGGTT CTGTAACAGA TGAAAAAGTT AGCAAATATA CAACTGATGA  | 660 |
| TCCTAAATTC GTCAAAGGTC TTGAAAAAGC AACTAGCTGG ATTAAAGACA ATTTGATCAA  | 720 |
| TAATGGTTCA CAATTTGACG GTGGGGCAGA TATCCAAAAC TTTGCCAACG GTCAAACATC  | 780 |

139

```

TTACACAATC CTTTGGGCAC CAGCTCAAAA TGGTATCCAA GCTAAACTTT TAGAAGCAAG      840
TAAGGTAGAA GTGGTAGAAG TACCATTCCC ATCAGACGAA GGTAAGCCAG CTCTTGAGTA      900
CCTTGTAAC GGGTTTGCAG TATTCAACAA TAAAGACGAC AAGAAAGTCG CTGCATCTAA      960
GAAATTCATC CAGTTTATCG CAGATGACAA GGAGTGGGGA CCTAAAGACG TAGTTCGTAC     1020
AGGTGCTTTC CCAGTCCGTA CTTCATTTGG AAAACTTTAT GAAGACAAAC GCATGGAAAC     1080
AATCAGCGGC TGGACTCAAT ACTACTCACC ATACTACAAC ACTATTGATG GATTTGCTGA     1140
AATGAGAACA CTTTGGTTCC CAATGTTGCA ATCTGTATCA AATGGTGACG AAAAACCAGC     1200
AGATGCTTTG AAAGCCTTCA CTGAAAAAGC GAACGAAACA ATCAAAAAAG CTATGAAACA     1260
A                                                                           1261

```

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 420 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

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

```

140

|                                                                 |     |     |     |     |     |
|-----------------------------------------------------------------|-----|-----|-----|-----|-----|
|                                                                 | 165 |     | 170 |     | 175 |
| Asp Lys Gly Tyr Thr Pro Gly Ser Leu Phe Ser Ser Gly Gln Gly Gly | 180 | 185 | 190 |     |     |
| Asp Gln Gly Thr Arg Ala Phe Ile Ser Asn Leu Tyr Ser Gly Ser Val | 195 | 200 | 205 |     |     |
| Thr Asp Glu Lys Val Ser Lys Tyr Thr Thr Asp Asp Pro Lys Phe Val | 210 | 215 | 220 |     |     |
| Lys Gly Leu Glu Lys Ala Thr Ser Trp Ile Lys Asp Asn Leu Ile Asn | 225 | 230 | 235 | 240 |     |
| Asn Gly Ser Gln Phe Asp Gly Gly Ala Asp Ile Gln Asn Phe Ala Asn | 245 | 250 | 255 |     |     |
| Gly Gln Thr Ser Tyr Thr Ile Leu Trp Ala Pro Ala Gln Asn Gly Ile | 260 | 265 | 270 |     |     |
| Gln Ala Lys Leu Leu Glu Ala Ser Lys Val Glu Val Val Glu Val Pro | 275 | 280 | 285 |     |     |
| Phe Pro Ser Asp Glu Gly Lys Pro Ala Leu Glu Tyr Leu Val Asn Gly | 290 | 295 | 300 |     |     |
| Phe Ala Val Phe Asn Asn Lys Asp Asp Lys Lys Val Ala Ala Ser Lys | 305 | 310 | 315 | 320 |     |
| Lys Phe Ile Gln Phe Ile Ala Asp Asp Lys Glu Trp Gly Pro Lys Asp | 325 | 330 | 335 |     |     |
| Val Val Arg Thr Gly Ala Phe Pro Val Arg Thr Ser Phe Gly Lys Leu | 340 | 345 | 350 |     |     |
| Tyr Glu Asp Lys Arg Met Glu Thr Ile Ser Gly Trp Thr Gln Tyr Tyr | 355 | 360 | 365 |     |     |
| Ser Pro Tyr Tyr Asn Thr Ile Asp Gly Phe Ala Glu Met Arg Thr Leu | 370 | 375 | 380 |     |     |
| Trp Phe Pro Met Leu Gln Ser Val Ser Asn Gly Asp Glu Lys Pro Ala | 385 | 390 | 395 | 400 |     |
| Asp Ala Leu Lys Ala Phe Thr Glu Lys Ala Asn Glu Thr Ile Lys Lys | 405 | 410 | 415 |     |     |
| Ala Met Lys Gln                                                 | 420 |     |     |     |     |

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 658 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

TTCACAAGAA AAAACAAAAA ATGAAGATGG AGAACTAAG ACAGAACAGA CAGCCAAAGC

60

141

```

TGATGGAACA GTCGGTAGTA AGTCTCAAGG AGCTGCCCAG AAGAAAGCAG AAGTGGTCAA      120
TAAAGGTGAT TACTACAGCA TTCAAGGGAA ATACGATGAA ATCATCGTAG CCAACAAACA      180
CTATCCATTG TCTAAAGACT ATAATCCAGG GGAAAATCCA ACAGCCAAGG CAGAGTTGGT      240
CAAACATCATC AAAGCGATGC AAGAGGCAGG TTTCCCTATT AGTGATCATT ACAGTGGTTT      300
TAGAAGTTAT GAAACTCAGA CCAAGCTCTA TCAAGATTAT GTCAACCAAG ATGGAAAGGC      360
AGCAGCTGAC CGTTACTCTG CCCGTCCTGG CTATAGCGAA CACCAGACAG GCTTGGCCTT      420
TGATGTGATT GGGACTGATG GTGATTTGGT GACAGAAGAA AAAGCAGCCC AATGGCTCTT      480
GGATCATGCA GCTGATTATG GCTTTGTGTG CCGTTATCTC AAAGGCAAGG AAAAGGAAAC      540
AGGCTATATG GCTGAAGAAT GGCACCTGCG TTATGTAGGA AAAGAAGCTA AAGAAATTGC      600
TGCAAGTGGT CTCAGTTTGG AAGAATACTA TGGCTTTGAA GGCGGAGACT ACGTCGAT      658

```

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 219 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

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

```

(ii) MOLECULE TYPE: protein

143

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

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

(2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 781 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

144

AAACTCAGAA AAGAAAGCAG ACAATGCAAC AACTATCAAA ATCGCAACTG TTAACCGTAG 60  
 CGGTTCTGAA GAAAAACGTT GGGACAAAAT CCAAGAATTG GTTAAAAAAG ACGGAATTAC 120  
 CTTGGAATTT ACAGAGTTCA CAGACTACTC ACAACCAAAC AAAGCAACTG CTGATGGCGA 180  
 AGTAGATTTG AACGCTTTCC AACACTATAA CTTCTTGAAC AACTGGAACA AAGAAAACGG 240  
 AAAAGACCTT GTAGCGATTG CAGATACTTA CATCTCTCCA ATCCGCCTTT ACTCAGGTTT 300  
 GAATGGAAGT GCCAACAAGT AACTAAAGT AGAAGACATC CCAGCAAACG GAGAAATCGC 360  
 TGTACCGAAT GACGCTACAA ACGAAAGCCG TCGCTTTAT TTGCTTCAAT CAGCTGGCTT 420  
 GATTAAATTG GATGTTTCTG GAACTGCTCT TGCAACAGTT GCCAACATCA AAGAAAATCC 480  
 AAAGAACTTG AAAATCACTG AATTGGACGC TAGCCAAACA GCTCGTTCAT TGTCATCAGT 540  
 TGACGCTGCC GTTGTAACA ATACCTTCGT TACAGAAGCA AAATTGGACT ACAAGAAATC 600  
 ACTTTTCAAA GAACAAGCTG ATGAAACTC AAAACAATGG TACAACATCA TTGTTGCAAA 660  
 AAAAGATTGG GAAACATCAC CTAAGGCTGA TGCTATCAAG AAAGTAATCG CAGCTTACCA 720  
 CACAGATGAC GTGAAAAAAG TTATCGAAGA ATCATCAGAT GGTGTTGGATC AACCAGTTTG 780  
 G 781

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 260 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Asn Ser Glu Lys Lys Ala Asp Asn Ala Thr Thr Ile Lys Ile Ala Thr  
 1 5 10 15  
 Val Asn Arg Ser Gly Ser Glu Glu Lys Arg Trp Asp Lys Ile Gln Glu  
 20 25 30  
 Leu Val Lys Lys Asp Gly Ile Thr Leu Glu Phe Thr Glu Phe Thr Asp  
 35 40 45  
 Tyr Ser Gln Pro Asn Lys Ala Thr Ala Asp Gly Glu Val Asp Leu Asn  
 50 55 60  
 Ala Phe Gln His Tyr Asn Phe Leu Asn Asn Trp Asn Lys Glu Asn Gly  
 65 70 75 80  
 Lys Asp Leu Val Ala Ile Ala Asp Thr Tyr Ile Ser Pro Ile Arg Leu  
 85 90 95  
 Tyr Ser Gly Leu Asn Gly Ser Ala Asn Lys Tyr Thr Lys Val Glu Asp  
 100 105 110

145

Ile Pro Ala Asn Gly Glu Ile Ala Val Pro Asn Asp Ala Thr Asn Glu  
 115 120 125  
 Ser Arg Ala Leu Tyr Leu Leu Gln Ser Ala Gly Leu Ile Lys Leu Asp  
 130 135 140  
 Val Ser Gly Thr Ala Leu Ala Thr Val Ala Asn Ile Lys Glu Asn Pro  
 145 150 155 160  
 Lys Asn Leu Lys Ile Thr Glu Leu Asp Ala Ser Gln Thr Ala Arg Ser  
 165 170 175  
 Leu Ser Ser Val Asp Ala Ala Val Val Asn Asn Thr Phe Val Thr Glu  
 180 185 190  
 Ala Lys Leu Asp Tyr Lys Lys Ser Leu Phe Lys Glu Gln Ala Asp Glu  
 195 200 205  
 Asn Ser Lys Gln Trp Tyr Asn Ile Ile Val Ala Lys Lys Asp Trp Glu  
 210 215 220  
 Thr Ser Pro Lys Ala Asp Ala Ile Lys Lys Val Ile Ala Ala Tyr His  
 225 230 235 240  
 Thr Asp Asp Val Lys Lys Val Ile Glu Glu Ser Ser Asp Gly Leu Asp  
 245 250 255  
 Gln Pro Val Trp  
 260

## (2) INFORMATION FOR SEQ ID NO: 33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 640 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| TTCGAAAGGG TCAGAAGGTG CAGACCTTAT CAGCATGAAA GGGGATGTCA TTACAGAACA  | 60  |
| TCAATTTTAT GAGCAAGTGA AAAGCAACCC TTCAGCCCAA CAAGTCTTGT TAAATATGAC  | 120 |
| CATCCAAAAA GTTTTGTAAA AACAAATATGG CTCAGAGCTT GATGATAAAG AGGTTGATGA | 180 |
| TACTATTGCC GAAGAAAAAA AACAAATATGG CGAAACTAC CAACGTGTCT TGTCACAAGC  | 240 |
| AGGTATGACT CTTGAAACAC GTAAAGCTCA AATTCGTACA AGTAAATTAG TTGAGTTGGC  | 300 |
| AGTTAAGAAG GTAGCAGAAG CTGAATTGAC AGATGAAGCC TATAAGAAAG CCTTTGATGA  | 360 |
| GTACACTCCA GATGTAACGG CTCAAATCAT CCGTCTTAAT AATGAAGATA AGGCCAAAGA  | 420 |
| AGTTCTCGAA AAAGCCAAGG CAGAAGGTGC TGATTTTGCT CAATTAGCCA AAGATAATTC  | 480 |
| AACTGATGAA AAAACAAAAG AAAATGGTGG AGAAATTACC TTTGATTCTG CTTCAACAGA  | 540 |
| AGTACCTGGA GCAAGTCCAA AAAAGCCGCT TTTCGCTTTT AGATGTGGGA TGGTGTCTTCT | 600 |
| GGATGTGGAT TACAGCAACT GGGGCACACC AAGCCTACAG                        | 640 |

146

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

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 213 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

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

```

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

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 631 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

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

```

GGGGATGGCA GCTTTTAAAA ATCCTAACAA TCAATACAAA GCTATTACAA TTGCTCAAAC      60
TCTAGGTGAT GATGCTTCTT CAGAGGAATT GGCTGGTAGA TATGGTTCTG CTGTTCACTG      120
TACAGAAGTG ACTGCCTCAA ACCTTTCAAC AGTTAAAACT AAAGCTACGG TTGTAGAAAA      180
ACCACTGAAA GATTTTtagag CGTCTACGTC TGATCAGTCT GGTGGGTGG AATCTAATGG      240
TAAATGGTAT TTCTATGAGT CTGGTGATGT GAAGACAGGT TGGGTGAAAA CAGATGGTAA      300
ATGGTACTAT TTGAATGACT TAGGTGTCAT GCAGACTGGA TTTGTAAAAT TTTCTGGTAG      360
CTGGTATTAC TTGAGCAATT CAGGTGCTAT GTTTACAGGC TGGGGAACAG ATGGTAGCAG      420
ATGGTTCTAC TTTGACGGCT CAGGAGCTAT GAAGACAGGC TGGTACAAGG AAAATGGCAC      480
TTGGTATTAC CTTGACGAAG CAGGTATCAT GAAGACAGGT TGGTTTAAAG TCGGACCACA      540
CTGGTACTAT GCCTACGGTT CAGGAGCTTT GGCTGTGAGC ACAACAACAC CAGATGGTTA      600
CCGTGTAAAT GGTAATGGTG AATGGGTAAA C                                     631

```

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 210 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

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

```

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

```

148

Ala Met Phe Thr Gly Trp Gly Thr Asp Gly Ser Arg Trp Phe Tyr Phe  
 130 135 140

Asp Gly Ser Gly Ala Met Lys Thr Gly Trp Tyr Lys Glu Asn Gly Thr  
 145 150 155 160

Trp Tyr Tyr Leu Asp Glu Ala Gly Ile Met Lys Thr Gly Trp Phe Lys  
 165 170 175

Val Gly Pro His Trp Tyr Tyr Ala Tyr Gly Ser Gly Ala Leu Ala Val  
 180 185 190

Ser Thr Thr Thr Pro Asp Gly Tyr Arg Val Asn Gly Asn Gly Glu Trp  
 195 200 205

Val Asn  
 210

## (2) INFORMATION FOR SEQ ID NO: 37:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1360 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                   |      |
|-------------------------------------------------------------------|------|
| AGACGAGCAA AAAATTAAGC AAGCAGAAGC GGAAGTTGAG AGTAAACAAG CTGAGGCTAC | 60   |
| AAGGTAAAA AAAATCAAGA CAGATCGTGA AGAAGCAGAA GAAGAAGCTA AACGAAGAGC  | 120  |
| AGATGCTAAA GAGCAAGGTA AACCAAAGGG GCGGGCAAAA CGAGGAGTTC CTGGAGAGCT | 180  |
| AGCAACACCT GATAAAAAAG AAAATGATGC GAAGTCTTCA GATTCTAGCG TAGGTGAAGA | 240  |
| AACTCTTCCA AGCCCATCCC TGAAACCAGA AAAAAAGGTA GCAGAAGCTG AGAAGAAGGT | 300  |
| TGAAGAAGCT AAGAAAAAG CCGAGGATCA AAAAGAAGAA GATCGCCGTA ACTACCCAAC  | 360  |
| CAATACTTAC AAAACGCTTG AACTTGAAAT TGCTGAGTCC GATGTGGAAG TTAAAAAGC  | 420  |
| GGAGCTTGAA CTAGTAAAAG AGGAAGCTAA GGAACCTCGA AACGAGGAAA AAGTTAAGCA | 480  |
| AGCAAAAGCG GAAGTTGAGA GTAAAAAAGC TGAGGCTACA AGGTTAGAAA AAATCAAGAC | 540  |
| AGATCGTAAA AAAGCAGAAG AAGAAGCTAA ACGAAAAGCA GCAGAAGAAG ATAAAGTTAA | 600  |
| AGAAAAACCA GCTGAACAAC CACAACCAGC GCCGGCTCCA AAAGCAGAAA AACCAGCTCC | 660  |
| AGCTCCAAAA CCAGAGAATC CAGCTGAACA ACCAAAAGCA GAAAAACCAG CTGATCAACA | 720  |
| AGCTGAAGAA GACTATGCTC GTAGATCAGA AGAAGAATAT AATCGCTTGA CTCAACAGCA | 780  |
| ACCGCCAAAA ACTGAAAAAC CAGCACAACC ATCTACTCCA AAAACAGGCT GGAAACAAGA | 840  |
| AAACGGTATG TGGTACTTCT ACAATACTGA TGGTTCAATG GCGACAGGAT GGCTCCAAAA | 900  |
| CAATGGCTCA TGGTACTACC TCAACAGCAA TGGCGCTATG GCGACAGGAT GGCTCCAAAA | 960  |
| CAATGGTTCA TGGTACTATC TAAACGCTAA TGGTTCAATG GCAACAGGAT GGCTCCAAAA | 1020 |

149

CAATGGTTCA TGGTACTACC TAAACGCTAA TGGTTCAATG GCGACAGGAT GGCTCCAATA 1080  
 CAATGGCTCA TGGTACTACC TAAACGCTAA TGGTTCAATG GCGACAGGAT GGCTCCAATA 1140  
 CAATGGCTCA TGGTACTACC TAAACGCTAA TGGTGATATG GCGACAGGTT GGGTGAAAGA 1200  
 TGGAGATACC TGGTACTATC TTGAAGCATC AGGTGCTATG AAAGCAAGCC AATGGTTCAA 1260  
 AGTATCAGAT AAATGGTACT ATGTCAATGG CTCAGGTGCC CTTGCAGTCA ACACAACTGT 1320  
 AGATGGCTAT GGAGTCAATG CCAATGGTGA ATGGGTAAAC 1360

## (2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 453 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

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

150

| 195 |     |     |     | 200 |     |     |     | 205 |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Pro | Ala | Pro | Ala | Pro | Lys | Ala | Glu | Lys | Pro | Ala | Pro | Ala | Pro | Lys | Pro |
| 210 |     |     |     | 215 |     |     |     |     |     | 220 |     |     |     |     |     |
| Glu | Asn | Pro | Ala | Glu | Gln | Pro | Lys | Ala | Glu | Lys | Pro | Ala | Asp | Gln | Gln |
| 225 |     |     |     | 230 |     |     |     |     |     | 235 |     |     |     | 240 |     |
| Ala | Glu | Glu | Asp | Tyr | Ala | Arg | Arg | Ser | Glu | Glu | Glu | Tyr | Asn | Arg | Leu |
|     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |     |
| Thr | Gln | Gln | Gln | Pro | Pro | Lys | Thr | Glu | Lys | Pro | Ala | Gln | Pro | Ser | Thr |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Pro | Lys | Thr | Gly | Trp | Lys | Gln | Glu | Asn | Gly | Met | Trp | Tyr | Phe | Tyr | Asn |
|     | 275 |     |     |     |     | 280 |     |     |     |     |     | 285 |     |     |     |
| Thr | Asp | Gly | Ser | Met | Ala | Thr | Gly | Trp | Leu | Gln | Asn | Asn | Gly | Ser | Trp |
|     | 290 |     |     |     |     | 295 |     |     |     | 300 |     |     |     |     |     |
| Tyr | Tyr | Leu | Asn | Ser | Asn | Gly | Ala | Met | Ala | Thr | Gly | Trp | Leu | Gln | Asn |
| 305 |     |     |     | 310 |     |     |     |     |     | 315 |     |     |     | 320 |     |
| Asn | Gly | Ser | Trp | Tyr | Tyr | Leu | Asn | Ala | Asn | Gly | Ser | Met | Ala | Thr | Gly |
|     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |     |
| Trp | Leu | Gln | Asn | Asn | Gly | Ser | Trp | Tyr | Tyr | Leu | Asn | Ala | Asn | Gly | Ser |
|     |     | 340 |     |     |     |     |     | 345 |     |     |     | 350 |     |     |     |
| Met | Ala | Thr | Gly | Trp | Leu | Gln | Tyr | Asn | Gly | Ser | Trp | Tyr | Tyr | Leu | Asn |
|     | 355 |     |     |     |     | 360 |     |     |     |     |     | 365 |     |     |     |
| Ala | Asn | Gly | Ser | Met | Ala | Thr | Gly | Trp | Leu | Gln | Tyr | Asn | Gly | Ser | Trp |
|     | 370 |     |     |     |     | 375 |     |     |     | 380 |     |     |     |     |     |
| Tyr | Tyr | Leu | Asn | Ala | Asn | Gly | Asp | Met | Ala | Thr | Gly | Trp | Val | Lys | Asp |
| 385 |     |     |     | 390 |     |     |     |     |     | 395 |     |     |     | 400 |     |
| Gly | Asp | Thr | Trp | Tyr | Tyr | Leu | Glu | Ala | Ser | Gly | Ala | Met | Lys | Ala | Ser |
|     |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |     |
| Gln | Trp | Phe | Lys | Val | Ser | Asp | Lys | Trp | Tyr | Tyr | Val | Asn | Gly | Ser | Gly |
|     |     | 420 |     |     |     |     |     | 425 |     |     |     | 430 |     |     |     |
| Ala | Leu | Ala | Val | Asn | Thr | Thr | Val | Asp | Gly | Tyr | Gly | Val | Asn | Ala | Asn |
|     | 435 |     |     |     |     | 440 |     |     |     |     |     | 445 |     |     |     |
| Gly | Glu | Trp | Val | Asn |     |     |     |     |     |     |     |     |     |     |     |
|     | 450 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

## (2) INFORMATION FOR SEQ ID NO: 39:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 412 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

CTGTGGTGAG GAAGAACTA AAAAGACTCA AGCAGCACAA CAGCCAAAAC AACAAACGAC

60

151

```

TGTACAACAA ATTGCTGTTG GAAAAGATGC TCCAGACTTC ACATTGCAAT CCATGGATGG      120
CAAAGAAGTT AAGTTATCTG ATTTTAAGGG TAAAAAGGTT TACTTGAAGT TTTGGGCTTC      180
ATGGTGTGGT CCATGCAAGA AAAGTATGCC AGAGTTGATG GAACTAGCGG CGAAACCAGA      240
TCGTGATTTT GAAATTCTTA CTGTCATTGC ACCAGGAATT CAAGGTGAAA AACTGTTGA      300
GCAATTCCCA CAATGGTTCC AGGAACAAGG ATATAAGGAT ATCCAGTTC TTTATGATAC      360
CAAAGCAACC ACTTCCAAGC TTATCAAATT CGAAGCATTC CTACAGAATA TT                412

```

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 137 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

Cys Gly Glu Glu Glu Thr Lys Lys Thr Gln Ala Ala Gln Gln Pro Lys
1              5              10              15

Gln Gln Thr Thr Val Gln Gln Ile Ala Val Gly Lys Asp Ala Pro Asp
                20              25              30

Phe Thr Leu Gln Ser Met Asp Gly Lys Glu Val Lys Leu Ser Asp Phe
35              40              45

Lys Gly Lys Lys Val Tyr Leu Lys Phe Trp Ala Ser Trp Cys Gly Pro
50              55              60

Cys Lys Lys Ser Met Pro Glu Leu Met Glu Leu Ala Ala Lys Pro Asp
65              70              75              80

Arg Asp Phe Glu Ile Leu Thr Val Ile Ala Pro Gly Ile Gln Gly Glu
85              90              95

Lys Thr Val Glu Gln Phe Pro Gln Trp Phe Gln Glu Gln Gly Tyr Lys
100             105             110

Asp Ile Pro Val Leu Tyr Asp Thr Lys Ala Thr Thr Ser Lys Leu Ile
115             120             125

Lys Phe Glu Ala Phe Leu Gln Asn Ile
130             135

```

(2) INFORMATION FOR SEQ ID NO: 41:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1462 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                    |      |
|--------------------------------------------------------------------|------|
| GACTTTTAAC AATAAACTA TTGAAGAGTT GCACAATCTC CTTGTCTCTA AGGAAATTTTC  | 60   |
| TGCAACAGAA TTGACCCAAG CAACACTTGA AAATATCAAG TCTCGTGAGG AAGCCCTCAA  | 120  |
| TTCATTGTGC ACCATCGCTG AGGAGCAAGC TCTTGTTCAA GCTAAAGCCA TTGATGAAGC  | 180  |
| tGGAATTGAT GCTGACAATG TCCTTTCAGG AATTCCACTT GCTGTTAAGG ATAACATCTC  | 240  |
| TACAGACGGT ATTCTCACAA CTGCTGCCTC AAAAATGCTC TACAACTATG AGCCAATCTT  | 300  |
| TGATGCGACa gCTgTTGCCA ATGCAAAAAC CAAGGGCATG ATTGTCGTTG GAAAGACCAA  | 360  |
| CATGGACGAA TTTGCTATGG GTGGTTCAGG tGAAACTTCA CACTACGGAG CAACTAAAAA  | 420  |
| CGCTTGGAAC CACAGCAAGG TTCCTGGTGG GTCATCAAGT GGTTCTGCCG CAGCTGTAGC  | 480  |
| CTCAGGACAA GTTCGCTTGT CACTTGGTTC TGATACTGGT GGTTCCATCC GCCAACCTGC  | 540  |
| TGCCTTCAAC GGAATCGTTG GTCTCAAACC AACCTACGGA ACAGTTTCAC GTTTCGGTCT  | 600  |
| CATTGCCTTT GGTAGCTCAT TAGACCAGAT TGGACCTTTT GCTCCTACTG TTAAGGAAAA  | 660  |
| TGCCCTCTTG CTCAACGCTA TTGCCAGCGA AGATGCTAAA GACTCTACTT CTGCTCCTGT  | 720  |
| CCGCATCGCC GACTTTACTT CAAAATCGG CCAAGACATC AAGGGTATGA AAATCGCTTT   | 780  |
| GCCTAAGGAA TACCTAGGCG AAGGAATTGA TCCAGAGGTT AAGGAAACAA TCTTAAACGC  | 840  |
| GGCCAAACAC TTTGAAAAAT TGGGTGCTAT CGTCGAAGAA GTCAGCCTTC CTCACTCTAA  | 900  |
| ATACGGTGTT GCCGTTTATT ACATCATCGC TTCATCAGAA GCTTCATCAA ACTTGCAACG  | 960  |
| CTTCGACGGT ATCCGTTACG GCTATCGCGC AGAAGATGCA ACCAACCTTG ATGAAATCTA  | 1020 |
| TGTAAACAGC CGAAGCCAAG GTTTTGGTGA AGAGGTAAAA CGTCGTATCA TGCTGGGTAC  | 1080 |
| TTTCAGTCTT TCATCAGGTT ACTATGATGC CTAATAAAA AAGGCTGGTC AAGTCCGTAC   | 1140 |
| CCTCATCATT CAAGATTTTCG AAAAAGTCTT CGCGGATTAC GATTTGATTT TGGGTCCAAC | 1200 |
| TGCTCCAAGT GTTGCCTATG ACTTGGATTC TCTCAACCAT GACCCAGTTG CCATGTACTT  | 1260 |
| AGCCGACCTA TTGACCATAC CTGTAAACTT GGCAGGACTG CCTGGAATTT CGATTCTCTG  | 1320 |
| TGGATTCTCT CAAGGTCTAC CTGTCGGACT CCAATTGATT GGTCCCAAGT ACTCTGAGGA  | 1380 |
| AACCATTTAC CAAGCTGCTG CTGCTTTTGA AGCAACAACA GACTACCACA AACAACAACC  | 1440 |
| CGTGATTTTT GGAGGTGACA AC                                           | 1462 |

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 487 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

153

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

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

154

Asp Glu Ile Tyr Val Asn Ser Arg Ser Gln Gly Phe Gly Glu Glu Val  
                   340                                  345                                  350  
 Lys Arg Arg Ile Met Leu Gly Thr Phe Ser Leu Ser Ser Gly Tyr Tyr  
                   355                                  360                                  365  
 Asp Ala Tyr Tyr Lys Lys Ala Gly Gln Val Arg Thr Leu Ile Ile Gln  
                   370                                  375                                  380  
 Asp Phe Glu Lys Val Phe Ala Asp Tyr Asp Leu Ile Leu Gly Pro Thr  
                   385                                  390                                  395                                  400  
 Ala Pro Ser Val Ala Tyr Asp Leu Asp Ser Leu Asn His Asp Pro Val  
                                   405                                  410                                  415  
 Ala Met Tyr Leu Ala Asp Leu Leu Thr Ile Pro Val Asn Leu Ala Gly  
                                   420                                  425                                  430  
 Leu Pro Gly Ile Ser Ile Pro Ala Gly Phe Ser Gln Gly Leu Pro Val  
                   435                                  440                                  445  
 Gly Leu Gln Leu Ile Gly Pro Lys Tyr Ser Glu Glu Thr Ile Tyr Gln  
                   450                                  455                                  460  
 Ala Ala Ala Ala Phe Glu Ala Thr Thr Asp Tyr His Lys Gln Gln Pro  
                   465                                  470                                  475                                  480  
 Val Ile Phe Gly Gly Asp Asn  
                                   485

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 433 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| CTTTACAGGT AAACAACCTAC AAGTCGGCGA CAAGGCGCTT GATTTTTCTC TTACTACAAC | 60  |
| AGATCTTTCT AAAAAATCTC TGGCTGATTT TGATGGCAAG AAAAAAGTCT TGAGTGTCGT  | 120 |
| TCCTTCTATC GATACAGGCA TCTGCTCAAC TCAACACGT CGTTTAAATG AAGAATTGGC   | 180 |
| TGGACTGGAC AACACGGTCG TATTGACTGT TTCAATGGAC CTACCTTTTG CTCAAAAACG  | 240 |
| TTGGTGCGGT GCTGAAGGCC TTGACAATGC CATTATGCTT TCAGACTACT TTGACCATTC  | 300 |
| TTTCGGGCGC GATTATGCCC TCTTGATCAA CGAATGGCAC CTATTAGCAC GCGCAGTCTT  | 360 |
| TGTCCTCGAT ACTGACAATA CGATTCGCTA CGTTGAATAC GTGGATAATA TCAATTCTGA  | 420 |
| GCCAAACTTC GAA                                                     | 433 |

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 144 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

155

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

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

```

(2) INFORMATION FOR SEQ ID NO: 45:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 724 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

```

CCAGGCTGAT ACAAGTATCG CAGACATTCA AAAAAGAGGC GAACTGGTTG TCGGTGTCAA      60
ACAAGACGTT CCCAATTTTG GTTACAAnGA TCCCAAGACC GGTACTTATT CTGGTATCGA      120
AaCCGACTTG GCCAAGATGG TAGCTGATGA ACTCAAGGTC AAGATTCGCT ATGTGCCGGT      180
TACAGCACAA ACCCGCGGCC CCCTTCTAGA CAATGAACAG GTCGATATGG ATATCGCGAC      240
CTTTACCATC ACGGACGAAC GCAAAAAACT CTACAACTTT ACCAGTCCCT ACTACACAGA      300
CGCTTCTGGA TTTTGGTCA ATAAATCTGC CAAAATCAAA AAGATTGAGG ACCTAAACGG      360
CAAAACCATC GGAGTCGCCC AAGGTTCTAT CACCCAACGC CTGATTACTG AACTGGGTAA      420
AAAGAAAGGT CTGAAGTTTA AATTCGTCGA ACTTGTTCC TACCCAGAAT TGATTACTTC      480

```

156

CCTGCACGCT CATCGTATCG ATACCTTTTC CGTTGACCGC TCTATTCTAT CTGGCTACAC 540  
 TAGTAAACGG ACAGCACTAC TAGATGATAG TTTCAAGCCA TCTGACTACG GTATTGTTAC 600  
 CAAGAAATCA AATACAGAGC TCAACGACTA TCTTGATAAC TTGGTTACTA AATGGAGCAA 660  
 GGATGGTAGT TTGCAGAAAC TTTATGACCG TTACAAGCTC AAACCATCTA GCCATACTGC 720  
 AGAT 724

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

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 241 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

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

157

Asp Tyr Leu Asp Asn Leu Val Thr Lys Trp Ser Lys Asp Gly Ser Leu  
 210 215 220

Gln Lys Leu Tyr Asp Arg Tyr Lys Leu Lys Pro Ser Ser His Thr Ala  
 225 230 235 240

Asp

(2) INFORMATION FOR SEQ ID NO: 47:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1279 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

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

|                                                                    |      |
|--------------------------------------------------------------------|------|
| GTCTGTATCA TTTGAAAACA AAGAAACAAA CCGTGGTGTC TTgACTTTCA CTATCTCTCA  | 60   |
| AGACCAAATC AAACCAGAAT TGGACCGTGT CTTCAAGtCA GTGAAGAAAT CTCTTAATGT  | 120  |
| TCCAGGTTTC CGTAAAGGTC ACCTTCCACG CCCTATCTTC GACCAAAAAT TTGGTGAAGA  | 180  |
| AGCTCTTTAT CAAGATGCAA TGAACGCACT TTTGCCAAAC GCTTATGAAG CAGCTGTAAA  | 240  |
| AGAAGCTGGT CTTGAAGTGG TTGCCCAACC AAAAATTGAC GTAAC TTCAA TGGAAAAAGG | 300  |
| TCAAGACTGG GTTATCACTG CTGAAGTCGT TACAAAACCT GAAGTAAAAT TGGGTGACTA  | 360  |
| CAAAAACCTT GAAGTATCAG TTGATGTAGA AAAAGAAGTA ACTGACGCTG ATGTCGAAGA  | 420  |
| GCGTATCGAA CGCGAACGCA ACAACCTGGC TGAATTGGTT ATCAAGGAAG CTGCTGCTGA  | 480  |
| AAACGGCGAC ACTGTTGTGA TCGACTTCGT TGGTTCATC GACGGTGTG AATTTGACGG    | 540  |
| TGGAAAAGGT GAAAACCTCT CACTTGGACT TGGTTCAGGT CAATTCATCC CTGGTTTCGA  | 600  |
| AGACCAATTG GTAGGTCACT CAGCTGGCGA AACCGTTGAT GTTATCGTAA CATTCCCAGA  | 660  |
| AGACTACCAA GCAGAAGACC TTGCAGGTAA AGAAGCTAAA TTCGTGACAA CTATCCACGA  | 720  |
| AGTAAAAGCT AAAGAAGTTC CGGCTCTTGA CGATGAACTT GCAAAAAGACA TTGATGAAGA | 780  |
| AGTTGAAACA CTTGCTGACT TGAAAGAAAA ATACAGCAAA GAATTGGCTG CTGCTAAAGA  | 840  |
| AGAAGCTTAC AAAGATGCAG TTGAAGGTGC AGCAATTGAT ACAGCTGTAG AAAATGCTGA  | 900  |
| AATCGTAGAA CTTCCAGAAG AAATGATCCA TGAAGAAGTT CACCGTTCAG TAAATGAATT  | 960  |
| CCTTGGGAAT TTGCAACGTC AAGGGATCAA CCCTGACATG TACTTCCAAA TCACTGGAAC  | 1020 |
| TACTCAAGAA GACCTTCACA ACCAATACCA AGCAGAAGCT GAGTCACGTA CTAAGACTAA  | 1080 |
| CCTTGTTATC GAAGCAGTTG CCAAAGCTGA AGGATTGAT GCTTCAGAAG AAGAAATCCA   | 1140 |
| AAAAGAAGTT GAGCAATTGG CAGCAGACTA CAACATGGAA GTTGCAACAAG TTCAAACTT  | 1200 |
| GCTTTCAGCT GACATGTTGA AACATGATAT CACTATCAAA AAAGCTGTTG AATTGATCAC  | 1260 |
| AAGCACAGCA ACAGTAAAA                                               | 1279 |

158

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

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 426 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

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

159

| 260                                                             | 265                                 | 270 |
|-----------------------------------------------------------------|-------------------------------------|-----|
| Lys Glu Leu Ala Ala Ala Lys                                     | Glu Glu Ala Tyr Lys Asp Ala Val Glu |     |
| 275                                                             | 280                                 | 285 |
| Gly Ala Ala Ile Asp Thr Ala Val Glu Asn Ala Glu Ile Val Glu Leu |                                     |     |
| 290                                                             | 295                                 | 300 |
| Pro Glu Glu Met Ile His Glu Glu Val His Arg Ser Val Asn Glu Phe |                                     |     |
| 305                                                             | 310                                 | 315 |
| Leu Gly Asn Leu Gln Arg Gln Gly Ile Asn Pro Asp Met Tyr Phe Gln |                                     |     |
|                                                                 | 325                                 | 330 |
| Ile Thr Gly Thr Thr Gln Glu Asp Leu His Asn Gln Tyr Gln Ala Glu |                                     |     |
|                                                                 | 340                                 | 345 |
| Ala Glu Ser Arg Thr Lys Thr Asn Leu Val Ile Glu Ala Val Ala Lys |                                     |     |
|                                                                 | 355                                 | 360 |
| Ala Glu Gly Phe Asp Ala Ser Glu Glu Glu Ile Gln Lys Glu Val Glu |                                     |     |
|                                                                 | 370                                 | 375 |
| Gln Leu Ala Ala Asp Tyr Asn Met Glu Val Ala Gln Val Gln Asn Leu |                                     |     |
| 385                                                             | 390                                 | 395 |
| Leu Ser Ala Asp Met Leu Lys His Asp Ile Thr Ile Lys Lys Ala Val |                                     |     |
|                                                                 | 405                                 | 410 |
| Glu Leu Ile Thr Ser Thr Ala Thr Val Lys                         |                                     |     |
|                                                                 | 420                                 | 425 |

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 490 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| TGGTCAAAAG GAAAGTCAGA CAGGAAAGGG GATGAAAATT GTGACCAGTT TTTATCCTAT | 60  |
| CTACGCTATG GTTAAGGAAG TATCTGGTGA CTTGAATGAT GTTCGGATGA TTCAGTCAAG | 120 |
| TAGTGGTATT CACTCCTTTG AACCTTCGGC AAATGATATC GCAGCCATCT ATGATGCAGA | 180 |
| TGTCTTTGTT TACCATTCTC ATACACTCGA ATCTTGGGCA GGAAGTCTGG ATCCAAATCT | 240 |
| AAAAAATCC AAAGTGAAGG TCTTAGAGGC TTCTGAGGGA ATGACCTTGG AACGTGTCCC  | 300 |
| TGGACTAGAG GATGTGGAAG CAGGGGATGG AGTTGATGAA AAAACGCTCT ATGACCCTCA | 360 |
| CACATGGCTA GATCCTGAAA AAGCTGGAGA AGAAGCCCAA ATTATCGCTG ATAAACTTTC | 420 |
| AGAGGTGGAT AGTGAGCATA AAGAGACTTA TCAAAAAAAT GCGCAACCTT TATCAAAAAA | 480 |
| GCTCAGGAAT                                                        | 490 |

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

160

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

(ii) MOLECULE TYPE: protein

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

```

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

```

(2) INFORMATION FOR SEQ ID NO: 51:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1006 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

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

```

GAAGGATAGA TATATTTTAG CATTTGAGAC ATCCTGTGAT GAGACCAGTG TCGCCGTCTT      60
GAAAAACGAC GATGAGCTCT TGTCCAATGT CATTGCTAGT CAAATTGAGA GTCACAAACG      120
TTTTGGTGGC GTAGTGCCCC AAGTAGCCAG TCGTCACCAT GTCGAGGTCA TTACAGCCTG      180

```

161

```

TATCGAGGAG GCATTGGCAG AAGCAGGGAT TACCGAAGAG GACGTGACAG CTGTTGCGGT      240
TACCTACGGA CCAGGCTTGG TCGGAGCCTT GCTAGTTGGT TTGTCAGCTG CCAAGGCCTT      300
TGCTTGGGCT CACGGACTTC CACTGATTCC TGTTAATCAC ATGGCTGGGC ACCTCATGGC      360
AGCTCAGAGT GTGGAGCCTT TGGAGTTTCC CTTGCTAGCC CTCTTGGTCA GCGGCGGACA      420
CACAGAGTTG GTTTATGTTT CGGAGGCAGG AGATTATAAG ATTGTTGGGG AAACCCGTGA      480
TGATGCGGTT GGTGAGGCTT ATGATAAGGT CGGCCGTGTC ATGGGCTTGA CCTATCCTGC      540
AGGTCGTGAG ATTGACGAGC TGGCTCATCA GGGGCAGGAT ATTTATGATT TCCCCCGTGC      600
CATGATTAAG GAAGATAATC TGGAGTTTCT CTTCTCAGGT TTGAAATCTG CCTTTATCAA      660
TCTTCATCAC AATGCCGAGC AAAAGGGAGA AAGCCTGTCT ACAGAAGATT TGTGTGCTTC      720
CTTCCAAGCA GCAGTTATGG ACATTCTCAT GGCAAAAACC AAGAAGGCTT TGGAGAAATA      780
TCCTGTAAAT ATCCTAGTTG TGGCAGGTGG TGTGGCAGCC AATAAAGGTC TCAGAGAACG      840
CCTAGCAGCC GAAATCACAG ATGTCAAGGT TATCATCCCC CCTCTGCGAC TCTGCGGAGA      900
CAATGCAGGT ATGATTGCCT ATGCCAGCGT CAGCNAGTGG AACAAAGAAA ACTTCGCAGG      960
CTGGGACCTC AATGCCAAAC CAAGTCTTGC CTTTGATACC ATGGAA                      1006

```

## (2) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 335 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

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

```

162

His Met Ala Gly His Leu Met Ala Ala Gln Ser Val Glu Pro Leu Glu  
 115 120 125  
 Phe Pro Leu Leu Ala Leu Leu Val Ser Gly Gly His Thr Glu Leu Val  
 130 135 140  
 Tyr Val Ser Glu Ala Gly Asp Tyr Lys Ile Val Gly Glu Thr Arg Asp  
 145 150 155 160  
 Asp Ala Val Gly Glu Ala Tyr Asp Lys Val Gly Arg Val Met Gly Leu  
 165 170 175  
 Thr Tyr Pro Ala Gly Arg Glu Ile Asp Glu Leu Ala His Gln Gly Gln  
 180 185 190  
 Asp Ile Tyr Asp Phe Pro Arg Ala Met Ile Lys Glu Asp Asn Leu Glu  
 195 200 205  
 Phe Ser Phe Ser Gly Leu Lys Ser Ala Phe Ile Asn Leu His His Asn  
 210 215 220  
 Ala Glu Gln Lys Gly Glu Ser Leu Ser Thr Glu Asp Leu Cys Ala Ser  
 225 230 235 240  
 Phe Gln Ala Ala Val Met Asp Ile Leu Met Ala Lys Thr Lys Lys Ala  
 245 250 255  
 Leu Glu Lys Tyr Pro Val Lys Ile Leu Val Val Ala Gly Gly Val Ala  
 260 265 270  
 Ala Asn Lys Gly Leu Arg Glu Arg Leu Ala Ala Glu Ile Thr Asp Val  
 275 280 285  
 Lys Val Ile Ile Pro Pro Leu Arg Leu Cys Gly Asp Asn Ala Gly Met  
 290 295 300  
 Ile Ala Tyr Ala Ser Val Ser Xaa Trp Asn Lys Glu Asn Phe Ala Gly  
 305 310 315 320  
 Trp Asp Leu Asn Ala Lys Pro Ser Leu Ala Phe Asp Thr Met Glu  
 325 330 335

## (2) INFORMATION FOR SEQ ID NO: 53:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1000 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| GGTAGTTAAA GTTGGTATTA ACGGTTTCGG ACGTATCGGT CGTCTTGCTT TCCGTCGTAT | 60  |
| CCAAAACGTA GAAGGTGTTG AAGTTACACG CATCAACGAC CTTACAGATC CAGTTATGCT | 120 |
| TGCACACTTG TTGAAATACG ACACAACTCA AGGTCGTTTC GACGGTACTG TTGAAGTTAA | 180 |
| AGAAGGTGGA TTTGAAGTTA ACGGTAAATT CATCAAAGTT TCTGCTGAAC GTGATCCAGA | 240 |
| ACAAATCGAC TGGGCTACTG ACGGTGTAGA AATCGTTCTT GAAGCTACTG GTTTCCTTGC | 300 |

163

```

TAAGAAAGAA GCAGCTGAAA AACACCTTAA AGGTGGAGCT AAAAAAGTTG TTATCACTGC      360
TCCTGGTGGA AACGACGTTA AAACAGTTGT ATTCAACACT AACCACGACG TTCTTGACGG      420
TACTGAAACA GTTATCTCAG GTGCTTCATG TACTACAAAC TGCTTGCTC CAATGGCTAA      480
AGCTCTTCAA GACAACCTTG GTGTTGTTGA AGGATTGATG ACTACTATCC ACGCTTACAC      540
TGGTGACCAA ATGATCCTTG ACGGACCACA CCGTGGTGGT GACCTTCGCC GTGCTCGCGC      600
TGGTGCTGCA AACATCGTTC CTAACCTAAC TGGTGCTGCA AAAGCTATCG GTCTTGTAAT      660
CCCAGAAATTG AATGGTAAAC TTGACGGATC TGCACAACGC GTTCCAACCTC CAACTGGATC      720
AGTTACTGAA TTGGTAGCAG TTCTTGAAAA GAACGTTACT GTTGATGAAG TGAACGCAGC      780
TATGAAAGCA GCTTCAAACG AATCATACGG TTACACAGAA GATCCAATCG TATCTTCAGA      840
TATCGTAGGT ATGTCTTACG GTTCATTGTT TGACGCAACT CAACTAAAG TTCTTGACGT      900
TGACGGTAAA CAATTGGTTA AAGTTGTATC ATGGTACGAC AACGAAATGT CATACTGTC      960
ACAACTTGTT CGTACTCTTG GAATACTTCG CAAAAATTGC      1000

```

## (2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 333 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

```

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

```

164

| 130                                                                                | 135 | 140 |
|------------------------------------------------------------------------------------|-----|-----|
| Ile Ser Gly Ala Ser Cys Thr Thr Asn Cys Leu Ala Pro Met Ala Lys<br>145 150 155 160 |     |     |
| Ala Leu Gln Asp Asn Phe Gly Val Val Glu Gly Leu Met Thr Thr Ile<br>165 170 175     |     |     |
| His Ala Tyr Thr Gly Asp Gln Met Ile Leu Asp Gly Pro His Arg Gly<br>180 185 190     |     |     |
| Gly Asp Leu Arg Arg Ala Arg Ala Gly Ala Ala Asn Ile Val Pro Asn<br>195 200 205     |     |     |
| Ser Thr Gly Ala Ala Lys Ala Ile Gly Leu Val Ile Pro Glu Leu Asn<br>210 215 220     |     |     |
| Gly Lys Leu Asp Gly Ser Ala Gln Arg Val Pro Thr Pro Thr Gly Ser<br>225 230 235 240 |     |     |
| Val Thr Glu Leu Val Ala Val Leu Glu Lys Asn Val Thr Val Asp Glu<br>245 250 255     |     |     |
| Val Asn Ala Ala Met Lys Ala Ala Ser Asn Glu Ser Tyr Gly Tyr Thr<br>260 265 270     |     |     |
| Glu Asp Pro Ile Val Ser Ser Asp Ile Val Gly Met Ser Tyr Gly Ser<br>275 280 285     |     |     |
| Leu Phe Asp Ala Thr Gln Thr Lys Val Leu Asp Val Asp Gly Lys Gln<br>290 295 300     |     |     |
| Leu Val Lys Val Val Ser Trp Tyr Asp Asn Glu Met Ser Tyr Thr Ala<br>305 310 315 320 |     |     |
| Gln Leu Val Arg Thr Leu Gly Ile Leu Arg Lys Asn Cys<br>325 330                     |     |     |

## (2) INFORMATION FOR SEQ ID NO: 55:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2389 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| TTCTTACGAG TTGGGACTGT ATCAAGCTAG AACGGTTAAG GAAAATAATC GTGTTTCCTA | 60  |
| TATAGATGGA AAACAAGCGA CGCAAAAAAC GGAGAATTG ACTCCTGATG AGGTTAGCAA  | 120 |
| GCGTGAAGGA ATCAATGCTG AGCAAATCGT CATCAAGATA ACAGACCAAG GCTATGTCAC | 180 |
| TTCACATGGC GACCACTATC ATTATTACAA TGGTAAGGTT CCTTATGACG CTATCATCAG | 240 |
| TGAAGAATTA CTCATGAAAG ATCCAACTA TAAGCTAAAA GATGAGGATA TGTTAATGA   | 300 |
| GGTCAAGGGT GGATATGTTA TCAAGGTAGA TGGAAAATAC TATGTTTACC TTAAGGATGC | 360 |
| TGCCCACGCG GATAACGTCC GTACAAAAGA GGAAATCAAT CGACAAAAAC AAGAGCATAG | 420 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| TCAACATCGT GAAGGTGGAA CTCCAAGAAA CGATGGTGCT GTTGCCTTGG CACGTTTCGCA | 480  |
| AGGACGCTAT ACTACAGATG ATGGTTATAT CTTTAATGCT TCTGATATCA TAGAGGATAC  | 540  |
| TGGTGATGCT TATATCGTTC CTCATGGAGA TCATTACCAT TACATTCCTA AGAATGAGTT  | 600  |
| ATCAGCTAGC GAGTTGGCTG CTGCAGAAGC CTTCTATCT GGTGAGGAA ATCTGTCAAA    | 660  |
| TTCAAGAACC TATCGCCGAC AAAATAGCGA TAACACTTCA AGAACAAACT GGGTACCTTC  | 720  |
| TGTAAGCAAT CCAGGAATA CAAATACTAA CACAAGCAAC AACAGCAACA CTAACAGTCA   | 780  |
| AGCAAGTCAA AGTAATGACA TTGATAGTCT CTTGAAACAG CTCTACAAAC TGCCTTTGAG  | 840  |
| TCAACGACAT GTAGAATCTG ATGGCCTTGT CTTTGATCCA GCACAAATCA CAAGTCGAAC  | 900  |
| AGCTAGAGGT GTTGACGTGC CACACGGAGA TCATTACCAC TTCATCCCTT ACTCTCAAAT  | 960  |
| GTCTGAATTG GAAGAACGAA TCGCTCGTAT TATTCCCCTT CGTTATCGTT CAAACCATTG  | 1020 |
| GGTACCAGAT TCAAGGCCAG AACCAACCAAG TCCACAACCG ACTCCGGAAC CTAGTCCAGG | 1080 |
| CCCGCAACCT GCACCAAATC TTAAAATAGA CTCAAATTCT TCTTTGGTTA GTCAGCTGGT  | 1140 |
| ACGAAAAGTT GGGGAAGGAT ATGTATTCTG AGAAAAGGGC ATCTCTCGTT ATGTCTTTGC  | 1200 |
| GAAAGATTTA CCATCTGAAA CTGTTAAAAA TCTTGAAAGC AAGTTATCAA AACAAAGAGAG | 1260 |
| TGTTTCACAC ACTTTAACTG CTAATAAAGA AAATGTTGCT CCTCGTGACC AAGAATTTTA  | 1320 |
| TGATAAAGCA TATAATCTGT TAACTGAGGC TCATAAAGCC TTGTTTGNA AATAAGGGTCG  | 1380 |
| TAATTCTGAT TTCCAAGCCT TAGACAAATT ATTAGAACGC TTGAATGATG AATCGACTAA  | 1440 |
| TAAAGAAAAA TTGGTAGATG ATTTATTGGC ATTCCTAGCA CCAATTACCC ATCCAGAGCG  | 1500 |
| ACTTGACAAA CCAAATTCTC AAATTGAGTA TACTGAAGAC GAAGTTCGTA TTGCTCAATT  | 1560 |
| AGCTGATAAG TATACAACGT CAGATGGTTA CATTTTTGAT GAACATGATA TAATCAGTGA  | 1620 |
| TGAAGGAGAT GCATATGTAA CGCCTCATAT GGGCCATAGT CACTGGATTG GAAAAGATAG  | 1680 |
| CCTTTCTGAT AAGGAAAAAG TTGCAGCTCA AGCCTATACT AAAGAAAAAG GTATCCTACC  | 1740 |
| TCCATCTCCA GACGCAGATG TTAAAGCAAA TCCAACCTGGA GATAGTCAG CAGCTATTTA  | 1800 |
| CAATCGTGTG AAAGGGGAAA AACGAATTCC ACTCGTTCGA CTTCCATATA TGTTGAGCA   | 1860 |
| TACAGTTGAG GTTAAAAACG GTAATTTGAT TATTCCTCAT AAGGATCATT ACCATAATAT  | 1920 |
| TAAATTTGCT TGGTTTGATG ATCACACATA CAAAGCTCCA AATGGCTATA CCTTGGAAGA  | 1980 |
| TTTGTGTCG ACGATTAAGT ACTACGTAGA ACACCCTGAC GAACGTCCAC ATTCTAATGA   | 2040 |
| TGGATGGGGC AATGCCAGTG AGCATGTGTT AGGCAAGAAA GACCACAGTG AAGATCCAAA  | 2100 |
| TAAGAACTTC AAAGCGGATG AAGAGCCAGT AGAGGAAACA CCTGCTGAGC CAGAAGTCCC  | 2160 |
| TCAAGTAGAG ACTGAAAAAG TAGAAGCCCA ACTCAAAGAA GCAGAAGTTT TGCTTGCGAA  | 2220 |
| AGTAACGGAT TCTAGTCTGA AAGCCAATGC AACAGAACT CTAGCTGGTT TACGAAATAA   | 2280 |
| TTTGACTCTT CAAATTATGG ATAACAATAG TATCATGGCA GAAGCAGAAA AATTACTTGC  | 2340 |

166

GTTGTTAAAA GGAAGTAATC CTTTCATCTGT AAGTAAGGAA AAAATAAAC

2389

## (2) INFORMATION FOR SEQ ID NO:56:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 796 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

```

Ser Tyr Glu Leu Gly Leu Tyr Gln Ala Arg Thr Val Lys Glu Asn Asn
1           5           10           15

Arg Val Ser Tyr Ile Asp Gly Lys Gln Ala Thr Gln Lys Thr Glu Asn
20          25          30

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

Ile Val Ile Lys Ile Thr Asp Gln Gly Tyr Val Thr Ser His Gly Asp
50          55          60

His Tyr His Tyr Tyr Asn Gly Lys Val Pro Tyr Asp Ala Ile Ile Ser
65          70          75          80

Glu Glu Leu Leu Met Lys Asp Pro Asn Tyr Lys Leu Lys Asp Glu Asp
85          90          95

Ile Val Asn Glu Val Lys Gly Gly Tyr Val Ile Lys Val Asp Gly Lys
100         105         110

Tyr Tyr Val Tyr Leu Lys Asp Ala Ala His Ala Asp Asn Val Arg Thr
115         120         125

Lys Glu Glu Ile Asn Arg Gln Lys Gln Glu His Ser Gln His Arg Glu
130         135         140

Gly Gly Thr Pro Arg Asn Asp Gly Ala Val Ala Leu Ala Arg Ser Gln
145         150         155         160

Gly Arg Tyr Thr Thr Asp Asp Gly Tyr Ile Phe Asn Ala Ser Asp Ile
165         170         175

Ile Glu Asp Thr Gly Asp Ala Tyr Ile Val Pro His Gly Asp His Tyr
180         185         190

His Tyr Ile Pro Lys Asn Glu Leu Ser Ala Ser Glu Leu Ala Ala Ala
195         200         205

Glu Ala Phe Leu Ser Gly Arg Gly Asn Leu Ser Asn Ser Arg Thr Tyr
210         215         220

Arg Arg Gln Asn Ser Asp Asn Thr Ser Arg Thr Asn Trp Val Pro Ser
225         230         235         240

Val Ser Asn Pro Gly Thr Thr Asn Thr Asn Thr Ser Asn Asn Ser Asn
245         250         255

```

167

Thr Asn Ser Gln Ala Ser Gln Ser Asn Asp Ile Asp Ser Leu Leu Lys  
 260 265 270  
 Gln Leu Tyr Lys Leu Pro Leu Ser Gln Arg His Val Glu Ser Asp Gly  
 275 280 285  
 Leu Val Phe Asp Pro Ala Gln Ile Thr Ser Arg Thr Ala Arg Gly Val  
 290 295 300  
 Ala Val Pro His Gly Asp His Tyr His Phe Ile Pro Tyr Ser Gln Met  
 305 310 315 320  
 Ser Glu Leu Glu Glu Arg Ile Ala Arg Ile Ile Pro Leu Arg Tyr Arg  
 325 330 335  
 Ser Asn His Trp Val Pro Asp Ser Arg Pro Glu Gln Pro Ser Pro Gln  
 340 345 350  
 Pro Thr Pro Glu Pro Ser Pro Gly Pro Gln Pro Ala Pro Asn Leu Lys  
 355 360 365  
 Ile Asp Ser Asn Ser Ser Leu Val Ser Gln Leu Val Arg Lys Val Gly  
 370 375 380  
 Glu Gly Tyr Val Phe Glu Glu Lys Gly Ile Ser Arg Tyr Val Phe Ala  
 385 390 395 400  
 Lys Asp Leu Pro Ser Glu Thr Val Lys Asn Leu Glu Ser Lys Leu Ser  
 405 410 415  
 Lys Gln Glu Ser Val Ser His Thr Leu Thr Ala Lys Lys Glu Asn Val  
 420 425 430  
 Ala Pro Arg Asp Gln Glu Phe Tyr Asp Lys Ala Tyr Asn Leu Leu Thr  
 435 440 445  
 Glu Ala His Lys Ala Leu Phe Xaa Asn Lys Gly Arg Asn Ser Asp Phe  
 450 455 460  
 Gln Ala Leu Asp Lys Leu Leu Glu Arg Leu Asn Asp Glu Ser Thr Asn  
 465 470 475 480  
 Lys Glu Lys Leu Val Asp Asp Leu Leu Ala Phe Leu Ala Pro Ile Thr  
 485 490 495  
 His Pro Glu Arg Leu Gly Lys Pro Asn Ser Gln Ile Glu Tyr Thr Glu  
 500 505 510  
 Asp Glu Val Arg Ile Ala Gln Leu Ala Asp Lys Tyr Thr Thr Ser Asp  
 515 520 525  
 Gly Tyr Ile Phe Asp Glu His Asp Ile Ile Ser Asp Glu Gly Asp Ala  
 530 535 540  
 Tyr Val Thr Pro His Met Gly His Ser His Trp Ile Gly Lys Asp Ser  
 545 550 555 560  
 Leu Ser Asp Lys Glu Lys Val Ala Ala Gln Ala Tyr Thr Lys Glu Lys  
 565 570 575  
 Gly Ile Leu Pro Pro Ser Pro Asp Ala Asp Val Lys Ala Asn Pro Thr  
 580 585 590  
 Gly Asp Ser Ala Ala Ala Ile Tyr Asn Arg Val Lys Gly Glu Lys Arg

168

|                                                                                    |     |     |
|------------------------------------------------------------------------------------|-----|-----|
| 595                                                                                | 600 | 605 |
| Ile Pro Leu Val Arg Leu Pro Tyr Met Val Glu His Thr Val Glu Val<br>610 615 620     |     |     |
| Lys Asn Gly Asn Leu Ile Ile Pro His Lys Asp His Tyr His Asn Ile<br>625 630 635 640 |     |     |
| Lys Phe Ala Trp Phe Asp Asp His Thr Tyr Lys Ala Pro Asn Gly Tyr<br>645 650 655     |     |     |
| Thr Leu Glu Asp Leu Phe Ala Thr Ile Lys Tyr Tyr Val Glu His Pro<br>660 665 670     |     |     |
| Asp Glu Arg Pro His Ser Asn Asp Gly Trp Gly Asn Ala Ser Glu His<br>675 680 685     |     |     |
| Val Leu Gly Lys Lys Asp His Ser Glu Asp Pro Asn Lys Asn Phe Lys<br>690 695 700     |     |     |
| Ala Asp Glu Glu Pro Val Glu Glu Thr Pro Ala Glu Pro Glu Val Pro<br>705 710 715 720 |     |     |
| Gln Val Glu Thr Glu Lys Val Glu Ala Gln Leu Lys Glu Ala Glu Val<br>725 730 735     |     |     |
| Leu Leu Ala Lys Val Thr Asp Ser Ser Leu Lys Ala Asn Ala Thr Glu<br>740 745 750     |     |     |
| Thr Leu Ala Gly Leu Arg Asn Asn Leu Thr Leu Gln Ile Met Asp Asn<br>755 760 765     |     |     |
| Asn Ser Ile Met Ala Glu Ala Glu Lys Leu Leu Ala Leu Leu Lys Gly<br>770 775 780     |     |     |
| Ser Asn Pro Ser Ser Val Ser Lys Glu Lys Ile Asn<br>785 790 795                     |     |     |

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1180 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| TACTGAGATG CATCATAATC TAGGAGCTGA AAAGCGTTCA GCAGTGGCTA CTACTATCGA  | 60  |
| TAGTTTTTAAG GAGCGAAGTC AAAAAGTCAG AGCACTATCT GATCCAAATG TCGTTTTTGT | 120 |
| TCCCTTCTTT GGCTCTAGTG AATGGCTTCG TTTTGACGGT GCTCATTCTG CGGTATTAGC  | 180 |
| TGAGAAATAC AATCGTTCCT ACCGTCCTTA TCTTTTAGGA CAGGGGGGAG CTGCATCGCT  | 240 |
| TAACCAATAT TTTGGAATGC AACAGATGTT ACCACAGCTG GAGAATAAAC AAGTTGTGTA  | 300 |
| TGTTATCTCA CCTCAGTGGT TCAGTAAAAA TGGCTATGAT CCAGCAGCCT TCCAGCAGTA  | 360 |
| TTTTAATGGA GACCAGTTGA CTAGTTTCT GAAACATCAA TCTGGGGATC AGGCTAGTCA   | 420 |

169

```

ATATGCAGCG ACTCGCTTAC TGCAACAGTT CCCAAACGTA GCTATGAAGG ACCTGGTTCA      480
GAAGTTGGCA AGTAAAGAAG AATTGTGCGAC AGCAGACAAT GAAATGATTG AATTATTGGC      540
TCGTTTTAAT GAACGCCAAG CTCCTTTTTT TGGTCAGTTT TCGGTTAGAG GCTATGTTAA      600
CTACGATAAG CATGTAGCTA AGTATTTAAA AATCTTGCCA GACCAGTTTT CTTATCAGGC      660
AATAGAAGAT GTTGTCAAAG CAGATGCTGA AAAAAATACT TCCAATAATG AGATGGGAAT      720
GGAAAATTAT TTCTATAATG AGCAGATCAA GAAGGATTTG AAGAAATTAA AGGATTCTCA      780
GAAAAGCTTT ACCTATCTCA AGTCGCCAGA GTATAATGNN TTGCAGTTGG TTTTAACACA      840
GTTTTCTAAA TCTAAGGTAA ACCCGATTTT TATCATTTCA CCTGTTAATA AAAAATGGAT      900
GNACTATGCT GGTCTACGAG AGGATATGTA CCAACAAACG GTGCAGAAGA TTCGCTACCA      960
GTTAGAAAGT CAAGGTTTTA CCAATATAGC AGATTTTCTT AAGGACGGCG GGGAGCCTTT     1020
CTTTATGAAG GACACCATTC ACCTTGGTTG GTTGGGTTGG TTGGCTTTTG ACAAGGCAGT     1080
TGATCCTTTC CTATCCAATC CCACACCAGC TCCGACTTAC CATCTGAATG AGCGCTTTTT     1140
CAGCAAAGAT TGGGCGACTT ATGATGGAGA TGTCAAAGAA                               1180

```

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 393 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

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

```

170

Phe Leu Lys His Gln Ser Gly Asp Gln Ala Ser Gln Tyr Ala Ala Thr  
 130 135 140  
 Arg Leu Leu Gln Gln Phe Pro Asn Val Ala Met Lys Asp Leu Val Gln  
 145 150 155 160  
 Lys Leu Ala Ser Lys Glu Glu Leu Ser Thr Ala Asp Asn Glu Met Ile  
 165 170 175  
 Glu Leu Leu Ala Arg Phe Asn Glu Arg Gln Ala Ser Phe Phe Gly Gln  
 180 185 190  
 Phe Ser Val Arg Gly Tyr Val Asn Tyr Asp Lys His Val Ala Lys Tyr  
 195 200 205  
 Leu Lys Ile Leu Pro Asp Gln Phe Ser Tyr Gln Ala Ile Glu Asp Val  
 210 215 220  
 Val Lys Ala Asp Ala Glu Lys Asn Thr Ser Asn Asn Glu Met Gly Met  
 225 230 235 240  
 Glu Asn Tyr Phe Tyr Asn Glu Gln Ile Lys Lys Asp Leu Lys Lys Leu  
 245 250 255  
 Lys Asp Ser Gln Lys Ser Phe Thr Tyr Leu Lys Ser Pro Glu Tyr Asn  
 260 265 270  
 Xaa Leu Gln Leu Val Leu Thr Gln Phe Ser Lys Ser Lys Val Asn Pro  
 275 280 285  
 Ile Phe Ile Ile Pro Pro Val Asn Lys Lys Trp Met Xaa Tyr Ala Gly  
 290 295 300  
 Leu Arg Glu Asp Met Tyr Gln Gln Thr Val Gln Lys Ile Arg Tyr Gln  
 305 310 315 320  
 Leu Glu Ser Gln Gly Phe Thr Asn Ile Ala Asp Phe Ser Lys Asp Gly  
 325 330 335  
 Gly Glu Pro Phe Phe Met Lys Asp Thr Ile His Leu Gly Trp Leu Gly  
 340 345 350  
 Trp Leu Ala Phe Asp Lys Ala Val Asp Pro Phe Leu Ser Asn Pro Thr  
 355 360 365  
 Pro Ala Pro Thr Tyr His Leu Asn Glu Arg Phe Phe Ser Lys Asp Trp  
 370 375 380  
 Ala Thr Tyr Asp Gly Asp Val Lys Glu  
 385 390

(2) INFORMATION FOR SEQ ID NO: 59:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1423 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

GGTTTGTGAGA AAGTATTTGC AGGGGGCCCT GATTGAGTCG ATTGAGCAAG TGGAAAATGA

171

```

CCGTATTGTG GAAATTACAG TTTCCAATAA AAACGAGATT GGAGACCATA TCCAGGCTAC      120
CTTGATTATC GAAATTATGG GGAAACACAG TAATATTCTA CTGGTCGATA AAAGCAGTCA      180
TAAAATCCTC GAAGTTATCA AACACGTCGG CTTTTCACAA AATAGCTACC GCACCTTACT      240
TCCAGGATCG ACCTATATCG CTCCGCCAAG TACAAAATCT CTCAATCCTT TTACTATCAA      300
GGATGAAAAG CTCTTTGAAA TCCTGCAAAC CCAAGAACTA ACAGCAAAAA ATCTTCAAAG      360
CCTCTTTCAA GGTCTGGGAC GCGATACGGC AAATGAATTG GAAAGGATAC TGGTTAGTGA      420
AAAACTTTCC GCTTTCGGAA ATTTTTTCAA TCAAGAAACC AAGCCATGCT TGA CTGAGAC      480
TTCCTTCAGT CCAGTTCCTT TTGCAAATCA GGTGGGAGAG CCTTTTGCAA ATCTTTCTGA      540
TTTGTGGGAC ACCTACTATA AGGATAAGGC TGAGCGCGAC CGCGTCAAAC AGCAGGCCAG      600
TGA ACTGATT CGTCGTGTTG AAAATGAACT TCAGAAAAAC CGACACAAAC TCAAAAAACA      660
GGAAAAAGAG TTACTGGCGA CAGACAACGC TGAAGAATTT CGTCAAAAAG GAGAATTGCT      720
GACAACCTTC CTCCACCAAG TGCCTAACGA CCAAGACCAG GTTATCCTAG ACAACTACTA      780
TACCAACCAA CCTATCATGA TTGCGCTTGA TAAGGCTCTG ACTCCCAACC AGAATGCCCA      840
ACGCTATTTT AAACGGTATC AGAAACTCAA AGAAGCTGTC AAATACTTGA CTGATTTGAT      900
TGAAGAAACC AAAGCCACTA TTCTCTATCT GGAAAGTGTA GAAACCGTCC TCAACCAAGC      960
TGGACTGGAA GAAATCGCTG AAATCCGTGA AGAATTGATT CAAACAGGTT TTATCCGCAG     1020
AAGACAACGG GAGAAAATCC AGAAACGCAA AAAACTAGAA CAATATCTAG CAAGCGATGG     1080
CAAAACCATC ATCTATGTCG GACGAAACAA TCTTCAAAAT GAGGAATTGA CCTTTAAAAT     1140
GGCCCGCAAG GAGGAACTTT GGTTCATGC TAAGGACATT CCTGGAAGCC ATGTTGTCAT     1200
CTCAGGAAAT CTTGACCCAT CTGATGCAGT CAAGACAGAC GCAGCAGAGT TAGCTGCCTA     1260
CTTCTCTCAA GGGCGCCTGT CGAATCTGGT GCAGGTAGAT ATGATTGAAG TCAAAAAACT     1320
CAATAAACCA ACTGGTGGAA AACCCGGCTT TGTCACCTAC ACAGGACAAA AGACCCTCCG     1380
CGTCACACCA GACTCCAAAA AAATTGCATC CATGAAAAAA TCC                          1423

```

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 474 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

Val Leu Arg Lys Tyr Leu Gln Gly Ala Leu Ile Glu Ser Ile Glu Gln
1           5           10           15

```

172

Val Glu Asn Asp Arg Ile Val Glu Ile Thr Val Ser Asn Lys Asn Glu  
                   20                                  25                                  30  
 Ile Gly Asp His Ile Gln Ala Thr Leu Ile Ile Glu Ile Met Gly Lys  
                   35                                  40                                  45  
 His Ser Asn Ile Leu Leu Val Asp Lys Ser Ser His Lys Ile Leu Glu  
                   50                                  55                                  60  
 Val Ile Lys His Val Gly Phe Ser Gln Asn Ser Tyr Arg Thr Leu Leu  
                   65                                  70                                  75                                  80  
 Pro Gly Ser Thr Tyr Ile Ala Pro Pro Ser Thr Lys Ser Leu Asn Pro  
                                   85                                  90                                  95  
 Phe Thr Ile Lys Asp Glu Lys Leu Phe Glu Ile Leu Gln Thr Gln Glu  
                                   100                                  105                                  110  
 Leu Thr Ala Lys Asn Leu Gln Ser Leu Phe Gln Gly Leu Gly Arg Asp  
                   115                                  120                                  125  
 Thr Ala Asn Glu Leu Glu Arg Ile Leu Val Ser Glu Lys Leu Ser Ala  
                   130                                  135                                  140  
 Phe Arg Asn Phe Phe Asn Gln Glu Thr Lys Pro Cys Leu Thr Glu Thr  
                   145                                  150                                  155                                  160  
 Ser Phe Ser Pro Val Pro Phe Ala Asn Gln Val Gly Glu Pro Phe Ala  
                                   165                                  170                                  175  
 Asn Leu Ser Asp Leu Leu Asp Thr Tyr Tyr Lys Asp Lys Ala Glu Arg  
                                   180                                  185                                  190  
 Asp Arg Val Lys Gln Gln Ala Ser Glu Leu Ile Arg Arg Val Glu Asn  
                   195                                  200                                  205  
 Glu Leu Gln Lys Asn Arg His Lys Leu Lys Lys Gln Glu Lys Glu Leu  
                   210                                  215                                  220  
 Leu Ala Thr Asp Asn Ala Glu Glu Phe Arg Gln Lys Gly Glu Leu Leu  
                   225                                  230                                  235                                  240  
 Thr Thr Phe Leu His Gln Val Pro Asn Asp Gln Asp Gln Val Ile Leu  
                                   245                                  250                                  255  
 Asp Asn Tyr Tyr Thr Asn Gln Pro Ile Met Ile Ala Leu Asp Lys Ala  
                   260                                  265                                  270  
 Leu Thr Pro Asn Gln Asn Ala Gln Arg Tyr Phe Lys Arg Tyr Gln Lys  
                   275                                  280                                  285  
 Leu Lys Glu Ala Val Lys Tyr Leu Thr Asp Leu Ile Glu Glu Thr Lys  
                   290                                  295                                  300  
 Ala Thr Ile Leu Tyr Leu Glu Ser Val Glu Thr Val Leu Asn Gln Ala  
                   305                                  310                                  315                                  320  
 Gly Leu Glu Glu Ile Ala Glu Ile Arg Glu Glu Leu Ile Gln Thr Gly  
                                   325                                  330                                  335  
 Phe Ile Arg Arg Arg Gln Arg Glu Lys Ile Gln Lys Arg Lys Lys Leu  
                   340                                  345                                  350  
 Glu Gln Tyr Leu Ala Ser Asp Gly Lys Thr Ile Ile Tyr Val Gly Arg

173

|                                                                 |     |     |     |     |
|-----------------------------------------------------------------|-----|-----|-----|-----|
| 355                                                             |     | 360 |     | 365 |
| Asn Asn Leu Gln Asn Glu Glu Leu Thr Phe Lys Met Ala Arg Lys Glu |     |     |     |     |
| 370                                                             |     | 375 |     | 380 |
| Glu Leu Trp Phe His Ala Lys Asp Ile Pro Gly Ser His Val Val Ile |     |     |     |     |
| 385                                                             |     | 390 |     | 400 |
| Ser Gly Asn Leu Asp Pro Ser Asp Ala Val Lys Thr Asp Ala Ala Glu |     |     |     |     |
|                                                                 | 405 |     | 410 | 415 |
| Leu Ala Ala Tyr Phe Ser Gln Gly Arg Leu Ser Asn Leu Val Gln Val |     |     |     |     |
|                                                                 | 420 |     | 425 | 430 |
| Asp Met Ile Glu Val Lys Lys Leu Asn Lys Pro Thr Gly Gly Lys Pro |     |     |     |     |
|                                                                 | 435 |     | 440 | 445 |
| Gly Phe Val Thr Tyr Thr Gly Gln Lys Thr Leu Arg Val Thr Pro Asp |     |     |     |     |
|                                                                 | 450 |     | 455 | 460 |
| Ser Lys Lys Ile Ala Ser Met Lys Lys Ser                         |     |     |     |     |
| 465                                                             |     | 470 |     |     |

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 544 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

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

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| GACAACATTT ACTATCCATA CAGTAGAGTC AGCACCAGCA GAAGTGAAAG AAATTCTTGA | 60  |
| AACAGTAGAA AAAGACAACA ATGGCTATAT TCCCAACCTA ATCGGTCTCT TGGCCAATGC | 120 |
| CCCGACTGTT TTAGAAGCCT ACCAAATTGT CTCATCTATC CACCGTCGCA ACAGCCTGAC | 180 |
| ACCCGTTGAG CGTGAAGTGG TGCAAATCAC GGCAGCCGTG ACCAATGGTT GTGCCTTCTG | 240 |
| TGTCGCAGGT CACACAGCCT TTTCCATCAA ACAAATCCAG ATGAATGATG ACTTGATTCA | 300 |
| AGCTCTTCGC AATCGTACTC CAATTGAAAC AGATCCTAAA TTGGATACCC TAGCTAAGTT | 360 |
| TACCTTGGA GTTATCAATA CCAAGGGTCG TGTAGGAGAT GAAGCCTTGT CTGAGTTTTT  | 420 |
| AGAAGCTGGC TAGACTCAAC AAAATGCCTT GGATGTGGTT TTTGGTGTC GCCTAGCAAT  | 480 |
| CCTCTGTAAC TATGCCAACA ACTTAGCTAA TACACCAATT AATCCAGAAT TGCAACCTTA | 540 |
| TGCC                                                              | 544 |

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

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 181 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

```

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

```

## (2) INFORMATION FOR SEQ ID NO: 63:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 811 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

```

GGCTAAGGAA AGAGTGGATG TACTAGCTTA TAAACAGGGG TTGTTTGAAA CGAGAGAGCA      60
GGCCAAGCGA GGTGTGATGG CTGGCCTAGT CGTAGCAGTC CTTAATGGAG AACGGTTTGA      120
CAAGCCAGGA GAGAAAATTC CAGATGACAC CGAATTAAAA CTCAAGGGGG AGAAACTCAA      180
GTATGTCAGC CGTGGTGGTT TGAAACTGGA AAAGGCCTTG CAGGTCTTTG ATTTGTCGGT      240
GGATGGCGCG ACTACGATTG ATATCGGGGC CTCTACTGGA GGTTTTACCG ATGTCATGCT      300
ACAGAATAGT GCCAAGTTGG TCTTTCAGT CGATGTTGGT ACCAATCAGT TGGCTTGGA      360

```

175

```

ATTACGCCAA GACCCACGAG TTGTCAGCAT GGAGCAGTTC AATTTCGCT ATGCTGAAAA      420
GACTGATTTC GAGCAGGAGC CGAGCTTTGC CAGTATTGAT GTGAGTTTCA TTTCCCTTAG      480
TCTGATTTTG CCAGCCTTGC ACCGTGTCTT GGCTGATCAA GGTCAGGTGG TAGCACTTGT      540
CAAACCTCAG TTTGAGGCAG GACGTGAGCA GATTGGGAAA AATGGAATTA TTCGAGATGC      600
TAAGGTTTCAT CAGAATGTCC TTGAATCTGT AACAGCTATG GCAGTAGAGG TAGGTTTTTC      660
AGTCCTTGGC TTGGACTTTT CTCCCATCCA AGGTGGACAT GGAAATATTG AATTTTTCAG      720
GTATTTGAAA AAAGAAAAGT CAGCAAGCAA TCAGATTCTT GCTGAGATTA AAGAAGCAGT      780
AGAGAGGGCG CATAGTCAAT TTAATAATGA A                                     811

```

(2) INFORMATION FOR SEQ ID NO:64:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 270 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

```

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

```

176

Val Ala Leu Val Lys Pro Gln Phe Glu Ala Gly Arg Glu Gln Ile Gly  
 180 185 190

Lys Asn Gly Ile Ile Arg Asp Ala Lys Val His Gln Asn Val Leu Glu  
 195 200 205

Ser Val Thr Ala Met Ala Val Glu Val Gly Phe Ser Val Leu Gly Leu  
 210 215 220

Asp Phe Ser Pro Ile Gln Gly Gly His Gly Asn Ile Glu Phe Leu Ala  
 225 230 235 240

Tyr Leu Lys Lys Glu Lys Ser Ala Ser Asn Gln Ile Leu Ala Glu Ile  
 245 250 255

Lys Glu Ala Val Glu Arg Ala His Ser Gln Phe Lys Asn Glu  
 260 265 270

## (2) INFORMATION FOR SEQ ID NO: 65:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2290 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

|                                                                    |      |
|--------------------------------------------------------------------|------|
| TTGTTTCCTAT GAACCTGGTC GTCACCAAGC TGGTCAGGTT AAGAAAGAGT CTAATCGAGT | 60   |
| TTCTTATATA GATGGTGATC AGGCTGGTCA AAAGGCAGAA AACTTGACAC CAGATGAAGT  | 120  |
| CAGTAAGAGG GAGGGGATCA ACGCCGAACA AATNGTNATC AAGATTACGG ATCAAGGTTA  | 180  |
| TGTGACCTCT CATGGAGACC ATTATCATT CTATAATGGC AAGGTTCCCTT ATGATGCCAT  | 240  |
| CATCAGTGAA GAGCTCCTCA TGAAAGATCC GAATTATCAG TTGAAGGATT CAGACATTGT  | 300  |
| CAATGAAATC AAGGGTGGTT ATGTCATTAA GGTAAACGGT AAATACTATG TNTACCTTAA  | 360  |
| GGATGCAGCT CATGCGGATA ATATTCGGAC AAAAGAAGAG ATTAACGTC AGAAGCAGGA   | 420  |
| ACGCAGTCAT AATCATAACT CAAGAGCAGA TAATGCTGTT GCTGCAGCCA GAGCCCAAGG  | 480  |
| ACGTTATACA ACGGATGATG GGTATATCTT CAATGCATCT GATATCATTG AGGACACGGG  | 540  |
| TGATGCTTAT ATCGTTCCTC ACGGCGACCA TTACCATTAC ATTCCTAAGA ATGAGTTATC  | 600  |
| AGCTAGCGAG TTAGCTGCTG CAGAAGCCTA TTGGAATGGG AAGCAGGGAT CTCGTCCTTC  | 660  |
| TTCAAGTTCT AGTTATAATG CAAATCCAGC TCAACCAAGA TTGTCAGAGA ACCACAATCT  | 720  |
| GACTGTCACT CCAACTTATC ATCAAAATCA AGGGGAAAAC ATTTCAAGCC TTTTACGTGA  | 780  |
| ATTGTATGCT AAACCCTTAT CAGAACGCCA TGTGGAATCT GATGGCCTTA TTTTCGACCC  | 840  |
| AGCGCAAATC ACAAGTCGAA CCGCCAGAGG TGTAGCTGTC CCTCATGGTA ACCATTACCA  | 900  |
| CTTTATCCCT TATGAACAAA TGTCTGAATT GGAAAAACGA ATTGCTCGTA TTATCCCCT   | 960  |
| TCGTTATCGT TCAAACCATT GGGTACCAGA TTCAAGACCA GAACAACCAA GTCCACAATC  | 1020 |

177

|                                                                    |      |
|--------------------------------------------------------------------|------|
| GACTCCGGAA CCTAGTCCAA GTCCGCAACC TGCACCAAAT CCTCAACCAG CTCCAAGCAA  | 1080 |
| TCCAATTGAT GAGAAATTGG TCAAAGAAGC TGTTGAAAA GTAGGCGATG GTTATGTCTT   | 1140 |
| TGAGGAGAAT GGAGTTTCTC GTTATATCCC AGCCAAGGAT CTTTCAGCAG AACAGCAGC   | 1200 |
| AGGCATTGAT AGCAAACCTGG CCAAGCAGGA AAGTTTATCT CATAAGCTAG GAGCTAAGAA | 1260 |
| AACTGACCTC CCATCTAGTG ATCGAGAATT TTACAATAAG GCTTATGACT TACTAGCAAG  | 1320 |
| AATTCACCAA GATTTACTTG ATAATAAAGG TCGACAAGTT GATTTTGAGG CTTTGGATAA  | 1380 |
| CCTGTTGGAA CGACTCAAGG ATGTCNCAAG TGATAAAGTC AAGTTAGTGG ANGATATTCT  | 1440 |
| TGCCTTCTTA GCTCCGATTC GTCATCCAGA ACGTTTAGGA AAACCAAATG CGCAAATTAC  | 1500 |
| CTACACTGAT GATGAGATTC AAGTAGCCAA GTTGGCAGGC AAGTACACAA CAGAAGACGG  | 1560 |
| TTATATCTTT GATCCTCGTG ATATAACCAG TGATGAGGGG GATGCCTATG TAACTCCACA  | 1620 |
| TATGACCCAT AGCCACTGGA TTAAAAAGA TAGTTTGTCT GAAGCTGAGA GAGCGGCAGC   | 1680 |
| CCAGGCTTAT GCTAAAGAGA AAGGTTTGAC CCCTCCTTCG ACAGACCATC AGGATTCAGG  | 1740 |
| AAATACTGAG GCAAAGGAG CAGAAGCTAT CTACAACCGC GTGAAAGCAG CTAAGAAGGT   | 1800 |
| GCCACTTGAT CGTATGCCTT ACAATCTTCA ATATACTGTA GAAGTCAAAA ACGGTAGTTT  | 1860 |
| AATCATACCT CATTATGACC ATTACCATAA CATCAAATTT GAGTGGTTTG ACGAAGGCCT  | 1920 |
| TTATGAGGCA CCTAAGGGGT ATACTCTTGA GGATCTTTTG GCGACTGTCA AGTACTATGT  | 1980 |
| CGAACATCCA AACGAACGTC CGCATTGAGA TAATGGTTT GGTAACGCTA GCGACCATGT   | 2040 |
| TCAAAGAAAC AAAAATGGTC AAGCTGATAC CAATCAAACG GAAAAACCAA GCGAGGAGAA  | 2100 |
| ACCTCAGACA GAAAAACCTG AGGAAGAAAC CCCTCGAGAA GAGAAACCGC AAAGCGAGAA  | 2160 |
| ACCAGAGTCT CCAAACCAA CAGAGGAACC AGAAGAATCA CCAGAGGAAT CAGAAGAACC   | 2220 |
| TCAGGTCGAG ACTGAAAAGG TTGAAGAAAA ACTGAGAGAG GCTGAAGATT TACTTGAAA   | 2280 |
| AATCCAGGAT                                                         | 2290 |

## (2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 763 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Cys | Ser | Tyr | Glu | Leu | Gly | Arg | His | Gln | Ala | Gly | Gln | Val | Lys | Lys | Glu |
| 1   |     |     |     | 5   |     |     |     | 10  |     |     |     |     | 15  |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Ser | Asn | Arg | Val | Ser | Tyr | Ile | Asp | Gly | Asp | Gln | Ala | Gly | Gln | Lys | Ala |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |

178

Glu Asn Leu Thr Pro Asp Glu Val Ser Lys Arg Glu Gly Ile Asn Ala  
 35 40 45  
 Glu Gln Xaa Val Ile Lys Ile Thr Asp Gln Gly Tyr Val Thr Ser His  
 50 55 60  
 Gly Asp His Tyr His Tyr Tyr Asn Gly Lys Val Pro Tyr Asp Ala Ile  
 65 70 75 80  
 Ile Ser Glu Glu Leu Leu Met Lys Asp Pro Asn Tyr Gln Leu Lys Asp  
 85 90 95  
 Ser Asp Ile Val Asn Glu Ile Lys Gly Gly Tyr Val Ile Lys Val Asn  
 100 105 110  
 Gly Lys Tyr Tyr Val Tyr Leu Lys Asp Ala Ala His Ala Asp Asn Ile  
 115 120 125  
 Arg Thr Lys Glu Glu Ile Lys Arg Gln Lys Gln Glu Arg Ser His Asn  
 130 135 140  
 His Asn Ser Arg Ala Asp Asn Ala Val Ala Ala Ala Arg Ala Gln Gly  
 145 150 155 160  
 Arg Tyr Thr Thr Asp Asp Gly Tyr Ile Phe Asn Ala Ser Asp Ile Ile  
 165 170 175  
 Glu Asp Thr Gly Asp Ala Tyr Ile Val Pro His Gly Asp His Tyr His  
 180 185 190  
 Tyr Ile Pro Lys Asn Glu Leu Ser Ala Ser Glu Leu Ala Ala Ala Glu  
 195 200 205  
 Ala Tyr Trp Asn Gly Lys Gln Gly Ser Arg Pro Ser Ser Ser Ser  
 210 215 220  
 Tyr Asn Ala Asn Pro Ala Gln Pro Arg Leu Ser Glu Asn His Asn Leu  
 225 230 235 240  
 Thr Val Thr Pro Thr Tyr His Gln Asn Gln Gly Glu Asn Ile Ser Ser  
 245 250 255  
 Leu Leu Arg Glu Leu Tyr Ala Lys Pro Leu Ser Glu Arg His Val Glu  
 260 265 270  
 Ser Asp Gly Leu Ile Phe Asp Pro Ala Gln Ile Thr Ser Arg Thr Ala  
 275 280 285  
 Arg Gly Val Ala Val Pro His Gly Asn His Tyr His Phe Ile Pro Tyr  
 290 295 300  
 Glu Gln Met Ser Glu Leu Glu Lys Arg Ile Ala Arg Ile Ile Pro Leu  
 305 310 315 320  
 Arg Tyr Arg Ser Asn His Trp Val Pro Asp Ser Arg Pro Glu Gln Pro  
 325 330 335  
 Ser Pro Gln Ser Thr Pro Glu Pro Ser Pro Ser Pro Gln Pro Ala Pro  
 340 345 350  
 Asn Pro Gln Pro Ala Pro Ser Asn Pro Ile Asp Glu Lys Leu Val Lys  
 355 360 365  
 Glu Ala Val Arg Lys Val Gly Asp Gly Tyr Val Phe Glu Glu Asn Gly

179

|                                                                 |     |         |
|-----------------------------------------------------------------|-----|---------|
| 370                                                             | 375 | 380     |
| Val Ser Arg Tyr Ile Pro Ala Lys Asp Leu Ser Ala Glu Thr Ala Ala |     |         |
| 385                                                             | 390 | 395 400 |
| Gly Ile Asp Ser Lys Leu Ala Lys Gln Glu Ser Leu Ser His Lys Leu |     |         |
|                                                                 | 405 | 410 415 |
| Gly Ala Lys Lys Thr Asp Leu Pro Ser Ser Asp Arg Glu Phe Tyr Asn |     |         |
|                                                                 | 420 | 425 430 |
| Lys Ala Tyr Asp Leu Leu Ala Arg Ile His Gln Asp Leu Leu Asp Asn |     |         |
|                                                                 | 435 | 440 445 |
| Lys Gly Arg Gln Val Asp Phe Glu Ala Leu Asp Asn Leu Leu Glu Arg |     |         |
|                                                                 | 450 | 455 460 |
| Leu Lys Asp Val Xaa Ser Asp Lys Val Lys Leu Val Xaa Asp Ile Leu |     |         |
| 465                                                             | 470 | 475 480 |
| Ala Phe Leu Ala Pro Ile Arg His Pro Glu Arg Leu Gly Lys Pro Asn |     |         |
|                                                                 | 485 | 490 495 |
| Ala Gln Ile Thr Tyr Thr Asp Asp Glu Ile Gln Val Ala Lys Leu Ala |     |         |
|                                                                 | 500 | 505 510 |
| Gly Lys Tyr Thr Thr Glu Asp Gly Tyr Ile Phe Asp Pro Arg Asp Ile |     |         |
|                                                                 | 515 | 520 525 |
| Thr Ser Asp Glu Gly Asp Ala Tyr Val Thr Pro His Met Thr His Ser |     |         |
|                                                                 | 530 | 535 540 |
| His Trp Ile Lys Lys Asp Ser Leu Ser Glu Ala Glu Arg Ala Ala Ala |     |         |
| 545                                                             | 550 | 555 560 |
| Gln Ala Tyr Ala Lys Glu Lys Gly Leu Thr Pro Pro Ser Thr Asp His |     |         |
|                                                                 | 565 | 570 575 |
| Gln Asp Ser Gly Asn Thr Glu Ala Lys Gly Ala Glu Ala Ile Tyr Asn |     |         |
|                                                                 | 580 | 585 590 |
| Arg Val Lys Ala Ala Lys Lys Val Pro Leu Asp Arg Met Pro Tyr Asn |     |         |
|                                                                 | 595 | 600 605 |
| Leu Gln Tyr Thr Val Glu Val Lys Asn Gly Ser Leu Ile Ile Pro His |     |         |
|                                                                 | 610 | 615 620 |
| Tyr Asp His Tyr His Asn Ile Lys Phe Glu Trp Phe Asp Glu Gly Leu |     |         |
| 625                                                             | 630 | 635 640 |
| Tyr Glu Ala Pro Lys Gly Tyr Thr Leu Glu Asp Leu Leu Ala Thr Val |     |         |
|                                                                 | 645 | 650 655 |
| Lys Tyr Tyr Val Glu His Pro Asn Glu Arg Pro His Ser Asp Asn Gly |     |         |
|                                                                 | 660 | 665 670 |
| Phe Gly Asn Ala Ser Asp His Val Gln Arg Asn Lys Asn Gly Gln Ala |     |         |
|                                                                 | 675 | 680 685 |
| Asp Thr Asn Gln Thr Glu Lys Pro Ser Glu Glu Lys Pro Gln Thr Glu |     |         |
|                                                                 | 690 | 695 700 |
| Lys Pro Glu Glu Glu Thr Pro Arg Glu Glu Lys Pro Gln Ser Glu Lys |     |         |
| 705                                                             | 710 | 715 720 |

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

|             |            |            |            |            |            |      |
|-------------|------------|------------|------------|------------|------------|------|
| GAATGTTTCAG | GCTCAAGAAA | GTTCAGGAAA | TAAATCCAC  | TTTATCAATG | TTCAAGAAGG | 60   |
| TGGCAGTGAT  | GCGATTATTC | TTGAAAGCAA | TGGACATTTT | GCCATGGTGG | ATACAGGAGA | 120  |
| AGATTATGAT  | TTCCCAGATG | GAAGTGATTC | TCGCTATCCA | TGGAGAGAAG | GAATTGAAAC | 180  |
| GTCTTATAAG  | CATGTTCTAA | CAGACCGTGT | CTTTCGTCGT | TTGAAGGAAT | TGGGTGTCCA | 240  |
| AAAACCTTGAT | TTTATTTTGG | TGACCCATAC | CCACAGTGAT | CATATTGGAA | ATGTTGATGA | 300  |
| ATTACTGTCT  | ACCTATCCAG | TTGACCGAGT | CTATCTTAAG | AAATATAGTG | ATAGTCGTAT | 360  |
| TACTAATTCT  | GAACGTCTAT | GGGATAATCT | GTATGGCTAT | GATAAGGTTT | TACAGACTGC | 420  |
| TGCAGAAAAA  | GGTGTTCAG  | TTATTCAAAA | TATCACACAA | GGGGATGCTC | ATTTTCAGTT | 480  |
| TGGGGACATG  | GATATTCAGC | TCTATAATTA | TGAAAATGAA | ACTGATTCAT | CGGGTGAATT | 540  |
| AAAGAAAATT  | TGGGATGACA | ATTCCAATTC | CTTGATTAGC | GTGGTGAAAG | TCAATGGCAA | 600  |
| GAAAATTTAC  | CTTGGGGGCG | ATTTAGATAA | TGTTCATGGA | GCAGAAGACA | AGTATGGTCC | 660  |
| TCTCATTGGA  | AAAGTTGATT | TGATGAAGTT | TAATCATCAC | CATGATACCA | ACAAATCAAA | 720  |
| TACCAAGGAT  | TTCATTAAAA | ATTTGAGTCC | GAGTTTGATT | GTTCAAACCT | CGGATAGTCT | 780  |
| ACCTTGGAAA  | AATGGTGTTG | ATAGTGAGTA | TGTTAATTGG | CTCAAAGAAC | GAGGAATTGA | 840  |
| GAGAATCAAC  | GCAGCCAGCA | AAGACTATGA | TGCAACAGTT | TTTGATATTC | GAAAAGACGG | 900  |
| TTTTGTCAAT  | ATTTCAACAT | CCTACAAGCC | GATTCCAAGT | TTTCAAGCTG | GTTGGCATAA | 960  |
| GAGTGCATAT  | GGGAACTGGT | GGTATCAAGC | GCCTGATTCT | ACAGGAGAGT | ATGCTGTCCG | 1020 |
| TTGGAATGAA  | ATCGAAGGTG | AATGGTATTA | CTTTAACCAA | ACGGGTATCT | TGTTACAGAA | 1080 |
| TCAATGGAAA  | AAATGGAACA | ATCATTTGGT | CTATTTGACA | GACTCTGGTG | CTTCTGCTAA | 1140 |
| AAATTGGAAG  | AAAATCGCTG | GAATCTGGTA | TTATTTTAAC | AAAGAAAACC | AGATGGAAAT | 1200 |
| TGGTTGGATT  | CAAGATAAAG | AGCAGTGGTA | TTATTTGGAT | GTTGATGGTT | CTATGAAGAC | 1260 |
| AGGATGGCTT  | CAATATATGG | GGCAATGGTA | TTACTTTGCT | CCATCAGGGG | AA         | 1312 |

182

## (2) INFORMATION FOR SEQ ID NO:70:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 437 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

```

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

```

183

|                                 |                                     |             |
|---------------------------------|-------------------------------------|-------------|
| 260                             | 265                                 | 270         |
| Trp Leu Lys Glu Arg Gly Ile     | Glu Arg Ile Asn Ala Ala             | Ser Lys Asp |
| 275                             | 280                                 | 285         |
| Tyr Asp Ala Thr Val Phe Asp     | Ile Arg Lys Asp Gly Phe Val Asn Ile |             |
| 290                             | 295                                 | 300         |
| Ser Thr Ser Tyr Lys Pro Ile     | Pro Ser Phe Gln Ala Gly Trp His Lys |             |
| 305                             | 310                                 | 315         |
| Ser Ala Tyr Gly Asn Trp Trp Tyr | Gln Ala Pro Asp Ser Thr Gly Glu     |             |
| 325                             | 330                                 | 335         |
| Tyr Ala Val Gly Trp Asn Glu Ile | Glu Gly Glu Trp Tyr Tyr Phe Asn     |             |
| 340                             | 345                                 | 350         |
| Gln Thr Gly Ile Leu Leu Gln Asn | Gln Trp Lys Lys Trp Asn Asn His     |             |
| 355                             | 360                                 | 365         |
| Trp Phe Tyr Leu Thr Asp Ser Gly | Ala Ser Ala Lys Asn Trp Lys Lys     |             |
| 370                             | 375                                 | 380         |
| Ile Ala Gly Ile Trp Tyr Tyr Phe | Asn Lys Glu Asn Gln Met Glu Ile     |             |
| 385                             | 390                                 | 395         |
| Gly Trp Ile Gln Asp Lys Glu Gln | Trp Tyr Tyr Leu Asp Val Asp Gly     |             |
| 405                             | 410                                 | 415         |
| Ser Met Lys Thr Gly Trp Leu Gln | Tyr Met Gly Gln Trp Tyr Tyr Phe     |             |
| 420                             | 425                                 | 430         |
| Ala Pro Ser Gly Glu             |                                     |             |
| 435                             |                                     |             |

## (2) INFORMATION FOR SEQ ID NO: 71:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1855 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| CTTGGGTGTA ACCCATATCC AGCTCCTTCC AGTCTTGTCT TACTACTTTG TCAATGAATT | 60  |
| GAAAAACCAT GAACGCTTGT CTGACTACGC TTCAAGCAAC AGCAACTACA ACTGGGGATA | 120 |
| TGACCCTCAA AACTACTTCT CCTTGACTGG TATGTACTCA AGCGATCCTA AGAATCCAGA | 180 |
| AAAACGAATC GCAGAATTTA AAAACCTCAT CAACGAAATC CACAAACGTG GTATGGGAGC | 240 |
| TATCCTAGAT GTCGTTTATA ACCACACAGC CAAAGTCGAT CTCTTTGAAG ATTTGGAACC | 300 |
| AAACTACTAC CACTTTATGG ATGCCGATGG CACACCTCGA ACTAGCTTTG GTGGTGAGC  | 360 |
| CTTGGGGACA ACCCACCATA TGACCAAACG GCTCCTAATT GACTCTATCA AATACCTAGT | 420 |
| TGATACCTAC AAAGTGATG GCTTCCGTTT CGATATGATG GGAGACCATG ACGCCGCTTC  | 480 |
| TATCGAAGAA GCTTACAAGG CTGCACGCGC CCTCAATCCA AACCTCATCA TGCTTGGTGA | 540 |

184

```

AGGTTGGAGA ACCTATGCCG GTGATGAAAA CATGCCTACT AAAGCTGCTG ACCAAGATTG      600
GATGAAACAT ACCGATACTG TCGCTGTCTT TTCAGATGAC ATCCGTAACA ACCTCAAATC      660
TGGTTATCCA AACGAAGGTC AACCTGCCTT TATCACAGGT GGCAAGCGTG ATGTCAACAC      720
CATCTTTAAA AATCTCATTG CTCAACCAAC TAACTTTGAA GCTGACAGCC CTGGAGATGT      780
CATCCAATAC ATCGCAGCCC ATGATAACTT GACCCTCTTT GACATCATTG CCCAGTCTAT      840
CAAAAAAGAC CCAAGCAAGG CTGAGAACTA TGCTGAAATC CACCGTCGTT TACGACTTGG      900
AAATCTCATG GTCTTGACAG CTCAAGGAAC TCCATTTATC CACTCCGGTC AGGAATATGG      960
ACGTACTAAA CAATTCCGTG ACCCAGCCTA CAAGACTCCA GTAGCAGAGG ATAAGGTTCC     1020
AAACAAATCT CACTTGTTGC GTGATAAGGA CGGCAACCCA TTTGACTATC CTTACTTCAT     1080
CCATGACTCT TACGATTCTA GTGATGCAGT CAACAAGTTT GACTGGACTA AGGCTACAGA     1140
TGGTAAAGCT TATCCTGAAA ATGTCAAGAG CCGTGA CTAT ATGAAAGGTT TGATTGCCCT     1200
TCGTCAATCT ACAGATGCCT TCCGACTTAA GAGTCTTCAA GATATCAAAG ACCGTGTCCA     1260
CCTCATCACT GTCCCAGGCC AAAATGGTGT GGAAAAAGAG GATGTAGTGA TTGGCTACCA     1320
AATCACTGCT CCAAAACGGCG ATATCTACGC AGTCTTTGTC AATGCGGATG AAAAAGCTCG     1380
CGAATTTAAT TTGGGAACTG CCTTTGCACA TCTAAGAAAT GCGGAAGTTT TGGCAGATGA     1440
AAACCAAGCA GGACCAAGTC GAATTGCCAA CCCGAAAGGA CTTGAATGGA CTGAAAAAGG     1500
CTTGAAATTG AATGCCCTTA CAGCTACTGT TCTTCGAGTC TCTCAAAATG GAACTAGCCA     1560
TGAGTCAACT GCAGAAGAGA AACCAGACTC AACCCTTCC AAGCCTGAAC ATCAAAATGA     1620
AGCTTCTCAC CCTGCACATC AAGACCCAGC TCCAGAAGCT AGACCTGATT CTAATAACC     1680
AGATGCCAAA GTAGCTGATG CGGAAAATAA ACCTAGCCAA GCTACAGCTG ATTCACAAGC     1740
TGAACAACCA GCACAAGAAG CACAAGCATC ATCTGTAAAA GAAGCGGTTT GAAACGAATC     1800
GGTAGAAAAC TCTAGCAAGG AAAATATACC TGCAACCCCA GATAACAAG CTGAA      1855

```

## (2) INFORMATION FOR SEQ ID NO:72:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 618 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

```

Leu Gly Val Thr His Ile Gln Leu Leu Pro Val Leu Ser Tyr Tyr Phe
1           5           10           15
Val Asn Glu Leu Lys Asn His Glu Arg Leu Ser Asp Tyr Ala Ser Ser
20           25           30

```

185

Asn Ser Asn Tyr Asn Trp Gly Tyr Asp Pro Gln Asn Tyr Phe Ser Leu  
 35 40 45  
 Thr Gly Met Tyr Ser Ser Asp Pro Lys Asn Pro Glu Lys Arg Ile Ala  
 50 55 60  
 Glu Phe Lys Asn Leu Ile Asn Glu Ile His Lys Arg Gly Met Gly Ala  
 65 70 75 80  
 Ile Leu Asp Val Val Tyr Asn His Thr Ala Lys Val Asp Leu Phe Glu  
 85 90 95  
 Asp Leu Glu Pro Asn Tyr Tyr His Phe Met Asp Ala Asp Gly Thr Pro  
 100 105 110  
 Arg Thr Ser Phe Gly Gly Gly Arg Leu Gly Thr Thr His His Met Thr  
 115 120 125  
 Lys Arg Leu Leu Ile Asp Ser Ile Lys Tyr Leu Val Asp Thr Tyr Lys  
 130 135 140  
 Val Asp Gly Phe Arg Phe Asp Met Met Gly Asp His Asp Ala Ala Ser  
 145 150 155 160  
 Ile Glu Glu Ala Tyr Lys Ala Ala Arg Ala Leu Asn Pro Asn Leu Ile  
 165 170 175  
 Met Leu Gly Glu Gly Trp Arg Thr Tyr Ala Gly Asp Glu Asn Met Pro  
 180 185 190  
 Thr Lys Ala Ala Asp Gln Asp Trp Met Lys His Thr Asp Thr Val Ala  
 195 200 205  
 Val Phe Ser Asp Asp Ile Arg Asn Asn Leu Lys Ser Gly Tyr Pro Asn  
 210 215 220  
 Glu Gly Gln Pro Ala Phe Ile Thr Gly Gly Lys Arg Asp Val Asn Thr  
 225 230 235 240  
 Ile Phe Lys Asn Leu Ile Ala Gln Pro Thr Asn Phe Glu Ala Asp Ser  
 245 250 255  
 Pro Gly Asp Val Ile Gln Tyr Ile Ala Ala His Asp Asn Leu Thr Leu  
 260 265 270  
 Phe Asp Ile Ile Ala Gln Ser Ile Lys Lys Asp Pro Ser Lys Ala Glu  
 275 280 285  
 Asn Tyr Ala Glu Ile His Arg Arg Leu Arg Leu Gly Asn Leu Met Val  
 290 295 300  
 Leu Thr Ala Gln Gly Thr Pro Phe Ile His Ser Gly Gln Glu Tyr Gly  
 305 310 315 320  
 Arg Thr Lys Gln Phe Arg Asp Pro Ala Tyr Lys Thr Pro Val Ala Glu  
 325 330 335  
 Asp Lys Val Pro Asn Lys Ser His Leu Leu Arg Asp Lys Asp Gly Asn  
 340 345 350  
 Pro Phe Asp Tyr Pro Tyr Phe Ile His Asp Ser Tyr Asp Ser Ser Asp  
 355 360 365  
 Ala Val Asn Lys Phe Asp Trp Thr Lys Ala Thr Asp Gly Lys Ala Tyr

186

|                                                                 |     |         |
|-----------------------------------------------------------------|-----|---------|
| 370                                                             | 375 | 380     |
| Pro Glu Asn Val Lys Ser Arg Asp Tyr Met Lys Gly Leu Ile Ala Leu |     |         |
| 385                                                             | 390 | 395 400 |
| Arg Gln Ser Thr Asp Ala Phe Arg Leu Lys Ser Leu Gln Asp Ile Lys |     |         |
|                                                                 | 405 | 410 415 |
| Asp Arg Val His Leu Ile Thr Val Pro Gly Gln Asn Gly Val Glu Lys |     |         |
|                                                                 | 420 | 425 430 |
| Glu Asp Val Val Ile Gly Tyr Gln Ile Thr Ala Pro Asn Gly Asp Ile |     |         |
|                                                                 | 435 | 440 445 |
| Tyr Ala Val Phe Val Asn Ala Asp Glu Lys Ala Arg Glu Phe Asn Leu |     |         |
|                                                                 | 450 | 455 460 |
| Gly Thr Ala Phe Ala His Leu Arg Asn Ala Glu Val Leu Ala Asp Glu |     |         |
| 465                                                             | 470 | 475 480 |
| Asn Gln Ala Gly Pro Val Gly Ile Ala Asn Pro Lys Gly Leu Glu Trp |     |         |
|                                                                 | 485 | 490 495 |
| Thr Glu Lys Gly Leu Lys Leu Asn Ala Leu Thr Ala Thr Val Leu Arg |     |         |
|                                                                 | 500 | 505 510 |
| Val Ser Gln Asn Gly Thr Ser His Glu Ser Thr Ala Glu Glu Lys Pro |     |         |
|                                                                 | 515 | 520 525 |
| Asp Ser Thr Pro Ser Lys Pro Glu His Gln Asn Glu Ala Ser His Pro |     |         |
|                                                                 | 530 | 535 540 |
| Ala His Gln Asp Pro Ala Pro Glu Ala Arg Pro Asp Ser Thr Lys Pro |     |         |
| 545                                                             | 550 | 555 560 |
| Asp Ala Lys Val Ala Asp Ala Glu Asn Lys Pro Ser Gln Ala Thr Ala |     |         |
|                                                                 | 565 | 570 575 |
| Asp Ser Gln Ala Glu Gln Pro Ala Gln Glu Ala Gln Ala Ser Ser Val |     |         |
|                                                                 | 580 | 585 590 |
| Lys Glu Ala Val Arg Asn Glu Ser Val Glu Asn Ser Ser Lys Glu Asn |     |         |
|                                                                 | 595 | 600 605 |
| Ile Pro Ala Thr Pro Asp Lys Gln Ala Glu                         |     |         |
| 610                                                             | 615 |         |

## (2) INFORMATION FOR SEQ ID NO: 73:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1774 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| TAGTGATGGT ACTTGCAAG GAAAACAGTA TCTGAAAGAA GATGGCAGTC AAGCAGCAAA  | 60  |
| TGAGTGGGTT TTNGATACTC ATTATCAATC TTGGTTCTAT ATAAAAGCAG ATGCTAACTA | 120 |
| TGCTGAAAAT GAATGGCTAA AGCAAGGTGA CGACTATTTT TACCTCAAAT CTGGTGGCTA | 180 |

187

|                                                                    |      |
|--------------------------------------------------------------------|------|
| TATGGCCAAA TCAGAATGGG TAGAAGACAA GGGAGCCTTT TATTATCTTG ACCAAGATGG  | 240  |
| AAAGATGAAA AGAAATGCTT GGGTAGGAAC TTCCTATGTT GGTGCAACAG GTGCCAAAGT  | 300  |
| AATAGAAGAC TGGGTCTATG ATTCTCAATA CGATGCTTGG TTTTATATCA AAGCAGATGG  | 360  |
| ACAGCACGCA GAGAAAGAAT GGCTCCAAAT TAAAGGGAAG GACTATTATT TCAAATCCGG  | 420  |
| TGGTTATCTA CTGACAAGTC AGTGGATTAA TCAAGCTTAT GTGAATGCTA GTGGTGCCAA  | 480  |
| AGTACAGCAA GGTGCGCTTT TTGACAAACA ATACCAATCT TGGTTTTACA TCAAAGAAAA  | 540  |
| TGGAACTAT GCTGATAAAG AATGGATTTT CGAGAATGGT CACTATTATT ATCTAAAATC   | 600  |
| CGGTGGCTAC ATGGCAGCCA ATGAATGGAT TTGGGATAAG GAATCTTGGT TTTATCTCAA  | 660  |
| ATTTGATGGG AAAATGGCTG AAAAAGAATG GGTCTACGAT TCTCATAGTC AAGCTTG GTA | 720  |
| CTACTTCAA TCCGGTGGTT ACATGACAGC CAATGAATGG ATTTGGGATA AGGAATCTTG   | 780  |
| GTTTTACCTC AAATCTGATG GGAAAATAGC TGAAAAGAA TGGGTCTACG ATTCTCATAG   | 840  |
| TCAAGCTTGG TACTACTTCA AATCTGGTGG CTACATGGCG AAAAATGAGA CAGTAGATGG  | 900  |
| TTATCAGCTT GGAAGCGATG GTAAATGGCT TGGAGGAAAA ACTACAAATG AAAATGCTGC  | 960  |
| TTACTATCAA GTAGTGCCTG TTACAGCCAA TGTTTTATGAT TCAGATGGTG AAAAGCTTTC | 1020 |
| CTATATATCG CAAGGTAGTG TCGTATGGCT AGATAAGGAT AGAAAAAGTG ATGACAAGCG  | 1080 |
| CTTGCTATT ACTATTTCTG GTTTGTCAGG CTATATGAAA ACAGAAGATT TACAAGCGCT   | 1140 |
| AGATGCTAGT AAGGACTTTA TCCCTTATTA TGAGAGTGAT GGCCACCGTT TTTATCACTA  | 1200 |
| TGTGGCTCAG AATGCTAGTA TCCAGTAGC TTCTCATCTT TCTGATATGG AAGTAGGCAA   | 1260 |
| GAAATATTAT TCGGCAGATG GCCTGCATTT TGATGGTTTT AAGCTTGAGA ATCCCTTCCT  | 1320 |
| TTTCAAAGAT TTAACAGAGG CTACAACTA CAGTGCTGAA GAATTGGATA AGGTATTTAG   | 1380 |
| TTTGCTAAAC ATTAACAATA GCCTTTTGA GAACAAGGCG GCTACTTTTA AGGAAGCCGA   | 1440 |
| AGAACATTAC CATATCAATG CTCTTTATCT CCTTGCCCAT AGTGCCCTAG AAAGTAACTG  | 1500 |
| GGGAAGAAGT AAAATTGCCA AAGATAAGAA TAATTTCTTT GGCATTACAG CCTATGATAC  | 1560 |
| GACCCCTTAC CTTTCTGCTA AGACATTTGA TGATGTGGAT AAGGGAATTT TAGGTGCAAC  | 1620 |
| CAAGTGGAAT AAGGAAAATT ATATCGATAG GGGAAGAACT TTCCTTGGA ACAAGGCTTC   | 1680 |
| TGGTATGAAT GTGGAATATG CTTAGACCC TTATTGGGGC GAAAAAATTG CTAGTGTGAT   | 1740 |
| GATGAAAATC AATGAGAAGC TAGGTGGCAA AGAT                              | 1774 |

(2) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 591 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

188

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

```

Ser Asp Gly Thr Trp Gln Gly Lys Gln Tyr Leu Lys Glu Asp Gly Ser
1          5          10          15

Gln Ala Ala Asn Glu Trp Val Xaa Asp Thr His Tyr Gln Ser Trp Phe
20          25          30

Tyr Ile Lys Ala Asp Ala Asn Tyr Ala Glu Asn Glu Trp Leu Lys Gln
35          40          45

Gly Asp Asp Tyr Phe Tyr Leu Lys Ser Gly Gly Tyr Met Ala Lys Ser
50          55          60

Glu Trp Val Glu Asp Lys Gly Ala Phe Tyr Tyr Leu Asp Gln Asp Gly
65          70          75          80

Lys Met Lys Arg Asn Ala Trp Val Gly Thr Ser Tyr Val Gly Ala Thr
85          90          95

Gly Ala Lys Val Ile Glu Asp Trp Val Tyr Asp Ser Gln Tyr Asp Ala
100         105         110

Trp Phe Tyr Ile Lys Ala Asp Gly Gln His Ala Glu Lys Glu Trp Leu
115         120         125

Gln Ile Lys Gly Lys Asp Tyr Tyr Phe Lys Ser Gly Gly Tyr Leu Leu
130         135         140

Thr Ser Gln Trp Ile Asn Gln Ala Tyr Val Asn Ala Ser Gly Ala Lys
145         150         155         160

Val Gln Gln Gly Trp Leu Phe Asp Lys Gln Tyr Gln Ser Trp Phe Tyr
165         170         175

Ile Lys Glu Asn Gly Asn Tyr Ala Asp Lys Glu Trp Ile Phe Glu Asn
180         185         190

Gly His Tyr Tyr Tyr Leu Lys Ser Gly Gly Tyr Met Ala Ala Asn Glu
195         200         205

Trp Ile Trp Asp Lys Glu Ser Trp Phe Tyr Leu Lys Phe Asp Gly Lys
210         215         220

Met Ala Glu Lys Glu Trp Val Tyr Asp Ser His Ser Gln Ala Trp Tyr
225         230         235         240

Tyr Phe Lys Ser Gly Gly Tyr Met Thr Ala Asn Glu Trp Ile Trp Asp
245         250         255

Lys Glu Ser Trp Phe Tyr Leu Lys Ser Asp Gly Lys Ile Ala Glu Lys
260         265         270

Glu Trp Val Tyr Asp Ser His Ser Gln Ala Trp Tyr Tyr Phe Lys Ser
275         280         285

Gly Gly Tyr Met Ala Lys Asn Glu Thr Val Asp Gly Tyr Gln Leu Gly
290         295         300

Ser Asp Gly Lys Trp Leu Gly Gly Lys Thr Thr Asn Glu Asn Ala Ala
305         310         315         320

```

189

Tyr Tyr Gln Val Val Pro Val Thr Ala Asn Val Tyr Asp Ser Asp Gly  
 325 330 335  
 Glu Lys Leu Ser Tyr Ile Ser Gln Gly Ser Val Val Trp Leu Asp Lys  
 340 345 350  
 Asp Arg Lys Ser Asp Asp Lys Arg Leu Ala Ile Thr Ile Ser Gly Leu  
 355 360 365  
 Ser Gly Tyr Met Lys Thr Glu Asp Leu Gln Ala Leu Asp Ala Ser Lys  
 370 375 380  
 Asp Phe Ile Pro Tyr Tyr Glu Ser Asp Gly His Arg Phe Tyr His Tyr  
 385 390 395 400  
 Val Ala Gln Asn Ala Ser Ile Pro Val Ala Ser His Leu Ser Asp Met  
 405 410 415  
 Glu Val Gly Lys Lys Tyr Tyr Ser Ala Asp Gly Leu His Phe Asp Gly  
 420 425 430  
 Phe Lys Leu Glu Asn Pro Phe Leu Phe Lys Asp Leu Thr Glu Ala Thr  
 435 440 445  
 Asn Tyr Ser Ala Glu Glu Leu Asp Lys Val Phe Ser Leu Leu Asn Ile  
 450 455 460  
 Asn Asn Ser Leu Leu Glu Asn Lys Gly Ala Thr Phe Lys Glu Ala Glu  
 465 470 475 480  
 Glu His Tyr His Ile Asn Ala Leu Tyr Leu Leu Ala His Ser Ala Leu  
 485 490 495  
 Glu Ser Asn Trp Gly Arg Ser Lys Ile Ala Lys Asp Lys Asn Asn Phe  
 500 505 510  
 Phe Gly Ile Thr Ala Tyr Asp Thr Thr Pro Tyr Leu Ser Ala Lys Thr  
 515 520 525  
 Phe Asp Asp Val Asp Lys Gly Ile Leu Gly Ala Thr Lys Trp Ile Lys  
 530 535 540  
 Glu Asn Tyr Ile Asp Arg Gly Arg Thr Phe Leu Gly Asn Lys Ala Ser  
 545 550 555 560  
 Gly Met Asn Val Glu Tyr Ala Ser Asp Pro Tyr Trp Gly Glu Lys Ile  
 565 570 575  
 Ala Ser Val Met Met Lys Ile Asn Glu Lys Leu Gly Gly Lys Asp  
 580 585 590

(2) INFORMATION FOR SEQ ID NO: 75:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1105 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

TGGGATTCAA TATGTCAGAG ATGATACTAG AGATAAAGAA GAGGGAATAG AGTATGATGA

60

190

```

CGCTGACAAT GGGGATATTA TTGTAAAAGT AGCGACTAAA CCTAAGGTAG TAACCAAGAA      120
AATTTCAAGT ACGCGAATTC GTTATGAAAA AGATGAAACA AAAGACCGTA GTGAAAATCC      180
TGTTACAATT GATGGAGAGG ATGGCTATGT AACTACGACA AGGACCTACG ATGTTAATCC      240
AGAGACTGGT TATGTTACCG AACAGGTTAC TGTTGATAGA AAAGAAGCCA CGGATACAGT      300
TATCAAAGTT CCAGCTAAAA GCAAGGTTGA AGAAGTTCTT GTTCCATTG CTACTAAATA      360
TGAAGCAGAC AATGACCTTT CTGCAGGACA GGAGCAAGAG ATTACTCTAG GAAAGAATGG      420
GAAAACAGTT ACAACGATAA CTTATAATGT AGATGGAAAG AGTGGACAAG TAACTGAGAG      480
TACTTTAAGT CAAAAAAG ACTCtCAAAC AAGAGTTGTT AAAAAAGaA CCArkCCCCA      540
AGTTCTTGTC CAAGAAATTC CAATCGAAAC AGAATATCTC GATGGCCCaA CTCTTGATAA      600
AaGTCAAGAA GTAGAAGAAG TAGGAGAAAT TGGTAAATTA CTCTTACTAC AATCTATACT      660
GGTAGATGAA CGTGATGGAA CAATTGAAGA AACTACTTCT CGTCAAATTA CTAAAGAGAT      720
GGTAAAAAGA CGTATAAGGA GAGGGACGAG AGAACCTGAA AAAGTTGTTG TTCCTGAGCA      780
ATCATCTATT CCTTCGTATC CTGTATCTGT TACATCTAAC CAAGGAACAG ATGTAGCAGT      840
AGAACCAGCT AAAGCAGTTG CTCCAACAAC AGACTGGAAA CAAGAAAATG GTATGTGGTA      900
TTTTTATAAT ACTGATGGTT CCATGGCAAC AGGTTGGGTA CAAGTTAATA GTTCATGGTA      960
CTACCTCAAC AGCAACGGTT CTATGAAAGT CAATCAATGG TTCCAAGTTG GTGGTAAATG     1020
GTATTATGTA AATACATCGG GTGAGTTAGC GGTCAATACA AGTATAGATG GCTATAGAGT     1080
CAATGATAAT GGTGAATGGG TCGCT                                           1105

```

## (2) INFORMATION FOR SEQ ID NO:76:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 368 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

```

Gly Ile Gln Tyr Val Arg Asp Asp Thr Arg Asp Lys Glu Glu Gly Ile
1           5           10           15
Glu Tyr Asp Asp Ala Asp Asn Gly Asp Ile Ile Val Lys Val Ala Thr
.           20           25           30
Lys Pro Lys Val Val Thr Lys Lys Ile Ser Ser Thr Arg Ile Arg Tyr
           35           40           45
Glu Lys Asp Glu Thr Lys Asp Arg Ser Glu Asn Pro Val Thr Ile Asp
50           55           60
Gly Glu Asp Gly Tyr Val Thr Thr Arg Thr Tyr Asp Val Asn Pro

```

191

|                                                                 |     |    |     |    |     |     |
|-----------------------------------------------------------------|-----|----|-----|----|-----|-----|
| 65                                                              |     | 70 |     | 75 |     | 80  |
| Glu Thr Gly Tyr Val Thr Glu Gln Val Thr Val Asp Arg Lys Glu Ala | 85  |    | 90  |    | 95  |     |
| Thr Asp Thr Val Ile Lys Val Pro Ala Lys Ser Lys Val Glu Glu Val | 100 |    | 105 |    | 110 |     |
| Leu Val Pro Phe Ala Thr Lys Tyr Glu Ala Asp Asn Asp Leu Ser Ala | 115 |    | 120 |    | 125 |     |
| Gly Gln Glu Gln Glu Ile Thr Leu Gly Lys Asn Gly Lys Thr Val Thr | 130 |    | 135 |    | 140 |     |
| Thr Ile Thr Tyr Asn Val Asp Gly Lys Ser Gly Gln Val Thr Glu Ser | 145 |    | 150 |    | 155 | 160 |
| Thr Leu Ser Gln Lys Lys Asp Ser Gln Thr Arg Val Val Lys Lys Arg | 165 |    | 170 |    | 175 |     |
| Thr Xaa Pro Gln Val Leu Val Gln Glu Ile Pro Ile Glu Thr Glu Tyr | 180 |    | 185 |    | 190 |     |
| Leu Asp Gly Pro Thr Leu Asp Lys Ser Gln Glu Val Glu Glu Val Gly | 195 |    | 200 |    | 205 |     |
| Glu Ile Gly Lys Leu Leu Leu Leu Gln Ser Ile Leu Val Asp Glu Arg | 210 |    | 215 |    | 220 |     |
| Asp Gly Thr Ile Glu Glu Thr Thr Ser Arg Gln Ile Thr Lys Glu Met | 225 |    | 230 |    | 235 | 240 |
| Val Lys Arg Arg Ile Arg Arg Gly Thr Arg Glu Pro Glu Lys Val Val | 245 |    | 250 |    | 255 |     |
| Val Pro Glu Gln Ser Ser Ile Pro Ser Tyr Pro Val Ser Val Thr Ser | 260 |    | 265 |    | 270 |     |
| Asn Gln Gly Thr Asp Val Ala Val Glu Pro Ala Lys Ala Val Ala Pro | 275 |    | 280 |    | 285 |     |
| Thr Thr Asp Trp Lys Gln Glu Asn Gly Met Trp Tyr Phe Tyr Asn Thr | 290 |    | 295 |    | 300 |     |
| Asp Gly Ser Met Ala Thr Gly Trp Val Gln Val Asn Ser Ser Trp Tyr | 305 |    | 310 |    | 315 | 320 |
| Tyr Leu Asn Ser Asn Gly Ser Met Lys Val Asn Gln Trp Phe Gln Val | 325 |    | 330 |    | 335 |     |
| Gly Gly Lys Trp Tyr Tyr Val Asn Thr Ser Gly Glu Leu Ala Val Asn | 340 |    | 345 |    | 350 |     |
| Thr Ser Ile Asp Gly Tyr Arg Val Asn Asp Asn Gly Glu Trp Val Arg | 355 |    | 360 |    | 365 |     |

## (2) INFORMATION FOR SEQ ID NO: 77:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 661 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

*What Is Claimed Is:*

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding any of the amino acid sequences of the polypeptides shown in Table 1; or

(b) a nucleotide sequence complementary to any of the nucleotide sequences in (a).

2. An isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide having a nucleotide sequence identical to a nucleotide sequence in (a) or (b) of claim 1 wherein said polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues.

3. An isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of a polypeptide having an amino acid sequence in (a) of claim 1.

4. The isolated nucleic acid molecule of claim 3, wherein said epitope-bearing portion of a polypeptide has an amino acid sequence listed in Table 2.

5. A method for making a recombinant vector comprising inserting an isolated nucleic acid molecule of claim 1 into a vector.

6. A recombinant vector produced by the method of claim 5.

7. A method of making a recombinant host cell comprising introducing the recombinant vector of claim 6 into a host cell.

8. A recombinant host cell produced by the method of claim 7.

9. A method of producing a polypeptide encoded by the nucleic acid molecule of claim 1 comprising culturing the host cell of claim 8 under conditions favoring expressing the heterologous polypeptide.

10. A polypeptide produced according to the method of claim 9.

11. An isolated polypeptide comprising an amino acid sequence at least 70% identical to a sequence selected from the group consisting of an amino acid sequence of any of the polypeptides described in Table 1.

12. An isolated polypeptide antigen comprising an amino acid sequence of an *S. pneumoniae* epitope shown in Table 2.

13. An isolated nucleic acid molecule comprising a polynucleotide with a nucleotide sequence encoding a polypeptide of claim 9.

14. An isolated antibody that binds specifically to a polypeptide of claim 11.

15. A hybridoma which produces an antibody according to claim 14.

16. A vaccine, comprising:

(1) one of more *S. pneumoniae* polypeptides selected from the group consisting of a polypeptide comprising an amino acid sequence identified in Table 1, or a fragment thereof; and

(2) a pharmaceutically acceptable diluent, carrier, or excipient; wherein said polypeptide is present, in an amount effective to elicit protective antibodies in an animal to a member of the *Streptococcus* genus.

17. A method of preventing or attenuating an infection caused by a member of the *Streptococcus* genus in an animal, comprising administering to said animal a polypeptide of claim 11, wherein said polypeptide is administered in an amount effective to prevent or attenuate said infection.

18. A method of detecting *Streptococcus* nucleic acids in a biological sample obtained from an animal involving assaying for one or more nucleic acid sequences encoding *Streptococcus* polypeptides in a sample comprising:

(a) contacting the sample with one or more of the above-described nucleic acid probes, under conditions such that hybridization occurs, and

(b) detecting hybridization of said one or more probes to the one or more *Streptococcus* nucleic acid sequences present in the biological sample.

19. A method of detecting *Streptococcus* nucleic acids in a biological sample obtained from an animal, comprising:

- 5 (a) amplifying one or more *Streptococcus* nucleic acid sequences in said sample using polymerase chain reaction, and  
(b) detecting said amplified *Streptococcus* nucleic acid.

20. A kit for detecting *Streptococcus* antibodies in a biological sample obtained from an animal, comprising

- 10 (a) a polypeptide of claim 12 attached to a solid support; and  
(b) detecting means.

21. A method of detecting *Streptococcus* antibodies in a biological sample obtained from an animal, comprising

- 15 (a) contacting the sample with a polypeptide of claim 12; and  
(b) detecting antibody-antigen complexes.