Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2321106 C 2013/07/23

(11)(21) **2 321 106**

(12) **BREVET CANADIEN**
CANADIAN PATENT

(13) **C**

(86) Date de dépôt PCT/PCT Filing Date: 1999/02/17
(87) Date publication PCT/PCT Publication Date: 1999/08/26
(45) Date de délivrance/Issue Date: 2013/07/23
(85) Entrée phase nationale/National Entry: 2000/08/18
(86) N° demande PCT/PCT Application No.: CA 1999/000114
(87) N° publication PCT/PCT Publication No.: 1999/042588
(30) Priorité/Priority: 1998/02/20 (US60/075,425)

(51) Cl.Int./Int.Cl. *C12N 15/31* (2006.01), *A61K 39/09* (2006.01), *C07K 14/315* (2006.01), *C12N 1/21* (2006.01), *A61K 39/00* (2006.01)

(72) Inventeurs/Inventors:
BRODEUR, BERNARD R., CA;
RIOUX, CLEMENT, CA;
BOYER, MARTINE, CA;
CHARLEBOIS, ISABELLE, CA;
HAMEL, JOSEE, CA;
MARTIN, DENIS, CA

(73) Propriétaire/Owner:
ID BIOMEDICAL CORPORATION OF QUEBEC, CA

(74) Agent: NORTON ROSE FULBRIGHT CANADA LLP/S.E.N.C.R.L., S.R.L.

(54) Titre : ANTIGENES DU STREPTOCOQUE DU GROUPE B
(54) Title: GROUP B STREPTOCOCCUS ANTIGENS

(57) Abrégé/Abstract:
Group B streptococcus (GBS) proteins and polynucleotides encoding them are disclosed. Said proteins are antigenic and therefore useful vaccine components for the prophylaxis or therapy of streptococcus infection in animals. Also disclosed are recombinant methods of producing the protein antigens as well as diagnostic assays for detecting streptococcus bacterial infection.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/31, C07K 14/315, A61K 39/09, C12N 1/21

A2

(11) International Publication Number: WO 99/42588

(43) International Publication Date: 26 August 1999 (26.08.99)

(21) International Application Number: PCT/CA99/00114

(22) International Filing Date: 17 February 1999 (17.02.99)

(30) Priority Data:
60/075,425 20 February 1998 (20.02.98) US

(71) Applicant *(for all designated States except US)*: BIOCHEM VACCINS INC. [CA/CA]; 2323 boulevard du Parc Technologique, Sainte-Foy, Québec G1P 4R8 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BRODEUR, Bernard, R. [CA/CA]; 2401 rue Maritain, Sillery, Québec G1T 1N6 (CA). RIOUX, Clément [CA/CA]; 1012 Jean-Charles Cantin, Ville de Cap Rouge, Québec G1Y 2X1 (CA). BOYER, Martine [CA/CA]; Apt. 204, 25 des Mouettes, Beauport, Québec G1E 7G1 (CA). CHARLEBOIS, Isabelle [CA/CA]; 410 Mirabel, St-Nicolas, Québec G7A 2L5 (CA). HAMEL, Josée [CA/CA]; 2401 rue Maritain, Sillery, Québec G1T 1N6 (CA). MARTIN, Denis [CA/CA]; 4728-G rue Gaboury, St-Augustin-de-Desmaures, Québec G3A 1E9 (CA).

(74) Agents: CÔTE, France et al.; Swabey Ogilvy Renault, Suite 1600, 1981 McGill College Avenue, Montréal, Québec H3A 2Y3 (CA).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: GROUP B STREPTOCOCCUS ANTIGENS

(57) Abstract

Group B streptococcus (GBS) proteins and polynucleotides encoding them are disclosed. Said proteins are antigenic and therefore useful vaccine components for the prophylaxis or therapy of streptococcus infection in animals. Also disclosed are recombinant methods of producing the protein antigens as well as diagnostic assays for detecting streptococcus bacterial infection.

GROUP B STREPTOCOCCUS ANTIGENS

FIELD OF THE INVENTION

The present invention is related to antigens, more particularly protein antigens of group B streptococcus (GBS) bacterial pathogen which are useful as vaccine components for therapy and/or prophylaxis.

BACKGROUND OF THE INVENTION

Streptococcus are gram (+) bacteria that are differentiated by group specific carbohydrate antigens A through O found on their cell surface. Streptococcus groups are further distinguished by type-specific capsular polysaccharide antigens. Several serotypes have been identified for the Group B streptococcus (GBS) : Ia, Ib, II, III, IV, V, VI, VII and VIII. GBS also contains antigenic proteins known as "C-proteins" (alpha, beta, gamma and delta), some of which have been cloned.

Although GBS is a common component of the normal human vaginal and colonic flora this pathogen has long been recognized as a major cause of neonatal sepsis and meningitis, late-onset meningitis in infants, postpartum endometritis as well as mastitis in dairy herds. Expectant mothers exposed to GBS are at risk of postpartum infection and may transfer the infection to their baby as the child passes through the birth canal. Although the organism is sensitive to antibiotics, the high attack rate and rapid onset of sepsis in neonates and meningitis in infants results in high morbidity and mortality.

To find a vaccine that will protect individuals from GBS infection, researches have turned to the type-specific antigens. Unfortunately these polysaccharides have proven to be poorly immunogenic in humans and are restricted to the particular serotype from which the polysaccharide originates. Further, capsular polysaccharide elicit a T cell independent response i.e. no IgG production. Consequently capsular polysaccharide antigens are unsuitable as a vaccine component for protection against GBS infection.

Others have focused on the C-protein beta antigen which demonstrated immunogenic properties in mice and rabbit models. This protein was found to be unsuitable as a human vaccine because of its undesirable property of interacting with high affinity and in a non-immunogenic manner with the Fc region of human IgA. The C-protein alpha antigen is rare in type III serotypes of GBS which is the serotype responsible for most GBS mediated conditions and is therefore of little use as a vaccine component.

Therefore there remains an unmet need for GBS antigens that may be used as vaccine components for the prophylaxis and/or therapy of GBS infection.

SUMMARY OF THE INVENTION

According to one aspect, the present invention provides an isolated polynucleotide encoding a polypeptide having at least 70% identity to a second polypeptide comprising a sequence selected from the group consisting of: SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19,

SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41 and SEQ ID NO:44 or fragments, analogs or derivatives thereof.

In other aspects, there is provided vectors comprising polynucleotides of the invention operably linked to an expression control region, as well as host cells transfected with said vectors and methods of producing polypeptides comprising culturing said host cells under conditions suitable for expression.

In yet another aspect, there is provided novel polypeptides encoded by polynucleotides of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1a is the DNA sequence of clone 1 (SEQ ID NO :1) with corresponding amino acid sequences for open reading frames;
figure 1b is the amino acid sequence SEQ ID NO: 2;
figure 1c is the amino acid sequence SEQ ID NO: 3;
figure 1d is the amino acid sequence SEQ ID NO: 4;
figure 1e is the amino acid sequence SEQ ID NO: 5;
figure 1f is the amino acid sequence SEQ ID NO: 6;

Figure 2a is the DNA sequence of clone 2 (SEQ ID NO :7) with corresponding amino acid sequences for open reading frames;
figure 2b is the amino acid sequence SEQ ID NO: 8;
figure 2c is the amino acid sequence SEQ ID NO: 9;
figure 2d is the amino acid sequence SEQ ID NO:10;
figure 2e is the amino acid sequence SEQ ID NO:11;
figure 2f is the amino acid sequence SEQ ID NO:12;

Figure 3a is the DNA sequence of clone 3 (SEQ ID NO :13) with corresponding amino acid sequences for open reading frames;
figure 3b is the amino acid sequence SEQ ID NO:14;
figure 3c is the amino acid sequence SEQ ID NO:15;
figure 3d is the amino acid sequence SEQ ID NO:16;
figure 3e is the amino acid sequence SEQ ID NO:17;
figure 3f is the amino acid sequence SEQ ID NO:18;
figure 3g is the amino acid sequence SEQ ID NO:19;
figure 3h is the amino acid sequence SEQ ID NO:20;
figure 3i is the amino acid sequence SEQ ID NO:21;

Figure 4a is the DNA sequence of clone 4 (SEQ ID NO :22) with corresponding amino acid sequences for open reading frames;
figure 4b is the amino acid sequence SEQ ID NO:23;
figure 4c is the amino acid sequence SEQ ID NO:24;
figure 4d is the amino acid sequence SEQ ID NO:25;
figure 4e is the amino acid sequence SEQ ID NO:26;

Figure 5a is the DNA sequence of clone 5 (SEQ ID NO :27) with corresponding amino acid sequences for open reading frames;
figure 5b is the amino acid sequence SEQ ID NO:28;
figure 5c is the amino acid sequence SEQ ID NO:29;
figure 5d is the amino acid sequence SEQ ID NO:30;
figure 5e is the amino acid sequence SEQ ID NO:31;

Figure 6a is the DNA sequence of clone 6 (SEQ ID NO :32) ;
figure 6b is the amino acid sequence SEQ ID NO:33;
figure 6c is the amino acid sequence SEQ ID NO:34;
figure 6d is the amino acid sequence SEQ ID NO:35;
figure 6e is the amino acid sequence SEQ ID NO:36;

Figure 7a is the DNA sequence of clone 7 (SEQ ID NO :37);
figure 7b is the amino acid sequence SEQ ID NO:38;

figure 7c is the amino acid sequence SEQ ID NO:39;
figure 7d is the amino acid sequence SEQ ID NO:40;
figure 7e is the amino acid sequence SEQ ID NO:41;

Figure 8 is the DNA sequence of a part of clone 7 including a signal sequence (SEQ ID NO :42);

Figure 9 is the DNA sequence of a part of clone 7 without a signal sequence (SEQ ID NO :43);
Figure 9a is the amino acid sequence (SEQ ID NO:44);

Figure 10 represents the distribution of anti-GBS ELISA titers in sera from CD-1 mice immunized with recombinant GBS protein corresponding to the SEQ ID NO:39.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to novel antigenic polypeptides of group B streptococcus (GBS) characterized by the amino acid sequence selected from the group consisting of:
SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41 and SEQ ID NO:44 or fragments, analogs or derivatives thereof.

A preferred embodiment of the invention includes SEQ ID NO :39 and SEQ ID NO:44.

A further preferred embodiment of the invention is SEQ ID NO :39.

A further preferred embodiment of the invention is SEQ ID NO :44.

As used herein, "fragments", "derivatives" or "analogs" of the polypeptides of the invention include those polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably conserved) and which may be natural or unnatural.

The terms «fragments», «derivatives» or «analogues» of polypeptides of the present invention also include polypeptides which are modified by addition, deletion,

substitution of amino acids provided that the polypeptides retain the capacity to induce an immune response.

By the term «conserved amino acid» is meant a substitution of one or more amino acids for another in which the antigenic determinant (including its secondary structure and hydropathic nature) of a given antigen is completely or partially conserved in spite of the substitution.

For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity, which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

Preferably, derivatives and analogs of polypeptides of the invention will have about 70% identity with those sequences illustrated in the figures or fragments thereof. That is, 70% of the residues are the same. More preferably polypeptides will have greater than 95% homology. In another preferred embodiment, derivatives and analogs of polypeptides of the invention will have fewer than about 20 amino acid residue substitutions, modifications or deletions and more preferably less than 10. Preferred substitutions are those known in the art as conserved i.e. the substituted residues share physical or chemical properties such as hydrophobicity, size, charge or functional groups.

Furthermore, in those situations where amino acid regions are found to be polymorphic, it may be desirable to vary one or more particular amino acids to more effectively mimic the different epitopes of the different GBS strains.

Also included are polypeptides which have fused thereto other compounds which alter the polypeptides biological or pharmacological properties i.e. polyethylene glycol (PEG) to increase half-life; leader or secretory amino acid sequences for ease of purification; prepro- and pro- sequences; and (poly)saccharides.

Moreover, the polypeptides of the present invention can be modified by terminal -NH_2 acylation (eg. by acetylation, or thioglycolic acid amidation, terminal carbosy amidation, e.g. with ammonia or methylamine) to provide stability, increased hydrophobicity for linking or binding to a support or other molecule.

Also contemplated are hetero and homo polypeptide multimers of the polypeptide fragments, analogues and derivatives. These polymeric forms include, for example, one or more polypeptides that have been cross-linked with cross-linkers such as avidin/biotin, gluteraldehyde or dimethyl-superimidate. Such polymeric forms also include polypeptides containing two or more tandem or inverted contiguous sequences, produced from multicistronic mRNAs generated by recombinant DNA technology.
Preferably, a fragment, analog or derivative of a polypeptide of the invention will comprise at least one antigenic region i.e. at least one epitope.

In order to achieve the formation of antigenic polymers (i.e. synthetic multimers), polypeptides may be utilized having bishaloacetyl groups, nitroarylhalides, or the like,

where the reagents being specific for thio groups. Therefore, the link between two mercapto groups of the different peptides may be a single bond or may be composed of a linking group of at least two, typically at least four, and not more than 16, but usually not more than about 14 carbon atoms.

In a particular embodiment, polypeptide fragments, analogs and derivatives of the invention do not contain a methionine (Met) starting residue. Preferably, polypeptides will not incorporate a leader or secretory sequence (signal sequence). The signal portion of a polypeptide of the invention may be determined according to established molecular biological techniques. In general, the polypeptide of interest may be isolated from a GBS culture and subsequently sequenced to determine the initial residue of the mature protein and therefor the sequence of the mature polypeptide.

According to another aspect, there is provided vaccine compositions comprising one or more GBS polypeptides of the invention in admixture with a pharmaceutically acceptable carrier diluent or adjuvant.

Suitable adjuvants include oils i.e. Freund's complete or incomplete adjuvant; salts i.e. $AlK(SO_4)_2$, $AlNa(SO_4)_2$, $AlNH_4(SO_4)_2$, $Al(OH)_3$, $AlPO_4$, silica, kaolin; saponin derivative; carbon polynucleotides i.e. poly IC and poly AU and also detoxified cholera toxin (CTB)and E.coli heat labile toxin for induction of mucosal immunity. Preferred adjuvants include QuilA™, Alhydrogel™ and Adjuphos™. Vaccines of the invention may be administered parenterally by injection, rapid infusion, nasopharyngeal absorption, dermoabsorption, or bucal or oral.

Vaccine compositions of the invention are used for the treatment or prophylaxis of *streptococcus* infection and/or diseases and symptoms mediated by *streptococcus* infection, in particular group A *streptococcus* (*pyogenes*), group B *streptococcus* (GBS or *agalactiae*), *dysgalactiae*, *uberis*, *nocardia* as well as *Staphylococcus aureus*. General information about Streptococcus is available in Manual of Clinical Microbiology by P.R.Murray et al.(1995, 6th Edition, ASM Press, Washington, D.C.). More particularly group B *streptococcus*, *agalactiae*. In a particular embodiment vaccines are administered to those individuals at risk of GBS infection such as pregnant women and infants for sepsis, meningitis and pneumonia as well as immunocompromised individuals such as those with diabetes, liver disease or cancer. Vaccines may also have veterinary applications such as for the treatment of mastitis in cattle which is mediated by the above mentioned bacteria as well as *E.coli*.

The vaccine of the present invention can also be used for the manufacture of a medicament used for the treatment or prophylaxis of *streptococcus* infection and/or diseases and symptoms mediated by *streptococcus* infection, in particular group A *streptococcus* (*pyogenes*), group B *streptococcus* (GBS or *agalactiae*), *dysgalactiae*, *uberis*, *nocardia* as well as *Staphylococcus aureus*. More particularly group B *streptococcus*, *agalactiae*.

Vaccine compositions are preferably in unit dosage form of about 0.001 to 100 µg/kg (antigen/body weight) and more preferably 0.01 to 10 µg/kg and most preferably 0.1 to 1 µg/kg 1 to 3 times with an interval of about 1 to 12 weeks intervals between immunizations, and more preferably 1 to 6

weeks.

According to another aspect, there is provided polynucleotides encoding polypeptides of group B streptococcus (GBS) characterized by the amino acid sequence selected from the group consisting of:
SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41 and SEQ ID NO:44 or fragments, analogs or derivatives thereof.

Preferred polynucleotides are those illustrated in figures 1a (SEQ ID NO: 1), 2a (SEQ ID NO: 7), 3a (SEQ ID NO: 13), 4a (SEQ ID NO: 22), 5a (SEQ ID NO: 27), 6a (SEQ ID NO: 32), 7a (SEQ ID NO: 37), 8 (SEQ ID NO : 42) and 9(SEQ ID NO : 43) which correspond to the open reading frames, encoding polypeptides of the invention.

Preferred polynucleotides are those illustrated in figures 1a (SEQ ID NO: 1), 2a (SEQ ID NO: 7), 3a (SEQ ID NO: 13), 4a (SEQ ID NO: 22), 5a (SEQ ID NO: 27), 6a (SEQ ID NO: 32), 7a (SEQ ID NO: 37), 8 (SEQ ID NO : 42) and 9(SEQ ID NO : 43) and fragments, analogues and derivatives thereof.

More preferred polynucleotides of the invention are those illustrated in Figures 7 (SEQ ID NO : 37), 8 (SEQ ID NO : 42) and 9(SEQ ID NO : 43).

Most preferred polynucleotides of the invention are those illustrated in Figures 8 (SEQ ID NO : 42) and 9(SEQ ID NO :

43).

It will be appreciated that the polynucleotide sequences illustrated in the figures may be altered with degenerate codons yet still encode the polypeptides of the invention.

Due to the degeneracy of nucleotide coding sequences, other polynucleotide sequences which encode for substantially the same polypeptides of the present invention may be used in the practice of the present invention. These include but are not limited to nucleotide sequences which are altered by the substitution of different codons that encode the same amino acid residue within the sequence, thus producing a silent change.

Accordingly the present invention further provides polynucleotides which hybridize to the polynucleotide sequences herein above described (or the complement sequences thereof) having 50% and preferably at least 70% identity between sequences. More preferably polynucleotides are hybridizable under stringent conditions i.e. having at least 95% identity and most preferably more than 97% identity.

By capable of hybridizing under stringent conditions is meant annealing of a nucleic acid molecule to at least a region of a second nucleic acid sequence (whether as cDNA, mRNA, or genomic DNA) or to its complementary strand under standard conditions, e.g. high temperature and/or low salt content, which tend to disfavor hybridization of noncomplementary nucleotide sequences. A suitable protocol is described in Maniatis T. et al., Molecular cloning : A Laboratory Manual, Cold Springs Harbor Laboratory, 1982.

In a further aspect, polynucleotides encoding polypeptides

of the invention, or fragments, analogs or derivatives thereof, may be used in a DNA immunization method. That is, they can be incorporated into a vector which is replicable and expressible upon injection thereby producing the antigenic polypeptide in vivo. For example polynucleotides may be incorporated into a plasmid vector under the control of the CMV promoter which is functional in eukaryotic cells. Preferably the vector is injected intramuscularly.

According to another aspect, there is provided a process for producing polypeptides of the invention by recombinant techniques by expressing a polynucleotide encoding said polypeptide in a host cell and recovering the expressed polypeptide product. Alternatively, the polypeptides can be produced according to established synthetic chemical techniques i.e. solution phase or solid phase synthesis of oligopeptides which are ligated to produce the full polypeptide (block ligation).

For recombinant production, host cells are transfected with vectors which encode the polypeptide, and then cultured in a nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes. Suitable vectors are those that are viable and replicable in the chosen host and include chromosomal, non-chromosomal and synthetic DNA sequences e.g. bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA. The polypeptide sequence may be incorporated in the vector at the appropriate site using restriction enzymes such that it is operably linked to an expression control region comprising a promoter, ribosome binding site (consensus region or Shine-Dalgarno sequence), and optionally an operator (control element). One can select individual components of the expression control region that are appropriate for a given

host and vector according to established molecular biology principles (Sambrook et al, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor, N.Y., 1989).

Suitable promoters include but are not limited to LTR or SV40 promoter, *E.coli* lac, tac or trp promoters and the phage lambda P_L promoter. Vectors will preferably incorporate an origin of replication as well as selection markers i.e. ampicillin resistance gene. Suitable bacterial vectors include pET, pQE70, pQE60, pQE-9, pbs, pD10 phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18A, pNH46A, ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 and eukaryotic vectors pBlueBacIII, pWLNEO, pSV2CAT, pOG44, pXT1, pSG, pSVK3, pBPV, pMSG and pSVL. Host cells may be bacterial i.e. *E.coli*, *Bacillus subtilis*, *Streptomyces*; fungal i.e. *Aspergillus niger*, *Aspergillus nidulins*; yeast i.e. *Saccharomyces* or eukaryotic i.e. CHO, COS.

Upon expression of the polypeptide in culture, cells are typically harvested by centrifugation then disrupted by physical or chemical means (if the expressed polypeptide is not secreted into the media) and the resulting crude extract retained to isolate the polypeptide of interest. Purification of the polypeptide from culture media or lysate may be achieved by established techniques depending on the properties of the polypeptide i.e. using ammonium sulfate or ethanol precipitation , acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, hydroxylapatite chromatography and lectin chromatography. Final purification may be achieved using HPLC.

The polypeptide may be expressed with or without a leader or secretion sequence. In the former case the leader may be removed using post-translational processing (see US

4,431,739; 4,425,437; and 4,338,397)
or be chemically removed subsequent to purifying the expressed polypeptide.

According to a further aspect, the GBS polypeptides of the invention may be used in a diagnostic test for streptococcus infection in particular GBS infection. Several diagnostic methods are possible, for example detecting streptococcus organism in a biological sample, the following procedure may be followed:

a) obtaining a biological sample from a patient;
b) incubating an antibody or fragment thereof reactive with a GBS polypeptide of the invention with the biological sample to form a mixture; and
c) detecting specifically bound antibody or bound fragment in the mixture which indicates the presence of streptococcus.

Alternatively, a method for the detection of antibody specific to a streptococcus antigen in a biological sample containing or suspected of containing said antibody may be performed as follows:

a) isolating a biological sample from a patient;
b) incubating one or more GBS polypeptides of the invention or fragments thereof with the biological sample to form a mixture; and
c) detecting specifically bound antigen or bound fragment in the mixture which indicates the presence of antibody specific to streptococcus.

One of skill in the art will recognize that this diagnostic test may take several forms, including an immunological test such as an enzyme-linked immunosorbent assay (ELISA), a radioimmunoassay or a latex agglutination assay, essentially to determine whether antibodies specific for the protein are present in an organism.

The DNA sequences encoding polypeptides of the invention may also be used to design DNA probes for use in detecting the presence of streptococcus in a biological sample suspected of containing such bacteria. The detection method of this invention comprises:

a) isolating the biological sample from a patient;
b) incubating one or more DNA probes having a DNA sequence encoding a polypeptide of the invention or fragments thereof with the biological sample to form a mixture; and
c) detecting specifically bound DNA probe in the mixture which indicates the presence of streptococcus bacteria.

The DNA probes of this invention may also be used for detecting circulating streptococcus i.e. GBS nucleic acids in a sample, for example using a polymerase chain reaction, as a method of diagnosing streptococcus infections. The probe may be synthesized using conventional techniques and may be immobilized on a solid phase, or may be labeled with a detectable label. A preferred DNA probe for this application is an oligomer having a sequence complementary to at least about 6 contiguous nucleotides of the GBS polypeptides of the invention.

Another diagnostic method for the detection of streptococcus in a patient comprises:

a) labeling an antibody reactive with a polypeptide of the invention or fragment thereof with a detectable label;
b) administering the labeled antibody or labeled fragment to the patient; and
c) detecting specifically bound labeled antibody or labeled fragment in the patient which indicates the presence of streptococcus.

A further aspect of the invention is the use of the GBS

polypeptides of the invention as immunogens for the production of specific antibodies for the diagnosis and in particular the treatment of streptococcus infection. Suitable antibodies may be determined using appropriate screening methods, for example by measuring the ability of a particular antibody to passively protect against streptococcus infection in a test model. One example of an animal model is the mouse model described in the examples herein. The antibody may be a whole antibody or an antigen-binding fragment thereof and may in general belong to any immunoglobulin class. The antibody or fragment may be of animal origin, specifically of mammalian origin and more specifically of murine, rat or human origin. It may be a natural antibody or a fragment thereof, or if desired, a recombinant antibody or antibody fragment. The term recombinant antibody or antibody fragment means antibody or antibody fragment which were produced using molecular biology techniques. The antibody or antibody fragments may be polyclonal, or preferably monoclonal. It may be specific for a number of epitopes associated with the GBS polypeptides but is preferably specific for one.

EXAMPLE 1 Murine model of lethal Group B Streptococcus (GBS) infection

The mouse model of GBS infection is described in detail in Lancefield et al (J Exp Med 142:165-179,1975). GBS strain C388/90 (Clinical isolate obtained in 1990 from the cephalorachidian fluid of a patient suffering from meningitis, Children's Hospital of Eastern Ontario, Ottawa, Canada) and NCS246 (National Center for Streptococcus, Provincial Laboratory of Public Health for Northern Alberta, Edmonton, Canada) were respectively serotyped as type Ia/c and type II/R.

To increase their virulence, the GBS strains C388/90 (serotype Ia/c) and NCS 246 (serotype II/R) were serially passaged through mice as described previously (Lancefield et al. J Exp Med 142:165-179, 1975). Briefly, the increase of virulence was monitored using intraperitoneal inoculations of serial dilutions of a subculture in Todd-Hewitt broth obtained from either the blood or spleen of infected mice. After the last passage, infected blood samples were used to inoculate Todd-Hewitt broth. After an incubation of 2 hours at 37°C with 7% CO_2, glycerol at a final concentration of 10% (v/v) was added to the culture. The culture was then aliquoted and stored at -80° C for use in GBS challenge experiments. The number of cfu of GBS present in these frozen samples was determined. The bacterial concentration necessary to kill 100% (LD100) of the 18 weeks old mice were determined to be $3.5X10^5$ and $1.1X10^5$ respectively for GBS strain C388/90 and NCS246, which corresponded to a significant increase in virulence for both strains. Indeed, the LD100 recorded before the passages for these two strains was higher than 10^9 cfu.

In a bacterial challenge, a freshly thawed aliquot of a virulent GBS strain was adjusted to the appropriate bacterial concentration using Todd-Hewitt broth and 1ml was injected intraperitoneally to each female CD-1 mouse. The mice used for the passive protection experiments were 6 to 8 weeks old, while the ones used for the active protection experiments were approximately 18 weeks old at the time of the challenge. All inocula were verified by colony counts. Animals were observed for any sign of infection four times daily for the first 48 h after challenge and then daily for the next 12 days. At the end of that period, blood samples were obtained from the survivors and frozen at -20°C. The spleen obtained from each mouse that survived the challenge was cultured in order to identify any remaining GBS.

EXAMPLE 2 Immunization and protection in mice with formaldehyde killed whole GBS cells

Formaldehyde killed GBS whole cells were prepared according to the procedures described in Lancefield et al (J Exp Med 142:165-179,1975). Briefly, an overnight culture on sheep blood agar plates (Quelab Laboratories, Montreal, Canada) of a GBS strain was washed twice in PBS buffer (phosphate buffered-saline, pH7.2), adjusted to approximately $3X10^9$ cfu/mL and incubated overnight in PBS containing 0.3% (v/v) formaldehyde. The killed GBS suspension was washed with PBS and kept frozen at -80°C.

Female CD-1 mice, 6 to 8 weeks old (Charles River, St-Constant, Québec, Canada), were injected subcutaneously three times at two weeks interval with 0.1 ml of formaldehyde killed cells of GBS strain C388/90 (~$6X10^7$ GBS), or 0.1 ml of PBS for the control group. On the day before the immunization, Alhydrogel™ (Superfos Biosector, Frederikssund, Denmark) at a final concentration of 0.14 mg or 0.21 mg of Al, was added to these preparations and incubated overnight at 4°C with agitation. Serum samples were obtained from each mouse before the beginning of the immunization protocol and two weeks after the last injection. The sera were frozen at -20°C.

Eight mice in each control group injected with PBS and the group immunized with formaldehyde killed whole cells GBS strain C388/90 (Ia/c) were challenged with $1.5X10^4$ cfu of GBS strain C388/90 (Ia/c) one week after the third injection. All mice immunized with the formaldehyde killed GBS whole cells survived the homologous challenge while, within 5 days after the challenge, only 4 out of the 8 mice injected with PBS survived from the infection. In order to increase the mortality rate in the control groups, the

bacterial suspension had to be adjusted according to the age of the mice at the time of the bacterial challenge. In subsequent challenge experiments, when mice were older than 15 weeks, the bacterial inoculum was increased to concentrations between $3.0X10^5$ and $2.5X10^6$ cfu.

Table 1 Immunization of CD1 mice with formaldehyde killed whole cells of GBS and subsequent homologous challenge [strain C388/90 (Ia/c)] and heterologous challenge [strain NCS246 (II/R)].

antigenic preparations used for immunization[1]	number of living mice 14 days after the bacterial challenge (% Survival)	
	homologous challenge: strain C388/90 (Ia/c)	heterologous challenge: strain NCS246 (II/R)
1st infection		
formaldehyde killed cells of GBS strain C388/90 (Ia/c)[2]	8/8 (100)[3]	n.d.[5]
control PBS	4/8 (50)	n.d.
2nd infection		
formaldehyde killed cells of GBS strain C388/90 (Ia/c)	6/6 (100)[4]	0/6 (0)[6]
control PBS	2/6 (33)	0/6 (0)

[1] alhydrogel™ at a final concentration of 0.14 mg or 0.21mg of Al was used;
[2] approximately $6X10^7$ cfu;
[3] intraperitoneal challenge with 1 mL Todd-Hewitt culture medium containing GBS C388/90 (Ia/c) suspension adjusted to $1.5X10^4$ cfu;
[4] intraperitoneal challenge with 1 mL Todd-Hewitt culture medium containing GBS C388/90 (Ia/c) suspension adjusted to $2.1X10^6$ cfu;
[5] not done;
[6] intraperitoneal challenge with 1 mL Todd-Hewitt culture medium containing GBS NCS246 (II/R) suspension adjusted to $1.2X10^5$ cfu.

In another experiment, one group of 12 mice corresponding to a control group was injected with PBS, while a second group of 12 mice was immunized with formaldehyde killed whole cells of GBS strain C388/90 (Ia/c). Six mice from each of these two groups were challenged with $2.1X10^6$ cfu of the GBS strain C388/90 (Ia/c) (Table I). As the first challenge experiment, all mice immunized with the GBS strain C388/90 (Ia/c) survived the homologous challenge. Only two out of the 6 mice injected with PBS survived the infection.

The remaining 6 mice in both groups were then used one week later to verify whether this antigenic preparation could confer cross protection against strain NCS246 (II/R) which produce a serologically distinct capsule. None of the mice infected with this second GBS strain survived the infection. The later result suggested that most of the protective immune response induced by formaldehyde killed strain C388/90 is directed against the capsular polysaccharide and that it could be restricted to strains of that particular serotype. These results clearly indicated that this particular model of infection can be efficiently used to study the protection conferred by vaccination.

EXAMPLE 3 Immunization of rabbit with formaldehyde killed whole GBS cells and passive protection in mice

A New Zealand rabbit (2.5 kg, Charles River, St-Constant, Québec, Canada) was immunized with formaldehyde killed cells of GBS strain C388/90 (Ia/c) to obtain hyperimmune serum. This rabbit was injected subcutaneously three times at three weeks interval with approximately $1.5X10^9$ cfu of formaldehyde killed whole cells of GBS strain C388/90 (Ia/c). Freund's complete adjuvant (Gibco BRL Life Technologies, Grand Island, New York) was used as the adjuvant for the first immunization, while Freund's incomplete adjuvant (Gibco BRL) was used for the following two injections. Serum samples were obtained before the beginning of the immunization protocol and two weeks after the last injection. The sera were frozen at -20°C.

The ability of this particular rabbit hyperimmune serum to passively protect mice against a lethal infection with GBS

was also evaluated. Intraperitoneal injection of mice with either 15 or 25 μL of hyperimmune rabbit serum 18 hours before the challenge protected 4 out of 5 mice (80%) against the infection. Comparatively, survival rates lower than 20% were recorded for mice in the control group injected with PBS or serum obtained from a rabbit immunized with meningococcal outer membrane preparation. This result clearly indicates that the immunization of another animal species with killed GBS cells can induce the production of antibodies that can passively protect mice. This reagent will also be used to characterize clones.

Table 2 Passive protection of CD-1 mice conferred by rabbit serum obtained after immunization with formaldehyde killed group B whole streptococci (strain C388/90 (Ia/c)) antigenic preparation

groups	number of living mice 14 days after the bacterial challenge with GBS strain C388/90 (Ia/c)[2]	% survival
rabbit hyperimmune serum[2] - 25 μl	4/5	80
rabbit hyperimmune serum[1] - 15 μl	4/5	80
control rabbit serum - 25 μl	1/5	20
control PBS	1/10	10

[1] Freund's complete adjuvant was used for first immunization, and Freund's incomplete adjuvant for the following two injections;
[2] intraperitoneal challenge with 1 ml Todd-Hewitt culture medium containing GBS C388/90 (Ia/c) suspension adjusted to $2X10^4$ cfu.

EXAMPLE 4 Recombinant production of His.Tag-GBS fusion protein

The coding region of a GBS gene was amplified by PCR (DNA Thermal Cycler GeneAmp PCR system 2400 Perkin Elmer, San Jose, CA) from the genomic DNA of GBS strain C388/90 (Ia/c) using the oligos that contained base extensions for the addition of the restriction sites BglII (AGATCT) and HindIII (AAGCTT), respectively. The PCR product was purified from agarose gel using a Qiaex II gel extraction kit from Qiagen (Chatsworth, CA), digested with the restriction enzymes BglII and HindIII (Pharmacia Canada Inc Baie d'Urfe, Canada), and extracted with phenol:chloroform before ethanol precipitation. The pET-32b(+) vector (Novagen, Madison, WI) containing the thioredoxin-His.Tag sequence was digested with the restriction enzymes BglII and HindIII, extracted with phenol:chloroform, and then ethanol precipitated. The BglII-HindIII genomic DNA fragment was ligated to the BglII-HindIII pET-32b(+) vector to create the coding sequence for thioredoxin-His.Tag-GBS fusion protein whose gene was under control of the T7 promoter. The ligated products were transformed into *E. coli* strain XLI Blue MRF'(Δ(*mcr*A)183Δ (*mcr*CB-*hsd*SMR-*mrr*)173 *end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 lac [F'*pro*AB *lac*I^{q}ZΔM15Tn10 (Tet^r)]^c) (Stratagene, La Jolla, CA) according to the method of Simanis (Hanahan, D. DNA Cloning, 1985, D.M. Glover (ed.), pp. 109-135). The recombinant pET plasmid was purified using a Qiagen kit (Qiagen, Chatsworth, CA) and the nucleotide sequence of the DNA insert was verified by DNA sequencing (Taq Dye Deoxy Terminator Cycle Sequencing kit, ABI, Foster City, CA). The recombinant pET plasmid was transformed by electroporation (Gene Pulser II apparatus, BIO-RAD Labs, Mississauga, Canada) into *E. coli* strain AD494 (DE3) (Δ*ara*$^-$*leu*7697 Δ*lac*X74 Δ*pho*A *Pvu*II *pho*R Δ*mal*F3 F'[*lac*$^+$(*lac*I^q) *pro*] *trx*B::Kan (DE3)) (Novagen, Madison, WI). In this strain of

E. coli, the T7 promoter controlling expression of the fusion protein, is specifically recognized by the T7 RNA polymerase (present on the λDE3 prophage) whose gene is under the control of the lac promoter which is inducible by isopropyl-β-D-thio-galactopyranoside (IPTG).

The transformant AD494(DE3)/rpET was grown at 37°C with agitation at 250 rpm in LB broth (peptone 10g/L, Yeast extract 5g/L, NaCl 10g/L) containing 100μg of ampicillin (Sigma-Aldrich Canada Ltd., Oakville, Canada) per mL until the A_{600} reached a value of 0.6. In order to induce the production of the thioredoxin-His.Tag-GBS fusion protein, the cells were incubated for 2 additional hours in the presence of IPTG at a final concentration of 1mM. The bacterial cells were harvested by centrifugation.

The recombinant fusion protein produced by AD494(DE3)/rpET32 upon IPTG induction for 2h was partially obtained as insoluble inclusion bodies which were purified from endogenous *E. coli* proteins by the isolation of insoluble aggregates (Gerlach, G.F. et al 1992, Infect. Immun. 60:892). Induced cells from a 500 mL culture were resuspended in 20 mL of 25% sucrose-50mM Tris-HCl buffer (pH8.0) and frozen at -70°C. Lysis of cells in thawed suspension was achieved by the addition of 5mL of a solution of lysozyme (10mg/mL) in 250mM Tris-HCl buffer (pH8.0) followed by an incubation of 10 to 15 min on ice, and the addition of 150mL of detergent mix (5 parts of 20mM Tris-HCl buffer [pH7.4]-300mM NaCl-2% deoxycholic acid-2% Nonidet P-40 and 4 parts of 100mM Tris-HCl buffer [pH8]-50mM EDTA-2% Triton X*-100) followed by 5 min incubation on ice. Upon sonication, protein aggregates were harvested by centrifugation for 30 min at 35,000 X g and a sample of the soluble cellular fraction was kept. The aggregated proteins were solubilized in 6M guanidine hydrochloride. The

*Trade-mark

presence of the fusion protein in both the soluble and insoluble fractions was shown by Western Blot analysis using the serum of a mouse injected with formaldehyde killed cells of GBS strain C388/90 (Ia/c) that survived a bacterial challenge with the corresponding GBS strain.

The purification of the fusion protein from the soluble fraction of IPTG-induced AD494(DE3)/rpET was done by affinity chromatography based on the properties of the His.Tag sequence (6 consecutive histidine residues) to bind to divalent cations (Ni^{2+}) immobilized on the His.Bind metal chelation resin (Novagen, Madison, WI). The purification method used are those described in the pET system Manual, 6th Edition (Novagen, Madison, WI). Briefly, the pelleted cells obtained from a 100mL culture induced with IPTG was resuspended in 4mL of Binding buffer (5mM imidazole-500mM NaCl-20mM Tris-HCl pH7.9), sonicated, and spun at 39,000 X g for 20 min to remove debris. The supernatant was filtered (0.45µm pore size membrane) and deposited on a column of His.Bind resin equilibrated in Binding buffer. The column was then washed with 10 column volumes of Binding buffer followed by 6 column volumes of Wash buffer (20mM imidazole-500mM NaCl-20mM Tris-HCl pH7.9). The thioredoxin-His.Tag-GBS fusion protein was eluted with Elute buffer (1M imidazole-500mM NaCl-20mM Tris-HCl pH7.9). The removal of the salt and imidazole from the sample was done by dialysis against 3 X 1 liter PBS at 4°C.

The quantities of fusion protein obtained from either the soluble or insoluble cytoplasmic fractions of *E. coli* were estimated by Coomassie staining of a sodium dodecyl sulfate (SDS)-polyacrylamide gel with serial dilutions of these proteins and a bovine serum albumin standard (Pierce Chemical Co. Rockford, Ill.).

EXAMPLE 5 Recombinant production of GBS protein under control of lambda P_L promoter

The DNA coding region of a GBS protein was inserted downstream of the promoter λP_L into the translation vector pURV22. This plasmid was derived from p629 (George et al, 1987, Bio/Technology 5:600) from which the coding region for a portion of the herpes simplex virus type I (HSV-I) glycoprotein (gD-1) was removed and the ampicillin resistance gene replaced by a kanamycin cassette obtained from the plasmid vector pUC4K (Pharmacia Biotech Canada Inc., Baie D'Urfe, Canada). The vector contained a cassette of the bacteriophage λ cI857 temperature sensitive repressor gene from which the functional P_R promoter had been deleted. The inactivation of the cI857 repressor by temperature increase from the ranges of 30-37°C to 37-42°C resulted in the induction of the gene under the control of λ P_L. The translation of the gene was controlled by the ribosome binding site cro followed downstream by a BglII restriction site (AGATCT) and the ATG: ACTAAGGAGGTTAGATCTATG.

Restriction enzymes and T4 DNA ligase were used according to suppliers (Pharmacia Biotech Canada Inc., Baie D'Urfe, Canada; and New England Biolabs Ltd., Mississauga, Canada). Agarose gel electrophoresis of DNA fragments was performed as described by Sambrook et al. (Molecular cloning : A laboratory Manual, 1989, Cold Spring Harbor Laboratory Press, N.Y). Chromosomal DNA of the GBS bacteria was prepared according to procedures described in Jayarao et al (J. Clin. Microbiol., 1991, 29:2774). DNA amplification reactions by polymerase chain reaction (PCR) were made using DNA Thermal Cycler GeneAmp PCR system 2400 (Perkin Elmer, San Jose, CA). Plasmids used for DNA sequencing were purified using plasmid kits from Qiagen (Chatsworth, CA). DNA fragments were purified from agarose gels using Qiaex II

gel extraction kits from Qiagen (Chatsworth, CA). Plasmid transformations were carried out by the method described by Hanahan (DNA Cloning, Glover (ed.) pp, 109-135, 1985). The sequencing of genomic DNA inserts in plasmids was done using synthetic oligonucleotides which were synthesized by oligonucleotide synthesizer model 394 (the Perkin-Elmer Corp., Applied Biosystems Div. (ABI), Foster City, CA). The sequencing reactions were carried out by PCR using the Taq Dye Deoxy Terminator Cycle Sequencing kit (ABI, Foster City, CA) and DNA electrophoresis was performed on automated DNA sequencer 373A (ABI, Foster City, CA). The assembly of the DNA sequence was performed using the program Sequencer 3.0 (Gene Codes Corporation, Ann Arbor, MI). Analysis of the DNA sequences and their predicted polypeptides was performed with the program Gene Works version 2.45 (Intelligenetics, Inc., Mountain View CA).

The coding region of the GBS gene was amplified by PCR from GBS strain C388/90 (Ia/c) genomic DNA using oligos that contained base extensions for the addition of restriction sites BglII (AGATCT) and XbaI(TCTAGA), respectively. The PCR product was purified from agarose gel using a Qiaex II gel extraction kit from Qiagen (Chatsworth, CA), digested with the restriction enzymes BglII and XbaI, and extracted with phenol:chloroform before ethanol precipitation. The pURV22 vector was digested with the restriction enzymes BglII and XbaI, extracted with phenol:chloroform, and ethanol precipitated. The BglII-XbaI genomic DNA fragment was ligated to the BglII-XbaI pURV22 vector in which the GBS gene was under the control of the λPL promoter. The ligated products were transformed into *E. coli* strain XLI Blue MRF' (Δ (*mcr*A)183Δ(*mcr*CB-*hsd*SMR-mrr)173 *end*A1 *sup*E44 *thi*-1 recA1 *gyr*A96 *rel*A1 lac[F' *pro*AB *lac*I^{q}ZΔM15 Tn10(Tet^r)]^c) (Stratagene, La Jolla CA) according to the methods described in Hanahan, supra. Transformants harboring plasmids with the

insert were identified by analysis of lysed cells submitted to electrophoresis on agarose gel (Sambrook et al, supra). The recombinant pURV22 plasmid was purified using a Qiagen kit (Qiagen, Chatsworth, CA) and the nucleotide sequence of the DNA insert was verified by DNA sequencing.

The transformant XLI Blue MRF'/rpURV22 was grown at 34°C with agitation at 250 rpm in LB broth containing 50μg of kanamycin per mL until the A_{600} reached a value of 0.6. In order to induce the production of the fusion protein, the cells were incubated for 4 additional hours at 39°C. The bacterial cells were harvested by centrifugation , resuspended in sample buffer, boiled for 10 min and kept at -20°C.

EXAMPLE 6 Subcloning GBS protein gene in CMV plasmid pCMV-GH

The DNA coding region of a GBS protein was inserted in phase downstream of the human growth hormone (hGH) gene which was under the transcriptional control of the cytomegalovirus (CMV) promoter in the plasmid vector pCMV-GH (Tang et al, Nature, 1992, 356:152). The CMV promoter is non functional in E. coli cells but active upon administration of the plasmid in eukaryotic cells. The vector also incorporated the ampicillin resistance gene.

The coding region of the gene was amplified by PCR from genomic DNA of GBS strain C388/90 (Ia/c) using the oligos that contained base extensions for the addition of the restriction sites BglII (AGATCT) and HindIII (AAGCTT). The PCR product was purified from agarose gel using a Qiaex II gel extraction kit from Qiagen (Chatsworth, CA), digested with the restriction enzymes BglII and HindIII, and extracted with phenol:chloroform before ethanol precipitation. The pCMV-GH vector (Laboratory of Dr. Stephen

A. Johnston, Department of Biochemistry, The University of Texas, Dallas, Texas) containing the human growth hormone to create fusion proteins was digested with the restriction enzymes BamHI and HindIII, extracted with phenol:chloroform, and ethanol precipitated. The 1.3-kb BglII-HindIII genomic DNA fragment was ligated to the BamHI -HindIII pCMV-GH vector to create the hGH-GBS fusion protein under the control of the CMV promoter. The ligated products were transformed into *E. coli* strain DH5α[φ80 *lacZ* ΔM15 *endA1* *recA1* *hsdR17* ($^{r}K^{-m}K^{+}$) *supE44* *thi-1*λ^{-} *gyrA96* *relA1* Δ(*lacZYA-argF*)U169] (Gibco BRL, Gaithersburg, MD) according to the methods described by Hanahan, supra. Transformants harboring plasmids with the insert were identified by analysis of lysed cells submitted to electrophoresis on agarose gel (Sambrook, J. et al , supra). The recombinant pCMV plasmid was purified using a Qiagen kit (Qiagen, Chatsworth, CA) and the nucleotide sequence of the DNA insert was verified by DNA sequencing.

EXAMPLE 7 Immunological activity of GBS protein to GBS challenge

Four groups of 12 female CD-1 mice (Charles River, St-Constant, Quebec, Canada) of 6 to 8 weeks were injected subcutaneously three times at three week intervals with 0.1mL of the following antigenic preparations: formaldehyde killed cells of GBS strain C388/90 (~$6X10^{7}$ cfu), 20μg of thioredoxin-His.Tag-GBS fusion protein obtained from the insoluble (inclusion bodies) or 20μg of the fusion protein, affinity purified (nickel column), from the soluble cytoplasmic fraction in *E.coli*, or 20μg of affinity purified (nickel column) thioredoxin-His.Tag control polypeptide. 20μg of QuilA™ (Cedarlane Laboratories Ltd, Hornby, Canada)

was added to each antigenic preparation as the adjuvant. Serum samples were obtained from each mouse before immunization (PB) and on days 20 (TB1), 41 (TB2) and 54 (TB3) during the immunization protocols. Sera were frozen at -20°C.

An increase of the ELISA titers was recorded after each injection of the fusion protein indicating a good primary response and a boost of the specific humoral immune response after each of the second and third administration. At the end of the immunization period, the means of reciprocal ELISA titers was 456,145 for the group immunized with 20μg of fusion protein obtained from inclusion bodies compared to 290,133 for the group of mice immunized with the protein from soluble fraction in *E.coli*. The latter result suggests that the protein obtained from inclusion bodies could be more immunogenic than the soluble protein. Analysis of mice sera in ELISA using the affinity purified thioredoxin-His.Tag to coat plates showed that negligible antibody titers are made against the thioredoxin-His.Tag portion of the fusion protein. The reactivity of the sera from mice injected with the recombinant fusion protein was also tested by ELISA against formaldehyde killed whole cells of GBS strain C388/90. The antibodies induced by immunization with recombinant fusion protein also recognized their specific epitopes on GBS cells indicating that their conformation is close enough to the native streptococcal protein to induce cross-reactive antibodies.

To verify whether the immune response induced by immunization could protect against GBS infection, mice were challenged with $3.5X10^5$ cfu of GBS strains C338/90(Ia/c) and $1.2X10^5$ cfu of strain NCS246(II/R) the results of which are illustrated in tables 3 and 4 respectively. Mice immunized with control thioredoxin-His.Tag peptide were not protected against challenge with either GBS strain while those

immunized with formaldehyde killed C388/90 whole cells only provided protection against homologous challenge. The thioredoxin-His.Tag-GBS fusion protein of the invention protected mice from challenge with both GBS strains. Blood and spleen culture of these mice did not reveal the presence of any GBS.

Table 3 Survival from GBS strain C388/90 (Ia/c) challenge[1]

immunizing agent	no. mice surviving challenge	% survival
thioredoxin-His.Tag[2]	1 / 6	17
formaldehyde killed C388/90 cells[3]	6 / 6	100
thioredoxin-His.Tag-GBS fusion (inclusion body preparation)[4]	6 / 6	100
thioredoxin-His.Tag-GBS fusion (cytoplasmic fraction)[4]	6 / 6	100

[1] intraperitoneal administration with 1 ml Todd-Hewitt culture medium adjusted to $3.5X10^5$ cfu;
[2] 20μg administered; posterior legs paralyzed in surviving mouse; GBS detected in blood and spleen;
[3] $6X10^7$ cfu administered;
[4] 20μg administered.

Table 4 Survival from GBS strain NCS246 (II/R) challenge[1]

immunizing agent	no. mice surviving challenge	% survival
thioredoxin-His.Tag[2]	0 / 6	0
formaldehyde killed C388/90 cells[3]	2 / 6	34
thioredoxin-His.Tag-GBS fusion (inclusion body preparation)[2]	5 / 5[4]	100
thioredoxin-His.Tag-GBS fusion (cytoplasmic fraction)[2]	6 / 6	100

[1] intraperitoneal administration with 1 ml Todd-Hewitt culture medium containing GBS NCS246(II/R) suspension adjusted to $1.2X10^5$ cfu.
[2] 20µg administered;
[3] $6X10^7$ cfu administered;
[4] one mouse died during immunization.

EXAMPLE 8 Immunization with recombinant GBS protein confers protection against experimental GBS infection

This example illustrates the protection of mice against fatal GBS infection by immunization with the recombinant protein corresponding to the SEQ ID NO:39.

Groups of 10 female CD-1 mice (Charles River) were immunized subcutaneously three times at three-week intervals with 20 µg of recombinant protein purified from E. coli strain BLR (Novagen) harboring the recombinant pURV22 plasmid vector containing the GBS gene corresponding to SEQ ID NO:42 in presence of 20 µg of QuilA™ adjuvant (Cedarlane Laboratories Ltd, Hornby, Canada) or, as control, with

QuilA™ adjuvant alone in PBS. Blood samples were collected from the orbital sinus on day 1, 22 and 43 prior to each immunization and fourteen days (day 57) following the third injection. One week later the mice were challenged with approximately 10^4 to 10^6 CFU of various virulent GBS strains. Samples of the GBS challenge inoculum were plated on TSA/5% sheep blood agar plates to determine the CFU and to verify the challenge dose. Deaths were recorded for a period of 14 days and on day 14 post-challenge, the surviving mice were sacrificed and blood and spleen were tested for the presence of GBS organisms. The survival data are shown in table 5.

Prechallenge sera were analyzed for the presence of antibodies reactive with GBS by standard immunoassays. Elisa and immunoblot analyses indicated that immunization with recombinant GBS protein produced in *E. coli* elicited antibodies reactive with both, recombinant and native GBS protein. Antibody responses to GBS are described in Example 9.

Table 5. Ability of recombinant GBS protein corresponding to SEQ ID NO: 39 to elicit protection against 8 diverse GBS challenge strains

	Challenge strain			
Immunogen	Designation	Type	No. alive:	No. dead [1]
rGBS protein none	C388/90	Ia/c	8 : 2 0 : 10	(P<0.0001)
rGBS protein none	NCS 246	II/R	10 : 0 3 : 7	(P=0.0012)
rGBS protein none	ATCC12401	Ib	10 : 0 3 : 7	(P=0.001)
rGBS protein none	NCS 535	V	10 : 0 5 : 5	(P=0.01)
rGBS protein none	NCS 9842	VI	10 : 0 0 : 10	(P<0.0001)
rGBS protein NCS 915-F[3] none	NCS 915	III	7 : 3 1 : 9 4 : 6	(P=0.0007)[2]
rGBS protein NCS 954-F none	NCS 954	III/R	7 : 3 4 : 6 1 : 9	(P=0.002)
rGBS protein COH1-F none	COH1	III	4 : 6 3 : 7 0 : 10	(P=0.0004)

[1] Groups of 10 mice per group were used, the number of mice surviving to infection and the number of dead mice are indicated. The survival curves corresponding to recombinant GBS protein-immunized animals were compared to the survival curves corresponding to mock-immunized animals using the log-rank test for nonparametric analysis.

[2] Comparison analysis to NCS915-F-immunized animals.

[3] Animals were immunized with formaldehyde-killed GBS in presence of QuilA™ adjuvant.

All hemocultures from surviving mice were negative at day 14 post-challenge. Spleen cultures from surviving mice were negative except for few mice from experiment MB-11.

EXAMPLE 9 Vaccination with the recombinant GBS protein elicits an immune response to GBS

Groups of 10 female CD-1 mice were immunized subcutaneously with recombinant GBS protein corresponding to SEQ ID NO:39 as described in Example 8. In order to assess the antibody response to native GBS protein, sera from blood samples collected prior each immunization and fourteen days after the third immunization were tested for antibody reactive with GBS cells by ELISA using plates coated with formaldehyde-killed GBS cells from type III strain NCS 954, type Ib strain ATCC12401, type V strain NCS 535 or type VI strain NCS 9842. The specificity of the raised antibodies for GBS protein was confirmed by Western blot analyses to GBS cell extracts and purified recombinant antigens. The results shown in Figure 10 clearly demonstrate that animals respond strongly to recombinant GBS protein used as immunogens with median reciprocal antibody titers varying between 12000 and 128000, for sera collected after the third immunization, depending of the coating antigen. All preimmune sera were negative when tested at a dilution of 1 :100. GBS-reactive antibodies were detectable in the sera of each animal after a single injection of recombinant GBS protein.

Example 10 Antigenic conservation of the GBS protein of the present invention

Monoclonal antibodies (MAbs) specific to the GBS protein of the present invention were used to demonstrate that this surface antigen is produced by all GBS and that it is also antigenically highly conserved.

A collection of 68 GBS isolates was used to evaluate the reactivity of the GBS-specific MAbs. These strains were obtained from the National Center for Streptococcus, Provincial Laboratory of Public Health for Northern Alberta, Canada; Centre Hospitalier Universitaire de Quebec, Pavillon CHUL, Quebec, Canada; American Type Culture Collection, USA; Laboratoire de Sante Publique du Quebec, Canada; and Dept. of Infectious Disease, Children's Hospital and Medical Center, Seattle, USA. All eight Mabs were tested against the following panel of strains: 6 isolates of serotype Ia or Ia/c, 3 isolates of serotype Ib, 4 isolates of serotype II, 14 isolates of serotype III, 2 isolates of serotype IV, 2 isolates of serotype V, 2 isolates of serotype VI, 2 isolates of serotype VII, 1 isolate of serotype VIII, 10 isolates that were not serotyped and 3 bovine *S. agalactiae* strains. MAb 3A2 was also reacted with additional GBS: 9 isolates of serotype Ia/c and 10 isolates of serotype V. The strains were grown overnight on blood agar plates at 37°C in an atmosphere of 5% CO_2. Cultures were stored at -70°C in heart infusion broth with 20% (v/v) glycerol.

To obtain the GBS protein-specific MAbs, mice were immunized three times at three-week intervals with 20 μg of purified recombinant GBS protein (SEQ ID NO :44) in the presence of 20% QuilA™ adjuvant. Hybridoma cell lines were generated by fusion of spleen cells recovered from immunized mice with the nonsecreting SP2/O myeloma cell line as described

previously (Hamel, J. et al. 1987. J. Med. Microbiol. 23:163-170). Hybrid clone supernatants were tested for specific antibody production by ELISA using formaldehyde inactivated GBS and purified recombinant GBS protein (SEQ ID NO :39 or 44) as coating antigen, as previously described (Hamel, J. et al. 1987. J. Med. Microbiol. 23:163-170). Specific hybrid were cloned by limiting dilutions, expanded, and frozen in liquid nitrogen. Production of recombinant GBS protein was presented in Examples 4 & 5. Purified recombinant GBS protein or formaldehyde inactivated GBS were resolved by electrophoresis by using the discontinuous buffer system of Laemmli as recommended by the manufacturer and then transfer onto nitrocellulose membrane for Western immunoblotting as described previously (Martin et al. 1992. Infect. Immun. 60:2718-2725).

Western immunoblotting experiments clearly indicated that all eight MAbs recognized a protein band that corresponded to the purified recombinant GBS protein (SEQ ID NO :39). These MAbs also reacted with a protein band present in every GBS isolates tested so far. The reactivity of these GBS-specific MAbs are presented in Table 6. Each MAb reacted well with all 46 GBS. In addition, these MAbs also recognized the 3 *S. agalactiae* strains of bovine origin that were tested. MAb 3A2 also recognized nineteen GBS; 9 isolates of serotype Ia/c and 10 of serotype V. The other MAbs were not tested against these additional strains.

These results demonstrated that the GBS protein (SEQ ID NO :39) was produced by all the 65 GBS and the three 3 *S. agalactiae* strains of bovine origin that were tested so far. More importantly, these results clearly demonstrated that the epitopes recognized by these eight GBS-specific MAbs were widely distributed and conserved among GBS. These results also indicated that these epitopes were not

restricted to serologically related isolates since representatives of all known GBS serotypes including the major disease causing groups were tested.

In conclusion, the data presented in this example clearly demonstrated that the GBS protein of the present invention is produced by all GBS and that it is antigenically highly conserved.

CA 02321106 2000-08-18

Table 6. Reactivity of eight GBS protein-specific MAbs with different *S. agalactiae* strains as evaluated by Western immunoblots.

Mabs	Number of each serotype of *s. agalactiae* strains recognized by the MAbs.											
	Ia or Ia/c (6)	Ib (3)	II (4)	III (4)	IV (2)	V (2)	VI (2)	VII (2)	VIII (1)	NT(10)[2]	TOTAL (26)	Bovine (3)
3A2[1]	6	3	4	4	2	2	2	2	1	10	46	3
5A12	6	3	4	4	2	2	2	2	1	10	46	3
6G11	6	3	4	4	2	2	2	2	1	10	46	2
8B9	6	3	4	4	2	2	2	2	1	10	46	3
8E11	6	3	4	4	2	2	2	2	1	10	46	3
12B12	6	3	4	4	2	2	2	2	1	10	46	3
18F11	6	3	4	4	2	2	2	2	1	10	46	3
20G2	6	3	4	4	2	2	2	2	1	10	46	3

1 Nine additional strains of serotype Ia/c and 10 strains of serotype V were recognized by MAb 3A2.
2 These strains were not serotyped

CLAIMS:

1. An isolated polynucleotide encoding a polypeptide comprising an amino acid sequence greater than 95% identical to the amino acid sequence set forth in SEQ ID NO:39 or SEQ ID NO:44, wherein the polypeptide induces an immune response against group B streptococcus, and wherein the polypeptide elicits antibodies that specifically bind to a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:44.

2. The isolated polynucleotide according to claim 1, wherein the encoded polypeptide comprises the amino acid sequence set forth in either SEQ ID NO:39 or SEQ ID NO:44.

3. An isolated polynucleotide that is complementary to the polynucleotide of claim 1 or 2.

4. The polynucleotide of claim 1 or 2, wherein said polynucleotide is DNA.

5. The polynucleotide of claim 1 or 2, wherein said polynucleotide is RNA.

6. The polynucleotide according to claim 1, wherein the polynucleotide comprises a nucleotide sequence at least 95% identical to the nucleotide sequence set forth in SEQ ID NO:42 or SEQ ID NO:43.

7. The polynucleotide according to claim 1, wherein the polynucleotide comprises the nucleotide sequence set forth in SEQ ID NO:42 or SEQ ID NO:43.

8. A vector comprising the polynucleotide of any one of claims 1, 2, and 4-6, wherein said polynucleotide is operably linked to an expression control region.

9. A host cell transfected with the vector of claim 8.

10. The host cell of claim 9, wherein the host cell is a bacterial cell.

11. A process for producing a polypeptide encoded by the polynucleotide of any one of claims 1, 2, and 4-6, said process comprising culturing the host cell according to either claim 9 or 10 under conditions suitable for expression of said polypeptide.

12. An isolated polypeptide comprising an amino acid sequence greater than 95% identical to the amino acid sequence selected from SEQ ID NO:39 or SEQ ID NO:44, wherein the polypeptide induces an immune response to group B streptococcus, and wherein the polypeptide elicits antibodies that specifically bind to a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:44.

13. The isolated polypeptide of claim 12, comprising the amino acid sequence set forth in SEQ ID NO:39.

14. The isolated polypeptide of claim 12, comprising the amino acid sequence set forth in SEQ ID NO:44.

15. An isolated polypeptide comprising an antigenic fragment of the polypeptide of claim 14, wherein the antigenic fragment is capable of generating antibodies having binding specificity for a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:44.

16. A vaccine composition comprising the isolated polypeptide according to any one of claims 12-14 and a pharmaceutically acceptable carrier or diluent, wherein the isolated polypeptide induces a therapeutic or prophylactic immune response to group B streptococcus.

17. The vaccine composition according to claim 16, further comprising a pharmaceutically acceptable adjuvant.

18. A composition comprising (a) the isolated polypeptide according to claim 15 and (b) a pharmaceutically acceptable carrier or diluent.

19. The composition according to claim 18, further comprising a pharmaceutically acceptable adjuvant.

20. Use of the vaccine composition of either claim 16 or 17, for therapeutic or prophylactic treatment of a group B streptococcus infection in an animal.

21. Use of the isolated polypeptide of any one of claims 12-14, for the manufacture of a vaccine composition for therapeutic or prophylactic treatment of a group B streptococcus infection in an animal, wherein the isolated polypeptide induces a therapeutic or prophylactic immune response to group B streptococcus.

22. Use of the isolated polypeptide of claim 15, for the manufacture of a composition for generating antibodies having binding specificity for a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:44.

23. Use according to 20 or 21, wherein said animal is a bovine.

24. Use according to claim 20 or 21, wherein said animal is a human.

25. An isolated antibody, or an antigen-binding fragment thereof, that specifically binds to a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:44.

26. The isolated antibody, or antigen-binding fragment thereof, of claim 25, wherein the antibody is a monoclonal antibody.

27. A method of detecting group B streptococcus in a biological sample, said method comprising:

(a) incubating the antibody, or antigen-binding fragment thereof, of claim 25 with the biological sample to form a mixture; and

(b) detecting bound antibody, or antigen-binding fragment thereof, in the mixture, indicating the presence of group B streptococcus in the biological sample.

28. A method of detecting group B streptococcus in a biological sample, said method comprising:

(a) incubating the polynucleotide according to any one of claims 1-7 and the biological sample to form a mixture; and

(b) detecting specifically bound polynucleotide in the mixture, indicating the presence of group B streptococcus.

29. A method of detecting in a biological sample an antibody that specifically binds to the polypeptide of any one of claims 12-15, said method comprising:

(a) incubating the biological sample and one or more of the polypeptides of any one of claims 12-15 to form a mixture; and

(b) detecting specifically bound polypeptide in the mixture, indicating the presence of the antibody in the biological sample.

```
TATCTGGCAA AGAGCCAGCT AATCGTTTTA GTTGGGCTAA AAATAAATTA TTAATCAATG     60
 S  G  K   E  P  A   N  R  F  S   W  A  K   N  K  L   L  I  N  G
 ---->
GATTCATTGC AACTCTAGCA GCAACTATCT TATTTTTTGC AGTTCAATTC ATAGGTCTTA    120
 F  I  A   T  L  A   A  T  I  L   F  F  A   V  Q  F   I  G  L  K

AACCAGATTA CCCTGGAAAA ACCTACTTTA TTATCCTATT GACAGCATGG ACTTTGATGG    180
 P  D  Y   P  G  K   T  Y  F  I   I  L  L   T  A  W   T  L  M  A

CATTAGTAAC TGCTTTAGTG GGATGGGATA ATAGGTATGG TTCCTTCTTG TCGTTATTAA    240
 L  V  T   A  L  V   G  W  D  N   R  Y  G   S  F  L   S  L  L  I

TATTATTATT CCAGCTTGGT TCAAGCGCAG GAACTTACCC AATAGAATTG AGTCCTAAGT    300
 L  L  F   Q  L  G   S  S  A  G   T  Y  P   I  E  L   S  P  K  F

TCTTTCAAAC AATTCAACCA TTTTTACCGA TGACTTACTC TGTTTCAGGA TTAAGAGAGA    360
 F  Q  T   I  Q  P   F  L  P  M   T  Y  S   V  S  G   L  R  E  T

CCATCTCGTT GACGGGAGAC GTTAACCATC AATGGAGAAT GCTAGTAATC TTTTTAGTAT    420
 I  S  L   T  G  D   V  N  H  Q   W  R  M   L  V  I   F  L  V  S

CATCGATGAT ACTTGCTCTT CTTATTTATC GTAAACAAGA AGATTAATAG AAAGTATCTA    480
 S  M  I   L  A  L   L  I  Y  R   K  Q  E   D

GTGATAGACT AACAGTATGA TATGGTATGT CAAAGTATTT AGGAGGAGAA GATATGTCTA    540
                                                         M  S  T
                                                         |---->
CTTTAACAAT AATTATTGCA ACATTAACTG CTTTGGAACA TTTTTATATT ATGTATTTGG    600
 L  T  I   I  I  A   T  L  T  A   L  E  H   F  Y  I   M  Y  L  E

AGACGTTAGC CACCCAGTCA AATATGACTG GGAAGATTTT TAGTATGTCT AAAGAAGAGT    660
 T  L  A   T  Q  S   N  M  T  G   K  I  F   S  M  S   K  E  E  L

TGTCATATTT ACCCGTTATT AAACTTTTTA AGAATCAAGG TGTATACAAC GGCTTGATTG    720
 S  Y  L   P  V  I   K  L  F  K   N  Q  G   V  Y  N   G  L  I  G

GCCTATTCCT CCTTTATGGG TTATATATTT CACAGAATCA AGAAATTGTA GCTGTTTTTT    780
 L  F  L   L  Y  G   L  Y  I  S   Q  N  Q   E  I  V   A  V  F  L

TAATCAATGT ATTGCTAGTT GCTATTTATG GTGCTTTGAC AGTTGATAAA AAAATCTTAT    840
 I  N  V   L  L  V   A  I  Y  G   A  L  T   V  D  K   K  I  L  L

TAAAACAGGG TGGTTTACCT ATATTAGCTC TTTTAACATT CTTATTTTAA TACTACTTAG    900
 K  Q  G   G  L  P   I  L  A  L   L  T  F   L  F

CCGTTCGATT TAGTTGAACG GCTTTTAGTA ATCATTTTTT TCTCATAATA CAGGTAGTTT    960

AAGTAATTTG TCTTTAAAAA TAGTATAATA TAACTACGAA TTCAAAGAGA GGTGACTTTG   1020

ATTATGACTG AGAACTGGTT ACATACTAAA GATGGTTCAG ATATTTATTA TCGTGTCGTT   1080
  M  T  E   N  W  L   H  T  K   D  G  S  D   I  Y  Y   R  V  V
  |---->
GGTCAAGGTC AACCGATTGT TTTTTTACAT GGCAATAGCT TAAGTAGTCG CTATTTTGAT   1140
G  Q  G  Q   P  I  V   F  L  H   G  N  S  L   S  S  R   Y  F  D

AAGCAAATAG CATATTTTTC TAAGTATTAC CAAGTTATTG TTATGGATAG TAGAGGGCAT   1200
K  Q  I  A   Y  F  S   K  Y  Y   Q  V  I  V   M  D  S   R  G  H

GGCAAAAGTC ATGCAAAGCT AAATACCATT AGTTTCAGGC AAATAGCAGT TGACTTAAAG   1260
G  K  S  H   A  K  L   N  T  I   S  F  R  Q   I  A  V   D  L  K
```

```
GATATCTTAG TTCATTTAGA GATTGATAAA GTTATATTGG TAGGCCATAG CGATGGTGCC   1320
D  I  L  V   H  L  E   I  D  K   V  I  L  V   G  H  S   D  G  A

AATTTAGCTT TAGTTTTTCA AACGATGTTT CCAGGTATGG TTAGAGGGCT TTTGCTTAAT   1380
N  L  A  L   V  F  Q   T  M  F   P  G  M  V   R  G  L   L  L  N

TCAGGGAACC TGACTATTCA TGGTCAGCGA TGGTGGGATA TTCTTTTAGT AAGGATTGCC   1440
S  G  N  L   T  I  H   G  Q  R   W  W  D  I   L  L  V   R  I  A

TATAAATTCC TTCACTATTT AGGGAAACTC TTTCCGTATA TGAGGCAAAA AGCTCAAGTT   1500
Y  K  F  L   H  Y  L   G  K  L   F  P  Y  M   R  Q  K   A  Q  V

ATTTCGCTTA TGTTGGAGGA TTTGAAGATT AGTCCAGCTG ATTTACAGCA TGTGTCAACT   1560
I  S  L  M   L  E  D   L  K  I   S  P  A  D   L  Q  H   V  S  T

CCTGTAATGG TTTTGGTTGG AAATAAGGAC ATAATTAAGT TAAATCATTC TAAGAAACTT   1620
P  V  M  V   L  V  G   N  K  D   I  I  K  L   N  H  S   K  K  L

GCTTCTTATT TTCCAAGGGG GGAGTTTTAT TCTTTAGTTG GCTTTGGGCA TCACATTATT   1680
A  S  Y  F   P  R  G   E  F  Y   S  L  V  G   F  G  H   H  I  I

AAGCAAGATT CCCATGTTTT TAATATTATT GCAAAAAAGT TTATCAACGA TACGTTGAAA   1740
K  Q  D  S   H  V  F   N  I  I   A  K  K  F   I  N  D   T  L  K

GGAGAAATTG TTGAAAAAGC TAATTGAAAA AGTCAAATCA CTGACTTCTG TGATTAAAAT   1800
G  E  I  V   E  K  A   N

TGTATTTTTT ATATCTGTTT TAGTGCTTAT TATTGTTGAA ATGATTCATT TGAAACGAAC   1860
                                            M  I  H  L   K  R  T
                                            |---->
TATTTCTGTT GAGCAACTAA AGAGTGTTTT TGGGCAATTA TCTCCAATGA ATCTTTTCTT   1920
 I  S  V   E  Q  L  K   S  V  F   G  Q  L   S  P  M  N   L  F  L

AATTATCCTT GTGGGGGTTA TCGCTGTCTT ACCGACAACC GGATATGACT TTGTACTGAA   1980
 I  I  L   V  G  V  I   A  V  L   P  T  T   G  Y  D  F   V  L  N

TGGACTTTTA CGTACAGATA AAAGCAAAAG GTATATTTTA CAGACTAGTT GGTGTATCAA   2040
 G  L  L  R   T  D  K   S  K  R   Y  I  L  Q   T  S  W   C  I  N

CACTTTTAAT AACTTGTCAG GATTCGGTGG CTTAATCGAT ATTGGGTTGC GCATGGCTTT   2100
 T  F  N  N   L  S  G   F  G  G   L  I  D  I   G  L  R   M  A  F

TTATGGTAAA AAAGGTCAAG AGAAGAGTGA CCTAAGAGAA GTGACTCGTT TTTTACCCTA   2160
 Y  G  K  K   G  Q  E   K  S  D   L  R  E  V   T  R  F   L  P  Y

TCTTATTTCT GGTCTGTCAT TTATTAGTGT GATTGCCTTA ATCATGAGCC ATATTTTTCA   2220
 L  I  S  G   L  S  F   I  S  V   I  A  L  I   M  S  H   I  F  H

TGCCAAAGCT AGTGTTGATT ACTATTATTT GGTATTAATT GGTGCTAGTA TGTATTTTCC   2280
 A  K  A  S   V  D  Y   Y  Y  L   V  L  I  G   A  S  M   Y  F  P

TGTTATTTAT TGGATTTCTG GTCATAAAGG AAGCCATTAT TTCGGAGATA TGCCATCTAG   2340
 V  I  Y  W   I  S  G   H  K  G   S  H  Y  F   G  D  M   P  S  S

TACTCGTATA AAATTAGGTG TTGTTTCTTT TTTTGAATGG GGATGTGCGG CCGCAGCATT   2400
 T  R  I  K   L  G  V   V  S  F   F  E  W  G   C  A  A   A  A  F

TATAATTATC GGTTATTTAA TGGGCATTCA TCTACCAGTT TATAAAATTT TACCACTATT   2460
 I  I  I  G   Y  L  M   G  I  H   L  P  V  Y   K  I  L   P  L  F
```

```
TTGTATTGGT TGTGCCGTCG GGATTGTATC CCTTATTCCC GGTGGATTAG GAAGTTTTGA   2520
 C  I  G  C   A  V  G  I   V  S  L   I  P  G   G  L  G  S   F  E

ATTAGTTCTA TTTACAGGGT TTGCTGCCGA GGGACTACCT AAAGAAACTG TGGTTGCATG   2580
 L  V  L  F   T  G  F  A   A  E  G   L  P  K   E  T  V  V   A  W

GTTATTACTT TATCGTTTAG CCTACTATAT TATTCCATTC TTTGCAGGTA TCTATTTCTT   2640
 L  L  L  Y   R  L  A  Y   Y  I  I   P  F  F   A  G  I  Y   F  F

TATCCATTAT TTAGGTAGTC AAATAAATCA ACGTTATGAA AATGTCCCGA AAGAGTTAGT   2700
 I  H  Y  L   G  S  Q  I   N  Q  R   Y  E  N   V  P  K  E   L  V

ATCAACTGTT CTACAAACCA TGGTGAGCCA TTTGATGCGT ATTTTAGGTG CATTCTTAAT   2760
 S  T  V  L   Q  T  M  V   S  H  L   M  R  I   L  G  A  F   L  I
                                     |---->
ATTTTCAACA GCATTTTTTG AAAATATTAC TTATATTATG TGGTTGCAGA AGCTAGGCTT   2820
 F  S  T  A   F  F  E  N   I  T  Y   I  M  W   L  Q  K  L   G  L

GGACCCATTA CAAGAACAAA TGTTATGGCA GTTTCCAGGT TTATTGCTGG GGGTTTGTTT   2880
 D  P  L  Q   E  Q  M  L   W  Q  F   P  G  L   L  L  G  V   C  F

TATTCTCTTA GCTAGAACTA TTGATCAAAA AGTGAAAAAT GCTTTTCCAA TTGCTATTAT   2940
 I  L  L  A   R  T  I  D   Q  K  V   K  N  A   F  P  I  A   I  I

CTGGATTACT TTGACATTGT TTTATCTTAA TTTAGGTCAT ATTAGTTGGC GACTATCTTT   3000
 W  I  T  L   T  L  F  Y   L  N  L   G  H  I   S  W  R  L   S  F

CTGGTTTATT TTACTATTGT TAGGCTTATT AGTCATTAAG CCAACTCTCT ATAAAAAACA   3060
 W  F  I  L   L  L  L  G   L  L  V   I  K  P   T  L  Y  K   K  Q

ATTTATTTAT AGCTGGGAAG AGCGTATTAA GGATGGAATC ATTATCGTTA GTTTAATGGG   3120
 F  I  Y  S   W  E  E  R   I  K  D   G  I  I   I  V  S  L   M  G

AGTTCTATTT TATATTGCAG GACTACTATT CCCTATCAGG GCTCATATTA CAGGTGGTAG   3180
 V  L  F  Y   I  A  G  L   L  F  P   I  R  A   H  I  T  G   G  S

TATTGAACGC CTGCATTATA TCATAGCATG GGAGCCGATA GCATTGGCTA CGTTGATTCT   3240
 I  E  R  L   H  Y  I  I   A  W  E   P  I  A   L  A  T  L   I  L

TACTCTCGTT TATTTATGTT TGGTTAAGAT TTTACAAGGA AAATCTTGTC AGATTGGTGA   3300
 T  L  V  Y   L  C  L  V   K  I  L   Q  G  K   S  C  Q  I   G  D

TGTGTTCAAT GTGGATCGTT ATAAAAAACT ACTTCAAGCT TACGGTGGTT CTTCGGATAG   3360
 V  F  N  V   D  R  Y  K   K  L  L   Q  A  Y   G  G  S  S   D  S

CGGTTTAGCC TTTTTAAATG ATAAAAGGCT CTACTGGTAC CAAAAAAATG GAGAAGATTG   3420
 G  L  A  F   L  N  D  K   R  L  Y   W  Y  Q   K  N  G  E   D  C

CGTTGCGTTC CAATTTGTAA TTGTCAATAA TAAATGTCTT ATTATGGGGG AACCAGCCGG   3480
 V  A  F  Q   F  V  I  V   N  N  K   C  L  I   M  G  E  P   A  G

TGATGACACT TATATTCGTG AAGCTATTGA ATCGTTTATT GATGATGCTG ATAAGCTAGA   3540
 D  D  T  Y   I  R  E  A   I  E  S   F  I  D   D  A  D  K   L  D

CTATGACCTT GTTTTTTACA GTATTGGACA GAAGTTGACA CTACTTTTAC ATGAGTATGG   3600
 Y  D  L  V   F  Y  S  I   G  Q  K   L  T  L   L  L  H  E   Y  G

TTTTGACTTT ATGAAAGTTG GTGAGGATGC TTTAGTTAAT TTAGAAACGT TTACTCTTAA   3660
 F  D  F  M   K  V  G  E   D  A  L   V  N  L   E  T  F  T   L  K
```

```
AGGGAATAAG TACAAACCTT TCAGAAATGC CCTAAATAGA GTTGAAAAGG ATGGTTTCTA   3720
 G  N  K  Y  K  P  F  R  N  A  L  N  R  V  E  K  D  G  F  Y

TTTCGAAGTT GTACAATCGC CACATAGTCA AGAGCTACTA AATAGTTTGG AAGAGATTTC   3780
 F  E  V  V  Q  S  P  H  S  Q  E  L  L  N  S  L  E  E  I  S

TAATACTTGG TTAGAAGGAC GTCCTGAAAA AGGTTTCTCA CTAGGATATT TTAATAAAGA   3840
 N  T  W  L  E  G  R  P  E  K  G  F  S  L  G  Y  F  N  K  D

TTATTTCCAA CAAGCCCCAA TAGCTTTGGT AAAAAATGCT GAACACGAAG TTGTTGCTTT   3900
 Y  F  Q  Q  A  P  I  A  L  V  K  N  A  E  H  E  V  V  A  F

TGCTAATATT ATGCCAAACT ATGAAAAGAG TATTATCTCT ATTGATTTAA TGCGTCACGA   3960
 A  N  I  M  P  N  Y  E  K  S  I  I  S  I  D  L  M  R  H  D

TAAACAGAAA ATTCCGAATG GCGTTATGGA TTTCCTCTTT TTATCATTAT TCTCTTATTA   4020
 K  Q  K  I  P  N  G  V  M  D  F  L  F  L  S  L  F  S  Y  Y

TCAAGAGAAG GGATACCACT ATTTTGATTT GGGGATGGCA CCTTTATCAG GAGTTGGTCG   4080
 Q  E  K  G  Y  H  Y  F  D  L  G  M  A  P  L  S  G  V  G  R

CGTTGAAACA AGTTTTGCTA AAGAGAGAAT GGCGTATCTT GTCTATCATT TCGGTAGTCA   4140
 V  E  T  S  F  A  K  E  R  M  A  Y  L  V  Y  H  F  G  S  H

TTTCTACTCA TTTAATGGTT TACACAAGTA TAAGAAGAAG TTTACACCAT TGTGGTCGGA   4200
 F  Y  S  F  N  G  L  H  K  Y  K  K  K  F  T  P  L  W  S  E

ACGTTATATT TCTTGTTCTC GTTCGTCCTG GTTAATTTGT GCTATTTGTG CCCTATTAAT   4260
 R  Y  I  S  C  S  R  S  S  W  L  I  C  A  I  C  A  L  L  M

GGAAGATAGT AAAATTAAGA TTGTTAAATA AGCTTTATTT GGCAATTAAA AAGAGCATGT   4320
 E  D  S  K  I  K  I  V  K

CATGCGACAT GCTCTTTTTA AATCATTTAA TACCATTGAT TGCTTGAATC TACTTTATAA   4380

TATGATGTGC TTTTAAATAT TGTTTAGCTA CTGTAGCTGC TGATTTATGC TTTACAGCTA   4440

CTTGGTAGTT CATTTCTTGC ATTTCTTTTT CAGTGATATG ACCAGCAAGT TTATTGAGAG   4500

CTTTTTTTAC TTGA  (SEQ ID NO:1)                                      4514
```

FIG. 1a
[clone1-dna/aa]

```
SGKEPANRFS WAKNKLLING FIATLAATIL FFAVQFIGLK PDYPGKTYFI   50
ILLTAWTLMA LVTALVGWDN RYGSFLSLLI LLFQLGSSAG TYPIELSPKF  100
FQTIQPFLPM TYSVSGLRET ISLTGDVNHQ WRMLVIFLVS SMILALLIYR  150
KQED  (SEQ ID NO:2)                                     154
```

FIG. 1b

```
MSTLTIIIAT LTALEHFYIM YLETLATQSN MTGKIFSMSK EELSYLPVIK   50
LFKNQGVYNG LIGLFLLYGL YISQNQEIVA VFLINVLLVA IYGALTVDKK  100
ILLKQGGLPI LALLTFLF  (SEQ ID NO:3)                      118
```

FIG. 1c

```
MTENWLHTKD GSDIYYRVVG QGQPIVFLHG NSLSSRYFDK QIAYFSKYYQ   50
VIVMDSRGHG KSHAKLNTIS FRQIAVDLKD ILVHLEIDKV ILVGHSDGAN  100
LALVFQTMFP GMVRGLLLNS GNLTIHGQRW WDILLVRIAY KFLHYLGKLF  150
PYMRQKAQVI SLMLEDLKIS PADLQHVSTP VMVLVGNKDI IKLNHSKKLA  200
SYFPRGEFYS LVGFGHHIIK QDSHVFNIIA KKFINDTLKG EIVEKAN     247
(SEQ ID NO:4)
```

FIG. 1d

```
MIHLKRTISV EQLKSVFGQL SPMNLFLIIL VGVIAVLPTT GYDFVLNGLL   50
RTDKSKRYIL QTSWCINTFN NLSGFGGLID IGLRMAFYGK KGQEKSDLRE  100
VTRFLPYLIS GLSFISVIAL IMSHIFHAKA SVDYYYLVLI GASMYFPVIY  150
WISGHKGSHY FGDMPSSTRI KLGVVSFFEW GCAAAAFIII GYLMGIHLPV  200
YKILPLFCIG CAVGIVSLIP GGLGSFELVL FTGFAAEGLP KETVVAWLLL  250
YRLAYYIIPF FAGIYFFIHY LGSQINQRYE NVPKELVSTV LQTMVSHLMR  300
ILGAFLIFST AFFENITYIM WLQKLGLDPL QEQMLWQFPG LLLGVCFILL  350
ARTIDQKVKN AFPIAIIWIT LTLFYLNLGH ISWRLSFWFI LLLLGLLVIK  400
PTLYKKQFIY SWEERIKDGI IIVSLMGVLF YIAGLLFPIR AHITGGSIER  450
LHYIIAWEPI ALATLILTLV YLCLVKILQG KSCQIGDVFN VDRYKKLLQA  500
YGGSSDSGLA FLNDKRLYWY QKNGEDCVAF QFVIVNNKCL IMGEPAGDDT  550
YIREAIESFI DDADKLDYDL VFYSIGQKLT LLLHEYGFDF MKVGEDALVN  600
LETFTLKGNK YKPFRNALNR VEKDGFYFEV VQSPHSQELL NSLEEISNTW  650
LEGRPEKGFS LGYFNKDYFQ QAPIALVKNA EHEVVAFANI MPNYEKSIIS  700
IDLMRHDKQK IPNGVMDFLF LSLFSYYQEK GYHYFDLGMA PLSGVGRVET  750
SFAKERMAYL VYHFGSHFYS FNGLHKYKKK FTPLWSERYI SCSRSSWLIC  800
AICALLMEDS KIKIVK  (SEQ ID NO:5)                        816
```

FIG. 1e

```
MRILGAFLIF STAFFENITY IMWLQKLGLD PLQEQMLWQF PGLLLGVCFI   50
LLARTIDQKV KNAFPIAIIW ITLTLFYLNL GHISWRLSFW FILLLLGLLV  100
IKPTLYKKQF IYSWEERIKD GIIIVSLMGV LFYIAGLLFP IRAHITGGSI  150
ERLHYIIAWE PIALATLILT LVYLCLVKIL QGKSCQIGDV FNVDRYKKLL  200
QAYGGSSDSG LAFLNDKRLY WYQKNGEDCV AFQFVIVNNK CLIMGEPAGD  250
DTYIREAIES FIDDADKLDY DLVFYSIGQK LTLLLHEYGF DFMKVGEDAL  300
VNLETFTLKG NKYKPFRNAL NRVEKDGFYF EVVQSPHSQE LLNSLEEISN  350
TWLEGRPEKG FSLGYFNKDY FQQAPIALVK NAEHEVVAFA NIMPNYEKSI  400
ISIDLMRHDK QKIPNGVMDF LFLSLFSYYQ EKGYHYFDLG MAPLSGVGRV  450
ETSFAKERMA YLVYHFGSHF YSFNGLHKYK KKFTPLWSER YISCSRSSWL  500
ICAICALLME DSKIKIVK  (SEQ ID NO:6)                      518
```

FIG. 1f

```
AATTTTGATA TCGAAACAAC AACTTTTGAG GCAATGAAAA AGCACGCGTC ATTATTGGAG     60
N  F  D  I   E  T  T   T  F  E   A  M  K  K   H  A  S   L  L  E
---->
AAAATATCTG TTGAGCGTTC TTTTATTGAA TTTGATAAAC TTCTATTAGC ACCTTATTGG    120
K  I  S  V   E  R  S   F  I  E   F  D  K  L   L  L  A   P  Y  W

CGTAAAGGAA TGCTGGCACT AATAGATAGT CATGCTTTTA ATTATCTACC ATGCTTAAAA    180
R  K  G  M   L  A  L   I  D  S   H  A  F  N   Y  L  P   C  L  K

AATAGGGAAT TACAATTAAG CGCCTTTTTG TCCCAGTTAG ATAAAGATTT TTTATTTGAG    240
N  R  E  L   Q  L  S   A  F  L   S  Q  L  D   K  D  F   L  F  E

ACATCAGAAC AAGCTTGGGC ATCACTCATC TTGAGTATGG AAGTTGAACA CACAAAGACT    300
T  S  E  Q   A  W  A   S  L  I   L  S  M  E   V  E  H   T  K  T

TTTTTAAAAA AATGGAAGAC ATCAACTCAC TTTCAAAAAG ATGTTGAGCA TATAGTGGAT    360
F  L  K  K   W  K  T   S  T  H   F  Q  K  D   V  E  H   I  V  D

GTTTATCGTA TTCGTGAACA AATGGGATTG GCTAAAGAAC ATCTTTATCG TTATGGAAAA    420
V  Y  R  I   R  E  Q   M  G  L   A  K  E  H   L  Y  R   Y  G  K

ACTATAATAA AACAAGCGGA AGGTATTCGC AAAGCAAGAG GCTTGATGGT TGATTTCGAA    480
T  I  I  K   Q  A  E   G  I  R   K  A  R  G   L  M  V   D  F  E

AAAATAGAAC AACTAGATAG TGAGTTAGCA ATCCATGATA GGCATGAGAT AGTTGTCAAT    540
K  I  E  Q   L  D  S   E  L  A   I  H  D  R   H  E  I   V  V  N

GGTGGCACCT TAATCAAGAA ATTAGGAATA AAACCTGGTC CACAGATGGG AGATATTATC    600
G  G  T  L   I  K  K   L  G  I   K  P  G  P   Q  M  G   D  I  I

TCTCAAATTG AATTAGCCAT TGTTTTAGGA CAACTGATTA ATGAAGAAGA GGCTATTTTA    660
S  Q  I  E   L  A  I   V  L  G   Q  L  I  N   E  E  E   A  I  L

CATTTTGTTA AGCAGTACTT GATGGATTAG AGAGGATTAT ATGAGCGATT TTTTAGTAGA    720
H  F  V  K   Q  Y  L   M  D                 M  S  D  F   L  V  D
                                            |----->
TGGATTGACT AAGTCGGTTG GTGATAAGAC GGTCTTTAGT AATGTTTCAT TTATCATCCA    780
 G  L  T   K  S  V  G   D  K  T   V  F  S   N  V  S  F   I  I  H

TAGTTTAGAC CGTATTGGGA TTATTGGTGT CAATGGAACT GGAAAGACAA CACTATTAGA    840
 S  L  D   R  I  G  I   I  G  V   N  G  T   G  K  T  T   L  L  D

TGTTATTTCG GGTGAATTAG GTTTTGATGG TGATCGTTCC CCTTTTTCAT CAGCTAATGA    900
 V  I  S   G  E  L  G   F  D  G   D  R  S   P  F  S  S   A  N  D

TTATAAGATT GCTTATTTAA AACAAGAACC AGACTTTGAT GATTCTCAGA CAATTTTGGA    960
 Y  K  I   A  Y  L  K   Q  E  P   D  F  D   D  S  Q  T   I  L  D

CACCGTACTT TCTTCTGACT TAAGAGAGAT GGCTTTAATT AAAGAATATG AATTATTGCT   1020
 T  V  L   S  S  D  L   R  E  M   A  L  I   K  E  Y  E   L  L  L

TAATCACTAC GAAGAAAGTA AGCAATCACG TCTAGAGAAA GTAATGGCAG AAATGGATTC   1080
 N  H  Y   E  E  S  K   Q  S  R   L  E  K   V  M  A  E   M  D  S

TTTAGATGCT TGGTCTATTG AGAGCGAAGT CAAAACAGTA TTATCCAAAT TAGGTATTAC   1140
 L  D  A   W  S  I  E   S  E  V   K  T  V   L  S  K  L   G  I  T

TGATTTGCAG TTGTCGGTTG GTGAATTATC AGGAGGATTA CGAAGACGTG TTCAATTAGC   1200
 D  L  Q   L  S  V  G   E  L  S   G  G  L   R  R  R  V   Q  L  A
```

```
GCAAGTATTA TTAAATGATG CAGATTTATT GCTCTTAGAC GAACCTACTA ACCACTTAGA   1260
 Q  V  L   L  N  D  A   D  L  L   L  L  D   E  P  T  N   H  L  D

TATTGACACT ATTGCATGGT TAACGAATTT TTTGAAAAAT AGTAAAAAGA CAGTGCTTTT   1320
 I  D  T   I  A  W  L   T  N  F   L  K  N   S  K  K  T   V  L  F

TATAACTCAT GATCGTTATT TTCTAGACAA TGTTGCAACA CGTATTTTTG AATTAGATAA   1380
 I  T  H   D  R  Y  F   L  D  N   V  A  T   R  I  F  E   L  D  K

GGCACAGATT ACAGAATATC AAGGCAATTA TCAGGATTAT GTCCGACTTC GTGCAGAACA   1440
 A  Q  I   T  E  Y  Q   G  N  Y   Q  D  Y   V  R  L  R   A  E  Q

AGACGAGCGT GATGCTGCTA GTTTACATAA AAAGAAACAG CTTTATAAAC AGGAACTAGC   1500
 D  E  R   D  A  A  S   L  H  K   K  K  Q   L  Y  K  Q   E  L  A

TTGGATGCGT ACTCAGCCAC AAGCTCGTGC AACGAAACAA CAGGCTCGTA TTAATCGTTT   1560
 W  M  R   T  Q  P  Q   A  R  A   T  K  Q   Q  A  R  I   N  R  F

TCAAAATCTA AAAAACGATT TACACCAAAC AAGCGATACA AGCGATTTGG AAATGACATT   1620
 Q  N  L   K  N  D  L   H  Q  T   S  D  T   S  D  L  E   M  T  F

TGAAACAAGT CGAATTGGGA AAAAGGTTAT TAATTTTGAA AATGTCTCTT TTTCTTACCC   1680
 E  T  S   R  I  G  K   K  V  I   N  F  E   N  V  S  F   S  Y  P

AGATAAATCT ATCTTGAAAG ACTTTAATTT GTTAATTCAA AATAAAGACC GTATTGGCAT   1740
 D  K  S   I  L  K  D   F  N  L   L  I  Q   N  K  D  R   I  G  I

CGTTGGAGAT AATGGTGTTG GAAAGTCAAC CTTACTTAAT TTAATTGTTC AAGATTTACA   1800
 V  G  D   N  G  V  G   K  S  T   L  L  N   L  I  V  Q   D  L  Q

GCCGGATTCG GGTAATGTCT CTATTGGTGA AACGATACGT GTAGGTTACT TTTCACAACA   1860
 P  D  S   G  N  V  S   I  G  E   T  I  R   V  G  Y  F   S  Q  Q

ACTTCATAAT ATGGATGGCT CAAAACGTGT TATTAATTAT TTGCAAGAGG TTGCAGATGA   1920
 L  H  N   M  D  G  S   K  R  V   I  N  Y   L  Q  E  V   A  D  E

GGTTAAAACT AGTGTCGGTA CAACAAGTGT GACAGAACTA TTGGAACAAT TTCTCTTTCC   1980
 V  K  T   S  V  G  T   T  S  V   T  E  L   L  E  Q  F   L  F  P

ACGTTCGACA CATGGAACAC AAATTGCAAA ATTATCAGGT GGTGAGAAAA AAAGACTTTA   2040
 R  S  T   H  G  T  Q   I  A  K   L  S  G   G  E  K  K   R  L  Y

CCTTTTAAAA ATCCTGATTG AAAAGCCTAA TGTGTTACTA CTTGATGAGC CGACAAATGA   2100
 L  L  K   I  L  I  E   K  P  N   V  L  L   L  D  E  P   T  N  D

CTTAGATATT GCTACATTAA CTGTTCTTGA AAATTTTTTA CAAGGCTTTG GTGGTCCTGT   2160
 L  D  I   A  T  L  T   V  L  E   N  F  L   Q  G  F  G   G  P  V

GATTACAGTT AGTCACGATC GTTACTTTTT AGATAAAGTG GCTAATAAAA TTATTGCGTT   2220
 I  T  V   S  H  D  R   Y  F  L   D  K  V   A  N  K  I   I  A  F

TGAAGATAAC GATATCCGTG AATTTTTTGG TAATTATACT GATTATTTAG ATGAAAAAGC   2280
 E  D  N   D  I  R  E   F  F  G   N  Y  T   D  Y  L  D   E  K  A

ATTTAATGAG CAAAATAATG AAGTTATCAG TAAAAAAGAG AGTACCAAGA CAAGTCGTGA   2340
 F  N  E   Q  N  N  E   V  I  S   K  K  E   S  T  K  T   S  R  E

AAAGCAAAGT CGTAAAAGAA TGTCTTACTT TGAAAAACAA GAATGGGCGA CAATTGAAGA   2400
 K  Q  S   R  K  R  M   S  Y  F   E  K  Q   E  W  A  T   I  E  D

CGATATTATG ATATTGGAAA ATACTATCAC TCGTATAGAA AATGATATGC AAACATGTGG   2460
```

```
 D  I  M   I  L  E  N   T  I  T   R  I  E   N  D  M  Q   T  C  G

TAGTGATTTT ACAAGGTTAT CTGATTTACA AAAGGAATTA GATGCAAAAA ATGAAGCACT   2520
 S  D  F   T  R  L  S   D  L  Q   K  E  L   D  A  K  N   E  A  L

TCTAGAAAAG TATGACCGTT ATGAGTACCT TAGTGAGTTA GACACATGAT TATCCGTCCG   2580
 L  E  K   Y  D  R  Y   E  Y  L   S  E  L   D  T  M  I   I  R  P
                                                  |---->
ATTATTAAAA ATGATGACCA AGCAGTTGCA CAATTAATTC GACAAAGTTT ACGCGCCTAT   2640
I  I  K  N   D  D  Q   A  V  A   Q  L  I  R   Q  S  L   R  A  Y

GATTTAGATA AACCTGATAC AGCATATTCA GACCCTCACT TAGATCATTT GACCTCATAC   2700
D  L  D  K   P  D  T   A  Y  S   D  P  H  L   D  H  L   T  S  Y

TACGAAAAAA TAGAGAAGTC AGGATTCTTT GTCATTGAGG AGAGAGATGA GATTATTGGC   2760
Y  E  K  I   E  K  S   G  F  F   V  I  E  E   R  D  E   I  I  G

TGTGGCGGCT TTGGTCCGCT GAAAAATCTA ATTGCAGAGA TGCAGAAGGT GTACATTGCA   2820
C  G  G  F   G  P  L   K  N  L   I  A  E  M   Q  K  V   Y  I  A

GAACGTTTCC GTGGTAAGGG GCTTGCTACT GATTTAGTGA AAATGATTGA AGTAGAAGCT   2880
E  R  F  R   G  K  G   L  A  T   D  L  V  K   M  I  E   V  E  A

CGAAAAATTG GGTATAGACA ACTTTATTTA GAGACAGCCA GTACTTTGAG TAGGGCAACT   2940
R  K  I  G   Y  R  Q   L  Y  L   E  T  A  S   T  L  S   R  A  T

GCGGTTTATA AGCATATGGG ATATTGTGCC TTATCGCAAC CAATAGCAAA TGATCAAGGT   3000
A  V  Y  K   H  M  G   Y  C  A   L  S  Q  P   I  A  N   D  Q  G

CATACAGCTA TGGATATTTG GATGATTAAA GATTTATAAG TTGAAAGTGG ATTAGTGAAC   3060
H  T  A  M   D  I  W   M  I  K   D  L

ATGGATTAAT TATTTTGAGA TAAGAGGAAA GAAAAGGAGA CATATATGGC ATATATTTGG   3120
                                                M  A   Y  I  W
                                                |---->
TCTTATTTGA AAAGGTACCC CAATTGGTTA TGGCTTGATT TACTAGGAGC TATGCTTTTT   3180
S  Y  L  K   R  Y  P   N  W  L   W  L  D  L   L  G  A   M  L  F

GTGACGGTTA TCCTAGGAAT GCCCACAGCC TTAGCGGGTA TGATTGATAA TGGCGTTACA   3240
V  T  V  I   L  G  M   P  T  A   L  A  G  M   I  D  N   G  V  T

AAAGGTGATC GGACTGGAGT TTATCTGTGG ACGTTCATCA TGTTTATATT TGTTGTACTA   3300
K  G  D  R   T  G  V   Y  L  W   T  F  I  M   F  I  F   V  V  L

GGTATTATTG GGCGTATTAC GATGGCTTAC GCATCTAGTC GCTTAACGAC AACAATGATT   3360
G  I  I  G   R  I  T   M  A  Y   A  S  S  R   L  T  T   T  M  I

AGAGATATGC GTAATGATAT GTATGCTAAG CTTCAAGAAT ACTCCCATCA TGAATATGAA   3420
R  D  M  R   N  D  M   Y  A  K   L  Q  E  Y   S  H  H   E  Y  E

CAGATAGGTG TATCTTCACT AGTGACACGT ATGACAAGCG ATACTTTTGT TTTGATGCAA   3480
Q  I  G  V   S  S  L   V  T  R   M  T  S  D   T  F  V   L  M  Q

TTTGCTGAAA TGTCTTTACG TTTAGGCCTA GTAACTCCTA TGGTAATGAT TTTTAGCGTG   3540
F  A  E  M   S  L  R   L  G  L   V  T  P  M   V  M  I   F  S  V

GTTATGATAC TAATTACGAG TCCATCTTTG GCTTGGCTTG TAGCGGTTGC GATGCCTCTT   3600
V  M  I  L   I  T  S   P  S  L   A  W  L  V   A  V  A   M  P  L

TTGGTAGGAG TCGTTTTATA TGTAGCTATA AAAACAAAAC CTTTATCTGA AAGACAACAG   3660
L  V  G  V   V  L  Y   V  A  I   K  T  K  P   L  S  E   R  Q  Q
```

```
ACTATGCTTG ATAAAATCAA TCAATATGTT CGTGAAAATT TAACAGGGTT ACGCGTTGTT   3720
T  M  L  D   K  I  N   Q  Y  V   R  E  N  L   T  G  L   R  V  V

AGAGCCTTTG CAAGAGAGAA TTTTCAATCA CAAAAATTTC AAGTCGCTAA CCAACGTTAC   3780
R  A  F  A   R  E  N   F  Q  S   Q  K  F  Q   V  A  N   Q  R  Y

ACAGATACTT CAACTGGTCT TTTTAAATTA ACAGGGCTAA CAGAACCACT TTTCGTTCAA   3840
T  D  T  S   T  G  L   F  K  L   T  G  L  T   E  P  L   F  V  Q

ATTATTATTG CAATGATTGT GGCTATCGTT TGGTTTGCTT TGGATCCCTT ACAAAGAGGT   3900
I  I  I  A   M  I  V   A  I  V   W  F  A  L   D  P  L   Q  R  G

GCTATTAAAA TAGGGGATTT AGTTGCTTTT ATCGAATATA GCTTCCATGC TCTCTTTTCA   3960
A  I  K  I   G  D  L   V  A  F   I  E  Y  S   F  H  A   L  F  S

TTTTTGCTAT TTGCCAATCT TTTTACTATG TATCCTCGTA TGGTGGTATC AAGCCATCGT   4020
F  L  L  F   A  N  L   F  T  M   Y  P  R  M   V  V  S   S  H  R

ATTAGAGAGG TGATGGATAT GCCAATCTCT ATCAATCCTA ATGCCGAAGG TGTTACGGAT   4080
I  R  E  V   M  D  M   P  I  S   I  N  P  N   A  E  G   V  T  D

ACGAAACTTA AAGGGCATTT AGAATTTGAT AATGTAACAT TCGCTTATCC AGGAGAAACA   4140
T  K  L  K   G  H  L   E  F  D   N  V  T  F   A  Y  P   G  E  T

GAGAGTCCCG TTTTGCATGA TATTTCTTTT AAAGCTAAGC CTGGAGAAAC AATTGCTTTT   4200
E  S  P  V   L  H  D   I  S  F   K  A  K  P   G  E  T   I  A  F

ATTGGTTCAA CAGGTTCAGG AAAATCTTCT CTTGTTAATT TGATTCCACG TTTTTATGAT   4260
I  G  S  T   G  S  G   K  S  S   L  V  N  L   I  P  R   F  Y  D

GTGACACTTG GAAAAATCTT AGTAGATGGA GTTGATGTAA GAGATTATAA CCTTAAATCA   4320
V  T  L  G   K  I  L   V  D  G   V  D  V  R   D  Y  N   L  K  S

CTTCGCCAAA AGATTGGATT TATCCCCCAA AAAGCTCTTT TATTTACAGG GACAATAGGA   4380
L  R  Q  K   I  G  F   I  P  Q   K  A  L  L   F  T  G   T  I  G

GAGAATTTAA AATATGGAAA AGCTGATGCT ACTATTGATG ATCTTAGACA AGCGGTTGAT   4440
E  N  L  K   Y  G  K   A  D  A   T  I  D  D   L  R  Q   A  V  D

ATTTCTCAAG CTAAAGAGTT TATTGAGAGT CACCAAGAAG CCTTTGAAAC GCATTTAGCT   4500
I  S  Q  A   K  E  F   I  E  S   H  Q  E  A   F  E  T   H  L  A

GAAGGTGGGA GCAATCTTTC TGGGGGTCAA AAACAACGGT TATCTATTGC TAGGGCTGTT   4560
E  G  G  S   N  L  S   G  G  Q   K  Q  R  L   S  I  A   R  A  V

GTTAAAGATC CAGATTTATA TATTTTTGAT GATTCATTTT CTGCTCTCGA TTATAAGACA   4620
V  K  D  P   D  L  Y   I  F  D   D  S  F  S   A  L  D   Y  K  T

GACGCTACTT TAAGAGCGCG TCTAAAAGAA GTAACCGGTG ATTCTACAGT TTTGATAGTT   4680
D  A  T  L   R  A  R   L  K  E   V  T  G  D   S  T  V   L  I  V

GCTCAAAGGG TGGGTACGAT TATGGATGCT GATCAGATTA TTGTCCTTGA TGAAGGCGAA   4740
A  Q  R  V   G  T  I   M  D  A   D  Q  I  I   V  L  D   E  G  E

ATTGTCGGTC GTGGTACCCA CGCTCAATTA ATAGAAAATA ATGCTATTTA TCGTGAAATC   4800
I  V  G  R   G  T  H   A  Q  L   I  E  N  N   A  I  Y   R  E  I

GCTGAGTCAC AACTGAAGAA CCAAAACTTA TCAGAAGGAG AGTGATTGTA TGAGAAAAAA   4860
A  E  S  Q   L  K  N   Q  N  L   S  E  G  E            M   R  K  K
                                                       |---->
```

```
ATCTGTTTTT TTGAGATTAT GGTCTTACCT AACTCGCTAC AAAGCTACTC TTTTCTTAGC   4920
 S  V  F   L  R  L  W   S  Y  L   T  R  Y   K  A  T  L   F  L  A

GATTTTTTTG AAAGTTTTAT CTAGTTTTAT GAGTGTTCTG GAGCCTTTTA TTTTAGGGTT   4980
 I  F  L   K  V  L  S   S  F  M   S  V  L   E  P  F  I   L  G  L

AGCGATAACA GAGTTGACTG CTAACCTTGT TGATATGGCT AAGGGAGTTT CTGGGGCAGA   5040
 A  I  T   E  L  T  A   N  L  V   D  M  A   K  G  V  S   G  A  E

ATTGAACGTT CCTTATATTG CTGGTATTTT GATTATTTAT TTTTTCAGAG GTGTTTTCTA   5100
 L  N  V   P  Y  I  A   G  I  L   I  I  Y   F  F  R  G   V  F  Y

TGAATTAGGT TCTTATGGCT CAAATT  (SEQ ID NO:7)                         5126
 E  L  G   S  Y  G  S   N
```

FIG. 2a

```
NFDIETTTFE AMKKHASLLE KISVERSFIE FDKLLLAPYW RKGMLALIDS   50
HAFNYLPCLK NRELQLSAFL SQLDKDFLFE TSEQAWASLI LSMEVEHTKT  100
FLKKWKTSTH FQKDVEHIVD VYRIREQMGL AKEHLYRYGK TIIKQAEGIR  150
KARGLMVDFE KIEQLDSELA IHDRHEIVVN GGTLIKKLGI KPGPQMGDII  200
SQIELAIVLG QLINEEEAIL HFVKQYLMD  (SEQ ID NO:8)          229
```

FIG. 2b

```
MSDFLVDGLT KSVGDKTVFS NVSFIIHSLD RIGIIGVNGT GKTTLLDVIS   50
GELGFDGDRS PFSSANDYKI AYLKQEPDFD DSQTILDTVL SSDLREMALI  100
KEYELLLNHY EESKQSRLEK VMAEMDSLDA WSIESEVKTV LSKLGITDLQ  150
LSVGELSGGL RRRVQLAQVL LNDADLLLLD EPTNHLDIDT IAWLTNFLKN  200
SKKTVLFITH DRYFLDNVAT RIFELDKAQI TEYQGNYQDY VRLRAEQDER  250
DAASLHKKKQ LYKQELAWMR TQPQARATKQ QARINRFQNL KNDLHQTSDT  300
SDLEMTFETS RIGKKVINFE NVSFSYPDKS ILKDFNLLIQ NKDRIGIVGD  350
NGVGKSTLLN LIVQDLQPDS GNVSIGETIR VGYFSQQLHN MDGSKRVINY  400
LQEVADEVKT SVGTTSVTEL LEQFLFPRST HGTQIAKLSG GEKKRLYLLK  450
ILIEKPNVLL LDEPTNDLDI ATLTVLENFL QGFGGPVITV SHDRYFLDKV  500
ANKIIAFEDN DIREFFGNYT DYLDEKAFNE QNNEVISKKE STKTSREKQS  550
RKRMSYFEKQ EWATIEDDIM ILENTITRIE NDMQTCGSDF TRLSDLQKEL  600
DAKNEALLEK YDRYEYLSEL DT  (SEQ ID NO:9)                 622
```

FIG. 2c

```
MIIRPIIKND DQAVAQLIRQ SLRAYDLDKP DTAYSDPHLD HLTSYYEKIE   50
KSGFFVIEER DEIIGCGGFG PLKNLIAEMQ KVYIAERFRG KGLATDLVKM  100
IEVEARKIGY RQLYLETAST LSRATAVYKH MGYCALSQPI ANDQGHTAMD  150
IWMIKDL  (SEQ ID NO:10)                                 157
```

FIG. 2d

MAYIWSYLKR YPNWLWLDLL GAMLFVTVIL GMPTALAGMI DNGVTKGDRT 50
GVYLWTFIMF IFVVLGIIGR ITMAYASSRL TTTMIRDMRN DMYAKLQEYS 100
HHEYEQIGVS SLVTRMTSDT FVLMQFAEMS LRLGLVTPMV MIFSVVMILI 150
TSPSLAWLVA VAMPLLVGVV LYVAIKTKPL SERQQTMLDK INQYVRENLT 200
GLRVVRAFAR ENFQSQKFQV ANQRYTDTST GLFKLTGLTE PLFVQIIIAM 250
IVAIVWFALD PLQRGAIKIG DLVAFIEYSF HALFSFLLFA NLFTMYPRMV 300
VSSHRIREVM DMPISINPNA EGVTDTKLKG HLEFDNVTFA YPGETESPVL 350
HDISFKAKPG ETIAFIGSTG SGKSSLVNLI PRFYDVTLGK ILVDGVDVRD 400
YNLKSLRQKI GFIPQKALLF TGTIGENLKY GKADATIDDL RQAVDISQAK 450
EFIESHQEAF ETHLAEGGSN LSGGQKQRLS IARAVVKDPD LYIFDDSFSA 500
LDYKTDATLR ARLKEVTGDS TVLIVAQRVG TIMDADQIIV LDEGEIVGRG 550
THAQLIENNA IYREIAESQL KNQNLSEGE (SEQ ID NO:11) 579

FIG. 2e

MRKKSVFLRL WSYLTRYKAT LFLAIFLKVL SSFMSVLEPF ILGLAITELT 50
ANLVDMAKGV SGAELNVPYI AGILIIYFFR GVFYELGSYG SN 92
(SEQ ID NO:12)

FIG. 2f

```
AATTTGGAAG TGCTCTATCA ACAGTTGAAG TAAAGGAGAT TATTAGTGAA GAAAACATAT    60
 F  G  S   A  L  S   T  V  E  V   K  E  I   I  S  E   E  N  I  W
 ---->
GGTTATATCG GCTCAGTTGC TGCCATTTTA CTAGCTACTC ATATTGGAAG TTACCAACTT   120
 L  Y  R   L  S  C   C  H  F  T   S  Y  S   Y  W  K   L  P  T  W

GGTAAGCATC ATATGGGTCT AGCAACAAAG GACAATCAGA TTGCCTATAT TGATGACAGC   180
            M  G  L   A  T  K   D  N  Q  I   A  Y  I   D  D  S
            |---->
AAAGGTAAGG CAAAAGCCCC TAAAACAAAC AAAACGATGG ATCAAATCAG TGCTGAAGAA   240
K  G  K  A   K  A  P   K  T  N   K  T  M  D   Q  I  S   A  E  E

GGCATCTCTG CTGAACAGAT CGTAGTCAAA ATTACTGACC AAGGCTATGT GACCTCACAC   300
G  I  S  A   E  Q  I   V  V  K   I  T  D  Q   G  Y  V   T  S  H

GGTGACCATT ATCATTTTTA CAATGGGAAA GTTCCTTATG ATGCGATTAT TAGTGAAGAG   360
G  D  H  Y   H  F  Y   N  G  K   V  P  Y  D   A  I  I   S  E  E

TTGTTGATGA CGGATCCTAA TTACCGTTTT AAACAATCAG ACGTTATCAA TGAAATCTTA   420
L  L  M  T   D  P  N   Y  R  F   K  Q  S  D   V  I  N   E  I  L
      |---->
GACGGTTACG TTATTAAAGT CAATGGCAAC TATTATGTTT ACCTCAAGCC AGGTAGTAAG   480
D  G  Y  V   I  K  V   N  G  N   Y  Y  V  Y   L  K  P   G  S  K

CGCAAAAACA TTCGAACCAA ACAACAAATT GCTGAGCAAG TAGCCAAAGG AACTAAAGAA   540
R  K  N  I   R  T  K   Q  Q  I   A  E  Q  V   A  K  G   T  K  E

GCTAAAGAAA AAGGTTTAGC TCAAGTGGCC CATCTCAGTA AAGAAGAAGT TGCGGCAGTC   600
A  K  E  K   G  L  A   Q  V  A   H  L  S  K   E  E  V   A  A  V

AATGAAGCAA AAAGACAAGG ACGCTATACT ACAGACGATG GCTATATTTT TAGTCCGACA   660
N  E  A  K   R  Q  G   R  Y  T   T  D  D  G   Y  I  F   S  P  T

GATATCATTG ATGATTTAGG AGATGCTTAT TTAGTACCTC ATGGTAATCA CTATCATTAT   720
D  I  I  D   D  L  G   D  A  Y   L  V  P  H   G  N  H   Y  H  Y

ATTCCTAAAA AGGATTTGTC TCCAAGTGAG CTAGCTGCTG CACAAGCCTA CTGGAGTCAA   780
I  P  K  K   D  L  S   P  S  E   L  A  A  A   Q  A  Y   W  S  Q

AAACAAGGTC GAGGTGCTAG ACCGTCTGAT TACCGCCCGA CACCAGCCCC AGGTCGTAGG   840
K  Q  G  R   G  A  R   P  S  D   Y  R  P  T   P  A  P   G  R  R

AAAGCCCCAA TTCCTGATGT GACGCCTAAC CCTGGACAAG GTCATCAGCC AGATAACGGT   900
K  A  P  I   P  D  V   T  P  N   P  G  Q  G   H  Q  P   D  N  G

GGCTATCATC CAGCGCCTCC TAGGCCAAAT GATGCGTCAC AAAACAAACA CCAAAGAGAT   960
G  Y  H  P   A  P  P   R  P  N   D  A  S  Q   N  K  H   Q  R  D

GAGTTTAAAG GAAAAACCTT TAAGGAACTT TTAGATCAAC TACACCGTCT TGATTTGAAA  1020
E  F  K  G   K  T  F   K  E  L   L  D  Q  L   H  R  L   D  L  K

TACCGTCATG TGGAAGAAGA TGGGTTGATT TTTGAACCGA CTCAAGTGAT CAAATCAAAC  1080
Y  R  H  V   E  E  D   G  L  I   F  E  P  T   Q  V  I   K  S  N

GCTTTTGGGT ATGTGGTGCC TCATGGAGAT CATTATCATA TTATCCCAAG AAGTCAGTTA  1140
A  F  G  Y   V  V  P   H  G  D   H  Y  H  I   I  P  R   S  Q  L

TCACCTCTTG AAATGGAATT AGCAGATCGA TACTTAGCTG GCCAAACTGA GGACAATGAC  1200
S  P  L  E   M  E  L   A  D  R   Y  L  A  G   Q  T  E   D  N  D

TCAGGTTCAG AGCACTCAAA ACCATCAGAT AAAGAAGTGA CACATACCTT TCTTGGTCAT  1260
```

```
S  G  S  E   H  S  K   P  S  D   K  E  V  T   H  T  F   L  G  H

CGCATCAAAG CTTACGGAAA AGGCTTAGAT GGTAAACCAT ATGATACGAG TGATGCTTAT  1320
R  I  K  A   Y  G  K   G  L  D   G  K  P  Y   D  T  S   D  A  Y

GTTTTTAGTA AAGAATCCAT TCATTCAGTG GATAAATCAG GAGTTACAGC TAAACACGGA  1380
V  F  S  K   E  S  I   H  S  V   D  K  S  G   V  T  A   K  H  G

GATCATTTCC ACTATATAGG ATTTGGAGAA CTTGAACAAT ATGAGTTGGA TGAGGTCGCT  1440
D  H  F  H   Y  I  G   F  G  E   L  E  Q  Y   E  L  D   E  V  A

AACTGGGTGA AAGCAAAAGG TCAAGCTGAT GAGCTTGCTG CTGCTTTGGA TCAGGAACAA  1500
N  W  V  K   A  K  G   Q  A  D   E  L  A  A   A  L  D   Q  E  Q

GGCAAAGAAA AACCACTCTT TGACACTAAA AAAGTGAGTC GCAAAGTAAC AAAAGATGGT  1560
G  K  E  K   P  L  F   D  T  K   K  V  S  R   K  V  T   K  D  G

AAAGTGGGCT ATATGATGCC AAAAGATGGT AAGGACTATT TCTATGCTCG TGATCAACTT  1620
K  V  G  Y   M  M  P   K  D  G   K  D  Y  F   Y  A  R   D  Q  L

GATTTGACTC AGATTGCCTT TGCCGAACAA GAACTAATGC TTAAAGATAA GAAGCATTAC  1680
D  L  T  Q   I  A  F   A  E  Q   E  L  M  L   K  D  K   K  H  Y

CGTTATGACA TTGTTGACAC AGGTATTGAG CCACGACTTG CTGTAGATGT GTCAAGTCTG  1740
R  Y  D  I   V  D  T   G  I  E   P  R  L  A   V  D  V   S  S  L

CCGATGCATG CTGGTAATGC TACTTACGAT ACTGGAAGTT CGTTTGTTAT CCCACATATT  1800
P  M  H  A   G  N  A   T  Y  D   T  G  S  S   F  V  I   P  H  I

GATCATATCC ATGTCGTTCC GTATTCATGG TTGACGCGCG ATCAGATTGC AACAGTCAAG  1860
D  H  I  H   V  V  P   Y  S  W   L  T  R  D   Q  I  A   T  V  K

TATGTGATGC AACACCCCGA AGTTCGTCCG GATGTATGGT CTAAGCCAGG GCATGAAGAG  1920
Y  V  M  Q   H  P  E   V  R  P   D  V  W  S   K  P  G   H  E  E

TCAGGTTCGG TCATTCCAAA TGTTACGCCT CTTGATAAAC GTGCTGGTAT GCCAAACTGG  1980
S  G  S  V   I  P  N   V  T  P   L  D  K  R   A  G  M   P  N  W

CAAATTATCC ATTCTGCTGA AGAAGTTCAA AAAGCCCTAG CAGAAGGTCG TTTTGCAACA  2040
Q  I  I  H   S  A  E   E  V  Q   K  A  L  A   E  G  R   F  A  T

CCAGACGGCT ATATTTTCGA TCCACGAGAT GTTTTGGCCA AAGAAACTTT TGTATGGAAA  2100
P  D  G  Y   I  F  D   P  R  D   V  L  A  K   E  T  F   V  W  K

GATGGCTCCT TTAGCATCCC AAGAGCAGAT GGCAGTTCAT TGAGAACCAT TAATAAATCT  2160
D  G  S  F   S  I  P   R  A  D   G  S  S  L   R  T  I   N  K  S

GATCTATCCC AAGCTGAGTG GCAACAAGCT CAAGAGTTAT TGGCAAAGAA AAATACTGGT  2220
D  L  S  Q   A  E  W   Q  Q  A   Q  E  L  L   A  K  K   N  T  G

GATGCTACTG ATACGGATAA ACCCAAAGAA AAGCAACAGG CAGATAAGAG CAATGAAAAC  2280
D  A  T  D   T  D  K   P  K  E   K  Q  Q  A   D  K  S   N  E  N

CAACAGCCAA GTGAAGCCAG TAAAGAAGAA AAAGAATCAG ATGACTTTAT AGACAGTTTA  2340
Q  Q  P  S   E  A  S   K  E  E   K  E  S  D   D  F  I   D  S  L

CCAGACTATG GTCTAGATAG AGCAACCCTA GAAGATCATA TCAATCAATT AGCACAAAAA  2400
P  D  Y  G   L  D  R   A  T  L   E  D  H  I   N  Q  L   A  Q  K

GCTAATATCG ATCCTAAGTA TCTCATTTTC CAACCAGAAG GTGTCCAATT TTATAATAAA  2460
A  N  I  D   P  K  Y   L  I  F   Q  P  E  G   V  Q  F   Y  N  K
```

```
AATGGTGAAT TGGTAACTTA TGATATCAAG ACACTTCAAC AAATAAACCC TTAACCAAAA  2520
N  G  E  L   V  T  Y   D  I  K   T  L  Q  Q   I  N  P

GAAGATCTCA TTGTTAAAGC ACTGCTTTGT CAAAGCAAGT TACGGTGATT TTGAAGTCAT  2580

TCTATGTAAC GAGTAGTGAT AAAAGTTGGA TAATAGCGGT TTTCTTTTGC AAAGAAATGG  2640

TATCCATGTT AGAATAGTAA AAAAAGAGGA GGATTCTTGG ACTAATGTCA AATAAGTAGA  2700

CAGAAAACTG TGTTATTTTA TTGCGTTAAA ATAATTTTCT TCTTTCTGAT TAGGGGTTAG  2760
                 .K   I  A  N  F   Y  N  E   E  K  Q   N  P  T  L

TCCTAGATTA GCCGTATGTG GGTTGTAATT GTTATAAAAA TTCTCAATGT ATTCAAAGCA  2820
  G  L  N   A  T  H   P  N  Y  N   N  Y  F   N  E  I   Y  E  F  C

GTCTAATTGA ACCTGTTTGA TATTTTGATA ATGTTTTCGG TTGATTTGTC TATGCTTTAA  2880
  D  L  Q   V  Q  K   I  N  Q  Y   H  K  R   N  I  Q   R  H  K  L

ATACTTGAAA AATGCTTCAG TTACGGCATT ATCATAAGGA TATCCAGGAT TAGAAAAAGA  2940
  Y  K  F   F  A  E   T  V  A  N   D  Y  P   Y  G  P   N  S  F  S

ATGCATGATA TTGGCACTGC ACCCTAATAG TGAGACGCAA GAAAAACACT TTTAGGCAAT  3000
  H  M                                                        A  I
<----|
CAGTTTTCTG TACTGTACAG GCGACTGGTC GTTTAATCTC TGTTGAATTC TAGTTTCATT  3060
  L  K  R   Y  Q  V   P  S  Q  D   N  L  R   Q  Q  I   R  T  E  N

ATAAAATGTA ATGTAATTTT TAACAATATT TGTTATACTA TCTTTGTTGT ATTTTCTCCT  3120
  Y  F  T   I  Y  N   K  V  I  N   T  I  S   D  K  N   Y  K  R  R

ATTATGGAAA TAAAAGGTTT CAGTCTTTAG GACGGTGTGA AACCATTCAA TACAGGCATT  3180
  N  H  F   Y  F  T   E  T  K  L   V  T  H   F  W  E   I  C  A  N

ATCTGCAGGT GTTCCTTTTC GAGACATTGA GCGGATAATG TCTTTTTCCG TGCAAGCCTG  3240
  D  A  P   T  G  K   R  S  M  S   R  I  I   D  K  E   T  C  A  Q

GTAGTAAGCC ATAGAAGTAT ACACTGAGCC TTGGTCACTG TGTAAGATTG CTCCTTTATT  3300
  Y  Y  A   M
       <----|
TAGGCAATTT TAACTGATTA AGGGTGTCTA GTACAAAATC CGTGTCCTGA CAATCTGAGA  3360
K  P  L  K   L  Q  N   L  T  D   L  V  F  D   T  D  Q   C  D  S

TAGTGTAAGC TATAATTTCT CGGTTATAGA GATTCATAAT TGATGAGAGA TACAATTTAC  3420
I  T  Y  A   I  I  E   R  N  Y   L  N  M  I   S  S  L   Y  L  K

AGTTACCGAA ATATAGGTAG GTAATATCTG TTACGAGCTT TTCCTTAGGC TTATCGGCAT  3480
C  N  G  F   Y  L  Y   T  I  D   T  V  L  K   E  K  P   K  D  A

GGAAATCCCG ACTCAATTTA TTATCTGTTA AATAATAAGC TTTACCCAAA TTGGGAACTT  3540
H  G  D  R   S  L  K   N  D  T   L  Y  Y  A   K  G  L   N  P  V

TCTTGGTACG TGTCCGACAA AGCCAGCCAT TATTTTTCAT GATACGATAG ACTTTCTTTG  3600
K  K  T  R   T  R  C   L  W  G   N  N  K  M   I  R  Y   V  K  K

TATTAACAGT CAATCCGTGG ATTTTTTTGA GCAATCGTGT AATGGTACGA TAGCCATAAA  3660
T  N  V  T   L  G  H   I  K  K   L  L  R  T   I  T  R   Y  G  Y

TAAAGTGATT CTCCATACAG AGCTGTTCAA TTAATTCAAT AAGGTCATCT TTTTTTGCGG  3720
I  F  H  N   E  M
          <----|
```

```
CTTCTCATAC TCCTTTTTCC AACGGTAATA GGTCGACCGC TTGACCTTAA AACAGTCTAG   3780

AATGAAAACT ATCGGGTAGT TGTTTTTATA GTCTTCCACA AGCTTGATAA GACTTACTTT   3840
                                                            . K
ATCGATTTCC TTATCAAGCC TCGATACTTT TTTAAGAGGT CAACCTGTAA TTGTAATTGT   3900
 I  S  K   R  I  L  G   R  Y  K   K  L  L   D  V  Q  L   Q  L  Q

TCCACTTCAG ACAGATGTTC CAAGCCTTTA CCGTAGGTAT ATTGCTTGCC AACACCTTGA   3960
 E  V  E   S  L  H  E   L  G  K   G  Y  T   Y  Q  K  G   V  G  Q

TGAAAACGAT AAAGCTCCTC GTTTTCGTAC CATTTCATCC AAGTATAGAT TTGACTATTA   4020
 H  F  R   Y  L  E  E   N  E  Y   W  K  M   W  T  Y  I   Q  S  N

TTTTTGATGC CTAAAGTCTC CATAATAACT CTGTTAGACT TGCCTGCTTT CTTCATATCG   4080
 N  K  I   G  L  T  E   M  I  V   R  N  S   K  G  A  K   K  M  D

ATGCAAGCCA GCTTAGTTTC CCATGAATAT GCTTTTTTAA CCATAATAAA ACATTCCTGT   4140
 I  C  A   L  K  T  E   W  S  Y   A  K  K   V  M
                                            <----|
TTCTAGTTTA CTAAATTTCA ACAGGAGTGT TTTTCTTTTG TCTCATTTTA GGGATTCAGT   4200

GCCTATTGTT GTCATCAATT ATTTTTCTAA ATTCCCCGGA CTTAAATTGT GACCCTTGGT   4260

CGGAATGAAA GAGAAGTGTT CCTTCAATCT TTCTTTTATT AAGTGAAAAG GCAACACTTT   4320

TCTGTACAAC ATTTATAAAG TGTTTTTCTA GGCAATTAAT CTTTTAGTCA TTGGTGTTTG   4380
                                  . A  I  L   R  K  T   M  P  T  Q

GTAGTTGAGA CTACCATGAA TGCGGTGGTA ATTCCACCAA TGAACATAGT CTTTAGTCTT   4440
 Y  N  L   S  G  H   I  R  H  Y   N  W  W   H  V  Y   D  K  T  K

AAGAGCTAGT TCTTCCAGCA ATTGAAAGGT TTCTTGATAA ACAAATTCAA TTTTGAAAGC   4500
 L  A  L   E  E  L   L  Q  F  T   E  Q  Y   V  F  E   I  K  F  A

ACGATACGTA CTTTCAGCTA CGGCATTGTC ATAAGGATAA CCAGCCTGAC TAAGCGAACG   4560
 R  Y  T   S  E  A   V  A  N  D   Y  P  Y   G  A  Q   S  L  S  R

TGTGATTCCA AAGGCTTCCA ATATTTCATC AATTAACTGA TTATCAAACT CTTTGCCACG   4620
 T  I  G   F  A  E   L  I  E  D   I  L  Q   N  D  F   E  K  G  R

ATCTGAATGG AACATCTTGA CTTTGGTCAG GGCGTAAGGG ATGCTTTGTA TGGCTTGCTT   4680
 D  S  H   F  M  K   V  K  T  L   A  Y  P   I  S  Q   I  A  Q  K

AACGAGTTCA GCGGTCTTGT GCCAACCAAG AGACAGGCCG ATGATTTCAC GGTTGTATAG   4740
 V  L  E   A  T  K   H  W  G  L   S  L  G   I  I  E   R  N  Y  L

GTCAATGATG AGGCAAACAT AAGCCCAACG ATTGCCTACA CGAACATAGG TTAAGTCAGT   4800
 D  I  I   L  C  V   Y  A  W  R   N  G  V   R  V  Y   T  L  D  T

GACTAAGGCT TGTAGTGGTC TTTCTTGCTT AAATTGCCTG TCTAAGTGGT TGGGAATAGG   4860
 V  L  A   Q  L  P   R  E  Q  K   F  Q  R   D  L  H   N  P  I  P

GGCTTCATTC TTGCCTCTAG AATGTGGTTT GAAGGTGGCT TTCTGATAAA CAGAAACCAA   4920
 A  E  N   K  G  R   S  H  P  K   F  T  A   K  Q  Y   V  S  V  L

ATTGAGTCGC TTCATAATGC GTCGAATCCG ACGACGTGAA AGTGTGATAC CTTCGTTATT   4980
 N  L  R   K  M  I   R  R  I  R   R  R  S   L  T  I   G  E  N  N

CAAGCATATT TTGATTTTTC TGGATCCGTA TCTAGACTCG CTATCGAGAA AAATTCTTTT   5040
 L  C  I   K  I  K   R  S  G  Y   R  S  E   S  D  L   F  I  R  K
```

```
AATAGTTTCT TCAAACTCCG TTTCAGATAC TGACTCCACG GCTTGATAGT AATAACTTGA  5100
 I  T  E   E  F  E   T  E  S  V   S  E  V   A  Q  Y   Y  Y  S  S

GTGTGGCATA TTCAGCCAGC GACACATCTT TGAAATGCTG TATTTATCCT TATTAGCAGT  5160
 H  P  M   N  L  W   R  C  M  K   S  I  S   Y  K  D   K  N  A  T

GATTATTTCC CTTTTTGTGC CATAATCACC GCTGCTTGCT TTAGGATATC TAATT       5215
 I  I  E   R  K  T   G  Y  D  G   S  S  A   K  P  Y   R  I
(SEQ ID NO:13)                                        <----|
```

FIG. 3a

```
FGSALSTVEV KEIISEENIW LYRLSCCHFT SYSYWKLPTW                        40
(SEQ ID NO:14)
```

FIG. 3b

```
MGLATKDNQI AYIDDSKGKA KAPKTNKTMD QISAEEGISA EQIVVKITDQ              50
GYVTSHGDHY HFYNGKVPYD AIISEELLMT DPNYRFKQSD VINEILDGYV             100
IKVNGNYYVY LKPGSKRKNI RTKQQIAEQV AKGTKEAKEK GLAQVAHLSK             150
EEVAAVNEAK RQGRYTTDDG YIFSPTDIID DLGDAYLVPH GNHYHYIPKK             200
DLSPSELAAA QAYWSQKQGR GARPSDYRPT PAPGRRKAPI PDVTPNPGQG             250
HQPDNGGYHP APPRPNDASQ NKHQRDEFKG KTFKELLDQL HRLDLKYRHV             300
EEDGLIFEPT QVIKSNAFGY VVPHGDHYHI IPRSQLSPLE MELADRYLAG             350
QTEDNDSGSE HSKPSDKEVT HTFLGHRIKA YGKGLDGKPY DTSDAYVFSK             400
ESIHSVDKSG VTAKHGDHFH YIGFGELEQY ELDEVANWVK AKGQADELAA             450
ALDQEQGKEK PLFDTKKVSR KVTKDGKVGY MMPKDGKDYF YARDQLDLTQ             500
IAFAEQELML KDKKHYRYDI VDTGIEPRLA VDVSSLPMHA GNATYDTGSS             550
FVIPHIDHIH VVPYSWLTRD QIATVKYVMQ HPEVRPDVWS KPGHEESGSV             600
IPNVTPLDKR AGMPNWQIIH SAEEVQKALA EGRFATPDGY IFDPRDVLAK             650
ETFVWKDGSF SIPRADGSSL RTINKSDLSQ AEWQQAQELL AKKNTGDATD             700
TDKPKEKQQA DKSNENQQPS EASKEEKESD DFIDSLPDYG LDRATLEDHI             750
NQLAQKANID PKYLIFQPEG VQFYNKNGEL VTYDIKTLQQ INP                    793
(SEQ ID NO:15)
```

FIG. 3c

MTDPNYRFKQ SDVINEILDG YVIKVNGNYY VYLKPGSKRK NIRTKQQIAE 50
QVAKGTKEAK EKGLAQVAHL SKEEVAAVNE AKRQGRYTTD DGYIFSPTDI 100
IDDLGDAYLV PHGNHYHYIP KKDLSPSELA AAQAYWSQKQ GRGARPSDYR 150
PTPAPGRRKA PIPDVTPNPG QGHQPDNGGY HPAPPRPNDA SQNKHQRDEF 200
KGKTFKELLD QLHRLDLKYR HVEEDGLIFE PTQVIKSNAF GYVVPHGDHY 250
HIIPRSQLSP LEMELADRYL AGQTEDNDSG SEHSKPSDKE VTHTFLGHRI 300
KAYGKGLDGK PYDTSDAYVF SKESIHSVDK SGVTAKHGDH FHYIGFGELE 350
QYELDEVANW VKAKGQADEL AAALDQEQGK EKPLFDTKKV SRKVTKDGKV 400
GYMMPKDGKD YFYARDQLDL TQIAFAEQEL MLKDKKHYRY DIVDTGIEPR 450
LAVDVSSLPM HAGNATYDTG SSFVIPHIDH IHVVPYSWLT RDQIATVKYV 500
MQHPEVRPDV WSKPGHEESG SVIPNVTPLD KRAGMPNWQI IHSAEEVQKA 550
LAEGRFATPD GYIFDPRDVL AKETFVWKDG SFSIPRADGS SLRTINKSDL 600
SQAEWQQAQE LLAKKNTGDA TDTDKPKEKQ QADKSNENQQ PSEASKEEKE 650
SDDFIDSLPD YGLDRATLED HINQLAQKAN IDPKYLIFQP EGVQFYNKNG 700
ELVTYDIKTL QQINP (SEQ ID NO:16) 715

FIG. 3d

MHSFSNPGYP YDNAVTEAFF KYLKHRQINR KHYQNIKQVQ LDCFEYIENF 50
YNNYNPHTAN LGLTPNQKEE NYFNAIK (SEQ ID NO:17) 77

FIG. 3e

MAYYQACTEK DIIRSMSRKG TPADNACIEW FHTVLKTETF YFHNRRKYNK 50
DSITNIVKNY ITFYNETRIQ QRLNDQSPVQ YRKLIA (SEQ ID NO:18) 86

FIG. 3f

MENHFIYGYR TITRLLKKIH GLTVNTKKVY RIMKNNGWLC RTRTKKVPNL 50
GKAYYLTDNK LSRDFHADKP KEKLVTDITY LYFGNCKLYL SSIMNLYNRE 100
IIAYTISDCQ DTDFVLDTLN QLKLPK (SEQ ID NO:19) 126

FIG. 3g

```
MVKKAYSWET KLACIDMKKA GKSNRVIMET LGIKNNSQIY TWMKWYENEE    50
LYRFHQGVGK QYTYGKGLEH LSEVEQLQLQ VDLLKKYRGL IRKSIK        96
(SEQ ID NO:20)
```

FIG. 3h

```
IRYPKASSGD YGTKREIITA NKDKYSISKM CRWLNMPHSS YYYQAVESVS    50
ETEFEETIKR IFLDSESRYG SRKIKICLNN EGITLSRRRI RRIMKRLNLV   100
SVYQKATFKP HSRGKNEAPI PNHLDRQFKQ ERPLQALVTD LTYVRVGNRW   150
AYVCLIIDLY NREIIGLSLG WHKTAELVKQ AIQSIPYALT KVKMFHSDRG   200
KEFDNQLIDE ILEAFGITRS LSQAGYPYDN AVAESTYRAF KIEFVYQETF   250
QLLEELALKT KDYVHWWNYH RIHGSLNYQT PMTKRLIA (SEQ ID NO:21)288
```

FIG. 3i

```
AATTTGAAAG CAGAATTATC TGTAGAAGAT GAGCAATATA CAGCAACAGT TTATGGTAAA    60
N  L  K  A   E  L  S   V  E  D   E  Q  Y  T   A  T  V   Y  G  K
---->
TCTGCTCATG GTTCAACACC ACAAGAAGGT GTTAATGGGG CGACTTATTT AGCTCTTTAT   120
S  A  H  G   S  T  P   Q  E  G   V  N  G  A   T  Y  L   A  L  Y

CTAAGTCAAT TTGATTTTGA AGGTCCTGCT CGTGCTTTCT TAGATGTTAC AGCCAACATT   180
L  S  Q  F   D  F  E   G  P  A   R  A  F  L   D  V  T   A  N  I

ATTCACGAAG ACTTCTCAGG TGAAAAACTT GGAGTAGCTT ATGAAGATGA CTGTATGGGA   240
I  H  E  D   F  S  G   E  K  L   G  V  A  Y   E  D  D   C  M  G

CCATTGAGCA TGAATGCAGG TGTCTTCCAG TTTGATGAAA CTAATGATGA TAATACTATC   300
P  L  S  M   N  A  G   V  F  Q   F  D  E  T   N  D  D   N  T  I

GCTCTTAATT TCCGTTACCC ACAAGGGACA GATGCTAAAA CTATCCAAAC TAAGCTTGAG   360
A  L  N  F   R  Y  P   Q  G  T   D  A  K  T   I  Q  T   K  L  E

AAACTTAACG GAGTTGAAAA AGTGACTCTT TCTGACCATG AACACACACC ACACTATGTA   420
K  L  N  G   V  E  K   V  T  L   S  D  H  E   H  T  P   H  Y  V

CCTATGGACG ATGAATTAGT ATCAACCTTA CTAGCTGTCT ATGAAAAGCA AACTGGTCTT   480
P  M  D  D   E  L  V   S  T  L   L  A  V  Y   E  K  Q   T  G  L

AAAGGACATG AACAGGTTAT TGGTGGTGGG ACATTTGGTC GCTTACTTGA ACGGGGTGTT   540
K  G  H  E   Q  V  I   G  G  G   T  F  G  R   L  L  E   R  G  V

GCATACGGTG CCATGTTCCC AGGAGATGAA AACACTATGC ATCAAGCTAA TGAGTACATG   600
A  Y  G  A   M  F  P   G  D  E   N  T  M  H   Q  A  N   E  Y  M

CCTTTAGAAA ATATTTTCCG TTCGGCTGCT ATCTACGCAG AAGCTATCTA TGAATTAATC   660
```

```
P  L  E  N   I  F  R   S  A  A   I  Y  A  E   A  I  Y   E  L  I

AAATAAAATA ATCCTTAAAC TAAATATGTG ATCAATGATA AAGGGTGGTG AAGACATGAA   720
K .

AGTGTCTTTG CCTCTTTTCA TAAGGTTAGA TTTGGAGACT TTATGACTGA CTTGGAAAAA   780
                                             M  T  D   L  E  K
                                             |---->
ATTATTAAAG CAATAAAAAG TGATTCACAG AATCAAAATT ATACAGAAAA TGGTATTGAT   840
I  I  K  A   I  K  S   D  S  Q   N  Q  N  Y   T  E  N   G  I  D

CCTTTGTTTG CTGCTCCTAA AACAGCTAGG ATCAATATTG TTGGCCAAGC ACCTGGTTTA   900
P  L  F  A   A  P  K   T  A  R   I  N  I  V   G  Q  A   P  G  L

AAAACTCAAG AAGCAAGACT CTATTGGAAA GATAAATCTG GAGATCGTCT ACGCCAGTGG   960
K  T  Q  E   A  R  L   Y  W  K   D  K  S  G   D  R  L   R  Q  W

CTTGGAGTTG ATGAAGAGAC ATTTTACCAT TCTGGAAAAT TTGCTGTTTT ACCTTTAGAT  1020
L  G  V  D   E  E  T   F  Y  H   S  G  K  F   A  V  L   P  L  D

TTTTATTACC CAGGCAAAGG AAAATCAGGA GATTTACCCC CTAGAAAAGG TTTTGCGGAG  1080
F  Y  Y  P   G  K  G   K  S  G   D  L  P  P   R  K  G   F  A  E

AAATGGCACC CTCTTATTTT AAAAGAAATG CCTAATGTTC AATTGACCTT GCTAGTTGGT  1140
K  W  H  P   L  I  L   K  E  M   P  N  V  Q   L  T  L   L  V  G

CAGTATGCTC AGAAATATTA TCTTGGAAGC TCCGCACATA AAAATCTAAC AGAAACAGTT  1200
Q  Y  A  Q   K  Y  Y   L  G  S   S  A  H  K   N  L  T   E  T  V

AAAGCTTACA AAGACTATCT ACCCGATTAT TTACCCCTGG TTCACCCATC ACCGCGAAAT  1260
K  A  Y  K   D  Y  L   P  D  Y   L  P  L  V   H  P  S   P  R  N

CAAATTTGGC TAAAGAAGAA TCCATGGTTT GAAAAAGATC TAATCGTTGA TTTACAAAAG  1320
Q  I  W  L   K  K  N   P  W  F   E  K  D  L   I  V  D   L  Q  K

ATAGTAGCAG ATATTTTAAA AGATTAAGGA TAGGAGTTGG TATGAGAGAT AATCATCTAC  1380
I  V  A  D   I  L  K   D .                   M  R  D   N  H  L  H
                                             |---->
ACACGTATTT TTCCTATGAT TGTCAAACGG CATTTGAGGA CTATATTAAT GGTTTTACAG  1440
 T  Y  F   S  Y  D   C  Q  T  A   F  E  D   Y  I  N   G  F  T  G

GTGAATTTAT CACGACAGAA CATTTTGATT TATCAAATCC TTACACCGGT CAAGACGATG  1500
 E  F  I   T  T  E   H  F  D  L   S  N  P   Y  T  G   Q  D  D  V

TTCCTGATTA TAGTGCTTAT TGTCAAAAAA TAGATTATCT TAATCAGAAA TATGGAAATC  1560
 P  D  Y   S  A  Y   C  Q  K  I   D  Y  L   N  Q  K   Y  G  N  R

GATTTAAAAA AGGAATTGAA ATCGGTTATT TTAAAGATAG GGAATCAGAT ATTTTAGATT  1620
 F  K  K   G  I  E   I  G  Y  F   K  D  R   E  S  D   I  L  D  Y

ATTTAAAAAA TAAAGAATTT GATTTAAAAC TATTGTCAAT CCATCATAAT GGTAGGTATG  1680
 L  K  N   K  E  F   D  L  K  L   L  S  I   H  H  N   G  R  Y  D

ATTATCTGCA AGAAGAAGCT CTGAAAGTAC CAACAAAGGG AGCTTTTAGC AGATTACTTT  1740
 Y  L  Q   E  E  A   L  K  V  P   T  K  G   A  F  S   R  L  L .

AATCGTATGG AATTTGCCAT AGGCCGTGTG GAAGCGCACG TTTTAGCTCA CTTTGATTAT  1800

GGTTTTCGTA AGTTAAACTT AGATGTAGAA GATTTAAAAC CGTTTGAAAC GCAATTGAAG  1860

CGCATTTTCA TAAAGATGTT ATCTAAGGGG TTAGCTTTTG AACTAAATAC CAAATCCCTT  1920
```

```
TATCTATATG GGAATGAAAA ACTTTATCGC TATGCTTTAG AGATACTCAA ACAGCTTGGT   1980
TGTAAACAAT ACTCTATAGG CTCTGACGGT CATATTCCTG AACATTTTTG TTATGAATTT   2040
GATAGACTTC AAGGTCTGCT AAAGGACTAT CAAATTGATG AAAATCATTT GATATGAGGA   2100
AATTTTTGAT AAAAAAGCTA GGCAATATTG CTTAGCTTTT TTGTAATGCT ATTGATAGTT   2160
TTAGTGAAAA TTTCAAAAAA ATAAAGAAAT CATTTACTTG TTGCAAGCGC TTGCGTAAAT   2220
TGTTATGATT TTATTGGTAA CAATTCATTA AAAAAGGAGA ATGATATGAA AAGAAAAGAC   2280
                                                M  K   R  K  D
                                                |---->
TTATTTGGTG ATAAACAAAC TCAATACACG ATTAGAAAGT TAAGTGTTGG AGTAGCTTCA   2340
L  F  G  D   K  Q  T   Q  Y  T   I  R  K  L   S  V  G   V  A  S

GTTACAACAG GGGTATGTAT TTTTCTTCAT AGTCCACAGG TATTTGCTGA AGAAGTAAGT   2400
V  T  T  G   V  C  I   F  L  H   S  P  Q  V   F  A  E   E  V  S

GTTTCTCCTG CAACTACAGC GATTGCAGAG TCGAATATTA ATCAGGTTGA CAACCAACAA   2460
V  S  P  A   T  T  A   I  A  E   S  N  I  N   Q  V  D   N  Q  Q

TCTACTAATT TAAAAGATGA CATAAACTCA AACTCTGAGA CGGTTGTGAC ACCCTCAGAT   2520
S  T  N  L   K  D  D   I  N  S   N  S  E  T   V  V  T   P  S  D

ATGCCGGATA CCAAGCAATT AGTATCAGAT GAAACTGACA CTCAAAAGGG AGTGACAGAG   2580
M  P  D  T   K  Q  L   V  S  D   E  T  D  T   Q  K  G   V  T  E

CCGGATAAGG CGACAAGCCT GCTTGAAGAA AATAAAGGTC CTGTTTCAGA TAAAAATACC   2640
P  D  K  A   T  S  L   L  E  E   N  K  G  P   V  S  D   K  N  T

TTAGATTTAA AAGTAGCACC ATCTACATTG CAAAATACTC CCGACAAAAC TTCTCAAGCT   2700
L  D  L  K   V  A  P   S  T  L   Q  N  T  P   D  K  T   S  Q  A

ATAGGTGCTC CAAGCCCTAC CTTGAAAGTA GCTAATCAAG CTCCACGGAT TGAAAATGGT   2760
I  G  A  P   S  P  T   L  K  V   A  N  Q  A   P  R  I   E  N  G

TACTTTAGGC TACATCTTAA AGAATTGCCT CAAGGTCATC CTGTAGAAAG CACTGGACTT   2820
Y  F  R  L   H  L  K   E  L  P   Q  G  H  P   V  E  S   T  G  L

TGGATATGGG GAGATGTTGA TCAACCGTCT AGTAATTGGC CAAATGGTGC TATCCCTATG   2880
W  I  W  G   D  V  D   Q  P  S   S  N  W  P   N  G  A   I  P  M

ACTGATGCTA AGAAAGATGA TTACGGTTAT TATGTTGATT TTAAATTATC TGAAAAACAA   2940
T  D  A  K   K  D  D   Y  G  Y   Y  V  D  F   K  L  S   E  K  Q

CGAAAACAAA TATCTTTTTT AATTAATAAC AAAGCAGGGA CAAATTTAAG CGGCGATCAT   3000
R  K  Q  I   S  F  L   I  N  N   K  A  G  T   N  L  S   G  D  H

CATATTCCAT TATTACGACC TGAGATGAAC CAAGTTTGGA TTGATGAAAA GTACGGTATA   3060
H  I  P  L   L  R  P   E  M  N   Q  V  W  I   D  E  K   Y  G  I

CATACTTATC AACCCCTCAA AGAAGGGTAT GTCCGTATTA ACTATTTGAG TTCCTCTAGT   3120
H  T  Y  Q   P  L  K   E  G  Y   V  R  I  N   Y  L  S   S  S  S

AACTATGACC ACTTATCAGC ATGGCTCTTT AAAGATGTTG CAACCCCYTC AACAACTTGG   3180
N  Y  D  H   L  S  A   W  L  F   K  D  V  A   T  P  S   T  T  W

CCAGATGGTA GTAATTTTGT GAATCAAGGA CTATATGGAA GGTATATTGA TGTATCACTA   3240
P  D  G  S   N  F  V   N  Q  G   L  Y  G  R   Y  I  D   V  S  L
```

```
AAAACTAACG CCAAAGAGAT TGGTTTTCTA ATCTTAGATG AAAGTAAGAC AGGAGATGCA   3300
K  T  N  A   K  E  I   G  F  L   I  L  D  E   S  K  T   G  D  A

GTGAAAGTTC AACCCAACGA CTATGTTTTT AGAGATTTAG CTAACCATAA CCAAATTTTT   3360
V  K  V  Q   P  N  D   Y  V  F   R  D  L  A   N  H  N   Q  I  F

GTAAAAGATA AGGATCCAAA GGTTTATAAT AATCCTTATT ACATTGATCA AGTGCAGCTA   3420
V  K  D  K   D  P  K   V  Y  N   N  P  Y  Y   I  D  Q   V  Q  L

AAGGATGCCC AACAAATTGA TTTAACAAGT ATTCAAGCAA GTTTTACAAC TCTAGATGGG   3480
K  D  A  Q   Q  I  D   L  T  S   I  Q  A  S   F  T  T   L  D  G

GTAGATAAAA CTGAAATTTT AAAAGAATTG AAAGTGACTG ATAAAAATCA AAATGCTATA   3540
V  D  K  T   E  I  L   K  E  L   K  V  T  D   K  N  Q   N  A  I

CAAATTTCTG ATATCACTCT CGATACTAGT AAATCTCTTT TAATAATCAA AGGCGACTTT   3600
Q  I  S  D   I  T  L   D  T  S   K  S  L  L   I  I  K   G  D  F

AATCCTAAAC AAGGTCATTT CAACATATCT TATAATGGTA ACAATGTCAT GACAAGGCAA   3660
N  P  K  Q   G  H  F   N  I  S   Y  N  G  N   N  V  M   T  R  Q

TCTTGGGAAT TTAAAGACCA ACTTTATGCT TATAGTGGAA ATTTAGGTGC AGTTCTCAAT   3720
S  W  E  F   K  D  Q   L  Y  A   Y  S  G  N   L  G  A   V  L  N

CAAGATGGTT CAAAAGTTGA AGCCAGCCTC TGGTCACCGA GTGCTGATAG TGTCACTATG   3780
Q  D  G  S   K  V  E   A  S  L   W  S  P  S   A  D  S   V  T  M

ATTATTTATG ACAAAGATAA CCAAAACAGG GTTGTAGCGA CTACCCCCCT TGTGAAAAAT   3840
I  I  Y  D   K  D  N   Q  N  R   V  V  A  T   T  P  L   V  K  N

AATAAAGGTG TTTGGCAGAC GATACTTGAT ACTAAATTAG GTATTAAAAA CTATACTGGT   3900
N  K  G  V   W  Q  T   I  L  D   T  K  L  G   I  K  N   Y  T  G

TACTATTATC TTTACGAAAT AAAAAGAGGT AAGGATAAGG TTAAGATTTT AGATCCTTAT   3960
Y  Y  Y  L   Y  E  I   K  R  G   K  D  K  V   K  I  L   D  P  Y

GCAAAGTCAT TAGCAGAGTG GGATAGTAAT ACTGTTAATG ATGATATTAA AACGGCTAAA   4020
A  K  S  L   A  E  W   D  S  N   T  V  N  D   D  I  K   T  A  K

GCAGCTTTTG TAAATCCAAG TCAACTTGGA CCTCAAAATT TAAGTTTTGC TAAAATTGCT   4080
A  A  F  V   N  P  S   Q  L  G   P  Q  N  L   S  F  A   K  I  A

AATTTTAAAG GAAGACAAGA TGCTGTTATA TACGAAGCAC ATGTAAGAGA CTTCACTTCT   4140
N  F  K  G   R  Q  D   A  V  I   Y  E  A  H   V  R  D   F  T  S

GATCGATCTT TGGATGGAAA ATTAAAAAAT CAATTTGGTA CCTTTGCAGC CTTTTCAGAG   4200
D  R  S  L   D  G  K   L  K  N   Q  F  G  T   F  A  A   F  S  E

AAACTAGATT ATTTACAGAA ATTAGGAGTT ACACACATTC AGCTTTTACC GGTATTGAGT   4260
K  L  D  Y   L  Q  K   L  G  V   T  H  I  Q   L  L  P   V  L  S

TATTTTTATG TTAATGAAAT GGATAAGTCA CGCTCAACAG CTTACACTTC CTCAGACAAT   4320
Y  F  Y  V   N  E  M   D  K  S   R  S  T  A   Y  T  S   S  D  N

AATTACAATT GGGGCTATGA CCCACAGAGC TATTTTGCTC TTTCTGGGAT GTATTCAGAG   4380
N  Y  N  W   G  Y  D   P  Q  S   Y  F  A  L   S  G  M   Y  S  E

AAACCAAAAG ATCCATCAGC ACGTATCGCC GAATTAAAAC AATTAATACA TGATATTCAT   4440
K  P  K  D   P  S  A   R  I  A   E  L  K  Q   L  I  H   D  I  H
```

```
AAACGTGGCA TGGGGGTTAT ACTTGATGTC GTCTATAATC ACACTGCAAA AACTTATCTC   4500
K  R  G  M   G  V  I   L  D  V   V  Y  N  H   T  A  K   T  Y  L

TTTGAGGATA TAGAACCTAA TTATTATCAC TTTATGAATG AAGATGGTTC ACCAAGAGAA   4560
F  E  D  I   E  P  N   Y  Y  H   F  M  N  E   D  G  S   P  R  E

AGTTTTGGAG GGGGACGTTT AGGAACCACT CATGCAATGA GTCGTCGTGT TTTGGTTGAT   4620
S  F  G  G   G  R  L   G  T  T   H  A  M  S   R  R  V   L  V  D

TCCATTAAAT ATCTTACAAG TGAATTTAAA GTTGATGGTT TCCGTTTTGA TATGATGGGA   4680
S  I  K  Y   L  T  S   E  F  K   V  D  G  F   R  F  D   M  M  G

GATCATGATG CGGCTGCGAT TGAATTAGCT TATAAAGAAG CTAAAGCTAT TAATCCTAAT   4740
D  H  D  A   A  A  I   E  L  A   Y  K  E  A   K  A  I   N  P  N

ATGATTATGA TTGGTGAGGG CTGGAGAACA TTCCAAGGCG ATCAAGGTCA GCCGGTTAAA   4800
M  I  M  I   G  E  G   W  R  T   F  Q  G  D   Q  G  Q   P  V  K

CCAGCTGACC AAGATTGGAT GAAGTCAACC GATACAGTTG GCGTCTTTTC AGATGATATT   4860
P  A  D  Q   D  W  M   K  S  T   D  T  V  G   V  F  S   D  D  I

CGTAATAGCT TGAAATCTGG TTTTCCAAAT GAAGGTACTC CAGCTTTCAT CACAGGTGGC   4920
R  N  S  L   K  S  G   F  P  N   E  G  T  P   A  F  I   T  G  G

CCACAATCTT TACAAGGTAT TTTTAAAAAT ATCAAAGCAC AACCTGGGAA TTTTGAAGCA   4980
P  Q  S  L   Q  G  I   F  K  N   I  K  A  Q   P  G  N   F  E  A

GATTCGCCAG GAGATGTGGT GCAGTATATT GCTGCACATG ATAACCTTAC CTTGCATGAT   5040
D  S  P  G   D  V  V   Q  Y  I   A  A  H  D   N  L  T   L  H  D

GTGATTGCAA AATCAATT    (SEQ ID NO:22)                               5058
V  I  A  K   S  I .
```

FIG. 4a

```
NLKAELSVED EQYTATVYGK SAHGSTPQEG VNGATYLALY LSQFDFEGPA   50
RAFLDVTANI IHEDFSGEKL GVAYEDDCMG PLSMNAGVFQ FDETNDDNTI  100
ALNFRYPQGT DAKTIQTKLE KLNGVEKVTL SDHEHTPHYV PMDDELVSTL  150
LAVYEKQTGL KGHEQVIGGG TFGRLLERGV AYGAMFPGDE NTMHQANEYM  200
PLENIFRSAA IYAEAIYELI K  (SEQ ID NO:23)                 221
```

FIG. 4b

```
MTDLEKIIKA IKSDSQNQNY TENGIDPLFA APKTARINIV GQAPGLKTQE   50
ARLYWKDKSG DRLRQWLGVD EETFYHSGKF AVLPLDFYYP GKGKSGDLPP  100
RKGFAEKWHP LILKEMPNVQ LTLLVGQYAQ KYYLGSSAHK NLTETVKAYK  150
DYLPDYLPLV HPSPRNQIWL KKNPWFEKDL IVDLQKIVAD ILKD        194
(SEQ ID NO:24)
```

FIG. 4c

MRDNHLHTYF SYDCQTAFED YINGFTGEFI TTEHFDLSNP YTGQDDVPDY 50
SAYCQKIDYL NQKYGNRFKK GIEIGYFKDR ESDILDYLKN KEFDLKLLSI 100
HHNGRYDYLQ EEALKVPTKG AFSRLL (SEQ ID NO:25) 126

FIG. 4d

MKRKDLFGDK QTQYTIRKLS VGVASVTTGV CIFLHSPQVF AEEVSVSPAT 50
TAIAESNINQ VDNQQSTNLK DDINSNSETV VTPSDMPDTK QLVSDETDTQ 100
KGVTEPDKAT SLLEENKGPV SDKNTLDLKV APSTLQNTPD KTSQAIGAPS 150
PTLKVANQAP RIENGYFRLH LKELPQGHPV ESTGLWIWGD VDQPSSNWPN 200
GAIPMTDAKK DDYGYYVDFK LSEKQRKQIS FLINNKAGTN LSGDHHIPLL 250
RPEMNQVWID EKYGIHTYQP LKEGYVRINY LSSSSNYDHL SAWLFKDVAT 300
PSTTWPDGSN FVNQGLYGRY IDVSLKTNAK EIGFLILDES KTGDAVKVQP 350
NDYVFRDLAN HNQIFVKDKD PKVYNNPYYI DQVQLKDAQQ IDLTSIQASF 400
TTLDGVDKTE ILKELKVTDK NQNAIQISDI TLDTSKSLLI IKGDFNPKQG 450
HFNISYNGNN VMTRQSWEFK DQLYAYSGNL GAVLNQDGSK VEASLWSPSA 500
DSVTMIIYDK DNQNRVVATT PLVKNNKGVW QTILDTKLGI KNYTGYYYLY 550
EIKRGKDKVK ILDPYAKSLA EWDSNTVNDD IKTAKAAFVN PSQLGPQNLS 600
FAKIANFKGR QDAVIYEAHV RDFTSDRSLD GKLKNQFGTF AAFSEKLDYL 650
QKLGVTHIQL LPVLSYFYVN EMDKSRSTAY TSSDNNYNWG YDPQSYFALS 700
GMYSEKPKDP SARIAELKQL IHDIHKRGMG VILDVVYNHT AKTYLFEDIE 750
PNYYHFMNED GSPRESFGGG RLGTTHAMSR RVLVDSIKYL TSEFKVDGFR 800
FDMMGDHDAA AIELAYKEAK AINPNMIMIG EGWRTFQGDQ GQPVKPADQD 850
WMKSTDTVGV FSDDIRNSLK SGFPNEGTPA FITGGPQSLQ GIFKNIKAQP 900
GNFEADSPGD VVQYIAAHDN LTLHDVIAKS I (SEQ ID NO:26) 931

FIG. 4e

```
AATTCAAAGT TTGACAGAAG GTCAACTTCG TTCTGATATC CCTGAGTTCC GTGCTGGTGA    60
 I  Q  S   L  T  E  G   Q  L  R   S  D  I   P  E  F  R   A  G  D
 ---->
TACTGTACGT GTTCACGCTA AAGTTGTTGA AGGTACTCGC GAACGTATTC AGATCTTTGA   120
 T  V  R   V  H  A  K   V  V  E   G  T  R   E  R  I  Q   I  F  E

AGGTGTTGTT ATCTCACGTA AAGGTCAAGG AATCTCAGAA ATGTACACAG TACGTAAAAT   180
 G  V  V   I  S  R  K   G  Q  G   I  S  E   M  Y  T  V   R  K  I

TTCTGGTGGT ATCGGTGTAG AGCGTACATT CCCAATTCAC ACTCCTCGTG TTGATAAAAT   240
 S  G  G   I  G  V  E   R  T  F   P  I  H   T  P  R  V   D  K  I

CGAAGTTGTT CGTTATGGTA AAGTACGTCG TGCTAAACTT TACTACTTAC GCGCATTGCA   300
 E  V  V   R  Y  G  K   V  R  R   A  K  L   Y  Y  L  R   A  L  Q

AGGTAAAGCT GCACGTATTA AAGAAATCCG TCGTTAATTT TGATGATCAG ATTTTAAAAA   360

TGCTTGGTTG TTTGAGGATA GTAACTATGT TTTAAAACTG GACAACCAAG ACGTAAAAAA   420

TCTGCCTGTG GGCAGTTTTT TTACTAGGTC CCCTTAGTTC AATGGATATA ACAACTCCCT   480
                                              .  H  I  Y   C  S  G

CCTAAGGAGT AATTGCTGGT TCGATTCCGG CAGGGGACAT ATTCATTGCA TGTAAATAGC   540
 G  L  S   Y  N  S  T   R  N  R   C  P  V   Y  E  N  C   T  F  L

GGTTTAGAGC TATTTTGCCC CAAATTTCTC TGATTAAGTT TATCGTTCCT ATCTTTTTGT   600
 P  K  S   S  N  Q  G   L  N  R   Q  N  L   K  D  N  R   D  K  Q

TCTTGTAATT GATGTGCGTA AACTTCTAAA GTGATATTTA AATTCTCGTG ATCTAAAACT   660
 E  Q  L   Q  H  A  Y   V  E  L   T  I  N   L  N  E  H   D  L  V

TGAGAGATGG AAATTAGATA GCTTGCAAAT GTATGCCTGA GAGAGTGCAC TCGTACCTCG   720
 Q  S  I   S  I  L  Y   S  A  F   T  H  R   L  S  H  V   R  V  E

CGACCAGTTA TTTTTCGGAT AGTTTTATTG ACTGCATTAT TTGAAAGTTT GTCGAATAAT   780
 R  G  T   I  K  R  I   T  K  N   V  A  N   N  S  L  K   D  F  L

CTGTCGTTTT TATTTTTTGT AAATTCATGC AAAAAAAATA ATGTATCATT GTCAATTGGT   840
 R  D  N   K  N  K  T   F  E  H   L  F  F   L  T  D  N   D  I  P

ATATTTCTGA TACTACTTTT GTTTTTTGTT GGCAGGTATC TTTGGTTGAA ATGATAATCC   900
 I  N  R   I  S  S  K   N  K  T   P  L  Y   R  Q  N  F   H  Y  D

CAAGTTTTAT TAATTGATAA ATATTTGTTA GTGTAATCAA TATCATTAAC TGTTAAACCT   960
 W  T  K   N  I  S  L   Y  K  N   T  Y  D   I  D  N  V   T  L  G

AAACATTCAG CGAAGCGCAT GCCAGTTTTA GCGATGAGGT ATAACGCTGC ATACGATTGA  1020
 L  C  E   A  F  R  M
              <----|
TGTTGTGATT TTTCTTTACA AATTTTTATC AAGCGTAAGT ATTCATTGGT TTCAAGAAAT  1080

TTTATCTCTA TTTACGCCCC TTATTTTTTG CTTTAACCTT AGTGAATAAA CAAAAATTTT  1140

TTTCTATATA TCCCTCGTGA ACAGCCATGG ATACGCAGGC TTTTACATGT ATGTTAAAAC  1200

GCTTTACTGT ATCTTGCACA TGCGTTTGAC TATAATGATT TATGACTTGT TGATATTTAG  1260
```

TGGAAGTAAT ATTGCAAAGT AATATATTTC CTATTATATG TTTATACGAT ATTCGATATT 1320
CCCACCCGTT GTCGCGTTTA CGGAAATACG CCATTGATAT ACTCCACATT AGCTAAAGAA 1380
CAGGGTGTTC AAGGCTACCT TGATGGAAAA GGCTCTCTTA GAGATATTTG TAAATGGTAT 1440
GATATCTCAA GTCGCTCTGT TCTCCAAAAG TGGATAAAAC GGTATACTAG TGGTGAAGAC 1500
TTGAAAGCCA CTAGTAGAGG ATATAGCCGT ATGAAACAAG GAAGGCAAGC CACATTTGAA 1560
GAACGTGTAG AGATTGTTAA CTACACCATT GCCCATGGGA AAGACTATCA AGCAGCTATT 1620
GAGAAGTTTG GTGTTTCCTA CCAACAAATT TATTCTTGGG TGCGTAAGCT TGAGAAGAAT 1680
GGCTCACAAG GTTTGGTTGA TAGACGTGTG AAAGGGTTGG AGAGTAGGCC TGATTTAACC 1740
GAGATTGAGC AACTTTAACT CAAGATTAAA CAATTGGAGG AACGTAATCG TCTCTTAGAA 1800
ATCGAGGTTA GTTTACTAAA AAAGTTAGAA GACATCAAAC GAGGAAACAG ACGGTAAGAC 1860
TAGGTAAGCA TTTAGCGGAG TTCCAAGTAA TCAAGAATTA TTACGATGAG GAATCTAATG 1920
TGCCTATTCA GGCCTTATGC CAACTCTTGA AGGGGTCTCG TTCAGGCTAT TACAAGTGGC 1980
TCAATCGTCA AAAAACAGAT TTTGAGACAA AAAATACAAA GCTAATGGCT AAAATCAAGG 2040
AACTTCGTAG ACTCTACAAT GGTATCTTAG GTTATCGCCG TATGACAACA TTTATTAATC 2100
GTCAACTTGG GACAACTTAA AACAAGAAAC GGATTCGTTG ATTGATGAAC ATTCTGGGGA 2160
TTAGTTCAGT CATTCGTCGT GTTAGCCATG CTTGTACAAA AGCTGGTGAC AGATTTTACG 2220
AAGAAAATAT TCTTAATCGT GAATTTACAG CCACAGCTCA TAACCAGAAA TGGTGCACAG 2280
ATGTCACCTA TCTTCAATAC GGTCTGGGAG CTAAAGCTTA TCTCAGTGCG ATTAAAGACC 2340
TGTATAACGG TTCTATTATC GCTTATGAGA TTAGTCACAA CAATGAAATC CACTTGTTAT 2400
GAAGACCATT AAAAAGGGGC TAGAGCTCAA TCCAGGAGCC ACACCTATCA TCCATAGCGA 2460
TTGAGGTAGT CAATATACTT CCAAAGAATA CCGTTATATC ATACAACAAG CTGGTCTGAC 2520
CTTATCCATG TCCCGGATTG GCAAATGTAT TGATAATGCA CCAACTGAAA GTTTCTTTGG 2580
GTTTTTCAAG ACTGAGTCTT ACCACCTTAA GAAATACAAC TCTTATGATG AGTTGGTCAA 2640
TGATGTGGCA CGTTATATCG AATTCTACAA CACACAACGT TATCAATCAA AATTAAACAA 2700
CCTGACTCCT CTAGAATTCA GGAATCAGGT TGCATAACTT ATCTTTTATT ATTTGACTGT 2760
CTACTTGACA GGGAGCCGTT CAGATTGCTT AACCTTTCTA AATTTGCTAA AATAGCTACA 2820
AGAAAACGAG CCATTTAATG CTTATTTCTT ATACTGTCTT GCCTCACGCT CTCCTCGACC 2880
AAAAATTGAG CGTGAGGCTT TTTGTTTCAT TAAACGATGA TATTTCCATA TTCATCAGTT 2940
TGTTTTCCGA GAGCCATCAA AGCTTCGATA AGGTCGATAA TTCCAGGAAT AAAGGTAATA 3000
CTAAAAATAA TATATAAAAA AACCTGGCCT ATTTTTCCTG CGTAAAATTT ATGCGCTCCA 3060
ATGCCGCCCA AAAGAACGTT AATAAAACAT AAACTACTAT GTTAGCATAA GACTTTATTT 3120

```
TTACAACTGA ATTTCATATA AATGGATTAG AGTAAGGGAT AAAAGAAATT AGCATAGCTC   3180
TTTTGAAAAT AAAAAAATTA ATATAATATG GAAAAAATTT TATTTCATAA ACGTTTCATA   3240
AAAGGTATGT AATCTAGTAT TTAGGCAACA CTATTTTGTC ACTGGTGTCT AGTAACTTAT   3300
AGATTGATAA TTTTACTAGT AAACGTAATT CTTCGCTTTA AGAGTTAAAT GTCTATTTAT   3360
TGTAAGCTAA ATTGGGAGGT GAACTTATGT AAAATTAGAT AGGTACTGTC AAGTACGGGA   3420
TGATTATTGA AACAGCCAGT ATGCATCATA AAATCTGTAT TGCTTAATAA CTATTTCCTT   3480
AACCAGACAT CAGTTCATTG TTTATCATCG CTACCCTAAG TCTAGTTTTT TCAATAGAGC   3540
ATTAGGTAGT TTTTGATAAT AAAACTATAT AAACATGAGA ATTAGATTTC GTATTGCATT   3600
CTTCATAATG AGTTATTTGA GATTTTCCTT TGAATAAATA GATACGAAAT TCAGTAACTT   3660
CATATATAAA CGGCTCTATC ATTGAGATAG TTTGTCAAAT GAAGAAATTT TTAATGGAAA   3720
TAGTTTTAAA AACATTAGTT GTAGGCGATG TAAAAATATT AATCCAGTGG ATGCAATAGT   3780
TGCGGAGTAA AAATAGAGAG GAGTAATTAG GAAGTGATAA AAAATGCTAT AGCATATATT   3840
ACCAGAAAAA AAAATAGAAC ACTTATTATA TTTGCTATTT TAACAATTGT TCTTTCTTGC   3900
TTGTATTCAT GTTTAACAAT AATGAAATCA AGTAATGAAA TAGAAAAGGC TTTATATGAA   3960
                       M  K  S   S  N  E  I   E  K  A   L  Y  E
                       |---->
AGTTCTAATT CTTCAATATC AATTACAAAA AAAGATGGTA AATATTTTAA TATTAATCAA   4020
S  S  N  S   S  I  S   I  T  K   K  D  G  K   Y  F  N   I  N  Q

TTTAAGAATA TTGAAAAAAT AAAAGAGGTT GAAGAAAAAA TATTTCAATA TGATGGATTA   4080
F  K  N  I   E  K  I   K  E  V   E  E  K  I   F  Q  Y   D  G  L

GCAAAATTGA AAGATCTTAA AGTAGTTAGT GGTGAGCAAA GTATAAATAG AGAAGATTTA   4140
A  K  L  K   D  L  K   V  V  S   G  E  Q  S   I  N  R   E  D  L

TCTGACGAAT TTAAAAATGT TGTTTCACTA GAAGCTACAA GTAATACTAA AAGAAATCTT   4200
S  D  E  F   K  N  V   V  S  L   E  A  T  S   N  T  K   R  N  L

TTATTTAGTA GTGGAGTATT TAGTTTTAAA GAAGGAAAAA ATATAGAAGA AAATGATAAG   4260
L  F  S  S   G  V  F   S  F  K   E  G  K  N   I  E  E   N  D  K

AATTCAATTC TTGTTCATGA AGAATTTGCT AAACAAAACA AACTAAAATT GGGTGATGAA   4320
N  S  I  L   V  H  E   E  F  A   K  Q  N  K   L  K  L   G  D  E

ATTGATCTTG AATTACTAGA TACGGAAAAA AGTGGAAAAA TAAAAAGTCA TAAATTTAAA   4380
I  D  L  E   L  L  D   T  E  K   S  G  K  I   K  S  H   K  F  K

ATTATAGGAA TCTTTTCTGG TAAAAAACAG GAAACATATA CAGGATTATC ATCTGATTTT   4440
I  I  G  I   F  S  G   K  K  Q   E  T  Y  T   G  L  S   S  D  F

AGCGAAAATA TGGTTTTTGT AGATTATTCA ACTAGCCAAG AAATATTAAA TAAATCAGAG   4500
S  E  N  M   V  F  V   D  Y  S   T  S  Q  E   I  L  N   K  S  E

AATAATAGAA TTGCAAATAA AATTTTAATG TATTCTGGTA GTTTAGAATC TACAGAGCTT   4560
N  N  R  I   A  N  K   I  L  M   Y  S  G  S   L  E  S   T  E  L

GCCTTAAACA AATTGAAAGA CTTTAAAATT GATAAGTCAA AGTATTCTAT TAAGAAAGAT   4620
```

```
A  L  N  K   L  K  D   F  K  I   D  K  S  K   Y  S  I   K  K  D

AATAAAGCAT TCGAAGAGTC TTTAGAGTCA GTGAGTGGAA TAAAACATAT AATTAAAATA  4680
N  K  A  F   E  E  S   L  E  S   V  S  G  I   K  H  I   I  K  I

ATGACTTATT CGATTATGTT AGGTGGAATA GTTGTTCTTT CATTAATCTT GATTCTATGG  4740
M  T  Y  S   I  M  L   G  G  I   V  V  L  S   L  I  L   I  L  W

TTAAGAGAAA GAATTTATGA AATAGGTATA TTTTTATCTA TTGGAACAAC TAAGATACAA  4800
L  R  E  R   I  Y  E   I  G  I   F  L  S  I   G  T  T   K  I  Q

ATTATAAGGC AATTTATATT TGAGTTAATA TTCATATCAA TACCAAGTAT AATATCCTCC  4860
I  I  R  Q   F  I  F   E  L  I   F  I  S  I   P  S  I   I  S  S

TTATTTTTAG GGAATCTACT ATTAAAAGTA ATTGTAGAAG GATTTATTAA CTCAGAGAAC  4920
L  F  L  G   N  L  L   L  K  V   I  V  E  G   F  I  N   S  E  N

TCAATGATTT TCGGTGGAAG TTTAATAAAT AAAAGCAGTT TTATGTTAAA CATAACAACA  4980
S  M  I  F   G  G  S   L  I  N   K  S  S  F   M  L  N   I  T  T

CTTGCAGAAA GTTATTTAAT ATTAATAAGT ATTATTGTTT TATCAGTTGT AATGGCCTCT  5040
L  A  E  S   Y  L  I   L  I  S   I  I  V  L   S  V  V   M  A  S

TCATTAATAT TATTTAAGAA ACCACAAGAA ATATTATCAA AAATAAGTTA GGAGCAAATA  5100
S  L  I  L   F  K  K   P  Q  E   I  L  S  K   I  S .

ATGGATATAT TAGAAATAAA GAATGTAAAT TACAGTTACG CAAATTCTAA AGAAAAAGTT  5160
M  D  I  L   E  I  K   N  V  N   Y  S  Y  A   N  S  K   E  K  V
|---->
TTGTCAGGAG TAAATCAAAA ATTTGAACTT GGAAAGTTTT ATGCGATAGT AGGGAAGTCA  5220
L  S  G  V   N  Q  K   F  E  L   G  K  F  Y   A  I  V   G  K  S

GGAACAGGAA AATCCACACT TCTTTCCTTA CTTGCAGGAC TTGATAAAGT TCAAACAGGA  5280
G  T  G  K   S  T  L   L  S  L   L  A  G  L   D  K  V   Q  T  G

AAAATCTTGT TTAAGAATGA AGATATAGAA AAGAAAGGAT ATAGTAATCA CAGAAAAAAT  5340
K  I  L  F   K  N  E   D  I  E   K  K  G  Y   S  N  H   R  K  N

AATATATCTT TGGTATTTCA AAATTATAAT TTAATAGATT ATTTATCGCC GATTGAAAAT  5400
N  I  S  L   V  F  Q   N  Y  N   L  I  D  Y   L  S  P   I  E  N

ATTAGACTAG TAAATAAATC AGTAGATGAG AGTATCTTGT TCGAATTAGG TTTAGATAAA  5460
I  R  L  V   N  K  S   V  D  E   S  I  L  F   E  L  G   L  D  K

AAACAAATAA AAAGAAATGT TATGAAATTA TCTGGTGGTC AGCAACAAAG GGTAGCTATT  5520
K  Q  I  K   R  N  V   M  K  L   S  G  G  Q   Q  Q  R   V  A  I

GCTAGGGCAC TGGTATCAGA TGCCCCAATA ATACTAGCTG ATGAGCCTAC CGGTAACCTA  5580
A  R  A  L   V  S  D   A  P  I   I  L  A  D   E  P  T   G  N  L

GACAGTGTTA CTGCTGGAGA AATAATT    (SEQ ID NO:27)                    5607
D  S  V  T   A  G  E   I  I .
```

FIG. 5a

IQSLTEGQLR SDIPEFRAGD TVRVHAKVVE GTRERIQIFE GVVISRKGQG 50
ISEMYTVRKI SGGIGVERTF PIHTPRVDKI EVVRYGKVRR AKLYYLRALQ 100
GKAARIKEIR R (SEQ ID NO:28) 111

FIG. 5b

MRFAECLGLT VNDIDYTNKY LSINKTWDYH FNQRYLPTKN KSSIRNIPID 50
NDTLFFLHEF TKNKNDRLFD KLSNNAVNKT IRKITGREVR VHSLRHTFAS 100
YLISISQVLD HENLNITLEV YAHQLQEQKD RNDKLNQRNL GQNSSKPLFT 150
CNEYVPCRNR TSNYSLGGSC YIH (SEQ ID NO:29) 173

FIG. 5c

MKSSNEIEKA LYESSNSSIS ITKKDGKYFN INQFKNIEKI KEVEEKIFQY 50
DGLAKLKDLK VVSGEQSINR EDLSDEFKNV VSLEATSNTK RNLLFSSGVF 100
SFKEGKNIEE NDKNSILVHE EFAKQNKLKL GDEIDLELLD TEKSGKIKSH 150
KFKIIGIFSG KKQETYTGLS SDFSENMVFV DYSTSQEILN KSENNRIANK 200
ILMYSGSLES TELALNKLKD FKIDKSKYSI KKDNKAFEES LESVSGIKHI 250
IKIMTYSIML GGIVVLSLIL ILWLRERIYE IGIFLSIGTT KIQIIRQFIF 300
ELIFISIPSI ISSLFLGNLL LKVIVEGFIN SENSMIFGGS LINKSSFMLN 350
ITTLAESYLI LISIIVLSVV MASSLILFKK PQEILSKIS 389
(SEQ ID NO:30)

FIG. 5d

MDILEIKNVN YSYANSKEKV LSGVNQKFEL GKFYAIVGKS GTGKSTLLSL 50
LAGLDKVQTG KILFKNEDIE KKGYSNHRKN NISLVFQNYN LIDYLSPIEN 100
IRLVNKSVDE SILFELGLDK KQIKRNVMKL SGGQQQRVAI ARALVSDAPI 150
ILADEPTGNL DSVTAGEII (SEQ ID NO:31) 169

FIG. 5e

CATATGACAA TATTTTTCAA AGTCTACATC ACTTACTCGC CTGTCGTGGA AAATCTGGCA 60
ATACATTAAT CGACCAATTA GTTGCTGATG GTTTACTTCA TGCAGATAAT CACTACCATT 120
TTTTCAATGG GAAGTCTCTG GCCACTTTCA ATACTAACCA ATTGATTCGC GAAGTTGTCT 180
ATGTTGAAAT ATCCTTAGAT ACTATGTCTA GTGGTGAACA TGATTTAGTA AAAGTTAACA 240
TTATCAGACC CACTACCGAG CATACTATCC CCACGATGAT GACAGCTAGC CCCTATCATC 300
AAGGTATCAA TGATCCTGCC GCAGACCAAA AAACATACCA AATGGAGGGT GCGCTAGCAG 360
TTAAACAGCC TAAACACATA CAAGTTGACA CAAAACCATT TAAAGAAGAA GTAAAACATC 420
CTTCAAAATT ACCCATCAGC CCTGCAACTG AAAGCTTCAC ACACATTGAC AGTTATAGTC 480
TCAATGACTA TTTTCTTTCT CGTGGTTTTG CTAATATATA CGTTTCAGGT GTGGGTACTG 540
CTGGCTCTAC GGGTTTCATG ACCAGTGGGG ATTACCAACA AATACAAAGC TTTAAAGCAG 600
TCATTGATTG GTTAAATGGT AAGGTTACTG CATTCACAAG TCATAAACGA GATAAACAAG 660
TCAAGGCTGA TTGGTCAAAC GGCCTTGTAG CAACCACAGG TAAATCTTAT CTCGGTACCA 720
TGTCAACTGG TTTAGCAACA ACTGGCGTTG AGGGGCTGAA AGTCATTATC GCTGAAGCCG 780
CAATCTCCAC ATGGTATGAT TATTATCGAG AAAATGGGCT TGTGTGTAGT CCAGGCGGCT 840
ACCCCGGTGA AGATTTAGAC GTTTTAACAG AATTAACATA CTCACGAAAC CTCTTAGCTG 900
GTGATTACAT CAAAAACAAC GATTGCTATC AAGCATTGTT AAATGAACAA TCAAAAGCAA 960
TTGACCGTCA AAGTGGGGAT TACAACCAAT ACTGGCATGA CCGTAATTAC CTAACTCACG 1020
TCAATAATGT CAAAAGTCGA GTAGTTTACA CTCATGGACT ACAGGATTGG AATGTTAAGC 1080
CAAGACATGT CTACAAAGTT TTCAATGCAT TGCCTCAAAC CATCAAAAAA CACCTTTTTT 1140
TACATCAAGG TCAACATGTG TATATGCATA ATTGGCAGTC GATTGATTTT CGTGAAAGCA 1200
TGAATGCCTT ACTAAGCCAA GAACTACTTG GCATTGACAA TCATTTCCAA TTAGAAGAGG 1260
TCATTTGGCA AGATAATACT ACTGAGCAAA CTTGGCAAGT TTTAGATGCT TTCGGAGGAA 1320
ACCATCAAGA GCAAATTGGT TTAGGTGATA GTAAAAAACT TATTGATAAC CATTATGACA 1380
AAGAAGCCTT TGATACTTAT TGTAAAGACT TCAATGTGTT CAAAAATGAT CTTTTCAAGG 1440
GAAATAATAA AACCAATCAA ATCACTATTA ATCTTCCTCT AAAGAAAAAT TATCTCCTGA 1500
ATGGACAGTG CAAACTCCAT CTACGTGTTA AAACTAGTGA CAAAAAGGCC ATTTTATCAG 1560
CCCAAATCTT AGACTATGGT CCTAAAAAAC GATTCAAAGA TACACCAACC ATCAAATTCT 1620
TAAACAGCCT TGATAATGGT AAAAATTTTG CCAGAGAAGC TTTACGTGAA CTCCCGTTTA 1680
CTAAAGATCA TTATCGTGTC ATCAGTAAAG GTGTCTTGAA CCTTCAAAAT CGTACAGACT 1740
TACTTACAAT TGAGGCTATC GAGCCAGAAC AATGGTTTGA TATCGAGTTT AGCCTCCAAC 1800
CAAGTATATA TCAATTGAGT AAAGGTGATA ATCTAAGGAT TATCCTTTAT ACAACTGATT 1860
TTGAACATAC CATTCGAGAT AATGCTAGTT ACTCTATAAC AGTAGATTTG AGTCAATCTT 1920
ATTTAACTAT CCCAACTAAT CAAGGAAATT AACTTATGAA ACTTCTTACT AAAGAACGGT 1980
TTGATGATTC TCAACACTTT TGGTACCAGA TCAATTTATT ACAAGAGAGT AACTTCGGAG 2040
CAGTTTTTGA CCATGATAAT AAAAACATTC CACAGGTTGT TGCAACTATT GTTGATGATT 2100
TACAAGGTTC CGGAAGTTCG AATCATTTCT GGTATTTTGG CAATACTACT GATACTTCCA 2160
TCCTTATGAT TGCTCATTTA AATCGAAAAT TCTATATTCA GGTTAATTTA AAGGACTTTG 2220
ACTTTGCACT CAATTTAATA GCTATAAATA ATTGGAAGAG TCTCCTCCAA ACTCAACTTG 2280
AAGCTCTAAA CGATACCCTA GCAATATTTC AATAAATAAG GTAGAATGGA GTGACAAAGC 2340
AACGCGAGGG AGACTGATTA ATGTCATCTT ATTGGAATAA CTATCCTGAA CTTAAAAAAA 2400

```
ATATTGATGA AACCAATCAA CTAATTCAAG AAAGAATACA GGTCAGAAAT AAAGATATTG  2460
AAGCGGCGCT AAGCCAACTC ACAGCTGCGG GAGGAAAACA GCTCAGACCA GCATTCTTTT  2520
ACCTTTTTTC TCAACTTGGT AATAAGGAGA ATCAAGATAC TCAGCAACTA AAGAAAATCG  2580
CTGCTTCTTT AGAAATCCTT CACGTTGCTA CATTAATCCA TGATGATGTC ATTGATGACT  2640
CACCACTAAG ACGTGGAAAT ATGACCATTC AAAGCAAGTT TGGCAAAGAC ATCGCAGTTT  2700
ATACTGGGGA TTTACTTTTC ACAGTCTTTT TCGATCTTAT TTTAGAATCT ATGACTGATA  2760
CACCATTTAT GAGGATTAAT GCAAAATCTA TGCGTAAAAT TCTCATGGGA GAATTGGACC  2820
AGATGCACCT TCGTTACAAT CAACAACAAG GTATCCATCA CTATTTACGT GCGATTTCAG  2880
GTAAGACAGC CGAACTCTTT AAATTAGCTA GCAAAGAAGG AGCTTACTTT GGTGGTGCAG  2940
AGAAGGAGGT TGTTCGTCTA GCAGGCCATA TCGGCTTTAA CATTGGTATG ACATTCCAAA  3000
TTTTGGATGA TATCCTGGAT TATACTGCAG ATAAAAAAAC ATTTAATAAG CCTGTCTTAG  3060
AGGATTTAAC ACAAGGCGTT TACAGCCTTC CTCTACTTCT TGCCATTGAA GAAAATCCTG  3120
ATATTTTCAA ACCTATTTTA GATAAAAAAA CAGATATGGC TACTGAAGAC ATGGAAAAAA  3180
TTGCTTATCT CGTCGTTTCC CATAGAGGTG TTGACAAAGC TCGCCATCTA GCTCGTAAAT  3240
TTACTGAGAA AGCTATTAGT GACATAAATA AGCTACCCCA GAACTCTGCA AAAAAACAGT  3300
TGCTACAATT AACTAATTAC CTTTTAAAAC GCAAAATTTA AATAATAAAA AAACATTCCA  3360
CAATGCTAGA AAAGCAGTTA GGGAATGTTT TTTTATTATC ATTTATTTAT CGCACCTATC  3420
AATCATCATA GATCACCATC ATCAGCGGCT TTCAGCTGAC GGTAACGTTG ACTACTTTGA  3480
GACAATTCTT GAGGAGAACC TTCCAACTCT AATTGCCCAT TTTCTATAAA TAAGATACGA  3540
TCAGCATGTT CAATACCTTT TAAGTGATGT GTAATCCAAA CTAAGGTCTT ACCTTCCAAT  3600
TCTTTCATAA ATACCCTTAG TAAGGCTTGT TCAGTAATAG GATCAAGTCC AACAGTTGGC  3660
TCATCTAAGA TAACAATTGG GACATCTTTT AGTAAGATTC TAGCCAAAGC AATTCTATGC  3720
CTTTCGCCAC CTGAAAACCT AAGTCCAGCT TCATCAACCA TTGTATAGAG ACCATCTGAT  3780
AAATCAGTGA CCATCTCTTT CAATCCAACT CGTTCAAGAA CTTTCCATAC ATCTTCTTCA  3840
CTAGCATCTT GGTTTCCAAT GCGAATGTTA TTTAGCAGGG TTGTATTAAA AAGGTAGGGC  3900
GCTTGTTGTA TCACTCCAAT ATAGTTAGAA ATGCAATCAC CAACTATTGA AACATCAGCA  3960
CCGCCTAGGG TAATCTTCCC TTGACTTGCT TTCAAGTCGC CACGAAGTAG ACTAGCTAAG  4020
GTACTCTTGC CAGAACCACT CCGCCCTAAA ATAGCAATTT TTTCTCCTTC TTTAATATCC  4080
AAATCTAAAT GATGCAAAAC CCATTTCTCT TGTGGCTTAT ACTGGAAACT TAAATTCTTG  4140
ACGGAAAAAT CATATGGCTT ATTAGGCAAT T   (SEQ ID NO:32)                 4171
```

FIG. 6a

YDNIFQSLHH LLACRGKSGN TLIDQLVADG LLHADNHYHF FNGKSLATFN 50
TNQLIREVVY VEISLDTMSS GEHDLVKVNI IRPTTEHTIP TMMTASPYHQ 100
GINDPAADQK TYQMEGALAV KQPKHIQVDT KPFKEEVKHP SKLPISPATE 150
SFTHIDSYSL NDYFLSRGFA NIYVSGVGTA GSTGFMTSGD YQQIQSFKAV 200
IDWLNGKVTA FTSHKRDKQV KADWSNGLVA TTGKSYLGTM STGLATTGVE 250
GLKVIIAEAA ISTWYDYYRE NGLVCSPGGY PGEDLDVLTE LTYSRNLLAG 300
DYIKNNDCYQ ALLNEQSKAI DRQSGDYNQY WHDRNYLTHV NNVKSRVVYT 350
HGLQDWNVKP RHVYKVFNAL PQTIKKHLFL HQGQHVYMHN WQSIDFRESM 400
NALLSQELLG IDNHFQLEEV IWQDNTTEQT WQVLDAFGGN HQEQIGLGDS 450
KKLIDNHYDK EAFDTYCKDF NVFKNDLFKG NNKTNQITIN LPLKKNYLLN 500
GQCKLHLRVK TSDKKAILSA QILDYGPKKR FKDTPTIKFL NSLDNGKNFA 550
REALRELPFT KDHYRVISKG VLNLQNRTDL LTIEAIEPEQ WFDIEFSLQP 600
SIYQLSKGDN LRIILYTTDF EHTIRDNASY SITVDLSQSY LTIPTNQGN 649
(SEQ ID NO:33)

FIG. 6b

MKLLTKERFD DSQHFWYQIN LLQESNFGAV FDHDNKNIPQ VVATIVDDLQ 50
GSGSSNHFWY FGNTTDTSIL MIAHLNRKFY IQVNLKDFDF ALNLIAINNW 100
KSLLQTQLEA LNDTLAIFQ (SEQ ID NO:34) 119

FIG. 6c

MSSYWNNYPE LKKNIDETNQ LIQERIQVRN KDIEAALSQL TAAGGKQLRP 50
AFFYLFSQLG NKENQDTQQL KKIAASLEIL HVATLIHDDV IDDSPLRRGN 100
MTIQSKFGKD IAVYTGDLLF TVFFDLILES MTDTPFMRIN AKSMRKILMG 150
ELDQMHLRYN QQQGIHHYLR AISGKTAELF KLASKEGAYF GGAEKEVVRL 200
AGHIGFNIGM TFQILDDILD YTADKKTFNK PVLEDLTQGV YSLPLLLAIE 250
ENPDIFKPIL DKKTDMATED MEKIAYLVVS HRGVDKARHL ARKFTEKAIS 300
DINKLPQNSA KKQLLQLTNY LLKRKI (SEQ ID NO:35) 326

FIG. 6d

LPNKPYDFSV KNLSFQYKPQ EKWVLHHLDL DIKEGEKIAI LGRSGSGKST 50
LASLLRGDLK ASQGKITLGG ADVSIVGDCI SNYIGVIQQA PYLFNTTLLN 100
NIRIGNQDAS EEDVWKVLER VGLKEMVTDL SDGLYTMVDE AGLRFSGGER 150
HRIALARILL KDVPIVILDE PTVGLDPITE QALLRVFMKE LEGKTLVWIT 200
HHLKGIEHAD RILFIENGQL ELEGSPQELS QSSQRYRQLK AADDGDL 247
(SEQ ID NO:36)

FIG. 6e

AATTCTATTT GGAGGTTTTT CTTGAATAAA TGGTTAGTTA AGGCAAGTTC CTTAGTTGTT 60
TTAGGTGGTA TGGTTTTATC TGCGGGTTCC CGAGTTTTAG CGGATACTTA TGTCCGTCCA 120
ATTGATAATG GTAGAATTAC AACAGGTTTC AATGGTTATC CTGGACATTG TGGGGTGGAT 180
TATGCTGTTC CGACTGGAAC GATTATTAGG GCAGTGGCAG ATGGTACTGT GAAATTTGCA 240
GGAGCTGGAG CCAACTTTTC TTGGATGACA GACTTAGCAG GAAATTGTGT CATGATTCAA 300
CATGCGGATG GAATGCATAG TGGTTACGCT CATATGTCAC GTGTGGTGGC TAGGACTGGG 360
GAAAAAGTCA AACAAGGAGA TATCATCGGT TACGTAGGAG CAACTGGTAT GGCGACGGGA 420
CCTCACCTTC ATTTTGAATT TTTACCAGCT AACCCTAATT TTCAAAATGG TTTCCATGGA 480
CGTATCAATC CAACGTCACT AATTGCTAAC GTTGCGACCT TTAGTGGAAA AACGCAAGCA 540
TCAGCTCCAA GCATTAAGCC ATTACAATCA GCTCCTGTAC AGAATCAATC TAGTAAATTA 600
AAAGTGTATC GAGTAGATGA ATTACAAAAG GTTAATGGTG TTTGGTTAGT CAAAAATAAC 660
ACCCTAACGC CGACTGGGTT TGATTGGAAC GATAATGGTA TACCAGCATC AGAAATTGAT 720
GAGGTTGATG CTAATGGTAA TTTGACAGCT GACCAGGTTC TTCAAAAAGG TGGTTACTTT 780
ATCTTTAATC CTAAAACTCT TAAGACTGTA GAAAAACCCA TCCAAGGAAC AGCTGGTTTA 840
ACTTGGGCTA AGACACGCTT TGCTAATGGT AGTTCAGTTT GGCTTCGCGT TGACAACAGT 900
CAAGAACTGC TTTACAAATA GTTTGAGGTA TTGATTCATT GTTTTAAATG ACAGTTTTGT 960
TACTAACTAA GTACAATTTC TTTAAACCGT CTGAAAATAA TTTTATAGTC CAGTAAAGTG 1020
TGATATTATA GTCTCGGACT AATAAAAAGG AAATAGGAAT TGAAGCAATG AAAATGAATA 1080
AAAAGGTACT ATTGACATCG ACAATGGCAG CTTCGCTATT ATCAGTCGCA AGTGTTCAAG 1140
CACAAGAAAC AGATACGACG TGGACAGCAC GTACTGTTTC AGAGGTAAAG GCTGATTTGG 1200
TAAAGCAAGA CAATAAATCA TCATATACTG TGAAATATGG TGATACACTA AGCGTTATTT 1260
CAGAAGCAAT GTCAATTGAT ATGAATGTCT TAGCAAAAAT TAATAACATT GCAGATATCA 1320
ATCTTATTTA TCCTGAGACA ACACTGACAG TAACTTACGA TCAGAAGAGT CATACTGCCA 1380
CTTCAATGAA AATAGAAACA CCAGCAACAA ATGCTGCTGG TCAAACAACA GCTACTGTGG 1440
ATTTGAAAAC CAATCAAGTT TCTGTTGCAG ACCAAAAAGT TTCTCTCAAT ACAATTTCGG 1500
AAGGTATGAC ACCAGAAGCA GCAACAACGA TTGTTTCGCC AATGAAGACA TATTCTTCTG 1560
CGCCAGCTTT GAAATCAAAA GAAGTATTAG CACAAGAGCA AGCTGTTAGT CAAGCAGCAG 1620
CTAATGAACA GGTATCAACA GCTCCTGTGA AGTCGATTAC TTCAGAAGTT CCAGCAGCTA 1680
AAGAGGAAGT TAAACCAACT CAGACGTCAG TCAGTCAGTC AACAACAGTA TCACCAGCTT 1740
CTGTTGCCGC TGAAACACCA GCTCCAGTAG CTAAAGTAGC ACCGGTAAGA ACTGTAGCAG 1800
CCCCTAGAGT GGCAAGTGTT AAAGTAGTCA CTCCTAAAGT AGAAACTGGT GCATCACCAG 1860
AGCATGTATC AGCTCCAGCA GTTCCTGTGA CTACGACTTC AACAGCTACA GACAGTAAGT 1920
TACAAGCGAC TGAAGTTAAG AGCGTTCCGG TAGCACAAAA AGCTCCAACA GCAACACCGG 1980
TAGCACAACC AGCTTCAACA ACAAATGCAG TAGCTGCACA TCCTGAAAAT GCAGGGCTCC 2040
AACCTCATGT TGCAGCTTAT AAAGAAAAAG TAGCGTCAAC TTATGGAGTT AATGAATTCA 2100
GTACATACCG TGCAGGTGAT CCAGGTGATC ATGGTAAAGG TTTAGCAGTC GACTTTATTG 2160
TAGGTAAAAA CCAAGCACTT GGTAATGAAG TTGCACAGTA CTCTACACAA AATATGGCAG 2220
CAAATAACAT TTCATATGTT ATCTGGCAAC AAAAGTTTTA CTCAAATACA AATAGTATTT 2280
ATGGACCTGC TAATACTTGG AATGCAATGC CAGATCGTGG TGGCGTTACT GCCAACCATT 2340
ATGACCATGT TCACGTATCA TTTAACAAAT AATATAAAAA AGGAAGCTAT TTGGCTTCTT 2400

```
TTTTATATGC CTTGAATAGA CTTTCAAGGT TCTTATCTAA TTTTTATTAA ATTGAGGAGA  2460
TTAAGCTATA AGTCTGAAAC TACTTTCACG TTAACCGTGA CTAAATCAAA ACGTTAAAAC  2520
TAAAATCTAA GTCTGTAAAG ATTATTGAAA ACGCTTTAAA AACAGATATA ATAAGGTTTG  2580
TAGATATCTA AAATTAAAAA AGATAAGGAA GTGAGAATAT GCCACATCTA AGTAAAGAAG  2640
CTTTTAAAAA GCAAATAAAA AATGGCATTA TTGTGTCATG TCAAGCTTTG CCTGGGGAGC  2700
CTCTTTATAC TGAAAGTGGA GGTGTTATGC CTCTTTTAGC TTTGGCAGCT CAAGAAGCAG  2760
GAGCGGTTGG TATAAGAGCC AATAGTGTCC GCGACATTAA GGAAATTCAA GAAGTTACTA  2820
ATTTACCTAT CATCGGCATT ATTAAACGTG AATATCCTCC ACAAGAACCA TTTATCACTG  2880
CTACGATGAC AGAGGTGGAT CAATTAGCTA GTTTAGATAT TGCAGTAATA GCCTTAGATT  2940
GTACACTTAG AGAGCGTCAT GATGGTTTGA GTGTAGCTGA GTTTATTCAA AAGATAAAAG  3000
GGAAATATCC TGAACAGTTG CTAATGGCTG ATATAAGTAC TTTTGAAGAA GGTAAAAATG  3060
CTTTTGAAGC AGGAGTTGAT TTTGTGGGTA CAACTCTATC TGGATACACA GATTACAGCC  3120
GCCAAGAAGA AGGACCGGAT ATAGAACTCC TTAATAAGCT TTGTCAAGCC GGTATAGATG  3180
TGATTGCGGA AGGTAAAATT CATACTCCTA AGCAAGCTAA TGAAATTAAT CATATAGGTG  3240
TTGCAGGAAT TGTAGTTGGT GGTGCTATCA CTAGACCAAA AGAAATAGCG GAGCGTTTCA  3300
TCTCAGGACT TAGTTAAAAG TGTTACTCAA AAATCAAAAT CAAAATAAAA AAGGGGAATA  3360
GTTATGAGTA TCAAAAAAAG TGTGATTGGT TTTTGCCTCG GAGCTGCAGC ATTATCAATG  3420
TTTGCTTGTG TAGACAGTAG TCAATCTGTT ATGGCTGCCG AGAAGGATAA AGTCGAAATT  3480
(SEQ ID NO:37)
```

FIG. 7a

```
NSIWRFFLNK WLVKASSLVV LGGMVLSAGS RVLADTYVRP IDNGRITTGF   50
NGYPGHCGVD YAVPTGTIIR AVADGTVKFA GAGANFSWMT DLAGNCVMIQ  100
HADGMHSGYA HMSRVVARTG EKVKQGDIIG YVGATGMATG PHLHFEFLPA  150
NPNFQNGFHG RINPTSLIAN VATFSGKTQA SAPSIKPLQS APVQNQSSKL  200
KVYRVDELQK VNGVWLVKNN TLTPTGFDWN DNGIPASEID EVDANGNLTA  250
DQVLQKGGYF IFNPKTLKTV EKPIQGTAGL TWAKTRFANG SSVWLRVDNS  300
QELLYK    (SEQ ID NO:38)                                306
```

FIG. 7b

MKMNKKVLLT STMAASLLSV ASVQAQETDT TWTARTVSEV KADLVKQDNK 50
SSYTVKYGDT LSVISEAMSI DMNVLAKINN IADINLIYPE TTLTVTYDQK 100
SHTATSMKIE TPATNAAGQT TATVDLKTNQ VSVADQKVSL NTISEGMTPE 150
AATTIVSPMK TYSSAPALKS KEVLAQEQAV SQAAANEQVS TAPVKSITSE 200
VPAAKEEVKP TQTSVSQSTT VSPASVAAET PAPVAKVAPV RTVAAPRVAS 250
VKVVTPKVET GASPEHVSAP AVPVTTTSTA TDSKLQATEV KSVPVAQKAP 300
TATPVAQPAS TTNAVAAHPE NAGLQPHVAA YKEKVASTYG VNEFSTYRAG 350
DPGDHGKGLA VDFIVGKNQA LGNEVAQYST QNMAANNISY VIWQQKFYSN 400
TNSIYGPANT WNAMPDRGGV TANHYDHVHV SFNK (SEQ ID NO:39) 434

FIG. 7c

MPHLSKEAFK KQIKNGIIVS CQALPGEPLY TESGGVMPLL ALAAQEAGAV 50
GIRANSVRDI KEIQEVTNLP IIGIIKREYP PQEPFITATM TEVDQLASLD 100
IAVIALDCTL RERHDGLSVA EFIQKIKGKY PEQLLMADIS TFEEGKNAFE 150
AGVDFVGTTL SGYTDYXRQE EGPDIELLNK LCQAGIDVIA EGKIHTPKQA 200
NEINHIGVAG IVVGGAITRP KEIAERFISG LS (SEQ ID NO:40) 232

FIG. 7d

MSIKKSVIGF CLGAAALSMF ACVDSSQSVM AAEKDKVEI 39
(SEQ ID NO:41)

FIG. 7e

```
ATGAAAATGA ATAAAAAGGT ACTATTGACA TCGACAATGG CAGCTTCGCT    50
ATTATCAGTC GCAAGTGTTC AAGCACAAGA AACAGATACG ACGTGGACAG   100
CACGTACTGT TTCAGAGGTA AAGGCTGATT TGGTAAAGCA AGACAATAAA   150
TCATCATATA CTGTGAAATA TGGTGATACA CTAAGCGTTA TTTCAGAAGC   200
AATGTCAATT GATATGAATG TCTTAGCAAA AATTAATAAC ATTGCAGATA   250
TCAATCTTAT TTATCCTGAG ACAACACTGA CAGTAACTTA CGATCAGAAG   300
AGTCATACTG CCACTTCAAT GAAAATAGAA ACACCAGCAA CAAATGCTGC   350
TGGTCAAACA ACAGCTACTG TGGATTTGAA AACCAATCAA GTTTCTGTTG   400
CAGACCAAAA AGTTTCTCTC AATACAATTT CGGAAGGTAT GACACCAGAA   450
GCAGCAACAA CGATTGTTTC GCCAATGAAG ACATATTCTT CTGCGCCAGC   500
TTTGAAATCA AAAGAAGTAT TAGCACAAGA GCAAGCTGTT AGTCAAGCAG   550
CAGCTAATGA ACAGGTATCA ACAGCTCCTG TGAAGTCGAT TACTTCAGAA   600
GTTCCAGCAG CTAAAGAGGA AGTTAAACCA ACTCAGACGT CAGTCAGTCA   650
GTCAACAACA GTATCACCAG CTTCTGTTGC CGCTGAAACA CCAGCTCCAG   700
TAGCTAAAGT AGCACCGGTA AGAACTGTAG CAGCCCCTAG AGTGGCAAGT   750
GTTAAAGTAG TCACTCCTAA AGTAGAAACT GGTGCATCAC CAGAGCATGT   800
ATCAGCTCCA GCAGTTCCTG TGACTACGAC TTCAACAGCT ACAGACAGTA   850
AGTTACAAGC GACTGAAGTT AAGAGCGTTC CGGTAGCACA AAAAGCTCCA   900
ACAGCAACAC CGGTAGCACA ACCAGCTTCA ACAACAAATG CAGTAGCTGC   950
ACATCCTGAA AATGCAGGGC TCCAACCTCA TGTTGCAGCT TATAAAGAAA  1000
AAGTAGCGTC AACTTATGGA GTTAATGAAT TCAGTACATA CCGTGCAGGT  1050
GATCCAGGTG ATCATGGTAA AGGTTTAGCA GTCGACTTTA TTGTAGGTAA  1100
AAACCAAGCA CTTGGTAATG AAGTTGCACA GTACTCTACA CAAAATATGG  1150
CAGCAAATAA CATTTCATAT GTTATCTGGC AACAAAAGTT TTACTCAAAT  1200
ACAAATAGTA TTTATGGACC TGCTAATACT TGGAATGCAA TGCCAGATCG  1250
TGGTGGCGTT ACTGCCAACC ATTATGACCA TGTTCACGTA TCATTTAACA  1300
AATAA                                                   1305
```

(SEQ ID NO:42)

FIG. 8

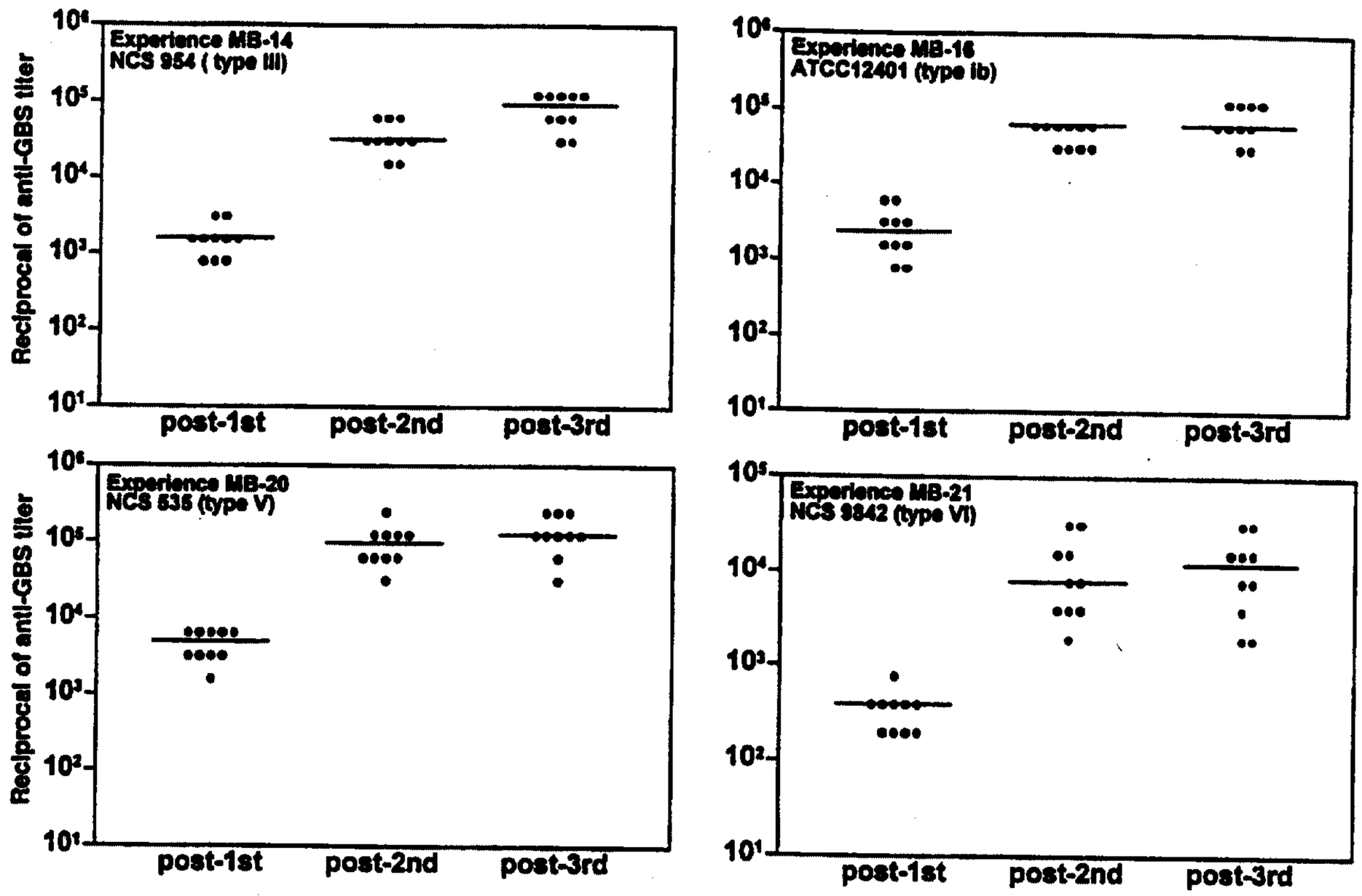
```
CAAGAAACAG ATACGACGTG GACAGCACGT ACTGTTTCAG AGGTAAAGGC   50
TGATTTGGTA AAGCAAGACA ATAAATCATC ATATACTGTG AAATATGGTG  100
ATACACTAAG CGTTATTTCA GAAGCAATGT CAATTGATAT GAATGTCTTA  150
GCAAAAATTA ATAACATTGC AGATATCAAT CTTATTTATC CTGAGACAAC  200
ACTGACAGTA ACTTACGATC AGAAGAGTCA TACTGCCACT TCAATGAAAA  250
TAGAAACACC AGCAACAAAT GCTGCTGGTC AAACAACAGC TACTGTGGAT  300
TTGAAAACCA ATCAAGTTTC TGTTGCAGAC CAAAAAGTTT CTCTCAATAC  350
AATTTCGGAA GGTATGACAC CAGAAGCAGC AACAACGATT GTTTCGCCAA  400
TGAAGACATA TTCTTCTGCG CCAGCTTTGA AATCAAAAGA AGTATTAGCA  450
CAAGAGCAAG CTGTTAGTCA AGCAGCAGCT AATGAACAGG TATCAACAGC  500
TCCTGTGAAG TCGATTACTT CAGAAGTTCC AGCAGCTAAA GAGGAAGTTA  550
AACCAACTCA GACGTCAGTC AGTCAGTCAA CAACAGTATC ACCAGCTTCT  600
GTTGCCGCTG AAACACCAGC TCCAGTAGCT AAAGTAGCAC CGGTAAGAAC  650
TGTAGCAGCC CCTAGAGTGG CAAGTGTTAA AGTAGTCACT CCTAAAGTAG  700
AAACTGGTGC ATCACCAGAG CATGTATCAG CTCCAGCAGT TCCTGTGACT  750
ACGACTTCAA CAGCTACAGA CAGTAAGTTA CAAGCGACTG AAGTTAAGAG  800
CGTTCCGGTA GCACAAAAAG CTCCAACAGC AACACCGGTA GCACAACCAG  850
CTTCAACAAC AAATGCAGTA GCTGCACATC CTGAAAATGC AGGGCTCCAA  900
CCTCATGTTG CAGCTTATAA AGAAAAAGTA GCGTCAACTT ATGGAGTTAA  950
TGAATTCAGT ACATACCGTG CAGGTGATCC AGGTGATCAT GGTAAAGGTT 1000
TAGCAGTCGA CTTTATTGTA GGTAAAAACC AAGCACTTGG TAATGAAGTT 1050
GCACAGTACT CTACACAAAA TATGGCAGCA AATAACATTT CATATGTTAT 1100
CTGGCAACAA AAGTTTTACT CAAATACAAA TAGTATTTAT GGACCTGCTA 1150
ATACTTGGAA TGCAATGCCA GATCGTGGTG GCGTTACTGC CAACCATTAT 1200
GACCATGTTC ACGTATCATT TAACAAATAA     (SEQ ID NO:43)    1230
```

FIG. 9

```
QETDTTWTAR TVSEVKADLV KQDNKSSYTV KYGDTLSVIS EAMSIDMNVL   50
AKINNIADIN LIYPETTLTV TYDQKSHTAT SMKIETPATN AAGQTTATVD  100
LKTNQVSVAD QKVSLNTISE GMTPEAATTI VSPMKTYSSA PALKSKEVLA  150
QEQAVSQAAA NEQVSTAPVK SITSEVPAAK EEVKPTQTSV SQSTTVSPAS  200
VAAETPAPVA KVAPVRTVAA PRVASVKVVT PKVETGASPE HVSAPAVPVT  250
TTSTATDSKL QATEVKSVPV AQKAPTATPV AQPASTTNAV AAHPENAGLQ  300
PHVAAYKEKV ASTYGVNEFS TYRAGDPGDH GKGLAVDFIV GKNQALGNEV  350
AQYSTQNMAA NNISYVIWQQ KFYSNTNSIY GPANTWNAMP DRGGVTANHY  400
DHVHVSFNK  (SEQ ID NO:44)                               409
```

FIG. 9a

Fig. 10


Experience MB-14
NCS 954 (type III)
Experience MB-16
ATCC12401 (type Ib)
Experience MB-20
NCS 535 (type V)
Experience MB-21
NCS 9842 (type VI)
Reciprocal of anti-GBS titer
post-1st
post-2nd
post-3rd