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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 May 2002 (16.05.2002)

PCT

(10) International Publication Number
WO 02/39431 A2

(51) International Patent Classification[7]: G11B 5/00

(21) International Application Number: PCT/US01/47175

(22) International Filing Date:
6 November 2001 (06.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/246,039 6 November 2000 (06.11.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/246,039 (CIP)
Filed on 6 November 2000 (06.11.2000)

(71) Applicant *(for all designated States except US)*: **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SUN, Yongming** [US/US]; Apartment 260, 869 S. Winchester Boulevard, San Jose, CA 95128 (US). **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **CHEN, Sei-Yu** [—/US]; 160 Mira Street, Foster City, CA 94404 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way, #14, San Jose, CA 95117 (US).

(74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, 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, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, 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, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/39431 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO PROSTATE SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic prostate cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and non-cancerous disease states in prostate tissue, identifying prostate tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered prostate tissue for treatment and research.

COMPOSITIONS AND METHODS RELATING TO PROSTATE SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application Serial No. 60/246,039 filed November 6, 2000, which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic prostate cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and non-cancerous disease states in prostate tissue, identifying prostate tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered prostate tissue for treatment and research.

BACKGROUND OF THE INVENTION

Prostate cancer is the most prevalent cancer in men and is the second leading cause of death from cancer among males in the United States. AJCC Cancer Staging Handbook 203 (Irvin D. Fleming et al. eds., 5th ed. 1998); Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 147 (1998). In 1999, it was estimated that 37,000 men in the United States would die as result of prostate cancer. Elizabeth A. Platz et al., & Edward Giovannucci, *Epidemiology of and Risk Factors for Prostate Cancer, in* Management of Prostate Cancer 21 (Eric A Klein, ed. 2000). Cancer of the prostate typically occurs in older males, with a median age of 74 years for clinical diagnosis.

Burdette, *supra* at 147. A man's risk of being diagnosed with invasive prostate cancer in his lifetime is one in six. Platz et al., *supra* at 21.

Although our understanding of the etiology of prostate cancer is incomplete, the results of extensive research in this area point to a combination of age, genetic and environmental/dietary factors. Platz et al., *supra* at 19; Burdette, *supra* at 147; Steven K. Clinton, *Diet and Nutrition in Prostate Cancer Prevention and Therapy, in* Prostate Cancer: A Multidisciplinary Guide 246-269 (Philip W. Kantoff et al. eds. 1997). Broadly speaking, genetic risk factors predisposing one to prostate cancer include race and a family history of the disease. Platz et al., *supra* at 19, 28-29, 32-34. Aside from these generalities, a deeper understanding of the genetic basis of prostate cancer has remained elusive. Considerable research has been directed to studying the link between prostate cancer, androgens, and androgen regulation, as androgens play a crucial role in prostate growth and differentiation. Meena Augustus et al., *Molecular Genetics and Markers of Progression, in* Management of Prostate Cancer 59 (Eric A Klein ed. 2000). While a number of studies have concluded that prostate tumor development is linked to elevated levels of circulating androgen (*e.g.*, testosterone and dihydrotestosterone), the genetic determinants of these levels remain unknown. Platz et al., *supra* at 29-30.

Several studies have explored a possible link between prostate cancer and the androgen receptor (AR) gene, the gene product of which mediates the molecular and cellular effects of testosterone and dihydrotestosterone in tissues responsive to androgens. *Id.* at 30. Differences in the number of certain trinucleotide repeats in exon 1, the region involved in transactivational control, have been of particular interest. Augustus et al., *supra* at 60. For example, these studies have revealed that as the number of CAG repeats decreases the transactivation ability of the gene product increases, as does the risk of prostate cancer. Platz et al., *supra* at 30-31. Other research has focused on the α-reductase Type 2 gene, the gene which codes for the enzyme that converts testosterone into dihydrotestosterone. *Id.* at 30. Dihydrotestosterone has greater affinity for the AR than testosterone, resulting in increased transactivation of genes responsive to androgens. *Id.* While studies have reported differences among the races in the length of a TA dinucleotide repeat in the 3' untranslated region, no link has been established between the length of that repeat and prostate cancer. *Id.*

Interestingly, while *ras* gene mutations are implicated in numerous other cancers, such mutations appear not to play a significant role in prostate cancer, at least among Caucasian males. Augustus, *supra* at 52.

Environmental/dietary risk factors which may increase the risk of prostate cancer include intake of saturated fat and calcium. Platz et al., *supra* at 19, 25-26. Conversely, intake of selenium, vitamin E and tomato products (which contain the carotenoid lycopene) apparently decrease that risk. *Id.* at 19, 26-28 The impact of physical activity, cigarette smoking, and alcohol consumption on prostate cancer is unclear. Platz et al., *supra* at 23-25.

Periodic screening for prostate cancer is most effectively performed by digital rectal examination (DRE) of the prostate, in conjunction with determination of the serum level of prostate-specific antigen (PSA). Burdette, *supra* at 148. While the merits of such screening are the subject of considerable debate, Jerome P. Richie & Irving D. Kaplan, *Screening for Prostate Cancer: The Horns of a Dilemma, in* Prostate Cancer: A Multidisciplinary Guide 1-10 (Philip W. Kantoff et al. eds. 1997), the American Cancer Society and American Urological Association recommend that both of these tests be performed annually on men 50 years or older with a life expectancy of at least 10 years, and younger men at high risk for prostate cancer. Ian M. Thompson & John Foley, *Screening for Prostate Cancer, in* Management of Prostate Cancer 71 (Eric A Klein ed. 2000). If necessary, these screening methods may be followed by additional tests, including biopsy, ultrasonic imaging, computerized tomography, and magnetic resonance imaging. Christopher A. Haas & Martin I. Resnick, *Trends in Diagnosis, Biopsy, and Imaging, in* Management of Prostate Cancer 89-98 (Eric A Klein ed. 2000); Burdette, *supra* at 148.

Once the diagnosis of prostate cancer has been made, treatment decisions for the individual are typically linked to the stage of prostate cancer present in that individual, as well as his age and overall health. Burdette, *supra* at 151. One preferred classification system for staging prostate cancer was developed by the American Urological Association (AUA). *Id.* at 148. The AUA classification system divides prostate tumors into four broad stages, A to D, which are in turn accompanied by a number of smaller substages. Burdette, *supra* at 152-153; Anthony V. D'Amico et al., *The Staging of*

Prostate Cancer, in Prostate Cancer: A Multidisciplinary Guide 41 (Philip W. Kantoff et al. eds. 1997).

Stage A prostate cancer refers to the presence of microscopic cancer within the prostate gland. D'Amico, *supra* at 41. This stage is comprised of two substages: A1, which involves less than four well-differentiated cancer foci within the prostate, and A2, which involves greater than three well-differentiated cancer foci or alternatively, moderately to poorly differentiated foci within the prostate. Burdette, *supra* at 152; D'Amico, *supra* at 41. Treatment for stage A1 preferentially involves following PSA levels and periodic DRE. Burdette, *supra* at 151. Should PSA levels rise, preferred treatments include radical prostatectomy in patients 70 years of age and younger, external beam radiotherapy for patients between 70 and 80 years of age, and hormone therapy for those over 80 years of age. *Id.*

Stage B prostate cancer is characterized by the presence of a palpable lump within the prostate. Burdette, *supra* at 152-53; D'Amico, *supra* at 41. This stage is comprised of three substages: B1, in which the lump is less than 2 cm and is contained in one lobe of the prostate; B2, in which the lump is greater than 2 cm yet is still contained within one lobe; and B3, in which the lump has spread to both lobes. Burdette, *supra*, at 152-53. For stages B1 and B2, the treatment again involves radical prostatectomy in patients 70 years of age and younger, external beam radiotherapy for patients between 70 and 80 years of age, and hormone therapy for those over 80 years of age. *Id.* at 151. In stage B3, radical prostatectomy is employed if the cancer is well-differentiated and PSA levels are below 15 ng/mL; otherwise, external beam radiation is the chosen treatment option. *Id.*

Stage C prostate cancer involves a substantial cancer mass accompanied by extraprostatic extension. Burdette, *supra* at 153; D'Amico, *supra* at 41. Like stage A prostate cancer, Stage C is comprised of two substages: substage Cl, in which the tumor is relatively minimal, with minor prostatic extension, and substage C2, in which the tumor is large and bulky, with major prostatic extension. *Id.* The treatment of choice for both substages is external beam radiation. Burdette, *supra* at 151.

The fourth and final stage of prostate cancer, Stage D, describes the extent to which the cancer has metastasized. Burdette, *supra* at 153; D'Amico, *supra* at 41. This stage is comprised of four substages: (1) D0, in which acid phophatase levels are

persistently high, (2) D1, in which only the pelvic lymph nodes have been invaded, (3) D2, in which the lymph nodes above the aortic bifurcation have been invaded, with or without distant metastasis, and (4) D3, in which the metastasis progresses despite intense hormonal therapy. *Id.* Treatment at this stage may involve hormonal therapy, chemotherapy, and removal of one or both testes. Burdette, *supra* at 151.

Despite the need for accurate staging of prostate cancer, current staging methodology is limited. The wide variety of biological behavior displayed by neoplasms of the prostate has resulted in considerable difficulty in predicting and assessing the course of prostate cancer. Augustus et al., *supra* at 47. Indeed, despite the fact that most prostate cancer patients have carcinomas that are of intermediate grade and stage, prognosis for these types of carcinomas is highly variable. Andrew A Renshaw & Christopher L. Corless, *Prognostic Features in the Pathology of Prostate Cancer, in* Prostate Cancer: A Multidisciplinary Guide 26 (Philip W. Kantoff et al. eds. 1997). Techniques such as transrectal ultrasound, abdominal and pelvic computerized tomography, and MRI have not been particularly useful in predicting local tumor extension. D'Amico, *supra* at 53 (editors' comment). While the use of serum PSA in combination with the Gleason score is currently the most effective method of staging prostate cancer, *id.*, PSA is of limited predictive value, Augustus et al., *supra* at 47; Renshaw et al., *supra* at 26, and the Gleason score is prone to variability and error, King, C. R. & Long, J. P., *Int'l. J. Cancer* 90(6): 326-30 (2000). As such, the current focus of prostate cancer research has been to obtain biomarkers to help better assess the progression of the disease. Augustus et al., *supra* at 47; Renshaw et al., *supra* at 26; Pettaway, C. A., *Tech. Urol.* 4(1): 35-42 (1998).

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop prostate cancer, for diagnosing prostate cancer, for monitoring the progression of the disease, for staging the prostate cancer, for determining whether the prostate cancer has metastasized and for imaging the prostate cancer. There is also a need for better treatment of prostate cancer.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto

that may be used to identify, diagnose, monitor, stage, image and treat prostate cancer and non-cancerous disease states in prostate; identify and monitor prostate tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered prostate tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to prostate cells and/or prostate tissue. These prostate specific nucleic acids (PSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the PSNA is genomic DNA, then the PSNA is a prostate specific gene (PSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to prostate. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 143 through 249. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 142. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding a PSP, or that selectively hybridize or exhibit substantial sequence similarity to a PSNA, as well as allelic variants of a nucleic acid molecule encoding a PSP, and allelic variants of a PSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes a PSP or that comprises a part of a nucleic acid sequence of a PSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a PSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a PSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of a PSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a PSNA.

Another object of the invention is to provided methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a PSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of a PSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and non-cancerous disease states in prostate. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring prostate tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered prostate tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat prostate cancer and non-cancerous disease states in prostate. The invention provides methods of using the polypeptides of the invention to identify and/or monitor prostate tissue, and to produce engineered prostate tissue.

The agonists and antagonists of the instant invention may be used to treat prostate cancer and non-cancerous disease states in prostate and to produce engineered prostate tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in

the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may

comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid

molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are

not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phoshoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence.

In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. *See* Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$T_m = 81.5°C + 16.6 (\log_{10}[Na^+]) + 0.41$ (fraction G + C) – 0.63 (% formamide) - (600/l)

where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5 (\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.35 (% formamide) - (820/l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5(\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.50 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping

the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:
$T_m = 81.5°C + 16.6(\log_{10}[Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$,
wherein N is change length and the $[Na^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The

various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAS. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra.*

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity

with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding a PSP or is a PSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR

reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g.*, Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression

vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a PSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally

associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid

derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial

identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$--, and –CH_2SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al., Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid

sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See* Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as -, -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino

acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, -N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted

upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.,* Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)

Cost to extend a gap:	1 (default
Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', $F(ab')_2$, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an $F(ab')_2$ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single

arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.,* Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.,* Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g.*, Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that

purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than1 μM, preferably less than100 nM and most preferably less than 10 nM.

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "prostate specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the prostate as compared to other tissues in the body. In a preferred embodiment, a "prostate specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "prostate specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the prostate or to prostate cells or tissue or that are derived from such nucleic acid molecules. These isolated prostate specific nucleic acids (PSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to prostate, a prostate-specific polypeptide (PSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 143 through 249. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 142.

A PSNA may be derived from a human or from another animal. In a preferred embodiment, the PSNA is derived from a human or other mammal. In a more preferred embodiment, the PSNA is derived from a human or other primate. In an even more preferred embodiment, the PSNA is derived from a human.

By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a PSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a PSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a PSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 143 through 249. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 142.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a PSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a PSP under moderate stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule

encoding a PSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 143 through 249. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 142. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a PSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human PSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 143 through 249. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a PSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 143 through 249, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a PSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a PSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a PSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 142. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a PSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 142,

more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a PSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a PSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to a PSNA or to a nucleic acid molecule encoding a PSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the PSNA or the nucleic acid molecule encoding a PSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 143 through 249 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 142. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the PSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a PSNA. Further, the substantially

similar nucleic acid molecule may or may not be a PSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a PSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a PSNA or a nucleic acid encoding a PSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a PSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a PSP comprising an amino acid sequence of SEQ ID NO: 143 through 249. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a PSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 142. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a PSP. However, in a preferred embodiment, the part encodes a PSP. In one aspect, the invention comprises a part of a PSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a PSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a PSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a PSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides.

The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ^{33}P, ^{32}P, and ^{35}S, such as -^{32}P-dATP, -^{32}P-dCTP, -^{32}P-dGTP, -^{32}P-dTTP, -^{32}P-3'dATP, -^{32}P-ATP, -^{32}P-CTP, -^{32}P-GTP, -^{32}P-UTP, -^{35}S-dATP, α-^{35}S-GTP, α-^{33}P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See* Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see* Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.,* Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties.

Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. *See* Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307;

5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The Tm of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the Tm of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the Tm by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the Tm by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.*, Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1):

3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a PSNA, such as deletions, insertions, translocations, and duplications of the PSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify PSNA in, and isolate PSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A^+- selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to PSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*;

Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a PSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 143 through 249. In another preferred embodiment, the probe or primer is derived from a PSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 142.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of

which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another

embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous

polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous

derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.,* YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as ura3-52, his3-D1, leu2-D1, trp1-D1 and lys2-201.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda* , *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between

these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.,* vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.,* bovine papillomavirus), and retroviral vectors (*e.g.,* murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally

contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast _-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the

Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the PSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β-globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PltetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline

(Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverley, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from

those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g,* Cormack *et al.*, *Gene* 173: 33–38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g,.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.,* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.,* Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See* also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable

expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,

and acylation, and it is an aspect of the present invention to provide PSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database http://www.abrf.org/ABRF/Research Committees/deltamass/deltamass.html (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) and http://www.glycosuite.com/ (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999) and http://www.cbs.dtu.dk/databases/OGLYCBASE/ (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. Nucleic Acids Res 27(1):237-239 (1999) and http://www.cbs.dtu.dk/databases/PhosphoBase/ (accessed October 19, 2001); or http://pir.georgetown.edu/pirwww/search/textresid.html (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur

in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g, p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the

desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors

typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae, Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (*e.g.*, from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and

readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from prostate are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human prostate cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See, e.g.*, Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli,* in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.*, Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli.*

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.*, with $CaCl_2$, or a solution of Mg^{2+}, Mn^{2+}, Ca^{2+}, Rb^{+} or K^{+}, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (*e.g.*, Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols

(BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with $CaPO_4$ or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for $CaPO_4$ transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent,

FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al., Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a prostate specific polypeptide (PSP). In an even more preferred embodiment, the

polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 143 through 249. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a PSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 143 through 249. A polypeptide that comprises only a fragment of an entire PSP may or may not be a polypeptide that is also a PSP. For instance, a full-length polypeptide may be prostate-specific, while a fragment thereof may be found in other tissues as well as in prostate. A polypeptide that is not a PSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-PSP antibodies. However, in a preferred embodiment, the part or fragment is a PSP. Methods of determining whether a polypeptide is a PSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g.*, Lerner, Nature 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for

the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, a PSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably a PSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a PSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be prostate-specific. In a preferred embodiment, the mutein is prostate-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 143 through 249. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 143 through 249. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 143 through 249.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is prostate-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo*

mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to a PSP. In an even more preferred embodiment, the polypeptide is homologous to a PSP selected from the group having an amino acid sequence of SEQ ID NO: 143 through 249. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a PSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 143 through 249. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 143 through 249. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 143 through 249. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 143 through 249. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a PSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a PSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the PSNA is selected from the group consisting of SEQ ID NO: 1 through 142. In another preferred embodiment, the

homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a PSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the PSP is selected from the group consisting of SEQ ID NO: 143 through 249.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 143 through 249. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the PSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a PSP. Further, the homologous protein may or may not encode polypeptide that is a PSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a PSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the

present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a PSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 143 through 249. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 142.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a PSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 143 through 249, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983);

Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591,

BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-PSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999),

incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a PSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 143 through 249. In a preferred embodiment, the analog is one that comprises one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to a PSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$-- and –CH_2SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of a PSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g.,* Kole *et al.*, *Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer

Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The FMOC and *t*BOC derivatives of dabcyl-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyl chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyl quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-

2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)-β-alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a PSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 143 through 249, or is a mutein, homologous polypeptide,

analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 142, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 142.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.*, Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation

of secretion signals and/or leader sequences. For example, a His^6 tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. *See* Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbrizio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in Drosophila with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by

reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β-galactosidase, biotin trpE, protein A, β-lactamase, -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast _ mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the PSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize PSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly PSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser

scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of PSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of PSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.,* Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is a PSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 143 through 249, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a PSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of a PSP may be indicative of cancer. Differential degradation of the C or N-terminus of a PSP may also be a marker or target for anticancer therapy. For example, a PSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-PSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human prostate.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the

present invention will be at least about 1 x 10^{-6} molar (M), typically at least about 5 x 10^{-7} M, 1 x 10^{-7} M, with affinities and avidities of at least 1 x 10^{-8} M, 5 x 10^{-9} M, 1 x 10^{-10} M and up to 1 X 10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are

hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. *See* Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular

advantage in detection of the proteins of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.,* Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*,

4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994).

Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.,* Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.,* Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, J. Biotechnol. 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,

Biotechnol. Appl. Biochem. 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock et al., *J. Immunol Methods.* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); Limonta et al., *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, $F(ab)'_2$, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated

nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA*.81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions

including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β-galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopryanoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate

reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g.*, Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I.

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th, ^{227}Ac, ^{225}Ac, ^{223}Ra, ^{213}Bi, ^{212}Pb, ^{212}Bi, ^{211}At, ^{203}Pb, ^{194}Os, ^{188}Re, ^{186}Re, ^{153}Sm, ^{149}Tb, ^{131}I, ^{125}I, ^{111}In, ^{105}Rh, ^{99m}Tc, ^{97}Ru, ^{90}Y, ^{90}Sr, ^{88}Y, ^{72}Se, ^{67}Cu, or ^{47}Sc.

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a PSP. In a preferred embodiment, the PSP comprises an amino acid

sequence selected from SEQ ID NO: 143 through 249, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a PSNA of the invention, preferably a PSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 142, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human PSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.*, Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.*, Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.,* Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.,* Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.,* Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.*, Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.,* Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,* Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e.*, a

nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of

the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g.,* Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g.*, Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.,* Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC

compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.,* U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 142 and SEQ ID NO: 143 through 249 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set

representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Prostate Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by

comparing expression of a PSNA or a PSP in a human patient that has or may have prostate cancer, or who is at risk of developing prostate cancer, with the expression of a PSNA or a PSP in a normal human control. For purposes of the present invention, "expression of a PSNA" or "PSNA expression" means the quantity of PSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a PSP" or "PSP expression" means the amount of PSP that can be measured by any method known in the art or the level of translation of a PSG PSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing prostate cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of PSNA or PSP in cells, tissues, organs or bodily fluids compared with levels of PSNA or PSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a PSNA or PSP in the patient versus the normal human control is associated with the presence of prostate cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing prostate cancer in a patient by analyzing changes in the structure of the mRNA of a PSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing prostate cancer in a patient by analyzing changes in a PSP compared to a PSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the PSP or subcellular PSP localization.

In a preferred embodiment, the expression of a PSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 143 through 249, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the PSNA expression that is measured is the level of expression of a PSNA mRNA selected from SEQ ID NO: 1 through 142, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. PSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative

or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. PSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a PSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, PSNA expression may be compared to a known control, such as normal prostate nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a PSP is measured by determining the level of a PSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 143 through 249, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of PSNA or PSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of prostate cancer. The expression level of a PSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the PSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See*, e.g, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the PSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a PSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-PSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a

protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the PSP will bind to the anti-PSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-PSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the PSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of a PSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure PSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-PSP antibody is attached to a solid support and an allocated amount of a labeled PSP and a sample of interest are incubated with the solid support. The amount of labeled PSP detected which is attached to the solid support can be correlated to the quantity of a PSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a PSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other

mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more PSNAs of interest. In this approach, all or a portion of one or more PSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of PSNA or PSP includes, without limitation, prostate tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, prostate cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary prostate cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediatinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in PSNAs or PSPs in cells in sputum samples may be particularly

useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a PSNA or PSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other PSNA or PSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular PSNA or PSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a sample from a patient suspected of having prostate cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a PSNA and/or PSP and then ascertaining whether the patient has prostate cancer from the expression level of the PSNA or PSP. In general, if high expression relative to a control of a PSNA or PSP is indicative of prostate cancer, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a PSNA or PSP is indicative of prostate cancer, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether prostate cancer has metastasized in a patient. One may identify whether the prostate cancer has

metastasized by measuring the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a variety of tissues. The presence of a PSNA or PSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a PSNA or PSP is associated with prostate cancer. Similarly, the presence of a PSNA or PSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a PSNA or PSP is associated with prostate cancer. Further, the presence of a structurally altered PSNA or PSP that is associated with prostate cancer is also indicative of metastasis.

In general, if high expression relative to a control of a PSNA or PSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a PSNA or PSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The PSNA or PSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with prostate cancers or other prostate related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of prostate disorders.

Staging

The invention also provides a method of staging prostate cancer in a human patient. The method comprises identifying a human patient having prostate cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more PSNAs or PSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more PSNAs or PSPs is determined for each stage to

obtain a standard expression level for each PSNA and PSP. Then, the PSNA or PSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The PSNA or PSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the PSNAs and PSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a PSNA or PSP to determine the stage of a prostate cancer.

Monitoring

Further provided is a method of monitoring prostate cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the prostate cancer. The method comprises identifying a human patient that one wants to monitor for prostate cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more PSNAs or PSPs, and comparing the PSNA or PSP levels over time to those PSNA or PSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a PSNA or PSP that are associated with prostate cancer.

If increased expression of a PSNA or PSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a PSNA or PSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a PSNA or PSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of a PSNA or PSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of PSNAs or PSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples.

Monitoring a patient for onset of prostate cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a PSNA and/or PSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more PSNAs and/or PSPs are detected. The presence of higher (or lower) PSNA or PSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly prostate cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more PSNAs and/or PSPs of the invention can also be monitored by analyzing levels of expression of the PSNAs and/or PSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in a PSG, thereby determining if a human with the genetic lesion is susceptible to developing prostate cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing prostate cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the PSGs of this invention, a chromosomal rearrangement of PSG, an aberrant modification of PSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a PSG. Methods to detect such lesions in the PSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Prostate Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a sample from a patient suspected of having or known to have a noncancerous prostate disease. In general, the method comprises the steps of obtaining a sample from the patient,

determining the expression level or structural alterations of a PSNA and/or PSP, comparing the expression level or structural alteration of the PSNA or PSP to a normal prostate control, and then ascertaining whether the patient has a noncancerous prostate disease. In general, if high expression relative to a control of a PSNA or PSP is indicative of a particular noncancerous prostate disease, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a PSNA or PSP is indicative of a noncancerous prostate disease, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a PSNA and/or PSP is associated with a particular noncancerous prostate disease by obtaining prostate tissue from a patient having a noncancerous prostate disease of interest and determining which PSNAs and/or PSPs are expressed in the tissue at either a higher or a lower level than in normal prostate tissue. In another embodiment, one may determine whether a PSNA or PSP exhibits structural alterations in a particular noncancerous prostate disease state by obtaining prostate tissue from a patient having a noncancerous prostate disease of interest and determining the structural alterations in one or more PSNAs and/or PSPs relative to normal prostate tissue.

Methods for Identifying Prostate Tissue

In another aspect, the invention provides methods for identifying prostate tissue. These methods are particularly useful in, *e.g.*, forensic science, prostate cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is prostate tissue or has prostate tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising prostate tissue or

having prostate tissue-like characteristics, determining whether the sample expresses one or more PSNAs and/or PSPs, and, if the sample expresses one or more PSNAs and/or PSPs, concluding that the sample comprises prostate tissue. In a preferred embodiment, the PSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 143 through 249, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the PSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 142, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a PSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a PSP is expressed. Determining whether a sample expresses a PSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the PSP has an amino acid sequence selected from SEQ ID NO: 143 through 249, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two PSNAs and/or PSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five PSNAs and/or PSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is prostate tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into prostate tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new prostate tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Prostate Tissue

In another aspect, the invention provides methods for producing engineered prostate tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a PSNA or a PSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of prostate tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal prostate tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered prostate tissue or cells comprises one of these cell types. In another embodiment, the engineered prostate tissue or cells comprises more than one prostate cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the prostate cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more PSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode PSPs having amino acid sequences selected from SEQ ID NO: 143 through 249, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 142, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a PSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial prostate tissue may be used to treat patients who have lost some or all of their prostate function.

Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a PSNA or part thereof. In a more preferred embodiment, the PSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 142, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity

thereto. In another preferred embodiment, the pharmaceutical composition comprises a PSP or fragment thereof. In a more preferred embodiment, the PSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 143 through 249, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-PSP antibody, preferably an antibody that specifically binds to a PSP having an amino acid that is selected from the group consisting of SEQ ID NO: 143 through 249, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium

carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin,

carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as

polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example PSP polypeptide, fusion protein, or fragments thereof, antibodies specific for PSP, agonists, antagonists or inhibitors of PSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention , *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of prostate function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV) , for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a PSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a PSP are administered, for example, to complement a deficiency in the native PSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes a PSP having the amino acid sequence of SEQ ID NO: 143 through 249, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a PSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in PSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a PSP having the amino acid sequence of SEQ ID NO: 143 through 249, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a PSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a PSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a PSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to PSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the PSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.,* Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a PSP, preferably a PSP comprising an amino acid sequence of SEQ ID NO: 143 through 249, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 142, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a PSP, a

fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant PSP defect.

Protein compositions are administered, for example, to complement a deficiency in native PSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to PSP. The immune response can be used to modulate activity of PSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate PSP.

In a preferred embodiment, the polypeptide is a PSP comprising an amino acid sequence of SEQ ID NO: 143 through 249, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 142, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of PSP, or to target therapeutic agents to sites of PSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a PSP comprising an amino acid sequence of SEQ ID NO: 143 through 249, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a PSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 142, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a PSP or have a modulatory effect on the expression or activity of a PSP. Modulators which decrease the expression or activity of PSP (antagonists) are believed to be useful in treating prostate cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small

molecules predicted via computer imaging to specifically bind to regions of a PSP can also be designed, synthesized and tested for use in the imaging and treatment of prostate cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the PSPs identified herein. Molecules identified in the library as being capable of binding to a PSP are key candidates for further evaluation for use in the treatment of prostate cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a PSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of PSP is administered. Antagonists of PSP can be produced using methods generally known in the art. In particular, purified PSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a PSP.

In other embodiments a pharmaceutical composition comprising an agonist of a PSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a PSP comprising an amino acid sequence of SEQ ID NO: 143 through 249, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a PSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 142, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Prostate Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the prostate or to specific cells in the prostate. In a preferred embodiment, an anti-PSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if prostate tissue needs to be selectively destroyed. This would be useful for targeting and killing prostate cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting prostate cell function.

In another embodiment, an anti-PSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring prostate function, identifying prostate cancer tumors, and identifying noncancerous prostate diseases.

EXAMPLES

Example 1: Gene Expression analysis

PSGs were identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer" disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the PSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), highest differential expression for a given cancer (CLASP 4); differential expression in cancer tissue (CLASP 5), and. cDNA libraries were divided

into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into "gene bins," where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with rigorous statistical significant testing taking into account variations in sample size and relative gene abundance in different libraries and within each library (for the equations used to determine statistically significant expression see Audic and Claverie "The significance of digital gene expression profiles," Genome Res 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were analyzed for statistical significance.

The selection of the target genes meeting the rigorous CLASP™ profile criteria were as follows:

(a) CLASP 1: tissue-specific expression: To qualify as a CLASP 1 candidate, a gene must exhibit statistically significant expression in the tissue of interest compared to all other tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 1 candidate.

(b) CLASP 2: detectable expression only in cancer tissue: To qualify as a CLASP 2 candidate, a gene must exhibit detectable expression in tumor tissues and undetectable expression in libraries from normal individuals and libraries from normal tissue obtained from diseased patients. In addition, such a gene must also exhibit further specificity for the tumor tissues of interest.

(c) CLASP 5: differential expression in cancer tissue: To qualify as a CLASP 5 candidate, a gene must be differentially expressed in tumor libraries in the tissue of interest compared to normal libraries for all tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 5 candidate.

The CLASP™ scores for SEQ ID NO: 1-142 are listed below:

SEQ ID NO: 1	DEX0263_1	CLASP5 CLASP1
SEQ ID NO: 2	DEX0263_2	CLASP2
SEQ ID NO: 3	DEX0263_3	CLASP1
SEQ ID NO: 4	DEX0263_4	CLASP2 CLASP1
SEQ ID NO: 5	DEX0263_5	CLASP2 CLASP1
SEQ ID NO: 6	DEX0263_6	CLASP2
SEQ ID NO: 7	DEX0263_7	CLASP2
SEQ ID NO: 8	DEX0263_8	CLASP2
SEQ ID NO: 9	DEX0263_9	CLASP2 CLASP1
SEQ ID NO: 10	DEX0263_10	CLASP2 CLASP1
SEQ ID NO: 11	DEX0263_11	CLASP2 CLASP1
SEQ ID NO: 12	DEX0263_12	CLASP2 CLASP1
SEQ ID NO: 13	DEX0263_13	CLASP2 CLASP1
SEQ ID NO: 14	DEX0263_14	CLASP5 CLASP1
SEQ ID NO: 15	DEX0263_15	CLASP5 CLASP1
SEQ ID NO: 16	DEX0263_16	CLASP2
SEQ ID NO: 17	DEX0263_17	CLASP1
SEQ ID NO: 18	DEX0263_18	CLASP1
SEQ ID NO: 19	DEX0263_19	CLASP5 CLASP1
SEQ ID NO: 20	DEX0263_20	CLASP5 CLASP1
SEQ ID NO: 21	DEX0263_21	CLASP5 CLASP1
SEQ ID NO: 22	DEX0263_22	CLASP5 CLASP1
SEQ ID NO: 23	DEX0263_23	CLASP2 CLASP1
SEQ ID NO: 24	DEX0263_24	CLASP2
SEQ ID NO: 25	DEX0263_25	CLASP2
SEQ ID NO: 26	DEX0263_26	CLASP2
SEQ ID NO: 27	DEX0263_27	CLASP2
SEQ ID NO: 28	DEX0263_28	CLASP2
SEQ ID NO: 29	DEX0263_29	CLASP2
SEQ ID NO: 30	DEX0263_30	CLASP2
SEQ ID NO: 31	DEX0263_31	CLASP2
SEQ ID NO: 32	DEX0263_32	CLASP2
SEQ ID NO: 33	DEX0263_33	CLASP2
SEQ ID NO: 34	DEX0263_34	CLASP2
SEQ ID NO: 35	DEX0263_35	CLASP2
SEQ ID NO: 36	DEX0263_36	CLASP2
SEQ ID NO: 37	DEX0263_37	CLASP2 CLASP1
SEQ ID NO: 38	DEX0263_38	CLASP2 CLASP1
SEQ ID NO: 39	DEX0263_39	CLASP2 CLASP1
SEQ ID NO: 40	DEX0263_40	CLASP2 CLASP1
SEQ ID NO: 41	DEX0263_41	CLASP2 CLASP1
SEQ ID NO: 42	DEX0263_42	CLASP5 CLASP1
SEQ ID NO: 43	DEX0263_43	CLASP5 CLASP1
SEQ ID NO: 44	DEX0263_44	CLASP2 CLASP1
SEQ ID NO: 45	DEX0263_45	CLASP2

SEQ ID NO: 46	DEX0263_46 CLASP2
SEQ ID NO: 47	DEX0263_47 CLASP2
SEQ ID NO: 48	DEX0263_48 CLASP2
SEQ ID NO: 49	DEX0263_49 CLASP2
SEQ ID NO: 50	DEX0263_50 CLASP2
SEQ ID NO: 51	DEX0263_51 CLASP2
SEQ ID NO: 52	DEX0263_52 CLASP2
SEQ ID NO: 53	DEX0263_53 CLASP2
SEQ ID NO: 54	DEX0263_54 CLASP5 CLASP1
SEQ ID NO: 55	DEX0263_55 CLASP5 CLASP1
SEQ ID NO: 56	DEX0263_56 CLASP2
SEQ ID NO: 57	DEX0263_57 CLASP5
SEQ ID NO: 58	DEX0263_58 CLASP2
SEQ ID NO: 59	DEX0263_59 CLASP2
SEQ ID NO: 60	DEX0263_60 CLASP2
SEQ ID NO: 61	DEX0263_61 CLASP2
SEQ ID NO: 62	DEX0263_62 CLASP2
SEQ ID NO: 63	DEX0263_63 CLASP2
SEQ ID NO: 64	DEX0263_64 CLASP2
SEQ ID NO: 65	DEX0263_65 CLASP1
SEQ ID NO: 66	DEX0263_66 CLASP1
SEQ ID NO: 67	DEX0263_67 CLASP5 CLASP1
SEQ ID NO: 68	DEX0263_68 CLASP5 CLASP1
SEQ ID NO: 69	DEX0263_69 CLASP2
SEQ ID NO: 70	DEX0263_70 CLASP2
SEQ ID NO: 71	DEX0263_71 CLASP2
SEQ ID NO: 72	DEX0263_72 CLASP2
SEQ ID NO: 73	DEX0263_73 CLASP2
SEQ ID NO: 74	DEX0263_74 CLASP2
SEQ ID NO: 75	DEX0263_75 CLASP2
SEQ ID NO: 76	DEX0263_76 CLASP2
SEQ ID NO: 77	DEX0263_77 CLASP2
SEQ ID NO: 78	DEX0263_78 CLASP2
SEQ ID NO: 79	DEX0263_79 CLASP2
SEQ ID NO: 80	DEX0263_80 CLASP2
SEQ ID NO: 81	DEX0263_81 CLASP2
SEQ ID NO: 82	DEX0263_82 CLASP2
SEQ ID NO: 83	DEX0263_83 CLASP2
SEQ ID NO: 84	DEX0263_84 CLASP2
SEQ ID NO: 85	DEX0263_85 CLASP2
SEQ ID NO: 86	DEX0263_86 CLASP2
SEQ ID NO: 87	DEX0263_87 CLASP2
SEQ ID NO: 88	DEX0263_88 CLASP2
SEQ ID NO: 89	DEX0263_89 CLASP2
SEQ ID NO: 90	DEX0263_90 CLASP2
SEQ ID NO: 92	DEX0263_92 CLASP2
SEQ ID NO: 93	DEX0263_93 CLASP2
SEQ ID NO: 94	DEX0263_94 CLASP2

SEQ ID NO: 95	DEX0263_95	CLASP1
SEQ ID NO: 96	DEX0263_96	CLASP1
SEQ ID NO: 97	DEX0263_97	CLASP2
SEQ ID NO: 98	DEX0263_98	CLASP2
SEQ ID NO: 99	DEX0263_99	CLASP2
SEQ ID NO: 100	DEX0263_100	CLASP2
SEQ ID NO: 101	DEX0263_101	CLASP2
SEQ ID NO: 102	DEX0263_102	CLASP2
SEQ ID NO: 103	DEX0263_103	CLASP2
SEQ ID NO: 104	DEX0263_104	CLASP2
SEQ ID NO: 105	DEX0263_105	CLASP2 CLASP1
SEQ ID NO: 106	DEX0263_106	CLASP2 CLASP1
SEQ ID NO: 107	DEX0263_107	CLASP1
SEQ ID NO: 108	DEX0263_108	CLASP2
SEQ ID NO: 109	DEX0263_109	CLASP1
SEQ ID NO: 110	DEX0263_110	CLASP1
SEQ ID NO: 111	DEX0263_111	CLASP1
SEQ ID NO: 112	DEX0263_112	CLASP1
SEQ ID NO: 113	DEX0263_113	CLASP2
SEQ ID NO: 114	DEX0263_114	CLASP2
SEQ ID NO: 115	DEX0263_115	CLASP2
SEQ ID NO: 116	DEX0263_116	CLASP2
SEQ ID NO: 118	DEX0263_118	CLASP2
SEQ ID NO: 119	DEX0263_119	CLASP2
SEQ ID NO: 120	DEX0263_120	CLASP2
SEQ ID NO: 121	DEX0263_121	CLASP2
SEQ ID NO: 122	DEX0263_122	CLASP2
SEQ ID NO: 123	DEX0263_123	CLASP2
SEQ ID NO: 124	DEX0263_124	CLASP2
SEQ ID NO: 126	DEX0263_126	CLASP2
SEQ ID NO: 127	DEX0263_127	CLASP2
SEQ ID NO: 128	DEX0263_128	CLASP2
SEQ ID NO: 129	DEX0263_129	CLASP2
SEQ ID NO: 130	DEX0263_130	CLASP5
SEQ ID NO: 131	DEX0263_131	CLASP2
SEQ ID NO: 132	DEX0263_132	CLASP2
SEQ ID NO: 133	DEX0263_133	CLASP2
SEQ ID NO: 134	DEX0263_134	CLASP2
SEQ ID NO: 135	DEX0263_135	CLASP2
SEQ ID NO: 136	DEX0263_136	CLASP2
SEQ ID NO: 137	DEX0263_137	CLASP2
SEQ ID NO: 138	DEX0263_138	CLASP2
SEQ ID NO: 139	DEX0263_139	CLASP5 CLASP1
SEQ ID NO: 140	DEX0263_140	CLASP2 CLASP1
SEQ ID NO: 141	DEX0263_141	CLASP2
SEQ ID NO: 142	DEX0263_142	CLASP2

DEX0263 CLASP expression Level

SEQ ID NO: 1	PRO .0023	INL .0004	OVR .0007	INS .001	
SEQ ID NO: 2	PRO .002				
SEQ ID NO: 3	PRO .0014				
SEQ ID NO: 4	PRO .0017	LNG .0004	UNC .0057		
SEQ ID NO: 5	PRO .0017	LNG .0004	UNC .0057		
SEQ ID NO: 6	PRO .0038	MAM .0007			
SEQ ID NO: 7	PRO .002				
SEQ ID NO: 8	PRO .002				
SEQ ID NO: 9	PRO .0063	BLO .0006			
SEQ ID NO: 10	PRO .0063	BLO .0006			
SEQ ID NO: 11	PRO .0063	BLO .0006			
SEQ ID NO: 12	PRO .0063	BLO .0006			
SEQ ID NO: 13	PRO .0031	LNG .0004	BON .0022		
SEQ ID NO: 14	PRO .0017	FTS .0001			
SEQ ID NO: 15	PRO .0017	FTS .0001			
SEQ ID NO: 16	PRO .0031	MAM .0007			
SEQ ID NO: 17	PRO .0021	TST .0012			
SEQ ID NO: 18	PRO .0021	TST .0012			
SEQ ID NO: 19	PRO .0017	BRN .0001			
SEQ ID NO: 20	PRO .1249	BLO .0003	MAM .0004	FTS .0008	FTS .0009
SEQ ID NO: 21	PRO .1249	BLO .0003	MAM .0004	FTS .0008	FTS .0009
SEQ ID NO: 22	PRO .1249	BLO .0003	MAM .0004	FTS .0008	FTS .0009
SEQ ID NO: 23	PRO .0038				
SEQ ID NO: 24	PRO .0013	BRN .0009			
SEQ ID NO: 25	PRO .0013	BRN .0009			
SEQ ID NO: 26	PRO .0038	BLD .0038			
SEQ ID NO: 27	PRO .0038				
SEQ ID NO: 28	PRO .0038				
SEQ ID NO: 29	PRO .0038				
SEQ ID NO: 30	PRO .0038				
SEQ ID NO: 31	PRO .0038				
SEQ ID NO: 32	PRO .0038				
SEQ ID NO: 33	PRO .0038				
SEQ ID NO: 34	PRO .0038				
SEQ ID NO: 35	PRO .0038				
SEQ ID NO: 36	PRO .0038				
SEQ ID NO: 37	PRO .0057				
SEQ ID NO: 38	PRO .0057				
SEQ ID NO: 39	PRO .004				
SEQ ID NO: 40	PRO .004				
SEQ ID NO: 41	PRO .003				
SEQ ID NO: 42	PRO .0079	MAM .0004	BLO .0006	FTS .0006	CON .0007
SEQ ID NO: 43	PRO .0079	MAM .0004	BLO .0006	FTS .0006	CON .0007
SEQ ID NO: 44	PRO .0031				
SEQ ID NO: 45	PRO .002				
SEQ ID NO: 46	PRO .0013	BRN .0022			

	SEQ ID NO: 47	PRO .0013	BRN .0022		
	SEQ ID NO: 48	PRO .0013			
	SEQ ID NO: 49	PRO .0013			
	SEQ ID NO: 50	PRO .0013			
5	SEQ ID NO: 51	PRO .0013			
	SEQ ID NO: 52	PRO .0013			
	SEQ ID NO: 53	PRO .0013			
	SEQ ID NO: 54	PRO .0017	FTS .0003		
	SEQ ID NO: 55	PRO .0017	FTS .0003		
10	SEQ ID NO: 56	PRO .0032			
	SEQ ID NO: 57	PRO .0011			
	SEQ ID NO: 58	PRO .0013			
	SEQ ID NO: 59	PRO .0013			
	SEQ ID NO: 60	PRO .0051	MAM .0007	TST .0262	
15	SEQ ID NO: 61	PRO .0051	MAM .0007	TST .0262	
	SEQ ID NO: 62	PRO .0038			
	SEQ ID NO: 63	PRO .0038			
	SEQ ID NO: 64	PRO .0065			
	SEQ ID NO: 65	PRO .0011			
20	SEQ ID NO: 66	PRO .0011			
	SEQ ID NO: 67	PRO .0113	INL .0012	LMN .002	
	SEQ ID NO: 68	PRO .0113	INL .0012	LMN .002	
	SEQ ID NO: 69	PRO .0013			
	SEQ ID NO: 70	PRO .0013			
25	SEQ ID NO: 71	PRO .0038			
	SEQ ID NO: 72	PRO .0038			
	SEQ ID NO: 73	PRO .0017			
	SEQ ID NO: 74	PRO .0013			
	SEQ ID NO: 75	PRO .0044			
30	SEQ ID NO: 76	PRO .0044			
	SEQ ID NO: 77	PRO .0038			
	SEQ ID NO: 78	PRO .0038			
	SEQ ID NO: 79	PRO .0038			
	SEQ ID NO: 80	PRO .0038			
35	SEQ ID NO: 81	PRO .0038			
	SEQ ID NO: 82	PRO .0029			
	SEQ ID NO: 83	PRO .0029			
	SEQ ID NO: 84	PRO .0044			
	SEQ ID NO: 85	PRO .0044			
40	SEQ ID NO: 86	PRO .0038			
	SEQ ID NO: 87	PRO .0038			
	SEQ ID NO: 88	PRO .0038			
	SEQ ID NO: 89	PRO .0013			
	SEQ ID NO: 90	PRO .0013			
45	SEQ ID NO: 92	PRO .0038			
	SEQ ID NO: 93	PRO .0038			
	SEQ ID NO: 94	PRO .0038			
	SEQ ID NO: 95	PRO .0011	BRN .0001	INL .0004	KID .0006	CON .0007

SEQ ID NO: 96	PRO .0011	BRN .0001	INL .0004	KID .0006	CON .0007	
SEQ ID NO: 97	PRO .002					
SEQ ID NO: 98	PRO .002					
SEQ ID NO: 99	PRO .002					
SEQ ID NO: 100	PRO .002	CON .0024				
SEQ ID NO: 101	PRO .002	CON .0024				
SEQ ID NO: 102	PRO .0042					
SEQ ID NO: 103	PRO .003					
SEQ ID NO: 104	PRO .0013	BRN .0008				
SEQ ID NO: 105	PRO .0032					
SEQ ID NO: 106	PRO .0032					
SEQ ID NO: 107	PRO .0018					
SEQ ID NO: 108	PRO .002					
SEQ ID NO: 109	PRO .0028	BRN .0003	FTS .0003	UTR .0004	MAM .0016	
SEQ ID NO: 110	PRO .0028	BRN .0003	FTS .0003	UTR .0004	MAM .0016	
SEQ ID NO: 111	PRO .0014					
SEQ ID NO: 112	PRO .0014					
SEQ ID NO: 113	PRO .0044					
SEQ ID NO: 114	PRO .0044					
SEQ ID NO: 115	PRO .0044					
SEQ ID NO: 116	PRO .0044					
SEQ ID NO: 118	PRO .0044					
SEQ ID NO: 119	PRO .0044					
SEQ ID NO: 120	PRO .0044					
SEQ ID NO: 121	PRO .0044					
SEQ ID NO: 122	PRO .0044					
SEQ ID NO: 123	PRO .0044					
SEQ ID NO: 124	PRO .0038					
SEQ ID NO: 126	PRO .0038					
SEQ ID NO: 127	PRO .0038					
SEQ ID NO: 128	PRO .0013					
SEQ ID NO: 129	PRO .002					
SEQ ID NO: 130	PRO .0215					
SEQ ID NO: 131	PRO .0065					
SEQ ID NO: 132	PRO .0065					
SEQ ID NO: 133	PRO .0065					
SEQ ID NO: 134	PRO .0065					
SEQ ID NO: 135	PRO .0065					
SEQ ID NO: 136	PRO .0065					
SEQ ID NO: 137	PRO .0065					
SEQ ID NO: 138	PRO .002					
SEQ ID NO: 139	PRO .0023	BRN .0001				
SEQ ID NO: 140	PRO .0065					
SEQ ID NO: 141	PRO .0038					
SEQ ID NO: 142	PRO .0038					

Abbreviation for tissues:
BLO Blood; BRN Brain; CON Connective Tissue; CRD Heart; FTS Fetus; INL Intestine, Large; INS Intestine, Small; KID Kidney; LIV Liver; LNG Lung; MAM Breast; MSL

Muscles; NRV Nervous Tissue; OVR Ovary; PRO Prostate; STO Stomach; THR Thyroid Gland; TNS Tonsil / Adenoids; UTR Uterus

The chromosomal locations were determined for several of the sequences. Specifically:

DEX0263_1	chromosome 20
DEX0263_15	chromosome 5
DEX0263_17	chromosome X
DEX0263_18	chromosome X
DEX0263_29	chromosome X
DEX0263_30	chromosome X
DEX0263_34	chromosome X
DEX0263_35	chromosome X
DEX0263_47	chromosome 2
DEX0263_55	chromosome 3
DEX0263_56	chromosome 10
DEX0263_57	chromosome 10
DEX0263_59	chromosome 9
DEX0263_63	chromosome X
DEX0263_78	chromosome 12
DEX0263_87	chromosome 12
DEX0263_96	chromosome 3
DEX0263_101	chromosome 1
DEX0263_106	chromosome 14
DEX0263_130	chromosome 11
DEX0263_131	chromosome X
DEX0263_136	chromosome 16
DEX0263_137	chromosome 16
DEX0263_138	chromosome X
DEX0263_142	chromosome 6

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied,

the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the PSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the PSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the PSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 142 being a diagnostic marker for cancer.

Sequences	Sequence ID NO	Gene ID	QPCR prostate code
DEX0100_7	DEX0263_9(SEQ ID NO:9) DEX0263_10(SEQ ID NO:10)	231877	Pro154
DEX0100_50	DEX0263_67(SEQ ID NO:67) DEX0263_68(SEQ ID NO:68)	29050	Pro133

DEX0263_9(SEQ ID NO:9) DEX0263_10(SEQ ID NO:10); Pro154; sqpro046

Experiments are underway to test primers and probes for QPCR.

Experimental results from SQ PCR analysis are included below.

The relative levels of expression of Sqpro046 in 12 normal samples from 12 different tissues were determined. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Tissue	Normal
Breast	0
Colon	0
Endometrium	0
Kidney	0
Liver	0
Lung	0
Ovary	0
Prostate	0
Small Intestine	0
Stomach	0
Testis	0
Uterus	0

Relative levels of expression in the table above show that expression of Sqpro046 is not detected in all 12 normal tissues.

The relative levels of expression of Sqpro046 in 12 cancer samples from 12 different tissues were determined. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression.

A positive reaction in the most dilute sample indicates the highest relative expression value.

Tissue	Cancer
Bladder	0
Breast	0
Colon	0
Kidney	0
Liver	0
Lung	0
Ovary	0
Pancreas	0
Prostate	0
Stomach	0
Testis	0
Uterus	0

Relative levels of expression in the table show that expression of Sqpro046 is not detected in all 12 carcinomas.

The relative levels of expression of Sqpro046 in 6 prostate cancer matching samples were determined. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Sample ID	Tissue	Cancer	NAT
845B/846B	Prostate	1	1
916B/917B	Prostate	10	1
1105B/1106B	Prostate	1	1
902B/903B	Prostate	100	1
1222B/1223B	Prostate	100	10
1291B/1292B	Prostate	1000	10

Relative levels of expression in the table above show that Sqpro046 is expressed in higher level in cancer sample compared with its normal adjacent tissue in four out of six prostate cancer matching samples.

DEX0263_67(SEQ ID NO:67) & DEX0263_68(SEQ ID NO:68); Pro133

The relative levels of expression of Pro133 in 24 normal different tissues were determined. All the values are compared to normal endometrium (calibrator). These RNA samples are commercially pools, originated by pooling samples of a particular tissue from different individuals.

Tissue	NORMAL
Adrenal Gland	0.01
Bladder	0.00
Brain	0.01
Cervix	0.11
Colon	0.09
Endometrium	1.00
Esophagus	0.03
Heart	0.00
Kidney	0.01
Liver	0.00
Lung	0.00
Mammary Gland	0.00
Muscle	0.00
Ovary	0.00
Pancreas	0.00
Prostate	112.99
Rectum	21.33
Small Intestine	0.00
Spleen	0.00
Stomach	0.00
Testis	0.03
Thymus	0.33
Trachea	0.13
Uterus	0.00

0=negative

The relative levels of expression in the table above show that Pro133 mRNA expression is highest in prostate.

The absolute numbers were obtained analyzing pools of samples of a particular tissue from different individuals. They can not be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in the table below.

The absolute numbers are relative levels of expression of Pro133 in 46 pairs of matching samples and 4 prostate normal, and 18 prostatisis & Benign Hyperplasia (BPH) samples and 3 cancer ovary and 1 normal ovary. All the values are compared to normal

prostate (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

Sample ID	Tissue	CANCER	PROSTATISIS & (BPH) BENIGH HYPERPLACIA	MATCHING NORMAL ADJACENT	NORMAL
ProC153	Prostate 1				0.90
Pro53P	Prostate 2				51.45
Pro73P	Prostate 3				9.99
Pro77P	Prostate 4				6.17
Pro12B	Prostate 5	27.57		1.08	
Pro84XB	Prostate 6	100.43		17.27	
Pro101XB	Prostate 7	69.59		79.07	
Pro91X	Prostate 8	58.28		16.51	
Pro78XB	Prostate 9	45.73		47.01	
Pro109XB	Prostate 10	7.89		10.45	
Pro13XB	Prostate 11	1.39		3.53	
Pro125XB	Prostate 12	3.42		5.12	
Pro110	Prostate 13	5.48		40.79	
Pro23B	Prostate 14	80.73		56.3	
Pro65XB	Prostate 15	20.68		40.79	
Pro34B	Prostate 16	95.34		59.71	
Pro90XB	Prostate 17	32.33		31.89	
Pro69XB	Prostate 18	8.51		3.81	
Pro326	Prostate 19	100.43		50.56	
Pro10R	Prostate 20		15.35		
Pro20R	Prostate 21		21.26		
Pro784P	Prostate 22		13.83		
Pro855P	Prostate 23		50.04		
ProC003P	Prostate 24		2.67		
ProC034P	Prostate 25 (prostatisis)		5.96		
Pro10P	Prostate 26 (prostatisis)		35.75		
Pro13P	Prostate 27 (BPH)		1.47		
Pro65P	Prostate 28 (BPH)		10.09		
Pro277P	Prostate 29 (BPH)		36.76		
Pro34P	Prostate 30 (BPH)		2.15		
Pro705P	Prostate 31 (BPH)		2.6		
Pro271A	Prostate 32 (BPH)		7.06		
Pro460Z	Prostate 33 (BPH)		30.7		
Pro258	Prostate 34 (BPH)		10.7		

Pro263C	Prostate 35 (BPH)		46.53		
Pro267A	Prostate 36 (BPH)		8.46		
ProC032	Prostate 37 (BPH)		4.81		
Bld32XK	Bladder 1	0.03		0.04	
Bld46XK	Bladder 2	0.00		0.04	
Bld66X	Bladder 3	0.04		0.15	
Endo 10479	Endometrium 1	4.64		0.00	
Endo 12XA	Endometrium 2	4.59		0.12	
Endo 28XA	Endometrium 3	0.49		0.07	
Endo 5XA	Endometrium 4	0.00		0.69	
Endo3AX	Endometrium 5	0		0	
ClnAC19	Colon 1	1.51		0.00	
ClnAS12	Colon 2	0.50		0.00	
ClnDC22	Colon 3	1.49		0.38	
CvxKS83	Cervix1	0.06		8.43	
CvxNK23	Cervix2	0.77		0.89	
Lng LC80	Lung 1	0.00		0.00	
Lng143L	Lung 2	0.00		0.00	
Lng205L	Lung 3	0.00		0.00	
Kid716K	Kidney1	1.18		0.01	
Kid106XD	Kidney2	0.00		0.00	
Kid107XD	Kidney3	0.00		0.00	
Kid109XD	Kidney4	1.59		0.00	
Mam19DN	Mammary 1	0.12		0.00	
Mam173M	Mammary 2	0.00		0.00	
Mam 162X	Mammary 3	0.47		0.00	
Ovr1118	Ovary1	0.00			
Ovr32RA	Ovary2				0.00
OvrG010	Ovary3	0.03		0.00	
OvrG021	Ovary4	0.49		0.00	
Ovr1005O	Ovary5	0.00			
OvrC057	Ovary6	0.00			
Sto 88S	Stomach1	0.09		0.00	
Sto 115S	Stomach2	0.92		0.00	
Sto15S	Stomach3	0.00		0.00	
Uterus 141XO	Uterus1	0.57		0.00	
Uterus 135XO	Uterus2	0.00		0.00	

0.00= Negative

We compared the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent). Table 2 shows overexpression of Pro133 in

40% of the prostate matching samples tested (6 out of total of 15 prostate matching samples).

Altogether, the tissue specificity, plus the mRNA differential expression in the prostate matching samples tested are believed to make Pro133 a good marker for diagnosing, monitoring, staging,imaging and treating prostate cancer.

Primers Used for QPCR Expression Analysis

In DEX0263_67(SEQ ID NO:67)

Primer Probe Oligo	Start From	End To	queryLength	sbjctDescript
Pr0133For	232	254	23	DEX0100_50
Pro133Rev	375	354	22	DEX0100_50
Pro133Probe	292	268	25	DEX0100_50

In DEX0263_68(SEQ ID NO:68)

Primer Probe Oligo	Start From	End To	queryLength	sbjctDescript
Pro133For	236	258	23	flexsednt DEX0100_50
Pro133Rev	379	358	22	flexsednt DEX0100_50
Pro133Probe	296	272	25	flexsednt DEX0100_50

Example 3: Protein Expression

The PSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the PSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the PSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH_2-terminus of the coding sequence of PSNA, and six histidines, flanking the COOH-terminus of the coding sequence of PSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of PSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column

volumes of wash buffer. PSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5'and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/1 of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, μg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

DEX0263_145	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
11-31	1.08	21
DEX0263_147	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
5-29	1.10	25
DEX0263_148	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
6-18	1.14	13
DEX0263_151	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
46-55	1.30	10
DEX0263_152	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
39-53	1.30	15
DEX0263_153	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
29-38	1.00	10
DEX0263_154	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
45-61	1.01	17
DEX0263_157	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
57-66	1.05	10
DEX0263_158	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length

	76-103	1.03	28
DEX0263_160	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	6-23	1.08	18
DEX0263_161	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	4-25	1.03	22
DEX0263_162	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	63-81	1.12	19
DEX0263_166	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	15-26	1.04	12
DEX0263_168	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	27-37	1.05	11
DEX0263_173	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	346-357	1.06	12
	273-306	1.05	34
	173-191	1.01	19
DEX0263_175	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	3-12	1.16	10
DEX0263_177	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	620-636	1.22	17
	11-37	1.19	27
	417-479	1.14	63
	648-665	1.12	18
	680-697	1.01	18
DEX0263_184	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	8-43	1.09	36
DEX0263_185	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	2-13	1.34	12
	31-47	1.09	17
DEX0263_191	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	11-27	1.15	17
	38-48	1.14	11
DEX0263_199	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	31-42	1.06	12
DEX0263_200	Antigenicity Index(Jameson-Wolf)		
	positions	AI	avg length
	36-50	1.13	15

DEX0263_221	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
63-75	1.05	13
DEX0263_222	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
39-52	1.13	14
DEX0263_228	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
6-19	1.07	14
DEX0263_232	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
95-113	1.03	19
DEX0263_238	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
7-27	1.04	21
DEX0263_240	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
38-48	1.03	11
DEX0263_248	Antigenicity Index(Jameson-Wolf)	
positions	AI	avg length
38-55	1.01	18

Examples of post-translational modifications (PTMs) of the PSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. Using the ProSite database (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by reference), the following PTMs were predicted for the PSPs of the invention (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html most recently accessed October 23, 2001). For full definitions of the PTMs see http://www.expasy.org/cgi-bin/prosite-list.pl most recently accessed October 23, 2001.

DEX0263_143	Camp_Phospho_Site 19-22; Pkc_Phospho_Site 8-10; Tyr_Phospho_Site 17-25;
DEX0263_144	Ck2_Phospho_Site 21-24; Pkc_Phospho_Site 21-23;38-40;55-57;
DEX0263_145	Myristyl 30-35; Pkc_Phospho_Site 2-4;
DEX0263_146	Ck2_Phospho_Site 55-58; Myristyl 24-29;27-32;28-33; Pkc_Phospho_Site 83-85; Prokar_Lipoprotein 23-33;
DEX0263_147	Asn_Glycosylation 36-39; Ck2_Phospho_Site 16-19;38-41; Pkc_Phospho_Site 32-34;38-40;
DEX0263_148	Camp_Phospho_Site 9-12;
DEX0263_149	Myristyl 3-8;
DEX0263_150	Myristyl 12-17;40-45;
DEX0263_151	Amidation 47-50; Camp_Phospho_Site 9-12; Ck2_Phospho_Site 27-30; Pkc_Phospho_Site 6-8;12-14;
DEX0263_152	Camp_Phospho_Site 31-34; Ck2_Phospho_Site 40-43;

DEX0263_153	Ck2_Phospho_Site 17-20;28-31; Pkc_Phospho_Site 66-68;
DEX0263_154	Amidation 20-23; Ck2_Phospho_Site 78-81;
DEX0263_155	Ck2_Phospho_Site 20-23; Pkc_Phospho_Site 24-26;
DEX0263_156	Pkc_Phospho_Site 17-19;30-32;
DEX0263_157	Myristyl 35-40; Pkc_Phospho_Site 58-60;
DEX0263_158	Ck2_Phospho_Site 80-83;97-100; Leucine_Zipper 36-57; Myristyl 29-34;73-78; Pkc_Phospho_Site 3-5;77-79;136-138;
DEX0263_159	Myristyl 10-15;
DEX0263_160	Amidation 14-17; Myristyl 11-16;43-48;
DEX0263_161	Pkc_Phospho_Site 13-15;
DEX0263_162	Ck2_Phospho_Site 76-79; Myristyl 46-51; Pkc_Phospho_Site 29-31;61-63;70-72;91-93;99-101;
DEX0263_163	Myristyl 2-7; Pkc_Phospho_Site 40-42;
DEX0263_165	Pkc_Phospho_Site 12-14;
DEX0263_166	Pkc_Phospho_Site 32-34;
DEX0263_168	Amidation 44-47; Myristyl 17-22; Pkc_Phospho_Site 10-12;44-46;
DEX0263_169	Pkc_Phospho_Site 3-5;
DEX0263_170	Ck2_Phospho_Site 3-6; Pkc_Phospho_Site 9-11;
DEX0263_172	Ck2_Phospho_Site 28-31;36-39;40-43;48-51; Myristyl 6-11;7-12;71-76;84-89;87-92; Pkc_Phospho_Site 28-30;
DEX0263_173	Amidation 321-324; Asn_Glycosylation 69-72;360-363; Camp_Phospho_Site 353-356; Ck2_Phospho_Site 264-267;305-308; Homeobox_1 355-378; Myristyl 24-29;37-42;52-57;67-72;100-105;146-151; Pkc_Phospho_Site 222-224;318-320;343-345;362-364; Tyr_Phospho_Site 224-231;335-341;
DEX0263_174	Pkc_Phospho_Site 12-14;
DEX0263_175	Myristyl 40-45; Pkc_Phospho_Site 4-6;
DEX0263_177	Asn_Glycosylation 62-65;384-387;490-493;593-596; Camp_Phospho_Site 150-153; Ck2_Phospho_Site 27-30;31-34;64-67;291-294;329-332;380-383;416-419;470-473;586-589;643-646;648-651; Myristyl 12-17;73-78;292-297;536-541;571-576; Pkc_Phospho_Site 18-20;66-68;109-111;149-151;329-331;345-347;359-361;404-406;445-447;464-466;478-480;491-493;586-588;
DEX0263_178	Pkc_Phospho_Site 11-13;
DEX0263_179	Asn_Glycosylation 68-71; Ck2_Phospho_Site 70-73; Myristyl 59-64;64-69; Tyr_Phospho_Site 10-17;
DEX0263_180	Camp_Phospho_Site 5-8; Ck2_Phospho_Site 18-21; Pkc_Phospho_Site 8-10;35-37;
DEX0263_181	Amidation 74-77; Asn_Glycosylation 55-58; Ck2_Phospho_Site 15-18;48-51; Myristyl 6-11;74-79; Pkc_Phospho_Site 26-28;61-63;
DEX0263_183	Ck2_Phospho_Site 4-7; Pkc_Phospho_Site 44-46;
DEX0263_184	Asn_Glycosylation 10-13; Camp_Phospho_Site 22-25; Ck2_Phospho_Site 31-34; Tyr_Phospho_Site 62-69;
DEX0263_185	Camp_Phospho_Site 56-59; Myristyl 31-36;48-53; Pkc_Phospho_Site 36-38;

DEX0263_186	Camp_Phospho_Site 161-164;241-244; Ck2_Phospho_Site 218-221; Myristyl 57-62;153-158; Pkc_Phospho_Site 70-72;225-227;
DEX0263_188	Myristyl 9-14;
DEX0263_189	Pkc_Phospho_Site 2-4;
DEX0263_191	Asn_Glycosylation 60-63; Ck2_Phospho_Site 2-5; Myristyl 25-30; Pkc_Phospho_Site 17-19;
DEX0263_192	Amidation 14-17; Myristyl 5-10; Rgd 8-10;
DEX0263_193	Ck2_Phospho_Site 65-68; Myristyl 17-22;54-59;
DEX0263_194	Ck2_Phospho_Site 14-17; Prokar_Lipoprotein 26-36;
DEX0263_196	Asn_Glycosylation 29-32; Myristyl 30-35; Pkc_Phospho_Site 31-33;35-37;38-40;
DEX0263_197	Myristyl 26-31;
DEX0263_198	Asn_Glycosylation 8-11; Pkc_Phospho_Site 7-9;22-24;
DEX0263_199	Camp_Phospho_Site 74-77; Ck2_Phospho_Site 31-34;47-50;65-68; Myristyl 9-14; Pkc_Phospho_Site 2-4;60-62;77-79;
DEX0263_200	Asn_Glycosylation 43-46;63-66; Ck2_Phospho_Site 23-26;47-50;70-73; Myristyl 29-34; Pkc_Phospho_Site 47-49;
DEX0263_201	Pkc_Phospho_Site 8-10;
DEX0263_204	Pkc_Phospho_Site 6-8;41-43;
DEX0263_205	Myristyl 42-47;
DEX0263_208	Pkc_Phospho_Site 10-12;
DEX0263_209	Asn_Glycosylation 16-19;
DEX0263_210	Pkc_Phospho_Site 32-34;
DEX0263_212	Myristyl 52-57;
DEX0263_213	Ck2_Phospho_Site 13-16; Myristyl 8-13;19-24; Pkc_Phospho_Site 12-14;
DEX0263_214	Asn_Glycosylation 16-19; Myristyl 14-19;15-20;23-28;26-31; Pkc_Phospho_Site 9-11;
DEX0263_216	Asn_Glycosylation 12-15;
DEX0263_220	Ck2_Phospho_Site 3-6;
DEX0263_221	Ck2_Phospho_Site 13-16; Myristyl 19-24;68-73;
DEX0263_222	Ck2_Phospho_Site 10-13; Myristyl 28-33;
DEX0263_223	Ck2_Phospho_Site 47-50; Myristyl 35-40; Pkc_Phospho_Site 47-49;
DEX0263_224	Asn_Glycosylation 13-16;54-57; Myristyl 12-17; Pkc_Phospho_Site 2-4;40-42;83-85;
DEX0263_225	Ck2_Phospho_Site 44-47; Pkc_Phospho_Site 34-36;61-63;
DEX0263_226	Pkc_Phospho_Site 46-48;
DEX0263_227	Ck2_Phospho_Site 5-8; Myristyl 13-18;14-19;46-51; Pkc_Phospho_Site 26-28; Prokar_Lipoprotein 34-44;41-51;
DEX0263_229	Camp_Phospho_Site 8-11; Pkc_Phospho_Site 7-9;
DEX0263_230	Camp_Phospho_Site 30-33; Pkc_Phospho_Site 28-30;
DEX0263_232	Asn_Glycosylation 143-146; Camp_Phospho_Site 103-106; Ck2_Phospho_Site 12-15;25-28;158-161;165-168;188-191;211-214;247-250;289-292; Myristyl 19-24;258-263;284-289;301-306; Pkc_Phospho_Site 45-47;121-123;135-137;247-249;
DEX0263_234	Camp_Phospho_Site 13-16; Pkc_Phospho_Site 18-20;
DEX0263_237	Amidation 2-5; Myristyl 19-24; Tyr_Phospho_Site 7-15;

DEX0263_238	Ck2_Phospho_Site 11-14;
DEX0263_239	Tyr_Phospho_Site 15-23;
DEX0263_240	Asn_Glycosylation 16-19; Ck2_Phospho_Site 36-39;44-47; Pkc_Phospho_Site 6-8;25-27;
DEX0263_241	Myristyl 40-45; Pkc_Phospho_Site 18-20;
DEX0263_242	Asn_Glycosylation 22-25; Ck2_Phospho_Site 28-31; Pkc_Phospho_Site 43-45;
DEX0263_243	Myristyl 6-11;
DEX0263_244	Pkc_Phospho_Site 12-14;
DEX0263_245	Asn_Glycosylation 25-28; Camp_Phospho_Site 2-5; Ck2_Phospho_Site 19-22;36-39; Myristyl 23-28;40-45;41-46; Pkc_Phospho_Site 5-7;
DEX0263_246	Ck2_Phospho_Site 3-6;7-10;50-53; Myristyl 11-16; Pkc_Phospho_Site 7-9;35-37;
DEX0263_247	Myristyl 7-12;
DEX0263_248	Asn_Glycosylation 57-60; Pkc_Phospho_Site 22-24;28-30;49-51;
DEX0263_249	Asn_Glycosylation 49-52; Myristyl 50-55;

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See,* Sambrook (2001), *supra.* The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 136. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al., Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al., Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al., Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Manheim), and FISH is performed as described in Johnson *et al., Methods Cell Biol.* 35: 73-99

(1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 μg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear

scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers

of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine

or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.

Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 μg/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine

sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5'and 3'end sequences respectively as set forth in Example 1. Preferably, the 5'primer contains an EcoRI site and the 3'primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 μg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard

recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene

transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon,"Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences

required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512

(1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered

cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 143 through 249;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 142;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 60% sequence identity to the nucleic acid molecule of (a) or (b).

2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.

3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.

4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.

5. The nucleic acid molecule according to claim 4, wherein the nucleic acid molecule is a human nucleic acid molecule.

6. A method for determining the presence of a prostate specific nucleic acid (PSNA) in a sample, comprising the steps of:

(a) contacting the sample with the nucleic acid molecule according to claim 1 under conditions in which the nucleic acid molecule will selectively hybridize to a prostate specific nucleic acid; and

(b) detecting hybridization of the nucleic acid molecule to a PSNA in the sample, wherein the detection of the hybridization indicates the presence of a PSNA in the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 143 through 249; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 142.

12. An antibody or fragment thereof that specifically binds to the polypeptide according to claim 11.

13. A method for determining the presence of a prostate specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under conditions in which the antibody will selectively bind to the prostate specific protein; and

(b) detecting binding of the antibody to a prostate specific protein in the sample, wherein the detection of binding indicates the presence of a prostate specific protein in the sample.

14. A method for diagnosing and monitoring the presence and metastases of prostate cancer in a patient, comprising the steps of:

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the prostate specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of prostate cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with prostate cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the prostate cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

SEQUENCE LISTING

<110> Sun, Yongming
Recipon, Herve
Chen, Sei-Yu
Liu, Chenghua
diaDexus, Inc.

<120> Compositions and Methods Relating to Prostate Specific Genes and Proteins

<130> DEX-0264

<140>
<141>

<150> 60/246,039
<151> 2000-11-06

<160> 249

<170> PatentIn Ver. 2.1

<210> 1
<211> 302
<212> DNA
<213> Homo sapiens

<400> 1
atgtaatgga aattgggttt taagtattga attatcatgc agtatacaaa ggggaggtac 60
tattcaatta attcaatagt aacttcccct tggtatactg catgataata tactttgagt 120
cttacaagta aaaatccttc aagatagtta aggcagttat tagagtaata cctttaaagt 180
gattgtggca tattgaaagt cctcaataaa tgccagtttc ctttctttcc ttaaaaatga 240
aaccgtggaa tttcatgctc attgaggcat cctgaaccat agaaaagcgc tcaaaagcct 300
aa 302

<210> 2
<211> 409
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (347)
<223> a, c, g or t

<220>
<221> unsure
<222> (369)
<223> a, c, g or t

<220>
<221> unsure
<222> (378)
<223> a, c, g or t

<220>
<221> unsure

<222> (402)
<223> a, c, g or t

<400> 2
tgtgcttact gctttcataa tccatctaat tgtccttgaa gcttccaagc tgtagggttt 60
ggcttcactg ataattttat ggtcacatta ataattgggc tgcagcatct gaatatgttc 120
tccaaataga gatgaaacag ataagttatt tctacaggac agggaacaca gagtgctgtt 180
atggaatttc cctatttttt ccttctcaac ttggaaacac atcagaaaac agtactttga 240
atgattctct atattgccat tgccataagg aagaacaacc taaaaacttt aaactgctag 300
agcagattct tctccagtga cagtagggct ggagatcttt tgtagantag aagtttagag 360
aggagaaant taggttantg atgtaaaacc tgcagtggaa gnacatttg 409

<210> 3
<211> 494
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (154)..(239)
<223> a, c, g or t

<400> 3
cttggaaaag tctttaaaat gctcatagaa gcacatataa ttaagtagct ataattacag 60
tgggttttag gatctttttc catactccaa gtattctata tgtaacagag aaatctaaca 120
gctgactcca tcttgcctct aacctcacga gctnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnng 240
tcccttccca aaaccaactc ctgagtatat aaggagggtg tacacacaag tagcaatgtt 300
atgctaaaga tttataggaa cactgtgaac taaccaaaga ccaaagaagt tctgcaaaat 360
tctctggacc cctctgctga tacccagatg tctgtggtca ctggttgcct cctgatcgca 420
acccccttgtt tgttcccctt tccccagtat aaaaagaagc ttgagagtca tgtcttttca 480
gatagttctt tagg 494

<210> 4
<211> 613
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (24)
<223> a, c, g or t

<220>
<221> unsure
<222> (327)
<223> a, c, g or t

<400> 4
ctggaggaca aacaaaggag gagntttggt gcaggagtta ggacaaagac tcctgggaca 60
gcagccctca cttggggcca tagtgaatga atggggccct tttgcagaag ggaaatggag 120
gacatgtgtg gtgaaggctt tgctgaagca tctcttataa aaggggcaaa agcacttgct 180
ctcctcccag ctctggctca gcctctctct cttcaagtcc ctggcatccc tggaaaacca 240
tgatttttcc attctgcagc cagcccaggc acctcctccc acccccgcac actgcagtct 300
ggtgtgtgtg gtgggcacac atagaanact tgctggtctt gaaaatgaag ccatcttttg 360
gggcccctag aaatcaaaga gacccactta gccaagatat cactcttgga caccagcaag 420
ggtgagatca cttgataggt attttgccgt gtgtttcttt ttaactatat aaaaaacaaa 480

agggtggatg tgcaaaccca acccaagaat accttggctc aggctcaact agatgaaact 540
tttaaaaatg atttaatgtt atgttagact gttgcatatt aaagcgatgt aatgccttct 600
aagaaacgag agt 613

<210> 5
<211> 1145
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (215)..(508)
<223> a, c, g or t

<400> 5
ggcaatggaa gcgacagctc cccatcttcc ccggctgcat acgtgaggga ttctgctgca 60
ttcactgagt caccaaagga cactggaagg ccatcaggat ttttgggtga ccaagacatg 120
gctcccctct tactgcctgt ggggagacag gcatgtgaac acccaagtct aacccaagct 180
gaattgactg agggccattt aaagattcca cagannnnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnat aaagaaaggt tccaaggatg gtctggagga 540
caaacaaagg aggagatttg gtgcaggagt taggacaaag actcctggga cagcagccct 600
cacttggggc catagtgaat gaatggggcc cttttgcaga agggaaatgg aggacatgtg 660
tggtgaaggc tttgctgaag catctcttat aaaaggggca aaagcacttg ctctcctccc 720
agctctggct cagcctctct ctcttcaagt ccctggcatc cctggaaaac catgattttt 780
ccattctgca gccagcccag gcacctcctc ccacccccgc acactgcagt ctggtgtgtg 840
tggtgggcac acatagaaca cttgctggtc ttgaaaatga agccatcttt tggggccccc 900
agaaatcaaa gagacccact tagccaagat atcactcttg gacaccagca agggtgagat 960
cacttgatag gtattttgcc gtgtgtttct ttttaactat ataaaaaaca aaagggtgga 1020
tgtgcaaacc caacccaaga ataccttggc tcaggctcaa ctagatgaaa cttttaaaaa 1080
tgatttaatg ttatgttaga ctgttgcata ttaaagcgat gtaatgcctt ctaagaaacg 1140
agagt 1145

<210> 6
<211> 879
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (554)
<223> a, c, g or t

<220>
<221> unsure
<222> (567)
<223> a, c, g or t

<400> 6
caccaagttc tccaatccct gactctgatt cctgacatca atctattgct ccttgtataa 60
tcagtcaatg aatcaaaaaa aatttttttg gatatctact atctcctcag gatgaggctc 120
tgtcagggga aaagtaagaa cccagagaac aagcaataag caaagcagcc taatgttgac 180
atgcgcttgg caagtattca cagaccacct cacacccagc caagcactgc aggtgagagc 240
aacacaggag tgaggaagcc ggggtactta ccatcagtga ggaccaatct tacagatagg 300

gaaaaactat atttcatcca gttgaagaca cccattttct acattttaaa attcctaaat 360
taggctttat ctttcaatga ttttggatca ttggatgaaa tacagtaaca acagcaccaa 420
agcagcattg tgccctgaga ggtgtgcaga gcctatagca attagaacag tggtccaagg 480
tcaggctaca cattagatgt atctgggaaa ctcaaaaata cttatgccca gaccccatcg 540
cagaccaaca gaancagtgt ggaaggntgt ggctgtacac tggtatttca aaaagcacct 600
catgggattc taatacttag ggcagagttg agaaccaggg atagcaattt caaatgagag 660
tacacattac cctgaggtta catgaagact ttccaagggg cacgccttgg aaagtttagg 720
gcacacatca gtaaactagg cctatggccc aaatctgttc tactgcctgt ttttgtacct 780
ccggggagct aagttggttt ttacagtttc aaatggtcag gggaaaaaat caaaagggaa 840
ataatatgtt tggacaaatg aaaattatat gaaatttaa 879

<210> 7
<211> 544
<212> DNA
<213> Homo sapiens

<400> 7
acagcttcat tattaaaatt gtggaattct gtagtctggg tataccaaat atacttaact 60
gtttgcccat cagtgaacat tttagatttt ttttaatatt acgatactac ctacaatgct 120
acagtgaaca aatacacagg tatctacaca cagagtttta cacacatgtg gaagtatttc 180
tgcaagctag cttttaaaat gtagaattgc tgagttaaaa gataggctca cagaattgta 240
aaatgactga ggtaagtctg aatcacttta gaagtttatt tttccaaggt tgaggagatg 300
cctgggaaga aaagacaagc cacagcagga tctgtgccct gtacttttc tgaagaggtt 360
tgaggccttc agcatttaaa gaggaaaagc gagcaggagg agaaaaggaa aagaaaaaat 420
gagaggatgt ggtcacattc ttgtaaggtt ttgattaggc ttactgaatc cacatgttgc 480
acatgaaaag gaaggggtag agggaacagt gaattttgta tttggagtta aagtaaacat 540
agag 544

<210> 8
<211> 1028
<212> DNA
<213> Homo sapiens

<400> 8
cttaaatttt ataattatca tatgcattgc acatccaata ttagcttact ttgactttcc 60
ttctggcctg catctgcttc attttatatt ccttaggagc caactttgcc tgagtaatta 120
gaaacaacat ctataaagag cagataaatg ttaagaacac ctgttttatt tcccacatac 180
aaatttacac aaataatatg tgaatttgtg ttatttttta gaagaaacaa agcacaaata 240
catacaggaa aaaggatggt catactctgt atctgtttgt ctttgcccat accgaattcc 300
actagccttc tctgtggatc acaagtatga agtttgattt atatcatcgt gacacctttt 360
taaaactgca tgtatttctc ctttgttatg taaataagaa tactatacct actctaaaag 420
ttagattttg ttgcctcaaa taatagatgt aagagatctt gttgccaaga tcagtattta 480
aagaacagct tcattattaa aattgtggaa ttctgtagtc tgggtatacc aaatatactt 540
aactgtttgc ccatcagtga acattttaga ttttttttaa tattacgata ctacctacaa 600
tgctacagtg aacaaataca caggtatcta cacacagagt tttacacaca tgtggaagta 660
tttctgcaag ctagctttta aaatgtagaa ttgctgagtt aaaagatagg ctcacagaat 720
tgtaaaatga ctgaggtaag tctgaatcac tttagaagtt tatttttcca aggttgagga 780
gatgcctggg aagaaaagac aagccacagc aggatctgtg ccctgtactt tttctgaaga 840
ggtttgaggc cttcagcatt taaagaggaa aagcgagcag gaggaaaaaa ggaaaagaaa 900
aaatgagagg atgtggtcac attcttgtaa ggttttgatt aggcttactg aatccacatg 960
ttgcacatga aaaggaaggg gtagagggaa cagtgaattt tgtatttgga gttaaagtaa 1020
acatagag 1028

<210> 9
<211> 798
<212> DNA

<213> Homo sapiens

<220>
<221> unsure
<222> (473)..(552)
<223> a, c, g or t

<400> 9
attgcctctc ttttgagcag atgccaggca ttcaagtcac tgtgaatacc ctttgggctt 60
tttgcaactg tgatcttgac cagaagaaaa ctaaagaggg cattaacatg aaactctata 120
ttcttctttt gctgttatgc acctgcctca gatttctctg agaaaattat aaaggatctt 180
ttcccattta tcaacccaac tctgttccaa accacataag ctttgctaat gtgattagaa 240
attgttgaaa acagatctgc cttgcacaga atttcctaaa tggataatta agcattcttg 300
tgacttagaa cccttcgtgc tgcctttgaa agataaattc attaaggaaa attcttcatg 360
acagcagtta ccatttatct cactctaatg agtgccaggc atgtgctttc acgttacttc 420
acttgatcct tactgcctca gaggtaggtt ttaattagta tccccacttc acnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nntaacccta acttaatacc acctctgtgg tagtgttcat ttcctattta 600
ggatactgtt cattagaata cagaggagca ggtgtgatgt ttgagtattt aaaaagacat 660
tttccccact ttggtggtag tggtcattac tgttctgata gctgtcactt acatagcatt 720
tgctgtgtgc caggcactgt tgtgaatact ttgtaggata tattaaatca gtcacacccc 780
ttttacaggt gaggaaac 798

<210> 10
<211> 1305
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (980)..(1059)
<223> a, c, g or t

<400> 10
actgaactcc agtgcctgct aatcccagga ctgcccaggt tctgtttgga aggcccttct 60
gtggtcttca cttggcttct ctttgtcact caggtctcaa ccaagatgtc accctctcag 120
agagcccttc cttgactcct ctccctcatc taaagctccc ccccaacccc agtcattatc 180
taatgccccc ccttactttc ttcaagccac ttaagatttg aaattaagga tttgttttca 240
taacctgtga aattatttgt ttatatggct atcatctgtc tctccatccc aaaagttcac 300
tccatgagag caggcagttt tcgcactccc tgctgtagtt ccccagtgtc tagaatcctg 360
cttggaccct agaaagcatt cactgcatca gtgtttactg aatgcctact gagatgggaa 420
ctgacagttt gtgactgtaa aatgatactt tacaaactag tacatcccaa accattgtgc 480
cctatatgtg agtcttccct taggcacatt gcctctcttt tgagcagatg ccaggcattc 540
aagtcactgt gaataccctt tgggcttttt gcaactgtga tcttgaccag aagaaaacta 600
aagagggcat taacatggaa ctctatattc ttcttttcct gttatgcacc tgcctcagat 660
ttctctgaga aaattataaa ggatcttttc ccatttatca acccaactct gttccaaacc 720
acataagctt tgctaatgtg attagaaatt gttgaaaaca gatctgcctt gcacagaatt 780
tcctaaatgg ataattaagc attcttgtga cttagaaccc ttcgtgctgc ctttgaaaga 840
taaattcatt aaggaaaatt cttcatgaca gcagttacca tttatctcac tctaatgagt 900
gccaggcatg tgctttcacg ttacttcact tgatccttac tgcctcagag gtaggtttta 960
attagtatcc ccacttcacn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnt aaccctaact taataccacc 1080
tctgtggtag tgttcatttc ctatttagga tactgttcat tagaatacag aggagcaggt 1140
gtgatgtttg agtatttaaa aagacatttt ccccactttg gtggtagtgg tcattactgt 1200
tctgatagct gtcacttaca tagcatttgc tgtgtgccag gcactgttgt gaatactttg 1260
taggatatat taaatcagtc acaccccttt tacaggtgag gaaac 1305

<210> 11
<211> 416
<212> DNA
<213> Homo sapiens

<400> 11
gtctttgatc ttttgtcagc cagcatcctc cagacaatga tggggccagc ctctgcgatg 60
ggaatacagg tgtggattcg agttcctgcc ttcaggaaga tcccagtcta tgaggaactc 120
tcaccatcta gttggggaag gagggtgcac agtcacagtc ggtctaagct tattggctag 180
gtttgtccaa aaggaatatt taccaacagc aacctttca cagacaggga ccagatctgc 240
atttctaatt tttatattat tgtgtgttaa tctcctccat ctagtgtatc acttagagag 300
agatggtcag gaaaggcctg cagcgggaga gaacctgtgc tttatagtcc aacagctgaa 360
ggtttgactg cctggtcgag aaagctgaga aagactgtta agaaatttgg caataa 416

<210> 12
<211> 582
<212> DNA
<213> Homo sapiens

<400> 12
gtctttgatc ttttgtcagc cagcatcctc cagacaatga tggggccagc ctctgcgatg 60
ggaatacagg tgtggattcg agttcctgcc ttcaggaaga tcccagtcta tgaggaactc 120
tcaccatcta gttggggaag gagggtgcac agtcacagtc ggtctaagct tattggctag 180
gtttgtccaa aaggaatatt taccaacagc aacctttca ctagacaggg acctagtatc 240
tgcatttcta atttttatat tattgtgtgt taatctcctc catctagtgt atcacttaga 300
gagagatggt caggaaaggc ctgcagcggg agagaacctg tgctttatag tccaacagct 360
gaaggtttga ctgcctggtc gagaaagctg agaaagactg ttaagaaatt tggcttataa 420
gtcccagcat ggtggctcta ctacgctgta tatccctagc actttgggac gcccaggctg 480
gcagatctac ttgtagccca ggagtttgag agccagcctg atagaacatg gtgaatacct 540
catctctact gaaaatgaat gatagccagc gctgtgatga gt 582

<210> 13
<211> 422
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (79)..(254)
<223> a, c, g or t

<400> 13
gctcgaggca gaaataggct agaattcgaa gatagatgtt gtatggcaaa gtggctgagc 60
ttcagctctg gagagagann nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnntggcat atattaagta ttcaatacat gttacctatg acttgaattt 300
aattattatt aggagtattg gtgtatacat cagtaattgc cacagttttt acttttgcat 360
taaaattcag attctggcca ctgcantcca ncatgcgtaa cagagagtga cctctctctc 420
tc 422

<210> 14
<211> 373
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (289)
<223> a, c, g or t

<220>
<221> unsure
<222> (316)
<223> a, c, g or t

<220>
<221> unsure
<222> (361)
<223> a, c, g or t

<400> 14
gagccaccac acccggccag aggttctaac tcctcaccaa cccattcccc ttcctctcct 60
tcctttcaat gcgctcagat aagggcttga ccctgactgg aggatttcaa gttcttcact 120
tctagattca gccttttttat ggggagccat gaaagacagg gatcttaaac tttgttttct 180
gcgacaagtt gagagaaagc acatgggcag gaggtgaggc caggccaggg actggggaga 240
gaatggactc aggcctggct ggaatctctg cctgttcact ttacacagnc tggaccagct 300
acacaccgag ccccgngtcc gagccggaga attcactgtc tctttaaaac catttagagg 360
naaaacaaaa taa 373

<210> 15
<211> 2764
<212> DNA
<213> Homo sapiens

<400> 15
cagataacac agggacttaa aggctgtggg aagcgtttgg atcttactgt tttattccaa 60
gtgctaaaag gaagccactc atctggctgc caagcggaga aggggctgga gtgaccagtg 120
gaagctggaa aggtacattt ttaaggtaca actttaagtt ctaatttcca ccacaatgcc 180
accagagttc tggcagtgac ctgccttaca cctcaccttg ccttttcacg caaaggctgc 240
taagaatggg gaagaatagc aacttaattt cttaaggtct gcagagtgga cagattcaat 300
cctagacaaa cttacaacat agtacttatt tcaggtcctc tcctagcact ttttttctgg 360
taattcttct gctatattct cctctcacat tcactggtgt tctcttccca cctctggcta 420
tctagaaatc ttccatccca aatagatggg ttcatgcttc ccatgatagg cagaataatt 480
attccatcag tgctttttga atgtcagtaa aacaaattaa gccagtcata ggaacagatt 540
aaaggtagct gcattttaag agggatttat ttatttttag ctcaaacaaa tctcaaaata 600
gattgggacg gtatagccca ttatttatgg gaatatacat ttggggaagg aaatatgctt 660
gccagcctga tacactgaca atgttagacc atccttggaa gtaagaggag gaggggttcg 720
cataagaagt agtagttgtg agttagtgtc attctcagag cctccacccg tgatgatggg 780
agatgtgtaa aacctaagtt ttacaaacta gtggatgtgg tgcccaggct ctggtaatga 840
gctgccctgt ttctccgtat gagacaagtg ataggggctc acaagacatt gttccattca 900
ttgactactg aaaatataat caggggatta taaattgcct ccaagtctcc atcactatcc 960
tgtttggaag agtgaggttg agagcctggt ttgggaattg ttaacaaaag taatgacact 1020
gatttgggtg aattatgcct aatattcatg aataaaatcg ttatgcatgt aaatgtcaga 1080
ctaactttgt atattaaagt ttaaatacaa actttatttt gcacagctat ttagagggaa 1140
catttattga atagcttttg gtattatgaa acttttgaat taactcagta tcttaaattt 1200
aagagcagct tggtttctga tgtaattaaa atacttcatt tggctgtttt cttctgaatc 1260
tttctacatt tttcaaatgt attttatttt agagaagaaa tgggtctaaa gatagcctga 1320
tggaagaaaa acctcagaca tctacaaaca acctggctgg aaaacacaca gcaaaaacaa 1380
taaaaactat acaagcttcc cgcctcaaga cagagacttg atcctgatga agggtcaagg 1440
gtaggggtgg gaaggttgtg tgcgccactg gtacttttga aactgtgaaa taggtatctt 1500
aattcaaatc tcagacctgc aagtatttct tcagcatgag aaaatacatt atcttttgct 1560
tctttttttt ttttttttga gatgttatca ctctgtcgcc caggctggag tgcagcggca 1620
ccgtgtcagc tcaccgcagc ctccacttac tgggttaagc gattctcctg tctcaggcta 1680

ccgagcagct gggattacag gcgtgcacca caacacccgg ctaattcttt ttgtattttt 1740
agtagagaca gggctttgcc atgttggagg ctggtctcga actcctgacc tcaagtgatc 1800
cgcctgcctc agcctcccaa agtgatgaga ttacaggcat gagccaccac acccggccag 1860
aggcttctaa ctcctcacca acccattccc cttcctctcc ttcctttcaa tgcgctcaga 1920
taagggcttg accctgactg gaggatttca agttcttcac ttctagattc agccttttta 1980
tggggagcca tgaaagacag ggatcttaaa ctttgttttc tgcgacaagt tgagagaaag 2040
cacatgggca ggaggtgagg ccaggccagg gactggggag agaatggact caggcctggc 2100
tggaatctct gcctgttcac tttacacagc ctggaccagc tacacaccga gccccgcgtc 2160
cgagccgaga attcactgtc tctttaaaac catttagagg aaaaacaaaa taactttagt 2220
gcattaattc tagatgaaga tgggtctcca gatgacagcc tgatccacat tattccctgc 2280
aagacaactt cccctctcct ccgccccctc ccgccttcct gcccctctgc ccagctcttt 2340
gtcctcctcc gtggttcttg ctttgtgaaa agcggacacc ggggtggcat gaagcctaat 2400
cgtgtgcttg atgtttgaga agtcaatgct tctgtctgtt tgggacatca cgttcatgtc 2460
atgggtgagg tattgacttg cagccgtgga ccctgtggtt gcaaagtgct gggagccagt 2520
gtgtctcgga accaccgagc agccgtggtg gggttgcggc tgggggcacg tgagaagacg 2580
gtgaggggtc cgtgtctcgg cgcttgctag ggttaggcac attaggtctt agaatagcac 2640
caaaatgttc tagatctctc ctggctttgg gggaaaaaac acacagatac gcacaaactg 2700
cccagtaagc aaaaaataat tcagatgaga ttcgaatcag gtttgttaat atataactat 2760
atat 2764

<210> 16
<211> 880
<212> DNA
<213> Homo sapiens

<400> 16
gtttacttgg aggaggaaga gactgacaga aaaggcaaaa ttttaatggc ctctctagta 60
agcttcacca caagattcct ttggggtagg gtcatgtctc tggtaccgtg catgtccaat 120
accttgtgca gtgcatggct tgtgtctgtc actgaataac ctatattgag tgaagaaatg 180
ctgtggagtg caggagaggg gcatctaacc tggggtgaga cagagcagcc cctaccacat 240
gtacatcaag ttaagcaatg ggaagaagat ctaggtagaa ggagcagcag atggcataga 300
gatgtgagag aaagtacaag ttgctctgta aactgacaat agtttagttt tgccagagtg 360
taggattcat aagagacagt gctgagagac tttagactag agagttagac ttggcttgac 420
aagtaagggt cataataaag agtttggatt ttgttttctt taaaagtaat agtgtttttt 480
tctaaacaat agatatcagg tttcctaatg cccaggccga cactcttcat attacatgat 540
gcatgcaatg tcaaaaaagg attattaggt cttagtttcc atatccgttc atgatagcag 600
gaccatcaaa attctgtaaa gccattggtg aatgttctgt cttttggaga gaaagaggat 660
ataggataaa tacctttttt cttgggagca aagaatatgt agttgaaact cattaagggc 720
aagcaaggtc tgagtgtctt tgaatttttg ttcctggaat gatgccagga atgtatagca 780
gatctctgta catgaataaa tgaaaaaacc tttctttgca tttgcatata ctgttgctag 840
cattatactt ggttctgaaa agtaatagtt aaatgagaaa 880

<210> 17
<211> 719
<212> DNA
<213> Homo sapiens

<400> 17
gcgctgaggt tcccaacctg gaccaagaag cacatattca ggttcaaagg ataaaccaaa 60
acaaaacctg agctctctct ctagaatgaa gtgcctggtc ccagttcacc atctaatatt 120
tatggaaagc ttatcagaga acagaagtgt cagcagagga ctgtaagtgc tatgtgattg 180
gacgacggtg gccatagaat ttcacctgcc cagcacgtat tctctcatcc tctcctcctc 240
tgccttcttt gggagggaga aatcagaagg agtggggatg cagttcagag agcagaggag 300
aagagaagaa aagaaggaag aaggaagagg actagggttg ggtgggggggg gaggacacca 360
atgggaagag ggacagatca actctataca caaaagtaaa tcaaaacacc aaaaacaggg 420
gtctatgtaa agaagcctct tcccgtgaat tgctcgttgc atagctgcag ggagggtgtt 480
taggggcata gagaatgaaa acatacctgt attttggtgt agggaaattg tttctgtcaa 540

ttcacaccgt ccacacacca cctcccaccc caaccccgcc actaccaaat tcctctaaat 600
aaaaataatt atgagataca ggccaacaaa aacgtcagcg ttaggctgtt atttagagag 660
aattggaaag cgtttgaatg tggccctgtt gtttaataaa cgataacaat gattaaaaa 719

<210> 18
<211> 824
<212> DNA
<213> Homo sapiens

<400> 18
gcgctgaggt tcccaacctg gaccaagaag cacatattca ggttcaaagg ataaaccaaa 60
acaaaacctg agctctctct ctagaatgaa gtgcctggtc ccagttcacc atctaatatt 120
tatggaaagc ttatcagaga acagaagtgt cagcagagga ctgtaagtgc tatgtgattg 180
gacgacggtg gccatagaat ttcacctgcc cagcacgtat tctctcatcc tctcctcctc 240
tgccttcttt gggagggaga aatcagaagg agtggggatg cagttcagag agcagaggag 300
aagagaagaa aagaaggaag aaggaagagg actagggttg ggtgggggggg gaggacacca 360
atgggaagag ggacagatca actctataca caaaagtaaa tcaaaacacc aaaaacaggg 420
gtctatgtaa agaagcctct tcccgtgaat tgctcgttgc atagctgcag ggagggtgtt 480
taggggcata gagaatgaaa acatacctgt attttggtgt agggaaattg tttctgtcaa 540
ttcacaccgt ccacacacca cctcccaccc caaccccgcc actaccaaat tcctctaaat 600
aaaaataatt atgagataca ggccaacaaa aacgtcagcg ttaggctgtt atttagagag 660
aattggaaag cgtttgaatg tggccctgtt gtttaataaa cgataacaat gattaaaaaa 720
caaaccgaac cctctttgac catgtcaaaa gggagctcaa acaagtctta agatgtagca 780
attttacatg tatgccgatt tgctatatgc atttttctgc tctg 824

<210> 19
<211> 500
<212> DNA
<213> Homo sapiens

<400> 19
attcaaaatt tactaagtcc ctattttatg ccaggcattg ggcaaggacc cattggatat 60
acagagatga ctgacacagc atttagtttt tcagaatctc atcggataga ggagacaatc 120
caggctgaga gcactatata aacaaaaggc tggagtcctg aaaatgcttt atgtgtttgc 180
aggggatata gtttatgaag gaggggatta ggcttaaaaa gagctagaca gtgaaaggtg 240
aaagtaggga acagtggaag gtgatatggc ctttaaactt tgctgcttac aggtctttca 300
ggttacagtt tgaacttgat cctggaagca gcagagaaat caatcactgg ctactcttaa 360
aacagctgag tgacacaatc aaatgtgtac ttcagaaatt ttccatgtgg caatgtggag 420
gatgggctgg attgccttgt tattgggaga caaattagaa ggctactgca atagtcctag 480
caagagatga tgaccaggct 500

<210> 20
<211> 271
<212> DNA
<213> Homo sapiens

<400> 20
agcctgcacg cgctggctcc gggtgacagc cgcgcgcctc ggccaggatc tgagtgatga 60
gacgtgtccc cactgaggtg ccccacagca gcaggtgttg agcatgggct gagaagctgg 120
accggcacca aagggctggc agaaatgggc gcctgctgcg gcaccggaaa gcccagctct 180
tgctggtcaa cctgctaacc tttggcctgg aggtgtgttt ggccgcagat tcacctatgt 240
gccgcctctg ctgctggaag tgggggtaga g 271

<210> 21
<211> 612

<212> DNA
<213> Homo sapiens

<400> 21
gagcatgggc tgagaagctg gaccggcaac aaaaggctgg cagaaatagg cgcctggctg 60
attcctaggc agttgcggcc agcaaggagg agaggccgca gctttctggg agcagagccg 120
agacgaagca gttctggagt gcctgaacgg ccccctgagc cctacccgcc tggcccacta 180
tggtccagag gctgtgggtg accgcctgct gcggcaccgg aaagcccagc tcttgctggt 240
caacctgcta acctttggcc tggaggtgtg tttgccgcag gcatcaccta tgtgccgcct 300
ctactctctc taggactggg ctgatgaagg cactgcccaa aatttccccc acccccaact 360
ttcccctacc cccaactttc cccaccagct ccacaaccct gtttggagct actgcaggac 420
cagaaggcac aaagtgcggt ttcccaagcc tttgtccatc tcagccccca gagtatatct 480
gtgcttgggg aatctcacac agaaactcag gagcacccc tgcctgagct aagggaggtc 540
ttatctctca gggggggttt aagtgccgtt tgcaataatg tcgtcttatt tatttagcgg 600
ggtgaatatt tt 612

<210> 22
<211> 828
<212> DNA
<213> Homo sapiens

<400> 22
cgcgtggggg gcaaggaagg gggggcggaa ccagcctgca cgcgctggct ccgggtgaca 60
gccgcgcgcc tcggccagga tctgagtgat gagacgtgtc cccactgagg tgccccacag 120
cagcaggtgt tgagcatggg ctgagaagct ggaccggcac caaagggctg gcagaaatgg 180
gcgcctggct gattcctagg cagttggcgg cagcaaggag gagaggccgc agctttctgg 240
agcagagccg agacgaagca gttctggagt gcctgaacgg ccccctgagc cctacccgcc 300
tggcccacta tggtccagag gctgtgggtg agccgcctgc tgcggcaccg gaaagcccag 360
ctcttgctgg tcaacctgct aacctttggc ctggaggtgt gtttaggccg caggcatcac 420
ctatgtgccg cctctactct ctctaggact gggctgatga aggcactgcc caaaatttcc 480
cctaccccca actttccccc acccccaact ttccccacca gctccacaac cctgtttgga 540
gctactgcag gaccagaagc acaaagtcga attggccaag cctttgtcca tctcagcccc 600
cagagtatat ctgtgcttgg ggaatctcac acagaaactc aggagcaccc cctgcctgag 660
ctaagggagg tcttatctct caggggggggt ttaagtgccg tttgcaataa tgtcgtctta 720
tttatttagc ggggtgaata ttttatactg taagtgagca tcagagtata atgtttatgg 780
tgacaaaatt aaaggctttc ttatatgttt aaaaaaaaaa agtcgacg 828

<210> 23
<211> 482
<212> DNA
<213> Homo sapiens

<400> 23
tctcgggaac acacacacac aaaaagaata tgtggtttta atgtgctttg atgagtactg 60
ccaaacttac tccacagaaa aggcctcttt ctgaacatcc tcgcctgcgt tctatttcac 120
ccaccgtgat gcctggcctg agggcggcgt gcttgcttgt agcatttctt gaggatttgc 180
tacttgttca cctgcctctt aggagcactg tgccctgcct ccatggaagg gctcttccgg 240
cagggatgca ggctcacagc gccctggggc tggacaccac cggccggagc atggcggaca 300
gcacacacgg cccggggcgg gaaccttgga aactttacac agatggggag ctcagccatt 360
ccacgtgtgc tttcgctcag cacaatgctt actacaaacc cacgtgtact tccttccagc 420
tggttgcttt ttattgttgc tgtcttaaac tccaaagttt taaggggaat ttattgaaac 480
gt 482

<210> 24
<211> 442
<212> DNA

<213> Homo sapiens

<400> 24
ctctttaaat actagacata ccgtctggcg catgtcgagg cgtgtattat tttgatccag 60
tgcgcgtgcc gtatgtgact tggcttggct tcccctgaaa gcagcggctg tggggagttg 120
attcggaagt gaagggccct gggcgacccg gcgagtagag gcaacaccaa cactcctcct 180
tagcgagggg tctccccgcc gcggtggctg cccggcccca aggacaggag ggatttgtgc 240
actgactcct gaccccgtcc tccagcgctg ctctgaaggg agagtctgtg cagtggcacc 300
tgcgcgaagc tggccaaagc ctgcccagac ggctcacctg tgcgggatgg aacaaaaggt 360
gagcccaggg ggcctgataa aatgacctca gtagccgcct gtgggagggg accctgagga 420
aagcaccata gtgactacca ag 442

<210> 25
<211> 954
<212> DNA
<213> Homo sapiens

<400> 25
ctctttaaat actagacata ccgtctggcg catgtcgagg cgtgtattat tttgatccag 60
tgcgcgtgcc gtatgtgact tggcttggct tcccctgaaa gcagcggctg tggggagttg 120
attcggaagt gaagggccct gggcgacccg gcgagtagag gcaacaccaa cactcctcct 180
tagcgagggg tctccccgcc gcggtggctg cccggcccca aggacaggag ggatttgtgc 240
actgactcct gaccccgtcc tccagcgctg ctctgaaggg agagtctgtg cagtggcacc 300
tgcgcgaagc tggccaaagc ctgcccagac ggctcacctg tgcgggatgg aacaaaaggt 360
gagcccaggg ggcctgataa aatgacctca gtagccgcct gtgggagggg accctgagga 420
aagcaccata gtgactacca agcctggatg gttttactgc ttactttgtt aatttatttt 480
attttattct aactttcata tctctatttt ccaggacgac cggggagtcc gtcgcttctt 540
aatttcaccc acgagcgggt tggcagttgg ttcggcatgg tcgccatggg ggcgccccgg 600
aaaccctcgg ggcttcaatc ttgggcaaac cctttcttgg gcctccccgc gggggtttcc 660
aaacaccaaa atttcctggg gccctccaag ttcctccccc gggagtgtta agacttcgcg 720
ggaaacttta cgagggcgtt tgccaccctt ccgtggggag aaaccggcct ggtggggctt 780
acaatttttt cgggggccgt gaccaacaag taacggggga accaggggggg gtttccccac 840
ccgtggtttg gtttggcccc ccaggggggc tgggtgtcct tcaaaaaatc ttaatcgtgg 900
aaggccttta agagggcaca aatttctggg ccccacaacc tttcggggct cttc 954

<210> 26
<211> 657
<212> DNA
<213> Homo sapiens

<400> 26
cccttctgga accttccaaa gaggctagcc tcctgagagg gcgtctcctg ggaatgcctg 60
tgatccaggc aaggcaccta aatcagccaa taaataatta gttttgagaa aaaagcctgc 120
acaaaagcag aactcacaga ttttctcatt gatcttcctg cctttgcctt ggccatggtc 180
agtctttggg gctattagct gcatgtacaa gtgcacaggg gatgggaggg gggctccgga 240
agatccctaa cccattgtgg acacaccatt tatgttagtg gaatctggta agagtgattg 300
aggagagagc attgggtgaa tgaccctgta atccttataa aaatagtagc aacggtaact 360
aatgatgttt atgaactgcc taagaatgcc atgactgtgt caagcacttc ccataactag 420
ctcatgtaaa gcctgtaacc atcttggggg tgaggacaat tatgtcccat tgtgttatgt 480
gtctgtgttt gcaggagttg ttcagtggta aggtgcagtc aggacttaaa cccgtgttcc 540
tgagagtagc tctgtgctgc ctgacacact ccttcctcct gaatgcacaa tcatatgttc 600
acgtgccaca tggttttcag aaagcactgc acgtgtagga agctcaagac agagacg 657

<210> 27
<211> 543
<212> DNA

<213> Homo sapiens

<400> 27
tgttgaatat gtggagagag cactctgtgc tcactgcaac atcgtggggg tgcactaatg 60
accgctgtga ccactaatga ccactgtgct atcccagtta agttttccac aggtggacct 120
tgaagatttt actctacagg ccctcctgtg cagtaaatac atatcacagc tggtattgta 180
agttactgaa attgtttttt ggattattct tcccacttgt ggaaaaaagt tactgtggag 240
gggagaaaac ccttttcttc ctcaaacttc tattggggta aaggattttt aatgcagcca 300
gaacactgag agacatgctg caattctaag aaaaggctta gatttcgact cttagaggac 360
cacagccaga ggcattgtgc ttcatgtggg ctttagacag gacaagctta gttctgagag 420
ggcaagacac tttattccac accattattt aaaaccttct gtggtaaatt gttttcagag 480
atgactgtag cattatctcc ctaagccttt ttgtattgtg actgccactt ctcccattaa 540
gag 543

<210> 28
<211> 385
<212> DNA
<213> Homo sapiens

<400> 28
gttccacagc gccgtgagac accttgcaga agcgagttgg agggtccgca gagacagcag 60
agaggggcag aagcagagtg aggtcagctc ctgcctggtt ccacagaccc tgcagctttt 120
cgggtctcga aaaatcctcg gagtccacac ccacgcccca ggggccacct gctcacaact 180
ctctgagcct gctcttatcc tgcccagtgc tggcagcagg aggggcctga cctttaagga 240
cagagagtgg ggacaaagct caggggcacc caggggccca gggggcaggt gaggctctgc 300
agatacactc aggatagctg gcatcgccaa gtgtgtaaga acatactcta tgccataagg 360
ccagcagaca ggtctcaata accta 385

<210> 29
<211> 653
<212> DNA
<213> Homo sapiens

<400> 29
tgttctttaa aaatgtcaag catatccatg tactagagtc gagagaatat atgaaggtta 60
gttggacagt actggatctc ctaagaattg tttgactgaa gccctttctc tggattgctg 120
actcaaccca ggctggagtt aagtgattgg aatccacacc ctcggaccag aatccaagtt 180
accttctgct aaagccatac atacacatta aaaccacact tccacagagt cagcttgtct 240
gggctcaatg ggctctcaaa ataatacagg gttccattac ctctgaccca gaggttctct 300
ccaagagaaa gagctgttga tgttacaatt ctgcctgtca ctgtccctgt tatagcaggg 360
agtgcgtact ggattagtac tccccagaag gatgtatggt cagttggatt ggtggtggat 420
ggcagtttgt accaaaacgt cacttgaact tcacctttga cagctgcaga cttccgcgtg 480
ctgataagga attaaagaaa gtggaaatgc aatgcagtcc aagataaact actttacaaa 540
actttgtctt atggaaaatc acaaacatac acaaagttag actagtaaaa tgaatcccca 600
tatacttatc atccagcttc aacaattaca gtgtctgctc atcatttgaa tgg 653

<210> 30
<211> 1437
<212> DNA
<213> Homo sapiens

<400> 30
ttaagtgagg gcggcggatg ggcgaaggtc cggtgactgc gactgtcgct gctttctgag 60
gccacaggaa aggggccgtc ggtcgccgcc atgacagcga gcgaggcgga ggaaccatgt 120
aagaagtctg actgccctgc tggagaaatc aaatggagag gtcaagagac aacatggaga 180
gagaacgcac agcccagctg agcccagctt tccagttatc cccaccaatg cgccacacat 240

atgacggaag tcattttggt ccctccaaaa caacctagcc agccagctgg ataccactga 300
gtgacctcag tagatgccac gtgggacagg agaattctct agtgagcttt gcccacattc 360
cttacccaca cagtcatgag tataattaaa tgattattgt tttaagcccc aagtttgggg 420
tggtttatta cacagcagta tataactgga acattgggca ctgaagtagt ggaaaccata 480
catgcacgtt atatttccgt taaacttgga aagggaaaga tctgatggag tgggaggatc 540
acataagatc aagggagaat ttttatgaag atgacagaga cttgggagtg tgcatatgct 600
gatggggaag agttaggaga aagagaatgt tgaagatagg ggagagagga gatggctgat 660
gcatgaggtt tatgagaagg taagggaagg gatgagatgc aaaattctgc tggacagact 720
gatcttgctt acaaggatag tgggatgctt cgccttgtaa aacaggaagt gtctgcccac 780
ctcctgttct ttaaaaatgt caagcatatc catgtactag agtcgagaga atatatgaag 840
gttagttgga cagtactgga tctcctaaga attgtttgac tgaagccctt tctctggatt 900
gctgactcaa cccaggctgg agttaagtga ttggaatcca caccctcgga ccagaatcca 960
agttaccttc tgctaaagcc atacatacac attaaaacca cacttccaca gagtcagctt 1020
gtctgggctc aatgggctct caaaataata cagggttcca ttacctctga cccagaggtt 1080
ctctccaaga gaaagagctg ttgatgttac aattctgcct gtcactgtcc ctgttatagc 1140
agggagtgcg tactggatta gtactcccca gaaggatgta tggtcagttg gattggtggt 1200
ggatggcagt ttgtaccaaa acgtcacttg aacttcacct ttgacagctg cagacttccg 1260
cgtgctgata aggaattaaa gaaagtggaa atgcaatgca gtccaagata aactacttta 1320
caaaactttg tcttatggaa aatcacaaac atacacaaag ttagactagt aaaatgaatc 1380
cccatatact tatcatccag cttcaacaat tacagtgtct gctcatcatt tgaatgg 1437

<210> 31
<211> 733
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (508)
<223> a, c, g or t

<220>
<221> unsure
<222> (522)
<223> a, c, g or t

<400> 31
atttgatctc aatgcaattt tcataaaaat caataaaaga ttacaaactc ttcatatcta 60
tagaaagtaa ggatcaattg gctcactttt actgtgtcac aaacaacctg aaaattcagt 120
agattaaaat gacaatcatt gatgatcact catgtgctgt agtttggctg gtgtttcagc 180
tgatataagc tgggctcggc tagatggtga gcactcaagc tgtgggcctg gtaggacttg 240
gccactcctg taggttgggc tcaaatctga tcaacatcta ttcattcagg ggctcaccct 300
aaagaaccat cacttactaa ggaaaaggta atctccatag tgaccacaag ggatgtggga 360
aggaaagact aactgcatgg tcctattgaa gcaattaact ccctattgtc tgctaacatc 420
ccattgtcca aatctaggga catcctgagc tgatatcaat ggagtgaagg gaaatgggga 480
actgcacctg aactgggagg acagtgantg atgtatggac antgaatgat gacatgatta 540
ggatcttgta aaagactgtc tggcaatata gcagcacaaa gtaaatattc ttgaacatta 600
aattaatctg ctaacaactg acaaacttac ctaaagggag acattaaaaa gaaaaatggg 660
aaagacacaa ataatgaata aattatttaa aaggagatct ccctaagacc cttcagccat 720
ataaaataag agt 733

<210> 32
<211> 404
<212> DNA
<213> Homo sapiens

<220>

<221> unsure
<222> (176)..(212)
<223> a, c, g or t

<400> 32
gttttgctgg tagctgttta acttttatat ctttgtattt ttaaccagat catttttaa 60
ccttagtttc agtttgtata ttaggactca gctttatttc ctatttatag aagttaataa 120
aatattcttt gaaattaaaa aatattcttt aatgacaagc caaaaatttt ttaaaannnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnctttgtca cttagcattt tttccaattt 240
catatagtct gtcatcgaca gtactcattt tagtcagtgt ctcaagtgct taaagagcgc 300
ttaaaaagtg ccccttttgg ggtagactgg tattactttt attacatata acctcaatta 360
tggtgagagt aaaacactta aagctaaggc caggcgtggt ggct 404

<210> 33
<211> 144
<212> DNA
<213> Homo sapiens

<400> 33
agttttgcag gaatttgtta cacagcaaaa tgaaactaat tcatcgggga aggaccacgt 60
gcctggtatg gtacggtgac tggaacagtt gttctccaac aaggctgcat gttggagtga 120
aaagctttaa aaagtactgc tgcc 144

<210> 34
<211> 156
<212> DNA
<213> Homo sapiens

<400> 34
caatctttta tacttatcta cataattatt tttaatagga ttctttgttt tgatgtttag 60
atttaaatta ttctatagtg tccctttctt tcagcctgaa gaactttctc tagtatttcc 120
tgtgaatagg aaatnctagc agcacagctc tgctag 156

<210> 35
<211> 554
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (533)
<223> a, c, g or t

<220>
<221> unsure
<222> (541)
<223> a, c, g or t

<400> 35
ttttttttt ttgacggact ccactctgtc accaggctgg agtgcagtgg tgcaatctcg 60
gcttactgca acctccactt cccaggttca agcaattctc ctgcctcagc ctcccgaata 120
gctgggacta caggcaagcg tcaccacgcc cagctaattt ttgtatttta agtagagaca 180
gggtttcacc atgttggcca ggatggtctt gatctcctga cctcgtgatc tgcctgcctc 240
ggcctcccaa agtgctgaga ttacaggcat gaaccaccac gcctggccgc gatatgttct 300
taactcagtt tacaaacaag gaatgataca atcttttata cttatctaca taattatttt 360
taataggatt ctttgttttg atgtttagat ttaaattatt ctatagtgtc cctttctttc 420

agcctgaaga actttctcta gtatttcctg tgaataggaa atactagcag cacagctctg 480
ctagtatgtt agtttcctat ggcttctgta acaaaattat cacaaaactt agnggcttaa 540
nacaacacaa atgt 554

<210> 36
<211> 607
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (75)..(203)
<223> a, c, g or t

<220>
<221> unsure
<222> (258)..(406)
<223> a, c, g or t

<220>
<221> unsure
<222> (432)
<223> a, c, g or t

<220>
<221> unsure
<222> (563)
<223> a, c, g or t

<400> 36
ccaaaacatt atatttattt agatgacctt gatatggaca ctatatccag ttggaaagtg 60
gagtacagac acaannnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnngtagttg agactagaga tacacacacc agtaatggtc 240
cagcccacta attagtgnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnacac aacctctaaa 420
aaatttttca tnacaccaaa ttggtaaaca tgttcaaaag atttaatgta cagaatcact 480
agatacggtg agtttacggt ttattgtaga agcaagtttg ttttaatcat cttaggagat 540
ttttttgtgt tgttttgttt ttntgagagt ctcactgcaa tctgccacct ccgggttcaa 600
ctgattc 607

<210> 37
<211> 113
<212> DNA
<213> Homo sapiens

<400> 37
tcgagcggct cgaggaaaca tataagaagt accagatgca gttttctgag ttgccataaa 60
tggttaaaaa attattttga tgatataaac atctgggtat actagacaat gtt 113

<210> 38
<211> 667
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (545)
<223> a, c, g or t

<400> 38
aaaaagcaca agatattgtt attttgataa ctatgattca gaaaggtggt gaattttatg 60
gtgtgcaaag aaaagtgagt tttcataagc aatacaaatt tagttctaac atcaggaaga 120
tcttttggtc tgtctgctag tttatacctg tatctgaaag ggcagcataa ctgatttttt 180
ttttaagaag gttgaatgac tacttttaaa agtctcctaa gagttgagaa tggaaaattt 240
ccttggacca ttttttaaaa gaaaatctag aagaatgtat aaaatgagtc ccagtaaaac 300
ttactgtaat atgaagcata aaaacatgtt ttaaagctac agatataaaa gaagcttaaa 360
atatttttcg tggaactacc tttacggtgc tggatggtaa agtttttatt ttaaataaag 420
acattccgtt ttccaggaca tgatttatac aagaaaacat ataatgtcct gttgaacaca 480
ggccaaatac attcccagac gctttgttac ccaaaccttt tgagagattt ctttatagat 540
gtttntggca gaagtggtga ttttagagga aacatataag aagtaccaga tgcagttttc 600
tgagttgcca taaatggtta aaaaattatt ttgatgatat aaacatctgg gtatactaga 660
caatgtt 667

<210> 39
<211> 210
<212> DNA
<213> Homo sapiens

<400> 39
ttggaaacac aatttttaa aatgaaaaat ctaatttaat tccttctttc cccatgtgca 60
taaaaaatct aatttatatt catgttaata tgtaataagt gtattatttg taaatgaata 120
aacaaacatc taaaattttt ctcagtttta atttctaata cagtatacat caatagatat 180
gactcatata aacaaaactt ctttggggtc 210

<210> 40
<211> 256
<212> DNA
<213> Homo sapiens

<400> 40
atgaaaataa tgctactctg cccactaatt ttttttcttg gaaacacaat tttttaaaat 60
gaaaaatcta atttaattcc ttctttcccc atgtgcataa aaaatctaat ttatattcat 120
gttaatatgt aataagtgta ttatttgtaa atgaataaac aaacatctaa aatttttctc 180
agttttaatt tctaatacag tatacatcaa tagatatgac tcatataaac aaaacttctt 240
tggggtctga tatggt 256

<210> 41
<211> 228
<212> DNA
<213> Homo sapiens

<400> 41
ttctgccttt ccctttgatc atcatttttt ttctccaaaa taaaggtcag ccacttttttc 60
cattaaaata ttttctcagg cttttagttc atcccagcct ttgccccttg ttccctctac 120
tgtgacaata tttcctgacc ataaggtcct tgtatgaatt attttggtct ttgtacctcc 180
cattttggtg tctaatgaaa tttggctgtg tctctgtgga atacctgc 228

<210> 42
<211> 3930

<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (1169)
<223> a, c, g or t

<220>
<221> unsure
<222> (1212)
<223> a, c, g or t

<220>
<221> unsure
<222> (3907)
<223> a, c, g or t

<400> 42
aaacgcaaga ggccattgta aacatctgct tgtccttctt aggtcgccat tccctttgca 60
tgttaagcgt ctgctcaggt aaatcttagt gaaattccta ccgttgttgt acgttctgca 120
aaacatttta tgtatagatt tagaggggaa acgagaaggt actgaaataa tgatcttgga 180
atatttgctg tgaagggaga aagggagaga aaactcttct gaggatcatt tgtcttggta 240
gtatagtaaa accaaccagc tgaacctttc aggctacaag agaacccggg tcggtaatgt 300
ctttttaaga ataatttta attgcttata acaagcatat tttgtggcat ttgaactata 360
tttactgctc caatatccgt tattttccaa aggattttgt atctttttga aaatgtttac 420
atcatcagat gatccacaga attcacttta tgtgagatct cccgagagtt tccatcccaa 480
cataatggac tttggtttga acacaattcg ttttttcatt tgaattggca tttcccaatt 540
atttgctaaa catttgctgg agaaatcatt tttcttttt ctttttaga aaactcagaa 600
tgaaaattca ttcccctgaa atatttaggt gtctatattc tatattttgg atctattaag 660
ggattagtat ttttccatgt ttattgtgtt atcagagtgc attagaaaga ttagtgattc 720
atcttcacag cacattttta atcaagcagt tatttcaacc atgcacattc gttttgttca 780
tattcactat agaatgatat cttgtaaata aagacattca gcacactgtg aaaatgtatt 840
tgtgcacctg ctttttaaat atttctacta aaaatgaaaa aaaaacccct tagacctgta 900
gatagtgata tcgtaatatt aattgttaat aaaatagtca ctgccattaa aatctgcaga 960
aaatactctt tactttccta tgctaaatgt gcactgcaca gaaagaaaag gagcaggttt 1020
atcttatata aaccaaggac cattctttat tattattcag tacttctttt gctaccaaat 1080
ttttaagaaa agggagaaat ccctcataac taaaatccta aaaaatgata agtgcttcaa 1140
ggagattaaa gtgtcaggta ttacaaaang gagagaaaag aaaggaggaa gagatagtat 1200
gaattcccag gntccctgtg gaaaatgaaa accattgttt tgtagcagca atgactaggc 1260
tgtcgtattt taaaatctgc tgttgggtct gcatgtgtct gagagggcag gttttgatat 1320
ttgttgaaat gacaggacta cgaaaggcct taaaccttga ccttgacgaa aaaaaagaca 1380
atttgtgacc ttgacttttg acagctcatg aattggcctt agctggatta gtagatcaag 1440
ggcgccacct cgcgttgaca agcctccttg taattcttca accaagggaa tcaagagctc 1500
ttacctcctt gactttagag attcagtact caattatttg gctccagctg gagcaaattt 1560
gaatatatgt aaattataca tgtccctgta tatatgtaaa tatatataat atgtgtacac 1620
atgtatacat ttgcatatta gtgcttgtgt gtgcttgtat atatatttgt agaatctgta 1680
cagatatcat atcattaatt gctaatgacg caagatattg taatttaagg tgatttctag 1740
aattgaggta caaaatgtac tttacaagat gaagccttgt tccttttctc taatggccaa 1800
atttgatttg tcttcacagt ggcttgctct gcatcagatt tctgcagatc tgctttaaag 1860
ctggtacatt tttgttacag tctaagatgt gttcttaaat caccattcct tcctggtccc 1920
ttcaccctcc agggtggtct cacactgtaa ttagggctat tgaggagtct ttacagcaaa 1980
ttaagattca gatgccttgc taagtctaga gttctagagt tatgtttcag aaagtctaag 2040
aaacccacct cttgagaggt cagtaaagag gacttaatat ttcatatcta caaaatgacc 2100
acaggattgg atacagaacg agagttatcc tggataactc agagctgagt actgctccag 2160
ggtggtgtgc aatcttatat tgatgcttgt gaatctgcca tttgatttgt aggataaata 2220
aatatgttta atattaacaa cttccatcaa aactataata ataatattat atctactgtt 2280
gacctctaac aacaatcagg tgctgtattc agagtcataa agaatacata tgtatttttg 2340
tatatttaat atacacatag gataatgtgt ttttgtcaaa tctatgctat gcgcatatgt 2400

attaaaatgt tggtggaatc agtagccgtg ggatgcagaa ttgtgagatg caggctaatt 2460
ttatctttac gttagaggaa agtaatagtg aatggaagaa ccagattcat cattttatgg 2520
aaaataatat tgattggtta caaaagtggc tgccggaaac ctctccaacc tctcctgagg 2580
actcacaact tctttataga tgagaattac ggaaagagtc tctaaagctg tctcaaaaac 2640
aggctggtgt tgaaacatac atatcggaaa ggtctcgaga aattcaatta tattttgctg 2700
ccctgaggat tttcactcca cttgttggaa acaaatgata atgttgcttc agacccacct 2760
ctccacctct tgttcccaaa cagtactcca tttccaaatg agctgccatc aagccaaagt 2820
aacaactttc tcccaaaatg tttaaagtgg tggaaaatcc cctcaaagct ggatcaaact 2880
gtgtgtttaa gaacagctcc aaaagggact cttctgttga caggggatat ttttctttag 2940
atgtaagaaa tattctgagc ctcgagccta ggaaggaggc caacagcagc tctccagccg 3000
ggatcccgcc cctttcgaaa gccatcacca ctacaggaac cctgggtaaa gtagtagcct 3060
tctttttttc tctctctagt ttttgaaact gtaatgacag aagacatttc agggttgttg 3120
tggggatttt ttccccccccc agttcctaaa atggtcattg atagtgtcga tcttacccta 3180
actctccctc cctatttaaa atttctgtgg atctaccaga taacctaaag tgggcatgga 3240
ttatttatta tggctccttt aatgcccaaa tagcaatcag tctaaaacct taaaacaaaa 3300
acaaaaacaa aaacaaaaat caaggagaga ccagccacct cctgaaaaat aggaataata 3360
gtatcagcca ctgacccatg gataatgttg ttgaaagcct gattaaagta ttttttttcta 3420
aaaagaaaac tcttcctgaa ttaagaaaat aaaaatggga ctttaaaaaa aaacgctact 3480
ttttctatct gctctcagta ctatcatatt tatttactta ttttaagtgg atgattgtta 3540
aatgtcgcta tgcatttcac tcctatttgt gactgatgca gaactgaagc cacccattca 3600
agttattcaa gttgtgggtt tttaaaaaaa attaacagaa aagggaagaa aaaaattaaa 3660
acaacctagt tgaaatcttt tcttaaaaat tcaccactag gcagtgaaat tcagctgcag 3720
atgcaacatt tgtgtgcgga gagttgacac agatttcact ttgtaataca tgtaagtgca 3780
gattgcactc tgctgcatgt gtttggaggc cttggtgaag atgcagtgtc ttgcttcctc 3840
tgtgaattcc cgagttcagg agcagagacg gaggttgccc tggactgcta atccccagac 3900
ccggtgngtt cagaggcgcc ctcctggatt 3930

<210> 43
<211> 5643
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (5471)
<223> a, c, g or t

<220>
<221> unsure
<222> (5501)
<223> a, c, g or t

<220>
<221> unsure
<222> (5634)
<223> a, c, g or t

<220>
<221> unsure
<222> (5636)
<223> a, c, g or t

<400> 43
atgacagcct ccgtgctcct ccacccccgc tggatcgagc ccaccgtcat gtttctctac 60
gacaacggcg gcggcctggt ggccgacgag ctcaacaaga acatggaagg ggcggcggcg 120
gctgcagcag cggctgcagc ggcggcggct gccggggccg ggggcggggg cttcccccac 180
ccggcggctg cggcggcagg gggcaacttc tcggtggcgg cggcggccgc ggctgcggcg 240
gcggccgcgg ccaaccagtg ccgcaacctg atggcgcacc cggcgccctt ggcgccagga 300

gccgcgtccg cctacagcag cgcccccggg gaggcgcccc cgtcggctgc cgccgctgct 360
gccgcggctg ccgctgcagc cgccgccgcc gccgccgcgt cgtcctcggg aggtcccggc 420
ccggcgggcc cggcgggcgc agaggccgcc aagcaatgca gcccctgctc ggcagcggcg 480
cagagctcgt cggggcccgc ggcgctgccc tatggctact tcggcagcgg ctactacccg 540
tgcgcccgca tgggcccgca ccccaacgcc atcaagtcgt gcgcgcagcc cgcctcggcc 600
gccgccgccg ccgccttcgc ggacaagtac atggataccg ccggcccagc tgccgaggag 660
ttcagctccc gcgctaagga gttcgccttc taccaccagg gctacgcagc cgggccttac 720
caccaccatc agcccatgcc tggctacctg gatatgccag tggtgccggg cctcgggggc 780
cccggcgagt cgcgccacga acccttgggt cttcccatgg aaagctacca gccctgggcg 840
ctgcccaacg gctggaacgg ccaaatgtac tgccccaaag agcaggcgca gcctccccac 900
ctctggaagt ccactctgcc cgacgtggtc tcccatccct cggatgccag ctcctatagg 960
agggggaga aagaagcgcg tgccttatac caaggtgcaa ttaaaagaac ttgaacggga 1020
atacgccacg aataaattca ttactaagga caaacggagg cggatatcag ccacgacgaa 1080
tctctctgag cggcaggtca caatctggtt ccagaacagg agggttaaag agaaaaaagt 1140
catcaacaaa ctgaaaacca ctagttaatg gattaaaaat agagcaagaa ggcaacttga 1200
agaaacgctt cagaactcgt tgctttgccc agataatgat aataatgctt aataataatt 1260
gaagaatggg aaagagaaag agacagagac tggcattttc ctctcccgaa ggagatctct 1320
ttctctttaa tggaatctac aactgtttta aaactttaag aaaggtaaag actgccagtt 1380
cttccgccaa ccccatcagc ccagcccgtt aaatgtcaaa cgtcaacccc caaaatacgc 1440
aatttcagat aagttacgca gttactgaaa tcttgtaagt atttaagtga tcgttacatt 1500
ttaggacact gcgttagatg gtaataatct ggaagttggt tacaacgcaa gaggccattg 1560
taaacatctg cttgtccttc ttaggtcgcc attcctttg catgttaagc gtctgctcag 1620
gtaaatctta gtgaaattcc taccgttgtt gtacgttctg caaaacattt tatgtataga 1680
tttagagggg aaacgagaag gtactgaaat aatgatcttg gaatatttgc tgtgaaggga 1740
gaaagggaga gaaaactctt ctgaggatca tttgtcttgg tagtatagta aaaccaacca 1800
gctgaacctt tcaggctaca agagaacccg ggtcggtaat gtctttttaa gaataatttt 1860
taattgctta taacaagcat attttgtggc atttgaacta tatttactgc tccaatatcc 1920
gttattttcc aaaggatttt gtatcttttt gaaaatgttt acatcatcag atgatccaca 1980
gaattcactt tatgtgagat ctcccgagag tttccatccc aacataatgg actttggttt 2040
gaacacaatt cgtttttttca tttgaattgg catttcccaa tatttgctaa acatttgctg 2100
gagaaatcat ttttcttttt tctttttttag aaaactcaga atgaaaattc attcccctga 2160
aatatttagg tgtctatatt ctatattttg atctattaag ggattagtat ttttccatgt 2220
ttattgtgtt atcagagtgc attagaaaga ttagtgattc atcttcacag cacattttta 2280
atcaagcagt tatttcaacc agcacattcg ttttgttcat attcactata gaatgatatc 2340
ttgtaaataa agacattcag cacactgtga aaatgtattt gtgcacctgc tttttaaata 2400
tttctactaa aaatgaaaaa aaaacccctt agacctgtag atagtgatat cgtaatatta 2460
attgttaata aaatagtcac tgccattaaa atctgcagaa aatactcttt actttcctat 2520
gctaaatgtg cactgcacag aaagaaaagg agcaggttta tcttatataa accaaggacc 2580
attctttatt attattcagt acttcttttg ctaccaaatt tttaagaaaa gggagaaatc 2640
cctcataact aaaatcctaa aaaatgataa gtgcttcaag gagattaaag tgtcaggtat 2700
tacaaaaggg agagaaaaga aaggaggaag agatagtatg aattcccagg ctccctgtgg 2760
aaaatgaaaa ccagttgttt tgtagcagca atgcctaggc tgtcgtattt taaaatctgc 2820
tgttgggtct gcatgtgtct gagagggcag gttttgatat ttgttgaaat gacaggacta 2880
cgaaaggcct taaaccttga ccttgacgaa aaaaaagaca atttgtgacc ttgacttttg 2940
acagctcatg aattggcctt agctggatta gtagatcaag ggcgccacct cgcgttgaca 3000
agcctccttg taattcttca accaagggaa tcaagagctc ttacctcctt gactttagag 3060
attcagtact caattatttg gctccagctg gagcaaattt gaatatatgt aaattataca 3120
tgtccctgta tatatgtaaa tatatataat atgtgtacac atgtatacat ttgcatatta 3180
gtgcttgtgt gtgcttgtat atatatttgt agaatctgta cagatatcat atcattaatt 3240
gctaatgacg caagatattg taatttaagg tgatttctag aattgaggta caaaatgtac 3300
tttacaagat gaagccttgt tccttttctc taatggccaa atttgatttg tcttcacagt 3360
ggcttgctct gcatcagatt tctgcagatc tgctttaaag ctgtacattt ttgttacagt 3420
ctaagatgtg ttcttaaatc accattcctt cctggtcctc accctccagg gtggtctcac 3480
actgtaatta gagctattga ggagtcttta cagcaaatta agattcagat gccttgctaa 3540
gtctagagtt ctagagttat gtttcagaaa gtctaagaaa cccacctctt gagaggtcag 3600
taaagaggac ttaatatttc atatctacaa aatgaccaca ggattggata cagaacgaga 3660
gttatcctgg ataactcaga gctgagtact gctccagggt ggtgtgcaat cttatattga 3720
tgcttgtgaa tctgccattt gatttgtagg ataaataaat atgtttaata ttaacaactt 3780
ccatcaaaac tataataata atattatatc tactgttgac ctctaacaac aatcaggtgc 3840

tgtattcaga gtcataaaga atacatatgt atttttgtat atttaatata cacataggat 3900
aatgtgtttt tgtcaaatct atgctatgcg catatgtatt aaaatgttgg tggaatcagt 3960
agccgtggga tgcagaattg tgagatgcag gctaatttta tctttacgtt agaggaaagt 4020
aatagtgaat ggaagaacca gattcatcat tttatggaaa ataatattga ttggttacaa 4080
aagtggctgc cggaaacctc tccaacctct cctgaggact cacaacttct ttatagatga 4140
gaattacgga aagagtctct aaagctgtct caaaaacagg ctggtgttga aacatacata 4200
tcggaaaggt ctcgagaaat tcaattatat tttgctgccc tgaggatttt cactccactt 4260
gttggaaaca aatgataatg ttgcttcaga cccacctctc cacctcttgt tcccaaacag 4320
tactccattt ccaaatgagc tgccatcaag ccaaagtaac aactttctcc caaaatgttt 4380
aaagtggtgg aaaatcccct caaagctgga tcaaactgtg tgtttaagaa cagctccaaa 4440
agggactctt ctgttgacag gggatatttt tctttagatg taagaaatat tctgagcctc 4500
gagcctagga aggaggccaa cagcagctct ccagccggga tcccgcccct ttcgaaagcc 4560
atcaccacta caggaaccct gggtaaagta gtagccttct ttttttctct ctctagtttt 4620
tgaaactgta atgacagaag acatttcagg gttgttgtgg ggatttttc ccccccagt 4680
tcctaaaatg gtcattgata gtgtcgatct taccctaact ctccctccct atttaaaatt 4740
tctgtggatc taccagataa cctaaagtgg gcatggatta tttattatgg ctcctttaat 4800
gcccaaatag caatcagtct aaaaccttaa aacaaaaaca aaaacaaaaa caaaaatcaa 4860
ggagagacca gccacctcct gaaaaatagg aataatagta tcagccactg acccatggat 4920
aatgttgttg aaagcctgat taaagtattt ttttctaaaa agaaaactct tcctgaatta 4980
agaaaataaa aatgggactt taaaaaaaaa cgctactttt tctatctgct ctcagtacta 5040
tcatatttat ttacttattt taagtggatg attgttaaat gtcgctatgc atttcactcc 5100
tatttgtgac tgatgcagaa ctgaagccac ccattcaagt tattcaagtt gtgggttttt 5160
aaaaaaaatt aacagaaaag ggaagaaaaa aattaaaaca acctagttga aatcttttct 5220
taaaaattca ccactaggca gtgaaattca gctgcagatg caacatttgt gtgcggagag 5280
ttgacacaga tttcactttg taatacatgt aagtgcagat tgcactctgc tgcatgtgtt 5340
tggaggcctt ggtgaagatg cagtgtcttg cttcctctgt gaattcccga gttcaggagc 5400
agagacggag gttgccctgg actgctaatc cccagacccg gtgggttcag aggcgccctc 5460
ctggatttgc ngacctctcc caaggagacg ctaggggctt nctgatttat gcgcattcct 5520
ttaaacttga tcgaatgcgt gtcagccgac caagacctnn agaatcagag aaatgtgcag 5580
cgagttgctc acatcccatt tctcataagc aaaagatata gcgttatcca aaangnagag 5640
gga 5643

<210> 44
<211> 228
<212> DNA
<213> Homo sapiens

<400> 44
tagggcccaa actcctcagc atagcataca aggcctagat ccctaaacta caactccatt 60
tcctatcagt tttcatctcc tattacttgt ccccagttta ttctattcgt taagccacaa 120
tgaatgtctt acttctagcc aagttctgtt tttcctctaa ggcccagttc aacattttag 180
tagtcagaaa agacttttt gaccctaaga agtagcatga atttcttt 228

<210> 45
<211> 528
<212> DNA
<213> Homo sapiens

<400> 45
atgagtgcaa gcactcgtta taagtcagct ttctcccagc cctctttgct gggtgctgag 60
gtcccagagc ttttgtccca gctcagcgca cagctgggtg agcagcctca tctcccgggc 120
cttggctcaa atgcacctgg aggtagcggt gagcccttca gagctccgga tgaggggaga 180
tgatggcttg agccagtgat gggaagaaga gagtggggct tgtgaagtgg tccctgatgg 240
atgtggactg cacctggccc catggggcag ggagccccag gtgagcagct tggccctccc 300
tgcgtggggt cagccagtgt ctctggagag actgtccaaa gcctctagga ctcttgagta 360
cagggcccct ccgaggggtg gaaggtgcag aacacccatc tcatctggcc acagggcctt 420
tttccaggga acatcaatgg tatcccagat cccgacgctg ctgtgagcag cagagtgact 480

caggccacat gctctgagag gggctgggag gccaccgtgg gccctgcc 528

<210> 46
<211> 695
<212> DNA
<213> Homo sapiens

<400> 46
ctcactatta cggcgcagtg tgctggcagc acagatcatg gaatcaggcc tctgaggtca 60
ttaatattca cgtagatata cttgcttatg agcctggaag gaaagatttg caattttata 120
tacgtataat atagtatttg ctcttggtat caacctaaat tcataaacag atgtgtatat 180
atctctattt atgtatactt gaagtctgta ttgtgagggt gaataattag aagttgactt 240
tatgtcaaac caggtttata gaattcagat gtgtccccaa aatttgtcct cttctagata 300
tgtccattta cttttagtat gactagtagg aaccataggg gttttagtgt ctgattttga 360
ggatataaac acaaaaaaag tatttttct cttttgtcta gattactaat ccagacaatg 420
taaacctcac ccaaaatttg tgaaatttga gaccggattt ctttttcatg gcattacaac 480
attattttaa aatagcgcta ctaaacagtt ttatcaaaaa attccaacaa atgacatgtg 540
gcccagtgga tgaaaaaaaa ttctgcaatt attaatagat gtttttactg ttcagagttt 600
tagctgttat ttgtttagta atgttggtgt ttgtatgtag gctttgaggc caggggacta 660
attagggatt ctctgtaata atttaagtga agctt 695

<210> 47
<211> 5350
<212> DNA
<213> Homo sapiens

<400> 47
cagtaaccta gaccagcccc agcttcagcc tcagctcccc tccttcctgg atcgagcgcc 60
cgcactcccg gccctgcagc cacccgagtc ccgctcgctg tcgcctgcac gcgagtcccc 120
cctggcacgc gctcccacat cccgggatcg tcccaacggc ccctgcgccc ttcctgggat 180
cactccgact gccccgcgcg ccctgggatc ggtccatcta ccccgcgtgg ccccagctgc 240
ttgcccggag cgccagctag cgctccccgc tctccgctcc ccggcactct cgggggggccc 300
gcccgccctg caccctggag ctccgggccg cgagcctctg ccaactcctc tggaccctcg 360
cggccgtggg cagcggctgc cgcgcctgtc tgcccgaggg aggaccttcg cctctgcatt 420
tgtccagtaa ctctggctgt gccggatact gcttgggtaa aacgggcacc ccaggaacat 480
ggcagacgaa gatctcatct tccgcctgga aggcgttgat ggcggccagt ccccccgagc 540
tggccatgat ggtgattctg atggggacag cgacgatgag gaaggttact tcatctgccc 600
catcacggat gacccaagct cgaaccagaa tgtcaattcc aaggttaata agtactacag 660
caacctaaca aaaagtgagc ggtatagctc cagcgggtcc ccggcaaact ccttccactt 720
caaggaagcc tggaagcacg caatccagaa ggccaagcac atgcccgacc cctgggctga 780
gttccacctg gaagatattg ccaccgaacg tgctactcga cacaggtaca acgccgtcac 840
cggggaatgg ctggatgatg aagttctgat caagatggca tctcagccct tcggccgagg 900
agcaatgagg gagtgcttcc ggacgaagaa gctctccaac ttcttgcatg cccagcagtg 960
gaagggcgcc tccaactacg tggcgaagcg ctacatcgag cccgtagacc gggatgtgta 1020
ctttgaggac gtgcgtctac agatggaggc caagctctgg ggggaggagt ataatcggca 1080
caagcccccc aagcaggtgg acatcatgca gatgtgcatc atcgagctga aggacagacc 1140
gggcaagccc ctcttccacc tggagcacta catcgagggc aagtacatca agtacaactc 1200
caactctggc tttgtccgcg atgacaacat ccgcctgacg ccgcaggcct tcagccactt 1260
cacttttgag cgttccggcc atcagctgat agtggtggac atccagggag ttggggatct 1320
ctacactgac ccacagatcc acacggagac gggcactgac tttggagacg gcaacctagg 1380
tgtccgcggg atggcgctct tcttctactc tcatgcctgc aaccggattt gcgagagcat 1440
gggccttgct ccctttgacc tctcgccccg ggagagggat gcagtgaatc agaacaccaa 1500
gctgctgcaa tcagccaaga ccatcttgag aggaacagag gaaaaatgtg ggagcccccg 1560
agtaaggacc ctctctggga gccggccacc cctgctccgt ccccttttcag agaactctgg 1620
agacgagaac atgagcgacg tgaccttcga ctctctccct tcttccccat cttcggccac 1680
accacacagc cagaagctag accacctcca ttggccagtg ttcagtgacc tcgataacat 1740
ggcatccaga gaccatgatc atctagacaa ccaccgggag tctgagaata gtggggacag 1800

cggataccc agtgagaagc ggggtgagct ggatgaccct gagccccgag aacatggcca 1860
ctcatacagt aatcggaagt acgagtctga cgaagacagc ctgggcagct ctggacgggt 1920
atgtgtagag aagtggaatc tcctcaactc ctcccgcctc cacctgccga gggcttcggc 1980
cgtggccctg gaagtgcaaa ggcttaatgc tctggacctc gaaaagaaaa tcgggaagtc 2040
cattttgggg aaggtccatc tggccatggt gcgctaccac gagggtgggc gcttctgcga 2100
gaagggcgag gagtgggacc aggagtcggc tgtcttccac ctggagcacg cagccaacct 2160
gggcgagctg gaggccatcg tgggcctggg actcatgtac tcacagttgc ctcatcacat 2220
cctagccgat gtctctctga aggagacaga agagaacaaa accaaaggat ttgattactt 2280
actaaaggcc gctgaagctg gcgacaggca gtccatgatc ctagtggcgc gagcttttga 2340
ctctggccag aacctcagcc cggacaggtg ccaagactgg ctagaggccc tgcactggta 2400
caacactgcc ctggagatga cggactgtga tgagggcggt gagtacgacg gaatgcagga 2460
cgagccccgg tacatgatgc tggccaggga ggccgagatg ctgttcacag gaggctacgg 2520
gctggagaag gacccgcaga gatcagggga cttgtatacc caggcagcag aggcagcgat 2580
ggaagccatg aagggccgac tggccaacca gtactaccaa aaggctgaag aggcctgggc 2640
ccagatggag gagtaaccag gaaaatcact gccggctagt cccaagcaaa cgggctagga 2700
ggaaagatta aaaaaacaac aacaacaact tatttagttt ggggagggga agcatttta 2760
agtgtgttgt aaaatcaaat tttatatttc atttttgac tcttgaaaaa tgtctttgct 2820
ccttggcagc taccagcaga gactctatag ctgtctctta gggcagtatt ttggggaagt 2880
ggggcttgaa gaagcagcct aatgaaccaa cataccgttt tgtgtgtggt ttttttggtt 2940
ggttggtttg tttgttttca gacagagtct tgctctgtca cacaggctgg agtgcagtga 3000
catgatctta gctcactgca acgtccacct cctgggttca agcgattctc ctgcctcagc 3060
ctcccaagta gctgggatta caggtgtgtg ccactaagct cagctaattt ttgtattttt 3120
agtagagacg gggtttcacc attttgacca ggatggtctt gatctcttga cctcatgatc 3180
tgtccacctc tgcccctcaa agtgctggga ttacaggcgt gagccaccac gcccagccgt 3240
acatttactt tttaaagcag cagactaggt acactaattc tcactcaaat attttcatgg 3300
gaatgtagtt atcaccaagt cctaaagtat tatttatgcc aaaaaaaatt tcattttaag 3360
gactacaaaa atgattctaa ttaaacattt tataatcaat agtaggttgg gtctttagcc 3420
attatatgtg tatatataca gacacatatg tatacactta cattttgaca gggtcttcat 3480
tgagtcttga tgcactttaa acccagctgg ctaccagaga tgcgaaggtg ggctctttga 3540
agattagcaa aatggacgtt tctgtcactt gagaaaagga aagttctttg cctttaaatt 3600
acacagtttt catcatgccc acaatctata ttattggctg gttaaacagc actgccctat 3660
tagcaatgtt aacaaaaatg aaattattta ttggcggtta tagattatct aattcaggaa 3720
atttctgagc tcaactttta cagcaactgt tatgccttct aatttagcaa ttgagttatg 3780
agtaagtttt gtgcttaact cctagaccct attgttgata accagatcaa atatagtctg 3840
tacagaggaa aacactggga acatttagta tttctaaagc ctcctttgga gttactactg 3900
attgtaattt ggaactgata ataggtagag attgctaaca ctgtttttt tcctggatct 3960
tttttatgcc agaaattaaa caggttctgc taactctttt ttttctcttg gttatcacca 4020
gaatgaaaat atttaaagtg atgactctag aaaagccatc tgtgcctggt taacattgag 4080
tttgagtctc ttcaatatat attgatcatg tattgattaa tctttatttt ttcatatttt 4140
ggctagacaa attcagatct atataatgga ataccccttc ttgagtgaac tatactacta 4200
atctacatga ttatatagta aggaaaaaag aagaaataac tgtaataggc atagtgtttg 4260
ttgttggttg tcttgtcatt catgtgatac tactcatttc caaaattcac acaaacttac 4320
atgaggtgga ttatttgttt tgttcattat ttagttccta tatgtttttt ttttcgagat 4380
ggagtctcac tctgtcaccc aggctggagt gcagtggcgc gatctcggct cactgcaacc 4440
ttcgcctccc aggttcacac cattctcctg cctcagcctc ccaagtagct gggactacag 4500
gtgcccacca ccacaacagg ctaatctttt gtatttttag taaagacggg gtttcaccat 4560
gttagccagg atggtctcga tctcctgacc tagtgatccg cccgcctcgg cctcccaaag 4620
tgctgggatt acaagattgc tctttttaat aatttaagct tcacttaaat tattacagag 4680
aatccctaat tagtcccctt gcctcaaagc ctacatacaa acaccaacat tactaaacaa 4740
ataacagcta aaactctgaa cagtaaaaac atctattaat aattgcagaa tttttttca 4800
tccactgggc cacatgtcat ttgttggaat tttttgataa aactgtttag tagcgctatt 4860
ttaaaataat gttgtaatgc catgaaaaag aaatccggtc tcaaatttca caaattttgg 4920
gtgaggttta cattgtctgg attagtaatc tagacaaaag agaaaaaata cttttttttgt 4980
gtttatatcc tcaaaatcag acactaaaac ccctatggtt cctactagtc atactaaaag 5040
taaatggaca tatctagaag aggacaaatt ttggggacac atctgaattc tataaacctg 5100
gtttgacata aagtcaactt ctaattattc accctcacaa tacagacttc aagtatacat 5160
aaatagagat atatacacat ctgtttatga atttaggttg ataccaagag caaatactat 5220
attatacgta tataaaattg caaatctttc cttccaggct cataagcaag tatatctacg 5280
tgaatattaa tgacctcaga ggcctgattc catgatctgt gctgccagca cactgcgccg 5340

taatagtgag 5350

<210> 48
<211> 53
<212> DNA
<213> Homo sapiens

<400> 48
cagaattaca atgtgcttag ctttccatga ctcccttgcc accctgaaaa tgt 53

<210> 49
<211> 513
<212> DNA
<213> Homo sapiens

<400> 49
gggtgacctt gtctccaggc ttggaagtca catgaccatc ttctttcccc tctgggccat 60
gcgtgcctga aactgcagac agtctctaga actcagagaa ctgggaagct tttgttcctc 120
tgtggcgagg tcgggtcccg atgggggatt gttttcgctc agctcaaagg gacactttag 180
agatagaata cttcaacctt aagaagcagc aacatttgct tgtagctgga agtcttcatt 240
tctggtctcc agctgttgtc tggagccacc aggcctccgc tgaatgggcc tatgcccagc 300
aactggtggg ggtgggggca gtgcctgccg gactgaacat gaaccagtct gtgcaggacg 360
cccatctcca ggacagcttg gctgcaagga caccctgtcc cctcccggtg gtggttgctg 420
gggctttgga ataaacccgt tagctctggg gtgactnaac cggggaaaaa ggggtccaca 480
agaagcggtt taaggggggaa agtagaccca agc 513

<210> 50
<211> 417
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (19)..(21)
<223> a, c, g or t

<220>
<221> unsure
<222> (34)
<223> a, c, g or t

<220>
<221> unsure
<222> (46)
<223> a, c, g or t

<400> 50
ggaaaagact tctattttnn ncccttntnt tgangcaggg tatganttca tcattgtatc 60
cataccctgc ctcaaacaaa gggagggtat ggatacaatg atgaattcaa ttttgaatgt 120
gtcaaatttg aaatgttatg agatatctaa ggaagatgtc cagtaagcag cttactatac 180
aaacctggag ctcaggagac ttaaatgtag aagtggatat aggagaaagt gtagcattat 240
cagaaaagaa ggcttgtagc ttggaaggag taggatctgg ataattacca atatgtaggg 300
aatgggcaga agaaaatgag gtcactaata accaaatccc ccaaaaccaa aatgtttcaa 360
gataaaggag gaaaaccaga agaaactggt acaactaaag ntaagctaca aagaaca 417

<210> 51
<211> 1049
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (746)
<223> a, c, g or t

<220>
<221> unsure
<222> (801)..(802)
<223> a, c, g or t

<220>
<221> unsure
<222> (835)
<223> a, c, g or t

<220>
<221> unsure
<222> (840)
<223> a, c, g or t

<220>
<221> unsure
<222> (872)
<223> a, c, g or t

<220>
<221> unsure
<222> (876)
<223> a, c, g or t

<220>
<221> unsure
<222> (881)
<223> a, c, g or t

<400> 51
ctcagctcag aagtagtgtg atgcctgttc cattgcattt gctcagttaa attacaaggc 60
cagcccagat tcagggagca gggaagttgc ttctgcctct tgatggtagt tgctactaag 120
tcatgttgca cggggcatgg gtccaggaag ggaatgggct gcagccattt gggtaaacag 180
tttactgctg aagcctaggg gctgacactg aagaggggag tctttcactg gcatctgcgc 240
atgtccagga accaaaatgg ccaccattct ccctttcgtc tgttaaaatg aaacttagca 300
caggggaaag aactctagaa caatttcgag tcccacttga aagaagtttg gccttgggca 360
gtgcaattct ccccccatgg ccccccgcca aagatgcttc agcgtctttc ctttgggact 420
ggattttact tgcctcatat tttgtaagtg tgaaaattaa gtaaaattat atgaaatgtc 480
taggtatgta gtangcttag aaagcttatt actgttcttg ctctcattgt cttagtaata 540
agagttctgg attctagtct tgttctgttt ctaactttac acatgagagt taacctgttt 600
ggccttttt ccccctctgt aaatgatgag tttggaccga gtgactttta aagacttatc 660
caaactcaag ttcccagttt ggtcttgtct ggtttccatc tgtgcccagg gtgggcagtc 720
caatgcccca gtgccttcga ccaagntggt cctgcctgtc tgaggacctt ggagtgcatt 780
ctcccaatgg gtactgaggc nncacgcctt agagccccca gcatccctac tcatnctggn 840
aaagatgcgc tcccctcaac cctgctgtgg cnacanagaa acccctggtg ncctcagagg 900
cacctctggc aagctccttt cacaaatctt gcgaaacttg gcctttagac aatccgagct 960
cttgactgag ctagattttt gtttttgttt tctttccccc tagagtttcc acaatatccc 1020
ttttgagcag ttagggttag aggactatg 1049

<210> 52
<211> 1420
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (746)
<223> a, c, g or t

<220>
<221> unsure
<222> (801)..(802)
<223> a, c, g or t

<220>
<221> unsure
<222> (835)
<223> a, c, g or t

<220>
<221> unsure
<222> (840)
<223> a, c, g or t

<220>
<221> unsure
<222> (872)
<223> a, c, g or t

<220>
<221> unsure
<222> (876)
<223> a, c, g or t

<220>
<221> unsure
<222> (891)
<223> a, c, g or t

<400> 52
ctcagctcag aagtagtgtg atgcctgttc cattgcattt gctcagttaa attacaaggc 60
cagcccagat tcagggagca gggaagttgc ttctgcctct tgatggtagt tgctactaag 120
tcatgttgca cggggcatgg gtccaggaag ggaatgggct gcagccattt gggtaaacag 180
tttactgctg aagcctaggg gctgacactg aagaggggag tctttcactg gcatctgcgc 240
atgtccagga accaaaatgg ccaccattct ccctttcgtc tgttaaaatg aaacttagca 300
caggggaaag aactctagaa caatttcgag tcccacttga aagaagtttg gccttgggca 360
gtgcaattct ccccccatgg ccccccgcca aagatgcttc agcgtctttc ctttgggact 420
ggattttact tgcctcatat tttgtaagtg tgaaaattaa gtaaaattat atgaaatgtc 480
taggtatgta gtaagcttag aaagcttatt actgttcttg ctctcattgt cttagtaata 540
agagttctgg attctagtct tgttctgttt ctaactttac acatgagagt taacctgttt 600
ggccttttt ccccctctgt aaatgatgag tttggaccga gtgactttta aagacttatc 660
caaactcaag ttcccagttt ggtcttgtct ggtttccatc tgtgcccagg gtgggcagtc 720
caatgcccca gtgccttcga ccaagntggt cctgcctgtc tgaggacctt ggagtgcatt 780
ctcccaatgg gtactgaggc nncacgcctt agagccccca gcatccctac tcatnctggn 840
aaagatgcgc tcccctcaac cctgctgtgg cnacanagaa acccctggtg ncctcagagg 900
cacctctggc aagctccttt cacaaatctt gcgaaacttg gcctttagac aatccgagct 960

cttgactgag ctagattttt gtttttgttt tctttcccct tagagtttcc acaatatccc 1020
ttttgagcag ttagggttag aggactatga atgaactgga ctgttcgcta caaaattaaa 1080
ttcgaggaca ttttaattaa gtttggttat gagtacgtca tagtttttat ccctgggaca 1140
tacaagagcc ttaagacatc ttgtgtctaa aacactgcat tcttagcaaa aagcccacca 1200
aggagcctga gggaaggacc ttataaggtc cttatgtcat taaaaggaga atacctcagg 1260
aaccattagt aatggccagg tttgggacat tcagcaccca gtatgccgga acctcagaga 1320
agggattcag tacagtgcca acatgtcctt catatatcct ttggttaaat tctgagatga 1380
gccccagaca tagagctttg gattactatc agagcggccg 1420

<210> 53
<211> 84
<212> DNA
<213> Homo sapiens

<400> 53
atgaaatctc ataaaacatt tgaataagat catgtcttct tagcgaagaa ttttatatta 60
ccagtaaatt tagaaaaaaa atag 84

<210> 54
<211> 696
<212> DNA
<213> Homo sapiens

<400> 54
ttggtggggg atcatatgat ttggaacata gatttttag tttttgtttt tttttgtggt 60
cttcaagaga gcagttcaga gaccagggtg catggtggtt tactgagtgg gttggaagaa 120
tatggaagca ataaatacag gaattgatta aagaagttta gtttgagaag gaagaacaac 180
aactcttatt ctaaaactgg aggcaagaag taacagatgg atgaagttac agcattttag 240
aagctgtgaa gaggatttga ttacaggtgg agaaggtgtg atttgaggga attttataga 300
agggttcgag tatttgttgg aattagggat ttaaatatga aaatgatttg gattagtcaa 360
tgagacggag agttgtattt agaataaatc tgttgtggaa gatttcatag ctttctggtg 420
gtgcttaaca cccagtgcgt gggcatggag aaaacagatg gtgaggattg tctatcactg 480
gggagatgca tagtgagaat aatggaaggt catgatattc tggagaggac agtgttaaaa 540
tggctgttgg acaggtttaa attatatagg gagacaataa agccaagtgg aggtaaagag 600
caggtctaca actagaacat aaagtagttg tggataaaga aaaggggggt ggtctctgaa 660
gttacaatcc tagtggattg taactatttt tttttt 696

<210> 55
<211> 1284
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (719)
<223> a, c, g or t

<400> 55
aaaaaaaaaa tagttacaat ccactaggat tgtaacttca gagaccaccc ccctttttctt 60
tatccacaac tactttatgt tctagttgta gacctgctct ttacctccac ttggctttat 120
tgtctcccta tataatttaa acctgtccaa cagccatttt aacactgtcc tctccagaat 180
atcatgacct tccattattc tcactatgca tctccccagt gatagacaat cctcaccatc 240
tgttttctcc atgcccacgc actgggtgtt aagcaccacc agaaagctat gaaatcttcc 300
acaacagatt tattctaaat acaactctcc gtctcattga ctaatccaaa tcattttcat 360
atttaaatcc ctaattccaa caaatactcg aacccttcta taaaattccc tcaaatcaca 420
ccttctccac ctgtaatcaa atcctcttca cagcttctaa aatgctgtaa cttcatccat 480

ctgttacttc ttgcctccag ttttagaata agagttgttg ttcttccttc tcaaactaaa 540
cttctttaat caattcctgt atttattgct tccatattct tccaacccac tcagtaaacc 600
accatgcacc ctggtctctg aactgctctc ttgaagacca caaaaaaaaa caaaaactaa 660
aaaatctatg ttccaaatca tatgatcccc cacccaagtc cttacccttt tttggtttng 720
tttttgtttt tttttttgag atgggctctt actctggtca cccaggctgg agtgcagtgg 780
ttcaatacca gctcactgca acctccgcct cccaagccca agcgatcctc ccacctcagc 840
cccccaagta gctgggacta caagtgcgcg ccaccacacc cagcaaagtt tgttatcttt 900
ggtagagatg aggtttcatc atgttgccca gcatggtctc gaactcctga gctcaagcaa 960
tatgcccacc tcagcctccc gaactgctga aataacaggt atgagctacc atgctcggcc 1020
agtccttacc ctctttttgg caatgacact taagctaccc accttgcttt atgtcactgt 1080
aatgattcta ctctgacctc tctggttgtt tcttctatct gaaatcactg ctgctcttcc 1140
tagcccaaat tgtggatatt ttcccaaggt ctgacctgga gcctcctata attctcagcc 1200
actgagggtc tcatccctca ccactgtttc tacattttca ttcttcatat ggctgtcttc 1260
ccccagcggt tggatgttga ttgg 1284

<210> 56
<211> 411
<212> DNA
<213> Homo sapiens

<400> 56
ttgcccaggg atacatagct agcaagtggc agcgctggat tgagtctggg ccttgtctga 60
ggctcgggtc ctgtcatgct ctgcggttgc tatgttgaca tgcaaaggga gaggcagctg 120
ctgggagtct aggtgggttt ctctttgaga atgctaacgt gaaccctcaa ggtgaatcag 180
aatccttttg caagtgaata atcagatgta ggttcctgtg tctccctgta aaatgaaagc 240
ctctttttt ccaaggtcca gtatagacct gaagctgggt tactctggaa tttccctctc 300
tggctggagt gactgaggcc ttgcacgtga cattggtgag gactcgcagc ctcaggtctg 360
gcttccctta gcaacccccc tttcctgtct ctgcctctgg agttcaccat t 411

<210> 57
<211> 970
<212> DNA
<213> Homo sapiens

<400> 57
cttctgtctt atgtaccact tcccttggct caaggcgtcc tttttatctt tcttccactt 60
gactaaatga gaatagtgtg ggtcactctc tacctgcctc ccatctgtgg ttccttttgg 120
agatgggccg agtgggccac tcacccttta attttctctt agtttccttt ctgtacacca 180
gtttgaacct tagtattatc actaatgcaa atatgagcct aggctcaatt tttccagtta 240
tgaaatgggg ctggcattat tccgtgatgt gcatgttaag agaggggaaa gctcacattt 300
ttgaggtcct cttgtggttc tattttgtgt aggaactcac gctttgttta ttcagcaatc 360
attcctccag aaataacctt aatagcaaca agaaaaaaga ataggtgttt tttgagctct 420
atctgccagt ttctctatat atgaacatta tatattgcaa cataacactc acaatgcctt 480
taaacatcat ccccgttata cagataagaa aacagaattt caaagaaggt aggggacttg 540
cccagggata catagctagc aagtggcagc gctggattga gtctgggcct tgtctgaggc 600
tcgggtcctg tcatgctctg cggttgctat gttgacatgc aaaggggagag gcagctgctg 660
ggagtctagg tgggtttctc tttgagaatg ctaacgtgaa ccctcaaggt gaatcagaat 720
ccttttgcaa gtgaataatc agatgtaggt tcctgtgtct ccctgtaaaa tgaaagcctc 780
ttttttttcca aggtccagta tagacctgaa gctgggttac tctggaattt ccctctctgg 840
ctggagtgac tgaggccttg cacgtgacat tggtgaggac tcgcagcctc aggtctggct 900
tcccttagca acccccctttt cctgtctctg cctctggagt tcaccattaa aaaaaaaaaa 960
aatttaaaag 970

<210> 58
<211> 117
<212> DNA

<213> Homo sapiens

<400> 58
tggcagtatt taattaaatt atatatacac atacctgttg acacagcaag caagcgcagg 60
gataaataag aatttatccc ttaagagtca cctccaggcc gctatgctag tggccct 117

<210> 59
<211> 2458
<212> DNA
<213> Homo sapiens

<400> 59
atgggcctcc ctatcctact gctgattgca ctgtttttgt gtcctggatt gttgctggta 60
accagggagt acgaattatg gaagcaggtt aagcttaatt taaagatctc tttaacagtt 120
aaaattgggc tgtttcaaga agaaataggt ggccttgatg gtggtggtct cctgctcccc 180
aagtctgaca gcacccccctg ctttgagatc cctcaggcca tggagagcaa gctcctcatc 240
gggggcagga acatcatgga tcacaccaac gaacagcaga agatgttgga actgaagagg 300
caggagattg ccgagcagaa acgtcgtgag cgggagatgc agcaggagat gatgctccgg 360
gacgaggaga ctatggagct ccggggcacc tacacatccc tgcagcagga ggtggaggtc 420
aaaaccaaga aactcaagaa gctctacgcc aagctgcagg cggtgaaggc ggagatccag 480
gaccagcatg atgagtatat ccgcgtgcgg caggacctgg aggaggcgca gaacgagcag 540
acccgcgaac tcaagctcaa tctgctccct ttggcttttc ttctgctttc cacatcagct 600
gaagatttag acccttttcct gctgcctgtg ttgagagctg tcttcctgag ttggaagcca 660
gccatggaga gcaagctcct catcgggggc aggaacatca tggatcacac caacgaacag 720
cagaagatgt tggaactgaa gaggcaggag attgccgagc agaaacgtcg tgagcgggag 780
atgcagcagg agatgatgct ccgggacgag gagactatgg agctccgggg cacctacaca 840
tccctgcagc aggaggtgga ggtcaaaacc aagaaactca agaagctcta cgccaagctg 900
caggcggtga aggcggagat ccaggaccag catgatgagt atatccgcgt gcggcaggac 960
ctggaggagg cgcagaacga gcagacccgc gaactcaagc tcaagtacct aatcatcgag 1020
aacttcatcc cgccggagga gaagaacaag atcatgaacc ggcttttcct ggactgtgag 1080
gaggagcagt ggaagttcca gccactggtg ccagccggcg tcagtagcag ccagatgaag 1140
aagcggccaa catctgcagt gggctacaag aggcctatca gccagtatgc tcgggttgcc 1200
atggcaatgg ggtcccaccc caggtacagg ctgtctttga gatggaattc tctcacgacc 1260
aagaacaaga ccctcgtgcg ctacacatcg gagaggctca tgcgattgga cagctttctg 1320
gaaagacctt ccacgtctaa agtccgaaag tccagatcct gcagtagcag ccagatgaag 1380
aagcggccaa catctgcagt gggctacaag aggcctatca gccagtatgc tcgggttgcc 1440
atggcaatgg ggtcccaccc caggtacagg ggggtggaca tccagattgt gggggatgac 1500
ctgacagtca ccaaccccaa gaggattgcc cagtcctttg agaagaaggt ctgcagctgt 1560
ctgctgctga aggtcaacca gatcggctcg gtgactgaat cgatccaggt tactcaaaac 1620
aagagaaacg taaattacaa ctacactgag ataccatttc tcacccatca gattgtcaaa 1680
aacttaaaag cttaaccgta ctctctacgg ataagactgt ggagaaacag gcattttcat 1740
acattgttgg tgggaatgca aaatgtttta agtcctatag agggaagttg gcagtattta 1800
attaaattat atatacacat acctgttgac acagcaagca agcgcaggga taaataagaa 1860
tttatccctt aagagtcacc tccaggccgc tatgctagtg gcccacgtct gtaatcctag 1920
cacttggcag gcccagatgg gcggattgcc tgagctcagg agttcgagac cagcttgggc 1980
aacatagtga aaccctgtct caactaaaat acaaaaaatg agccaggtgt ggtggcagtt 2040
gcctgtattc ccagctactc aggaggctga ggtatgagaa tgcttgaacc tgggaggcag 2100
aggttgcagt gagctgagat catgtcacta aattccagct gggcaaaagg gattataggc 2160
gtgagcacag ccccgctaaa atcatattca agaagcaatt cagtttcttt ctaagcttgt 2220
agtcaggggt caatgatttt ctagcctgaa ttaaccagtt taagtttgag gaaagtcctc 2280
ttcagtgggt tcaaatgaca tgggaagcaa aaagagaatt aaaaaaaaga caaagaaaag 2340
ggggaaaata ccggaggcag aagacacatt ggaaggccat gagattaaag aggggaaaaa 2400
gaaacgttga gctgggtgta aaaagaaaat atggtggtga taataaaaat aaaagaac 2458

<210> 60
<211> 133
<212> DNA

<213> Homo sapiens

<400> 60
ctcgagcagg catgagccac tgcacccagc ccaatgttaa tttttaagaa tggaaaaatg 60
ctttttaact tgaaataaaa ctacggggaa aattatagaa ttgaatgaaa aattaaacaa 120
tatctaaata aaa 133

<210> 61
<211> 501
<212> DNA
<213> Homo sapiens

<400> 61
ctcgagcagg catgagccac tgcacccagc ccaatgttaa tttttaagaa tggaaaaatg 60
ctttttaact tgaaataaaa ctacggggaa aattatagaa ttgaatgaaa aattaaacaa 120
tatctaaata aagaaagaca cttgttattt ttctggccag ggtaactcca cacgtaaaga 180
cgtccatcct ccccagactc cgaagaatta agaagaaaat tctgtgaaat ctgacaagct 240
gagtgttgca tacagacctg caaaggccca tttgccagga cagacacctg aaggaaaaga 300
ggagagtgag gtccaccaga taccaagaca gaatgatgat gctgcagtgg gaggaagaca 360
gacaacgggc ccagagtggg gagcccatcc ccagaccccc acagacacag gacccacacg 420
tgggaccaga gacagaggtg gaggctcagg accctggaga aagaatgcag tatccccaga 480
tgggacagga tacttaatgg g 501

<210> 62
<211> 308
<212> DNA
<213> Homo sapiens

<400> 62
gccaacctca ttcagaggaa aatgtgttta aatggccttc aagaagaatt ctgtggaaag 60
ataaaatcat catgggggac ttatattctg gtggatttgg caatgtggta gtgatcccat 120
tagagtttat gattttatgc atataaactg caggtcttta attattttaa attcagccaa 180
caaacattta ggcatttatt gaacacccac tgtgttgtag gcaatgtggc tatgctgtgt 240
gggtaaaaag tatacaggga cggttgcggt ggctccacac ttgtaatccc cagcactttg 300
gggaggct 308

<210> 63
<211> 442
<212> DNA
<213> Homo sapiens

<400> 63
gtgagaactg tgtagaatca tactactgtt gtatttattt cctaggtaac atttttggta 60
gttttgtcac acagttaatg gatttgtttc tattaacatg tagggccttc ttgacctaat 120
catttttcac acgtgttcct atgtaaattc taaatccaaa cgtgctaggt caatagctaa 180
attgagttta tagaagaaag gaccactgca aaatcaatta ttccagtttg acgctggctc 240
atttacttgc ctttgggagg acttttaata cccatttacc ccaaatgcaa ataacccaaa 300
agctgaaatg attaaagttt agtgggctat tatagtagaa atctggacac aaaaattgct 360
actaaaagga actgtggttg gtcatcttag taaccaaaaa cttttaattc caaacattca 420
ggaaatgggc aaggagacag ag 442

<210> 64
<211> 279
<212> DNA
<213> Homo sapiens

<400> 64
cttggaaaaa tatttattag gctctctcct tctatttgcc agaaatagag gcaaaggctg 60
cttttccata taatccaaac tagcttatat tctgtcttgg tgaagtagcc tctaaagaaa 120
aaaccattac tgaggggcaa atgctatcta tgtcatggta atataccttg taaataaggg 180
ataaatacct tgtaataggg ataaaataaa ggctggatct gccctattca ccacaatcac 240
tagcacagtg cttggaacat aggaggggct tagtacaca 279

<210> 65
<211> 846
<212> DNA
<213> Homo sapiens

<400> 65
ccgactccca aatagtggca ttgattttct ccaacttatg acaagacatg ggtcttgacc 60
agttaccctc ccaaagggag attttccgaa aatttccagg caggcaaaag ttgggcttac 120
aataaaactt tccatcttag aatgtagctc gcaaaagtca acactataac ttatttacct 180
tgagccagac tcatcacttg tactaattaa aagaaggatt gtctaaactc cagaagccca 240
tgttttggat agaatttaat tcaagttcct atggggacat gctgatggaa attgaaaaat 300
atttattggt aactagtata tcacaagcta gttataatgc aattcttgaa aatccatctt 360
tcaaggaatc taattgttct gtgtactcta gatttcagct ttaactcact tgccaaactt 420
tgcagataat cctaagagga actttatgta ttctgaatta gtgaacccct gaaacgacag 480
catttaagtg aaattgctta tactcagctt ctcatgtcat ttggaaggaa cttccattaa 540
cataatattg ttgctgctct ataaccccag gagaagagga ataacttatc tgattagtgg 600
gttctgcaga aagaaatggg catacacagc tcaatctgca gccttgatta tttatgtatt 660
tgtctgtttc aggtcttcct gcctgtaaga aggaggcact ctctgcttct gtcatgtgct 720
aagcttcgtc aaacagatta tctctctaaa ggtagaggcc tgtgattgac cgctgtacca 780
atgactgagc aggtgagctt cacttattca ggaatgaaat actagcaggt ccacttgggc 840
aacagg 846

<210> 66
<211> 1021
<212> DNA
<213> Homo sapiens

<400> 66
ccgactccca aatagtggca ttgattttct ccaacttatg acaagacatg ggtcttgacc 60
agttaccctc ccaaagggag attttccgaa aatttccagg caggcaaaag ttgggcttac 120
aataaaactt tccatcttag aatgtagctc gcaaaagtca acactataac ttatttacct 180
tgagccagac tcatcacttg tactaattaa aagaaggatt gtctaaactc cagaagccca 240
tgttttggat agaatttaat tcaagttcct atggggacat gctgatggaa attgaaaaat 300
atttattggt aactagtata tcacaagcta gttataatgc aattcttgaa aatccatctt 360
tcaaggaatc taattgttct gtgtactcta gatttcagct ttaactcact tgccaaactt 420
tgcagataat cctaagagga actttatgta ttctgaatta gtgaacccct gaaacgacag 480
catttaagtg aaattgctta tactcagctt ctcatgtcat ttggaaggaa cttccattaa 540
cataatattg ttgctgctct ataaccccag gagaagagga ataacttatc tgattagtgg 600
gttctgcaga aagaaatggg catacacagc tcaatctgca gccttgatta tttatgtatt 660
tgtctgtttc aggtcttcct gcctgtaaga aggaggcact ctctgcttct gtcatgtgct 720
aagcttcgtc aaacagatta tctctctaaa ggtagaggcc tgtgattgac cgctgtacca 780
atgactgagc aggtgagctt cacttattca ggaatgaaat actagcaggt ccacttgggc 840
aacaggccca gagccataaa acccaagaga tttaacagta gaaaaaatga atgaatgatt 900
tttccctttt tactcccgtg tgtaactctc aactggttct aaacacaggt gggaaaacgg 960
atctcattta ttggatcaac attgcgtcct cagtggtgtc agaacagaag gttaatttag 1020
c 1021

<210> 67

<211> 415
<212> DNA
<213> Homo sapiens

<400> 67
gtttgctctg aatttattgc gagtgaaaaa cagagaaaat cctcaagttt aagtttctga 60
tagcagagtg tgggagttag agcatgggga gtccagaggt tccagacccc caaaggtctc 120
taccagggcc atctccgtta gtggcggtgg cagcccctct tgtggccttt ttcctctctc 180
caaggggtca ccccgcacca tgccgctccc cctcatctat cttgcccgat cgttggtggg 240
tttgagctta tagaggcaga ggagtaagaa cctgcgatat tgaaagctac ccacatgggg 300
cttccttgaa ggaggacgtg gaaggcagaa agtgacctgc tctgagcggc gcatgtaacc 360
gaggacctta agctggacca cggggcttgg acgatttttt aaatcaggaa atcga 415

<210> 68
<211> 458
<212> DNA
<213> Homo sapiens

<400> 68
ttttgtttgc tctgaattta ttgcgagtga aaaacagaga aaatcctcaa gtttaagttt 60
ctgatagcag agtgtgggag ttagagcatg gggagtccag aggttccaga cccccaaagg 120
tctctaccag ggccatctcc gttagtggcg gtggcagccc ctcttgtggc ctttttcctc 180
tctccaaggg gtcaccccgc accatgccgc tccccctcat ctatcttgcc cgatcgttgg 240
tgggtttgag cttatagagg cagaggagta agaacctgcg atattgaaag ctacccacat 300
ggggcttcct tgaaggagga cgtggaaggc agaaagtgac ctgctctgag cggcgcatgt 360
aaccgaggac cttaagctgg accacggggc ttggacgatt ttttaaatca ggaaatcgac 420
ctcatcttcc tcctcctcgt cctcttcccc tgaacccc 458

<210> 69
<211> 1033
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (14)
<223> a, c, g or t

<220>
<221> unsure
<222> (479)
<223> a, c, g or t

<220>
<221> unsure
<222> (851)..(904)
<223> a, c, g or t

<400> 69
ccgccggttt cgangctggt tgatgacaaa atgtcggcag cgatcactgc ccctgcaaca 60
caagtgcttg tggtctgggg ctgctggtac acaggcagcc cagccttggc ccagggtctg 120
agctctgtgc ctgggtgcag gtgaggggtc ccagctcttg atccagaaca gacctgcctg 180
acctggggcc actgtacccc acttggagcc atggtgtgtt catcaggaag ctacggagag 240
gttttcaaac cgtggagccc tgggatcctg ggaagtacct aagcctgctc tggtggagtc 300
agggagagca cggctgtgac tggagtgagg caagtgaggc actcatctta ggtgcaaaat 360
ttaaaggggc accaaaaaac tcaataaaga aaactaataa cgcagtattt tagaaaatca 420
aaatatatga aaaaaaatcc acaatgaaca aaacaccaaa gttttaaata aagacaggnt 480

ccaaccctgc acctgtacaa ctcaacctca ccctactccc caccctgctg caatgatgga 540
gttccagctc ccacccccctc ttcggcctgt aaagtcccac cctaaaatcc taccctcttc 600
atctcccttt ttcctagaag aataacctct acacagtgat gtgtgtacat tataaatgtg 660
cagcttgatg aatttccata taggaaccct cccatgtaac tgccactcag gtcaagatac 720
aaaaccttc cagcccccag aagacctact tgcgctccca tccagtcaat gccccctaaa 780
ggtagccacc attccgacgc ctatcagcat agattagtct tgcccattag agaacttcta 840
taatacttct nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 900
nnnnccaagt agaattagtt tctttttttt ctcatttata tgtagtattc tagttaatga 960
aaaatccacc atgtacctat tctgatgata gacatttagg tagtttccgg atttgggctg 1020
ttataaataa agc 1033

<210> 70
<211> 1075
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (521)
<223> a, c, g or t

<220>
<221> unsure
<222> (893)..(946)
<223> a, c, g or t

<400> 70
cagccagggt gctgtggctc tgggtgtctc ctgagatgat gtaacgccgg tttcgaagct 60
ggttgatgac aaaatgtcgg cagcgatcac tgcccctgca acacaagtgc ttgtggtctg 120
gggctgctgg tcacaggcag cccagccttg gcccagggtc tgagctctgt gcctgggtgc 180
aggtgagggg tcccagctct tgatccagaa cagacctgcc tgacctgggg ccactgtacc 240
ccacttggag ccatggtgtg ttcatcagga agctacggag aggttttcaa accgtggagc 300
cctgggatcc tgggaagtac ctaagcctgc tctggtggag tcagggagag cacggctgtg 360
actggagtga ggcaagtgag gcactcatct taggtgcaaa atttaaaggg gcaccaaaaa 420
actcaataaa gaaaactaat aacgcagtat tttagaaaat caaaatatat gaaaaaaaat 480
ccacaatgaa caaaacacca aagttttaaa taaagacagg ntccaaccct gcacctgtac 540
aactcaacct caccctactc cccaccctgc tgcaatgatg gagttccagc tcccaccccc 600
tcttcggcct gtaaagtccc accctaaaat cctaccctct tcatctccct ttttcctaga 660
agaataacct ctacacagtg atgtgtgtac attataaatg tgcagcttga tgaatttcca 720
tataggaacc ctcccatgta actgccactc aggtcaagat acaaaaccct tccagccccc 780
agaagaccta cttgcgctcc catccagtca atgcccccta aaggtagcca ccattccgac 840
gcctatcagc atagattagt cttgcccatt agagaacttc tataatactt ctnnnnnnnn 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnccaa gtagaattag 960
tttctttttt ttctcattta tatgtagtat tctagttaat gaaaaatcca ccatgtacct 1020
attctgatga tagacattta ggtagtttcc ggatttgggc tgttataaat aaagc 1075

<210> 71
<211> 549
<212> DNA
<213> Homo sapiens

<400> 71
caaaacagtg ccccctcccc cctgcacatg aaagtcaaat ttgctaaaag catgtcattt 60
cttgtttcgg gttttgaaga caacgatttt tatttccggt gtgttctagg accagctgca 120
tcattttatt cttgcttaaa gtgctttatt ttaggtaagc tttttgactt gccccagagt 180
aaacttaaaa acctcaaggt gtgcctaaaa gcaacgattg aaaaaatttg aggaatgctt 240
tggatgagtg tctgtagcct ttgataattt agaaaatatc tatttcatgt aatgtattgt 300

ttctgttttt ttctttgcca ctttgactta tgactgcagg accaagtgta tttctgtcct 360
gtattctacc gagtccgagg gtggtactgc gtaggcagac ttctgttata gtttgtagtg 420
ttcatttttt tgttacacgt tttatttaat ttttttcaat ccaattcatc aagcaagaca 480
cacattatag attaagagct gataatagat gtttagtttt aaaaaggatt atgtgcggga 540
gactctctg 549

<210> 72
<211> 574
<212> DNA
<213> Homo sapiens

<400> 72
caaaacagtg ccccctcccc cctgcacatg aaagtcaaat ttgctaaaag catgtcattt 60
cttgtttcgg gttttgaaga caacgatttt tatttccggt gtgttctagg accagctgca 120
tcattttatt cttgcttaaa gtgctttatt ttaggtaagc tttttgactt gccccagagt 180
aaacttaaaa acctcaaggt gtgcctaaaa gcaacgattg aaaaaatttg aggaatgctt 240
tggatgagtg tctgtagcct ttgataattt agaaaatatc tatttcatgt aatgtattgt 300
ttctgttttt ttctttgcca ctttgactta tgactgcagg accaagtgta tttctgtcct 360
gtattctacc gagtccgagg gtggtactgc gtaggcagac ttctgttata gtttgtagtg 420
ttcatttttt tgttacacgt tttatttaat ttttttcaat ccaattcatc aagcaagaca 480
cacattatag attaagagct gataatagat gtttagtttt aaaaaggatt atgtgcggga 540
gactctctgt tagaggagcc tttttcatct gaca 574

<210> 73
<211> 299
<212> DNA
<213> Homo sapiens

<400> 73
aactattttc ccatagtgta tattagtcta ctacctgtaa gcatacactg tacatgcctc 60
agatatagga gtctcacaac gaggttcagc taatgggact gattattttg ataaacatat 120
gagaagaaca ttcatgttca gtgagtatat atttaaaagt aggtatcttg gtatactgtg 180
tccttttttt tttcctttaa agttaattac aaaccatatg agaagtaatc tgatataata 240
gatataatgt tagatttgaa gtcagaaaac ctttctgatc aagaattagc tcctctggc 299

<210> 74
<211> 898
<212> DNA
<213> Homo sapiens

<400> 74
attttttaa tgaaatggta gatgttctaa ggaagttcag tgtttaaaag ggaggttgta 60
aactcaaatg cttacaggag tcaagtgtag gtaacaggaa tatacctggg atagccaatg 120
cgcaggtcca gctggctgtt ggtattcacg taatacaggt gcaacattgc tattataagc 180
agaaagaacc aggaagcatt atgcacatac acataaagtc tttttatgac atctagaaaa 240
gacagactga aaattccttg agtattttcc ctctaatata ttttaacaca ttttttgaca 300
tgggaggtga tgtacattga aggaagcaga cacaggaaac tttattgtag gaaatggtga 360
tttaagtgca tgccatcatg acaccagagt gaaataaaga agtaagcctt tgagtgttgc 420
catgtgcctg gtactatgct cactactaca cagacttcta atctcaccaa cagtcctact 480
aattgggaat tattatatac attatcaagg aaaatgagct cagaaatact taacgtaatt 540
tgccgctagg gtgacacagt tactaagggg gagagccagg attataacct agatttgtct 600
gactcctttt caccatatca tttaacaaac tagacaaaca aaaatgaata aataagcacg 660
cttttcagaa gtgccaaaga tattaaaata ctttatgtgc ttatataaac atattactaa 720
ctgttcttgt aactattaac aattcaagct atactaataa tcatctaact ggaatatggc 780
ttttgaaata tattctataa ccacattatt atgccttgct tttctccagt gtcagctgca 840
ggtagatgaa tctaaagtaa atggtacaga aaagacatcc tcacggtccg ggcgtggt 898

<210> 75
<211> 846
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (49)..(50)
<223> a, c, g or t

<220>
<221> unsure
<222> (479)
<223> a, c, g or t

<220>
<221> unsure
<222> (515)
<223> a, c, g or t

<220>
<221> unsure
<222> (600)
<223> a, c, g or t

<220>
<221> unsure
<222> (623)
<223> a, c, g or t

<220>
<221> unsure
<222> (625)
<223> a, c, g or t

<220>
<221> unsure
<222> (632)
<223> a, c, g or t

<220>
<221> unsure
<222> (640)
<223> a, c, g or t

<220>
<221> unsure
<222> (685)
<223> a, c, g or t

<220>
<221> unsure
<222> (695)
<223> a, c, g or t

<220>
<221> unsure

<222> (729)
<223> a, c, g or t

<220>
<221> unsure
<222> (763)
<223> a, c, g or t

<220>
<221> unsure
<222> (773)
<223> a, c, g or t

<220>
<221> unsure
<222> (777)
<223> a, c, g or t

<220>
<221> unsure
<222> (791)
<223> a, c, g or t

<220>
<221> unsure
<222> (812)
<223> a, c, g or t

<220>
<221> unsure
<222> (840)
<223> a, c, g or t

<400> 75
aaaatactga atatgcattg caaattagca tgccaacctt gctaaatcnn aataacagcc 60
ctgacacgta cttaagtaga aagtggaact acacaaaaat ggaggaatat cagtagataa 120
catgtaaacc aggacgatga aaaaagagga cacaagtaag atcactgcgg gtatattatt 180
cttagctaca tatagctaac tgagtcacca ttttctaagg aggaaaaatt agaaggagtg 240
ctgaaaggga ataactccag actgttaatt acttacagtg aatatatgtt tgcatatatt 300
tgcttaaagt agctgttaag gaagctctat ttaatttatt tgtcagtgtg ggctgagtca 360
tcaacaacta tacttttcac ttttttcata gaatccagga caagaagaat ctttcaagtg 420
ttgctaacca gaactgagca gatccaagta gcaatcgtag ccaagccatt agtccatgna 480
tggagagacc agagagggca tccccacgtg gtgangcctc aagcttcagg agcactcgac 540
cagagtcagt gtagcccttg gcagctggtg aagcacaacc tgggacaaat ggaagctttn 600
gcaatgagcc gataggtcaa ggntnaccag anagaatgcn gctgggtaat gcaagctatc 660
ctatcttgta attaaaaggg ttntntgtgc ggttncctga cgtagttact aaatggcttg 720
catgaaatna catgcagcat tctgcagtta ctgtgcaatt acnttatatc atnaccntac 780
agtcaaaaga naaaaagaaa attcagggtg angcttttaa ccgcaatttg tagcaaagan 840
gtttgg 846

<210> 76
<211> 880
<212> DNA
<213> Homo sapiens

<400> 76
aaaatactga atatgcattg caaattagca tgccaacctt gctaaatgaa ataacagccc 60
tgacacgtac ttaagtagaa agtggaacta cacaaaaatg gaggaatatc agtagataac 120

atgtaaacca ggacgatgaa aaaagaggac acaagtaaga tcactgcggg tatattattc 180
ttagctacat atagctaact gagtcaccat tttctaagga ggaaaaatta gaaggagtgc 240
tgaaagggaa taactccaga ctgttaatta cttacagtga atatatgttt gcatatattt 300
gcttaaagta gctgttaagg aagctctatt taatttattt gtcagtgtgg gctgagtcat 360
caacaactat acttttcact tttttcatag aatccaggac aagaagaatc tttcaagtgt 420
tgctaaccag aactgagcag atccaagtag caatcgtagc caagccatta gtccatgtat 480
ggagagacca gagagggcat ccccacgtgg tgatgcctca agcttcagga gcactcgacc 540
agagtcagtg tagcccttgg cagctggtga agcacaacct gggacaaatg gaagctttag 600
caatgagccg ataggtcaag gctaaccaga aagaatgcag ctgggtaatg caagctatcc 660
tatcttgtaa ttaaaagggt tttctgtgcg gttacctgac gtagttacta aatggcttgc 720
atgaaataac atgcagcatt ctgcagttac tgtgcaatta ccttatatca tcaccctaca 780
gtcaaaagac aaaaagaaaa ttcagggtga agcttttaac cgcaatttgt agcaaagatg 840
tttggaataa aaacacattg cttgttaaaa aaaaaaaaaa 880

<210> 77
<211> 637
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (90)
<223> a, c, g or t

<400> 77
ttgcacagac ttttaaaaca aaagtcttgt ttccggatgt tttgttttgt actatgagtt 60
aatttgagtt ggctgtgaag agcgtaagtn attcttctca agtccctgtg ttttttgcat 120
cagaactggc atatggaata tgtactgtga aagggtttag aagacagtgc ttaatttcca 180
tttcggcaag atgagtttca gagattaaag agctgaggtg gtgagtggtt gtgatgtaaa 240
tgccatcatt ttctcaatac tggtggccag cgttagaggg aaagaggctg aaggcctgac 300
cttgctgatg tccagccttc tctgtcatgg ctctgctggt ctgtgtcctc caccctgtgg 360
ccacacccca ccttttcaag agcccttctt tgaaagtcag gagacctaag gttgaggact 420
ggtcctacct ttgtctttga atagttttta gcctgaagcg tcttatcccc tggtgggtgc 480
ttggatattt gtgggggaca attttggttg tcacaattat ggttattata ttatttttga 540
gttttgtttt atttgaagat aataatgatg gcattataaa tattaattat aaaacgaggg 600
tgctgggtgg gcgtggcgac tgacgcctat aatccca 637

<210> 78
<211> 874
<212> DNA
<213> Homo sapiens

<400> 78
ttttttttt gagatactgt cttgctgtgt cgcccaggct ggagtgcagg ggcatgatct 60
cagctcaccg caaactccac ctcctgggca agcaattctc ctgtctcagc ctcctgagta 120
gctgggatta caggcaccca ccaccacacc cagctaactc ttgtatcttc agtagagaca 180
gggcttcacc gtgttggcca ggctagtctt aaactcctaa cttcaagtga tccacactgg 240
gattataggc gtcagtcgcc acgcccaccc agcaccctcg ttttataatt aatatttata 300
atgccatcat tattatcttc aaataaaaca aaactcaaaa ataatataat aaccataatt 360
gtgacaacca aaattgtccc ccacaaatat ccaagcaccc accaggggat aagacgcttc 420
aggctaaaaa ctattcaaag acaaaggtag gaccagtcct caaccttagg tctcctgact 480
ttcaaagaag ggctcttgaa aaggtggggt gtggccacag ggtggaggac acagaccagc 540
agagccatga cagagaaggc tggacatcag caaggtcagg ccttcagcct ctttccctct 600
aacgctggcc accagtattg agaaaatgat ggcatttaca tcacaaccac tcaccacctc 660
agctctttaa tctctgaaac tcatcttgcc gaaatggaaa ttaagcactg tcttctaaac 720
cctttcacag tacatattcc atatgccagt tctgatgcaa aaaacacagg gacttgagaa 780
gaatcactta cgctcttcac agccaactca aattaactca tagtacaaaa caaaacatcc 840

ggaaacaaga cttttgtttt aaaagtctgt gcaa 874

<210> 79
<211> 1021
<212> DNA
<213> Homo sapiens

<400> 79
gcagctttga agacttattt gcatcttggt tcttggtgtt ttctggataa gttgaggttg 60
cagcatatga cttctgtctt tgataggcct ccccctttt tcctcctact cttgcctgaa 120
tcaattttgt tatggaaaga gtgaaagttt tctcagggaa aatgagaaga agctcaacgg 180
aggaagaccc agggcctgag caagccttgg ccggtccgtc agccactgtt tgtacagcct 240
aaatatactg agccctgcaa aaggttcagt gggatgtcat ggagcctgcc atcaagaagc 300
ctagtctgtg gttgggagag agaacaaaca gacctgaaac tcccgattaa aagcagtggt 360
tttgcatctt gtgtgagttc agagaagagt gatttaagtc aggggcttgg tgtcgggagg 420
gtgggatcca agtgggattc aagtaagtgg ccatggatga atttgtgaat caccagttag 480
gaagtggcag aggttaggga cagatgtagc agcacaactg aggaactcat ttcaaaggaa 540
gaatcgtaaa tgtatgaccg gttcacagac tgttgagttg tgtggctgat gaaaatgcag 600
tctcttgggc tccaggtgtg attgcccatg ggctgcggga gaggggaagt gcccaggagc 660
agcccaggtt gctctagtac agatgggata cactttggga aacattggta aaggatgggt 720
aaggcaaagc agatgaattt ccttcccatg agagacagaa attggggaag ccatccagac 780
tcaacagcct ggtggttgtc cctctattac agaagagcag gtagcaagca ggtattccta 840
agaatgacgg gacatgggag tgacaatagg gttagcacct cagatggagc cggtgttccc 900
gcaggcttca ccccaggttc tttcttatcc agcctctgtc acatctgaaa acatcatgat 960
gctcatgaaa tcaaggtcac attcaacctg catgtgcgcc agctggtttt ataccaagga 1020
a 1021

<210> 80
<211> 566
<212> DNA
<213> Homo sapiens

<400> 80
tcgagctcct ggctggaggc tgtaagcgga agtgacgcaa gcgaggcgcc accctctttt 60
ccggtactgg ctccgcgagc ctgagcgagg tttacccatg gttgtccttg atttcggagg 120
gagcaagtgt actggtagtg gaacgggtgg tgagagttga agtttatccg aagctgccga 180
ggggccttct tagagatact tggcctgctg tgctcgtgtc aggtaggtct gcgaggccgc 240
tcgggctgtc agtcctcgcg aaagtcgggg tcgataattg ccgccctcac ccatggagct 300
cccttcaaaa gcctccaaaa agaccattgt gagtttttc tatgaggaaa agaatttcct 360
ccacttgagc cacgttaact tgtcccccag cgtcgtcctg ccttacagac cctgcgattc 420
ccgcgcattc cggtgagcac tgggtgaggg atggttcaag ggggcatctt gttatacgaa 480
cgaatgtttc aactgagaat gttcctttgt ttatgcgtca taaacgtatt tttgacgcca 540
tacattctgt tataaagaca ctttaa 566

<210> 81
<211> 706
<212> DNA
<213> Homo sapiens

<400> 81
tcgagctcct ggctggaggc tgtaagcgga agtgacgcaa gcgaggcgcc accctctttt 60
ccggtactgg ctccgcgagc ctgagcgagg tttacccatg gttgtccttg atttcggagg 120
gagcaagtgt actggtagtg gaacgggtgg tgagagttga agtttatccg aagctgccga 180
ggggccttct tagagatact tggcctgctg tgctcgtgtc aggtaggtct gcgaggccgc 240
tcgggctgtc agtcctcgcg aaagtcgggg tcgataattg ccgccctcac ccatggagct 300
cccttcaaaa gcctccaaaa agaccattgt gagtttttc tatgaggaaa agaatttcct 360

ccacttgagc cacgttaact tgtcccccag cgtcgtcctg ccttacagac cctgcgattc 420
ccgcgcattc cggtgagcac tgggtgaggg atggttcaag ggggcatctt gttatacgaa 480
cgaatgtttc aactgagaat gttccttttgt ttatgcgtca taaacgtatt tttgacgcca 540
tacattctgt tataaagaca ctttaataag tatctaggga gtgcatattt ctctacggaa 600
atataaaatt atgtagacca caaacaggtc tttctactga agaaaaataa gaatgctaag 660
ctattttgat ctgaacttcg taggcctgga cctccccaca gatttt 706

<210> 82
<211> 378
<212> DNA
<213> Homo sapiens

<400> 82
caatttatca ataacagatt ccccgaaaaa aggaacttgt gtacataagg aaaggaagaa 60
agatgtgaga aggaatgccg gaaaacctta gaaaccattg gatctgtcta aactgtcatt 120
gactgtgtaa agcaataata acagtgacta acgtatagag tagcgacaaa aggcgcaact 180
gaagtactag acaacatcct ggtgttcatg cctttcaagg tcactatgct atttgggagg 240
agattcggct aaagtctacg agggccacgt atttatatat aatttctaga ccagtggttg 300
gaaaagggtt gaaagaaaac tttaacagat taatagaaag aaggaaaaga gaaggaggcg 360
tagaaaaatc aggaggca 378

<210> 83
<211> 391
<212> DNA
<213> Homo sapiens

<400> 83
gacaagtata aaaacattta tcaataacag attccccgaa aaaaggaact tgtgtacata 60
aggaaaggaa gaaagatgtg agaaggaatg ccggaaaacc ttagaaacca ttggatctgt 120
ctaaactgtc attgactgtg taaagcaata ataacagtga ctaacgtata gagtagcgac 180
aaaaggcgca actgaagtac tagacaacat cctggtgttc atgcctttca aggtcactat 240
gctatttggg aggagattcg gctaaagtct acgagggcca cgtatttata tataatttct 300
agaccagtgg ttggaaaagg gttgaaagaa aactttaaca gattaataga aagaaggaaa 360
agagaaggag gcgtagaaaa atcaggaggc a 391

<210> 84
<211> 384
<212> DNA
<213> Homo sapiens

<400> 84
caagttgtgt tcgtttatga ttctttaaat gttttccaat acttagatac atcaaaatta 60
taggacttct caattccatc ctattgttac agaatataaa tttaatcaag ataggaagac 120
cctcaaaaga tctttctcat gagttcagat attccaaata ataattacag aatttcattt 180
gtacatttga actcttatca ttgaatttgt ttaattcctt agtgtcttcc tgttttcagg 240
cttacttttc aattaatttc agtctgcaaa aagcttcaaa aatagatggt agcttttata 300
tggttcctaa tgttgagtga tttgattaaa gttttccaac tgattttgaa caaaatgtaa 360
tgaaagctta gaagactagt ttac 384

<210> 85
<211> 389
<212> DNA
<213> Homo sapiens

<400> 85

caagttgtgt tcgtttatga ttctttaaat gttttccaat acttagatac atcaaaatta 60
taggacttct caattccatc ctattgttac agaatataaa tttaatcaag ataggaagac 120
cctcaaaaga tctttctcat gagttcagat attccaaata ataattacag aatttcattt 180
gtacatttga actcttatca ttgaatttgt ttaattcctt agtgtcttcc tgttttcagg 240
cttacttttc aattaatttc agtctgcaaa aagcttcaaa aatagatggt agcttttata 300
tggttcctaa tgttgagtga tttgattaaa gttttccaac tgattttgaa caaaatgtaa 360
tgaaagctta gaagactagt ttacaaaaa 389

<210> 86
<211> 739
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (358)
<223> a, c, g or t

<400> 86
gaaagttcat acctttcaaa agaaaaagga acgtttgctt ttttacatct ttgttgttca 60
tctgactcat gaaagaacat gatcggttcg agtttatttt taggatatac tggtactggc 120
ttttagtttt agtaaatgtt aagttggaca agttaggggc ctagcttggg agctgcagaa 180
attggctgag ccccacaggt gatttataga taatctttcc agtaagaaca ttgaagggct 240
acacacaatg acacttagaa aaagaaggga aatgaagctg ttccttgact actaccccag 300
tttctgttga ggtttattac ttctagatga taaggtttac acgaagttta cattatgntt 360
tttcagttct caagtttcag caaatacctg aaccaagttt ttttctgtta ttctaagaac 420
tgccctggag tgccttttaa cttttgtacc accacgcaaa gtgtactgtc aattcatgtc 480
ctttagctct tctattcttc aatgcatttc tcccattcct gtaggtatgg cggggatcaa 540
cttttcatac caccaagagt caccctatt ccctttgaag tactgcccta tggcataagc 600
ttgttcatac ggtgttcaaa cagctaccgt tcacttctat gagggtcacc ttactggaaa 660
ccaaggtatg acgagtaact taaatcttct catcaagcag aagggagctg gactttagaa 720
atggagcctg ggccacgca 739

<210> 87
<211> 902
<212> DNA
<213> Homo sapiens

<400> 87
actgtcccct gcttaattaa agatcttttt ttttttttt ttttgtattt tttttgtaga 60
aacagggttt caccatgttg cccaggctgg tctcaaactc ctgggctcaa gcgatctgcc 120
cacctcggcc tcccaaagtc ctgggattac aggcgtgacg actctgcgtg gcccaggctc 180
catttctaaa gtccagctcc cttctgcttg atgagaagat ttagttactc gtcatacctt 240
ggtttccagt aaggtgaccc tcatagaagt gaacggtagc tgtttgaaca ccgtatgaac 300
aagcttatgc catagggcag tacttcaaag ggaatagggg tgactcttgg tggtatgaaa 360
agttgatccc cgccatacct acaggaatgg gagaaatgca ttgaagaata gaagagctaa 420
aggacatgaa ttgacagtac actttgcgtg gtggtacaaa agttaaaagg cactccaggg 480
cagttcttag aataacagaa aaaaacttgg ttcaggtatt tgctgaaact tgagaactga 540
aaaaacataa tgtaaacttc gtgtaaacct tatcatctag aagtaataaa cctcaacaga 600
aactggggta gtagtcaagg aacagcttca tttcccttct ttttctaagt gtcattgtgt 660
gtagcccttc aatgttctta ctggaaagat tatctataaa tcacctgtgg ggctcagcca 720
atttctgcag ctcccaagct aggcccctaa cttgtccaac ttaacattta ctaaaactaa 780
aagccagtac cagtatatcc taaaaataaa ctcgaaccga tcatgttctt tcatgagtca 840
gatgaacaac aaagatgtaa aaaagcaaac gttccttttt cttttgaaag gtatgaactt 900
tc 902

<210> 88
<211> 489
<212> DNA
<213> Homo sapiens

<400> 88
aaaccataaa gcccctctgt gccatgtact tggccagagg tctaattagg ggagcatgtg 60
gaaagattta aggggatctc ccaaatttag agattaggat tggcttatta caacctgctt 120
cagaactaga gttcaccaca caccaagaaa ccccagagaa agagcaaagc agatgacaaa 180
gccatgttat tattccttta caaattatac cctcctggcc ctttggtagt gtttttccaa 240
gaatgacttt tcatgaatgc agtttgttgg ggctagagtt ctttaggtct ctcatgttct 300
agttccacac aagattcagg gatttttcaa aaagaaaatt attaattatt gtcaacatag 360
tggaatccta attattaaag gtggagagac ttttgaagcc atttaaatat ttggtccctc 420
tcatgttatt atttatgtaa cttttaatat atttgtaaac taacagcaat agtaaaataa 480
tagtagcac 489

<210> 89
<211> 555
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (465)..(490)
<223> a, c, g or t

<400> 89
caaaaagcat gttacgcatc aaattgcttt ttaaattcaa tcttctgata ggtataaatg 60
aactcagtcc aagagattga attatcttaa aaactatata taagggctca acacattcag 120
aaaagtacct ggacaaattt atctggttga attataaaat tctttaaaga aaagattcta 180
aacatcattc agtagagcaa attaggaatg tatcggatgt gttggtttac agggagcact 240
ctgctagcac aggcatctca tattcttagt aaattcagtt aagaatttta ctgattgtat 300
catagctttg ctttgactag gaatatcctg aactgctaac taaattaatt ccaattttat 360
ccttgctact cctgccagct gggatgccag catggacatc cttcagtgga cttcgctctg 420
tgccaggaat ttgttcatat tacttttgaa aaacaagcat tcttnnnnnn nnnnnnnnnn 480
nnnnnnnnnn attataaact agagaataaa cctcttccaa agtggcacca atctggcaca 540
aaggcaaatt tttaa 555

<210> 90
<211> 490
<212> DNA
<213> Homo sapiens

<400> 90
ggcttgactt catttcctag tctggtatat ctaagacact atttgctctg aataggaaag 60
tctgacagtt ctgggtttga acccagctga gctactcact tgctgtttga ctcttgctta 120
cttagccttt cagggcttca ttttcttcac atgaaaaata ataataacag atttatatcg 180
tttagaagag aagccagcaa atattttcta tgaaaagcta gagaataaat attttaggcc 240
ttggagtcct agaatctcta ttgtacctgc tcaactctgc cactgaagca caaaagcagc 300
cacagatgat agataagtgg atgtgcaata ctgaattcca ataaaacttt atttacaaag 360
catacggaga gacagatttg gctcaagagc ctttgtttgc catcctttgg gttgaatata 420
gactacacaa tgacaaagag atcccaaaat atattgcctt aaataagaca gaagtggcca 480
ggtgtggtgg 490

<210> 91
<211> 277

<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (109)
<223> a, c, g or t

<400> 91
gaattcaatg cccatatcct aactatactt cagttgttaa gaagagatag aactacagct 60
aaacgtattt tatagcctct tcattgtaga aagatgtaca aaaatacanc acctgacggt 120
gctgtataat tgaatgaaaa gtccctcata tataaatagt ctttaaaaat tggcatgatt 180
tcagtatatt gaatgctaag ttttttaag tcgcggtttt ggttaatacc tgttttggga 240
agctggaaat ttttttaaag gcattaggaa cctggca 277

<210> 92
<211> 438
<212> DNA
<213> Homo sapiens

<400> 92
gcaggacata ttcgggtaga ttttattgta tctcaataaa agtttatttt aaatattttt 60
aaatggcaca taaaagaaaa ataaacaaac tggtcaataa gccaaagttc aaaaacattt 120
agtcacataa tagttctact acagggaaga aattctactt acttctcaaa aaaaatcact 180
gtttatgaga ttttgctttg aatcttctca atgtgtggaa atacaattac tactgcatca 240
gaattacttt catctctgca caacatggtt aaaaactact gacaggcaag aatcatgaca 300
gatacccgtt tttacccaat gcaagattca gtgtaaggct gaaaccccag ctcaaagctt 360
cttttctata aagcctaact tcatcctctg caatggaggg aatatttctc ctctctaaat 420
cagtgcaaga cttgacct 438

<210> 93
<211> 486
<212> DNA
<213> Homo sapiens

<400> 93
aattgagggg gtgatggaaa tgttccatag cttgattgtg gttatgtagc aggacatatt 60
cgggtagatt ttattgtatc tcaataaaag tttattttaa atatttttaa atggcacata 120
aaagaaaaat aaacaaactg gtcaataagc caaagttcaa aaacatttag tcacataata 180
gttctactac agggaagaaa ttctacttac ttctcaaaaa aaatcactgt ttatgagatt 240
ttgctttgaa tcttctcaat gtgtggaaat acaattacta ctgcatcaga attactttca 300
tctctgcaca acatggttaa aaactactga caggcaagaa tcatgacaga tacccgtttt 360
tacccaatgc aagattcagt gtaaggctga aaccccagct caaagcttct tttctataaa 420
gcctaacttc atcctctgca atggagggaa tatttctcct ctctaaatca gtgcaagact 480
tgacct 486

<210> 94
<211> 310
<212> DNA
<213> Homo sapiens

<400> 94
aaaagaaagc aacaggcaat aaaaacatac cacaaaggat ccaaataatg aagttttcag 60
acatataatt taaagtaact atacttaata tattcaagga attaagattg agagtttact 120
ggagaacttg aaactgaaaa aaaggaatca aattctagaa ctgaaaaata taataaattg 180
aaattaagaa tataatcaaa tttatcaaaa gttttataga aaacactttg tgtcttgtct 240

ttttatacct caaggattaa aatgtttacc aacatagtct accaaaagtt tttaaaatgt 300
tgattcactt 310

<210> 95
<211> 963
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (124)..(173)
<223> a, c, g or t

<400> 95
gccttttctt ataaactctg tgatttaaaa acttgaagca ccagatgaag atctttgtaa 60
ttgtttcaca gttgttccag cctcaaatag acgattggct tccagctttt atctgctgcg 120
ttannnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnctgtagt 180
gttgttgacc ctggggaaag tgggtcttaa ctactggtgc atctccagta agatccaaag 240
aatcagtgag aatcattgtg acttaagcaa tgtagcccac tgctgtcggg ttatttctaa 300
ggcagcagct ttgatcatat taccttctgt accccaaaac tgtgtttggg tccctactgc 360
ttccagaaag tttaggtaag ttccagagag tctggcattc caggtttatt gtatgtgggc 420
gtggagctgt ttccaaccct aagtaaccca tcgtttccca ccttgcccac accccattat 480
gcaccctatg ttgtagctac acggatgtat ctaagaccct tagtctggaa cccctgtctg 540
ttccttcccc aatctcttcg ttctgtgcac acgaattctt ccttatttca cagacctagc 600
ttcagtgctg gctctgccat aagccttctc tgatatgctc aggtggaaat aataccactt 660
atttcctata acatctaaaa tgttgtatct gtatataaag gtgtaattct atcacccagc 720
tcacaatggc tctgctctac tacatttcta attaattttt aaagtttcca gatagggcat 780
cataataaaa agtgaactgg taggtttata tatttttact tgatgcatct cagaatacac 840
cagtatactg agaagatgac cagataggac tcatagacct aaattctagt tccacatctc 900
tgttaaagct tagataactc aagatctctt agccttgttt cctctataaa agaaaaaagt 960
tca 963

<210> 96
<211> 2646
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (1013)..(1062)
<223> a, c, g or t

<400> 96
aattcgcggc cgcgtcgact ttttttttt ttttttgagg tatggtctca ctcggtcaac 60
caggctggag tgcagtggca caatcatggc tcactgcagc cttgacctcc caggctcaag 120
tgattgtcca gccttaacct cctgagtagc tggaaccaaa gatgtgtgct aacacagcca 180
gctaattttt taaaattatt ttttgtagag atggggtctc cttatgttat cgaggctggt 240
attgaattcc tggcctaaag tagtcctccc atcttggcct cccaaagtgc tgggctaaaa 300
ggcatgagcc atcacatctg gctgaacttt tttcttttat agaggaaaca aggctaagag 360
atcttgagtt atctaagctt taacagagat gtggaactag aatttaggtc tatgagtcct 420
atctggtcat cttctcagta tactggtgta ttctgagatg catcaagtaa aaatatataa 480
acctaccagt tcacttttta ttatgatgcc ctatctggaa actttaaaaa ttaattagaa 540
atgtagtaga gcagagccat tgtgagctgg gtgatagaat tacaccttta tatacagata 600
caacatttta gatgttatag gaaataagtg gtattatttc cacctgagca tatcagagaa 660
ggcttatggc agagccagca ctgaagctag gtctgtgaaa taaggaagaa ttcgtgtgca 720
cagaacgaag agattgggga aggaacagac aggggttcca gactaagggt cttagataca 780
tccgtgtagc tacaacatag ggtgcataat ggggtgtggg caaggtggga aacgatgggt 840

tacttagggt tggaaacagc tccacgccca catacaataa acctggaatg ccagactctc 900
tggaacttac ctaaactttc tggaagcagt agggacccaa acacagtttt ggggtacaga 960
aggtaatatg atcaaagctg ctgccttaga aataacccga cagcagtggg ctacattgct 1020
taagtcacaa tgattctcac tgattctttg gatcttactg gagatgcacc agtagttaag 1080
acccactttc cccagggtca acaacactac agnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1140
nnnnnnnnnn nnnnnnnnnn nntaacgcag cagataaaag ctggaaagcc aatcgtctat 1200
ttgaagctgg aacaactgtg aaacaattac aaagatcttc atctggtgct tcaagttttt 1260
aaatcacaga gtttataaga aaaggcatgc gctggtcagg cacagggctc acatctgtaa 1320
tcccagcacc aaaaatgtaa ggttttaaaa gagataacac aggtaaaaat agcctcagtg 1380
ctgatgaagg tatgagacac cactacatac cgttgggagt ataaattgtt agaaacttcc 1440
tggaaagcac atgaaaacaa gttttcaaaa tgtaaaaagt acatatgaat ggcctagaaa 1500
tcccatcggt ataaatatat accaaggaaa taaatgtgca aaaatgtaca tgaaaagggt 1560
gaatataggg aaacctgtct aaaaatcttt agtaacatgg aaagatgccc acaatatact 1620
cagaatataa aacaacactt tatgatcctt ttgttttaaa aaaaatgtta tacacaattc 1680
agtaacatat ataccagaat gttaatagta tctctgaggt atgatgacat ctggcagcaa 1740
gcccatacag tcccttcaag agaattaaaa aaagaatcct tcctggcaga catatcccca 1800
gaggcacaga gctctccaaa ctgatcctgc ctttgcaaga attagaacaa ggtgcccctt 1860
ctctaggcag acccaaccat ggtctaggat acacagcctg gcaagctgaa tatggatttc 1920
ggcctcctct cacgactccc tcacatatcc ttagagggat acccttacat tgggacagga 1980
ttatggttaa ttttcaccta tatatgccat tcttcactga cttactatta ttacaataaa 2040
caaaaaattt actgataaga gtaatactta tcgattttga agagattgac ttaatccttc 2100
caaaatgaat cccaacttca cttgtactag ttttacagtc cttacacatt gctctctatt 2160
tatttgtgca tgttatctat ttaaatgcat tgactcaata aaataaaatc tctaccctag 2220
gtcatacaat acaaaggttg aagagcacta tctgcagtaa gaaatgtcat tctaacacaa 2280
agtggagtta ctccattctg gaaatgagga agcactgggg aagggagaaa gggaggaagg 2340
tagcctatca aaattaatct ctctccacca ctcaaaaatt tcttcagcaa aaagttagta 2400
cactctagaa acaagcaatt aagctttata aatgttcatc ttttgttccc atgtaagtct 2460
ggatgcaaga tttatttctt catcacacta gtgggttccc aaaagaaaac agcaatttaa 2520
ataaaaatca tgagtgatta ccatgcagtt aggcaaggca tctgtaaaat aatacgaccg 2580
ctggactagg attaaaggag aactgggttc ttttcttggc tttgtggcat atgtaatttt 2640
acacaa 2646

<210> 97
<211> 266
<212> DNA
<213> Homo sapiens

<400> 97
gccgcggcta tattcgccgc ggcgtcactg ccttcctggc ctggtggtga gaggaagccg 60
ggccgcgaaa gcttcctgag gagaaaatgg agggcccttc tctcacaccg acgagaaaag 120
ttcgaggggg aaatacgagt tcctttctga agggacagga cggctgcttt tccacagccg 180
cgacgtgatt gagaaatggt ggctggcaag ggtagccctg ccttcgcccc tccaaagtaa 240
aaatcgggag ttgagaccaa aaaaaa 266

<210> 98
<211> 300
<212> DNA
<213> Homo sapiens

<400> 98
ctcaagatgg cgaaacctcg gccgccgcgg ctatattcgc cgcggcgtca ctgccttcct 60
ggcctggtgg tgagaggaag ccgggccgcg aaagcttcct gaggagaaaa tggagggccc 120
ttctctcaca ccgacgagaa aagttcgagg gggaaatacg agttcctttc tgaagggaca 180
ggacggctgc ttttccacag ccgcgacgtg attgagaaat ggtggctggc aagggtagcc 240
ctgccttcgc ccctccaaag taaaaatcgg gagttgagac caaaaaaaaa gtcgtatcga 300

<210> 99
<211> 805
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (692)
<223> a, c, g or t

<220>
<221> unsure
<222> (705)
<223> a, c, g or t

<220>
<221> unsure
<222> (726)
<223> a, c, g or t

<220>
<221> unsure
<222> (734)
<223> a, c, g or t

<220>
<221> unsure
<222> (788)
<223> a, c, g or t

<400> 99
gtgaaaatta gtgttttatg aaaattggta ttttacccca cagtaaaata gtagtaagga 60
aagtgtattg ttacctttta gtaacttagt gagacactta agtttctgtc ttgcataaaa 120
tactctttta agataattgt tggatttggg atttgcagat ggtactataa agttagcatt 180
cttctttctt ggcaattagc taaaaacaac aaagagaggg aaaaactgac acacaaaccc 240
agcattccac taaactagaa gacagttaaa aaccccaagc cacaaaatag gtggaaatgt 300
tgccaggagc agtggccatc aggctggctg catgtgggaa gaagtaggta ggatgccgtg 360
gctgtggagc tcagagacaa caagaaacac ccatgaaact gcaaggcccc cctgaggtat 420
aaaacctaaa ccccttgttt acaaaagtga agtgggtttc ttcagctcaa ggcaggagct 480
ttcctgcacc tggtttaatt ctaagaagtc ggggaaatga agccaaacaa ggagtcacat 540
aaacaagcca tttttgcagg aagtaatatg tgtctatctc tatagcagtt ggggaggaga 600
tgactgcatt gtgaaaaacc tccagactgc ctgtctgcat tggctgttga gaagtaaaat 660
gctaaactgt ctcttaggta aatctctgaa tnacgaaaaa gttgncagtc cagaactgga 720
ggcacnctgg cctnccgcat aaacctccta caaatatcag gttctggaca accagcactt 780
cttcaganat gaataatcaa aagga 805

<210> 100
<211> 158
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (49)
<223> a, c, g or t

<400> 100
attaaggtat gtgttaaatt gtatatggct gctgcatagc cattataant catcttctta 60

aataatattt agggtcttga gaaaacagtg tattattaat tgaaaaaagc aatatgtaat 120
agcaaaaaca caatgttgta ataccaattt agtatttt 158

<210> 101
<211> 454
<212> DNA
<213> Homo sapiens

<400> 101
cacgagggtt ggaggagtgg gtgatatgca ttaagagcct taaggatgtt catttcattt 60
ccctcagtaa tttcacttct gggaatgtat cctatggaac taattagaga tacccacaaa 120
aatgtgctga ccaacatact cattgaagca ttatttataa tatgaaaaaa ctgaaaacct 180
aaaacttgaa tattaaggta tgtgttaaat tgtatatggc tgctgcatag ccattataaa 240
tcatcttctt aaataatatt tagggtcttg agaaaacagt gtattattaa ttgaaaaaag 300
caatatgtaa tagcaaaaac acaatgttgt aataccaatt tagtattttt aaaaatatgt 360
gtatagccag gtgcggtagt gcatgcctgt agttccagct acttgggaga ctgaggtagg 420
aggatcgctt gagcccagga gtcctgggct gtta 454

<210> 102
<211> 273
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (118)..(198)
<223> a, c, g or t

<400> 102
ggcgtgagcc actgtgctac accaaatgtt gacaaggctg tgtgcaacct gaattgtcat 60
acatttgtgg tgaaaggata aaatgataca atcactttga aaaaaggtct taaccagnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnta tgtcaacaaa gaaacaactt gtacaacccc aagcttgaaa 240
caacccaagt ttctatcaat aggacatgaa tac 273

<210> 103
<211> 833
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (17)..(18)
<223> a, c, g or t

<220>
<221> unsure
<222> (22)..(23)
<223> a, c, g or t

<220>
<221> unsure
<222> (73)
<223> a, c, g or t

<220>

<221> unsure
<222> (93)
<223> a, c, g or t

<220>
<221> unsure
<222> (120)
<223> a, c, g or t

<220>
<221> unsure
<222> (137)
<223> a, c, g or t

<220>
<221> unsure
<222> (159)
<223> a, c, g or t

<220>
<221> unsure
<222> (775)
<223> a, c, g or t

<220>
<221> unsure
<222> (822)
<223> a, c, g or t

<220>
<221> unsure
<222> (826)..(827)
<223> a, c, g or t

<400> 103
aaaaggcaca cgcttgnnct gnngaatgag gacaatatac acatggtagc tgagaatatg 60
gnctgaaatc agngatctgt atacaaatac tgatcctgcc agntactggc gtggctgggn 120
gatgatggaa aaattgntga acctgtcttc tcatctgtna aataggttaa tgacccttta 180
atatggttgt gagcattaaa tgaaatgata tataatgaca tatagtgccc aaaagatacc 240
acttgataaa tagattttat gataatatat atggtattgt ctaccagctg atgagtatga 300
gataagttgt taacattttc ctttctaaac tacttggcta tatgaaacct gacaagtcca 360
acttgtttta ttttgtattg tgtagtatct gtcagtgttg cttggccatc gctagttaaa 420
aaaaaaatct gtagctttta gggtatctgg gtgattattt catgccaaag tactacaaga 480
ttttgtttgt ctgatctgag atttcattct tattcagttc aagaatgtgg taggaagtaa 540
atgtttgttt tagttctgta ttagatatgc ttacgtataa aaacaatgtt acagaagagt 600
gttatagaaa aggtgtcatt tggctaaata actatataaa agtgtatgga aggataaata 660
acattactgg tgatttttag ggaatacatt tggaactgaa tttgcaggag ggagtgttgg 720
ataaggacct tttacttttt atttcttata ctttcatata atttgaatta ttctnacaac 780
aagtatgtaa cacttttatg attaaaactt ttttaaattg anaaannngaa aaa 833

<210> 104
<211> 820
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (143)..(423)

<223> a, c, g or t

<400> 104
agacaaatgt tttttgaatg gataatctct aactgaaagt taatttctac tttttccact 60
aaatattgat tagttaactg tacagtcccg agtcacctaa catcagggat tcattctgag 120
aactgcattg tcaggcaagt tcnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnttattct tatagtatgt agatttcagt tattctagag ttattttaga aaatgactgt 480
gtgacaatat tgaatactaa gaaaactgtc aattggggaa tgacttcaaa atatttaaag 540
tgcgtttgag ttgtctttaa atttagattc atcacactta catacttacg agtgagagta 600
cctcatattt taagggaaat gcttgagaaa tttacatgat ttttgaaaag aatgagaaca 660
cagtaagaat tctcttgcct gctgctttct tggaaatttc tatttgaaca gtctgttcag 720
agcgatgggg caaagatgag gatgaggctg attaatcaaa atagtcttcc gtcttccaac 780
tttatttgaa acagtttaag acagagtgag actccgtctc 820

<210> 105
<211> 548
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (73)
<223> a, c, g or t

<400> 105
caagaatggt cttcccatgt taaatttcaa agatgcaaat atttgacaaa acattagaat 60
aataaaacac tgnatttctt tattgatcat gaggaatatt acttaaccat tagctgtatt 120
actagagaaa accatgttag ataaaattag agataaataa tgtttgttca tgagagttgt 180
ttatttttat caattcttat ttatctttga atcttataag agacacttta aaatgtttag 240
ctaaatcaaa atagtcctta cccaatgaaa tcacaacatg atagagtcat gagaagtatg 300
gataataaac tgtggcaatg ggtgaggggt agggagataa ggagactgat tgtaaaacat 360
ctgagcttta tcaatatata ttaatcatat aataatttac aaataatgta cagtacaatg 420
aaaggtaaaa gtaatgctct cagaggaaat gtttccattc tagttctatt tggaagtcct 480
tgggacaatt tgtgttttga ttatatttag aaatcttcag tattttgtct tggccaggtg 540
tagttgct 548

<210> 106
<211> 856
<212> DNA
<213> Homo sapiens

<400> 106
tttaaaattt tggtcccagc ccccaaaaat ttttttttct ttttttgaga cggagtttct 60
cttggtcacc aggctggagt gcagtggcgc aatcaagact cactgcaacc ttcgcctcct 120
aggtccaagc gattctcttg cttcagcttc ctaagtagct tggactacag gcacccacca 180
ccacatccag ctaatttctg tatttttagt agagacaggg tttcaccatt ttggccaggg 240
tgattcaaac tcctgacctc aggtaatcca cccgccttag cctcccaaag tgctgcgatt 300
acaagcctga gcaactacac ctggccaaga caaaatactg aagatttcta aatataatca 360
aaacacaaat tgtcccaagg acttccaaat agaactagaa tggaaacatt tcctctgaga 420
gcattacttt tacctttcat tgtactgtac attatttgta aattattata tgattaatat 480
atattgataa agctcagatg ttttacaatc agtctcctta tctccctacc cctcacccat 540
tgccacagtt tattatccat acttctcatg actctatcat gttgtgattt cattgggtaa 600
ggactatttt gatttagcta aacattttaa agtgtctctt ataagattca aagataaata 660

agaattgata aaaataaaca actctcatga acaaacatta tttatctcta attttatcta 720
acatggtttt ctctagtaat acagctaatg gttaagtaat attcctcatg atcaataaag 780
aaattcagtg ttttattatt ctaatgtttt gtcaaatatt tgcatctttg aaatttaaca 840
tgggaagacc attctt 856

<210> 107
<211> 612
<212> DNA
<213> Homo sapiens

<400> 107
aaataacaat atgcaggtag agtgaaatat gattaagcag caaaaaatgt taatagaaag 60
aaaaatgtaa gggttgatgt gctgcagccc gctcatacag atctcaagag tggaatgtgt 120
ccatcagttc ccaactctca gttcaaccac ttcacctggg cagctccaat gtggcaggag 180
tattttcacc aaagaaatta aatgctacaa atcctaccac cacaccactg gtttgggctc 240
atagaaagtt gttaagagtc tgtgacatga ggtggcctct aatacagtga gttcaatatt 300
tgaacttctg taaagaaaag gattagattt attcagtatg atcttaaaga gggatgctag 360
gggcagcagg taaaaatttc agggagacag attttcgttc agtgttggga aactcctgag 420
gtaagaagtg cccattaggc tgggtggcca ctcaccttag aagagagata ctacagagaa 480
ggctgaaaca tggaagattg taggggttga gtggctcatt agtttggcca ccaaatctag 540
caaataaaag tacaggatgc cgagataaat ttgacaaaag gctgtaactg ccaagttttt 600
gtaattcatt gg 612

<210> 108
<211> 648
<212> DNA
<213> Homo sapiens

<400> 108
atggattcta cttattgccc cgccgtgatc tcccttaaag aactgctgtg agaaattaca 60
tagctatttg cagagccaaa tccacatgac atgtgtgaaa tacagaagtg ataggcagca 120
gaggttgaga gcatagattg aagagctacc acactggcag cattgaattc tggctccatc 180
agttaccaac cctgtggcta ggacatattc tctaacaact cagtgtttgt ttcctcaccc 240
ataaatggga tgatagcaga ccctacctgc aatatagagc gattatgagg attcaattgt 300
cagacataca ttgcttataa cagtgccctg tacacagtaa agtatagata tgtggtagtg 360
aggcagatag cctatgtaac cacgtgttag tttccaattc tgcagttaac taccaccatg 420
accaagtttg aatttatgtt acttcacatt aatgtcatta gattacttgt tgttatacat 480
taataaacag ttgttttgca tattcgtggt tctaaattca ctcataactt taaatgtgca 540
atataatatc aatgtttttt atgctctgtc ttaatttgtt gttgtatttt taaggaatct 600
gaagattttt gagtttctaa gtaacattgt tctgagagga tacagtcc 648

<210> 109
<211> 1003
<212> DNA
<213> Homo sapiens

<400> 109
tttttaaact ggaaagcatt tttgtcagtg tgaatgaggg tcaatagtgc agccagtggt 60
gacatttttc tttattttgc aaaatgcttt taaaaccaaa ggctgctcta gttgatggac 120
agtatcagtc ttgatctaaa ttgtaggaca ctttttcatg taacataaca tttggggatt 180
gggtttattt agtgtaatga agataatttg atataaaaat attttgtgta tatatatatt 240
tttactttgt tttctaaatt gctgtttgca gtaacagtaa gcgcaaagca aaatatataa 300
gttatgactg tatgatcaga tgaagtatga gttcttttgg tttgcatcct taaatagtta 360
gagatctctg ataaaaactt tggaatcttt gcaaaacaat acaaaaatgc caaaatgtga 420
gcatgtcaat gaaaactaaa gacaaatact tcactctttt tcatactatt ataagttatt 480
ctggtattaa atatgttaat aaaagtgttt ttgttttgac atatttcagt taaatgaatg 540

aatgctggtt gtattttatt tgaatgagtc atgattcatg tttgccatct ttttaaaaaa 600
atcagcaaat ttcttctatg ttataaatta tagatgacaa ggcaatatag gacaactatt 660
cacatgattt tttttaatac caaaggttgg aagattttat aattaacatg tcaagaagac 720
tttatagtaa gcacatcctt ggtaatatct ccaattgcaa tgacttttta atttatttt 780
tcttttgctg ctttaacatt ttctggatat taaaatcccc ccagtccttt aaaagaatct 840
tgaacaatgc tgagccggca gctgaaaatc taactcataa tttatgttgt agagaaatag 900
aattacctct attctttgtt ttgccatatg taatcatttt aataaaatta ataactgcca 960
ggagttcttg acagatttaa aataaaagtt aatttctaga aaa 1003

<210> 110
<211> 1301
<212> DNA
<213> Homo sapiens

<400> 110
aaattcggca cgaggtcgat tgaaagaaaa cattttgttt ctaaattagt ctaccattga 60
gtgagaataa tcaatatcaa gaaagaagac tatctttctc aactaaacaa taatattcca 120
atcagcttgg taagacctga aacttgaata agcagtggaa atgccaaata taacagaggg 180
tatgtgctac agagaagtaa aaagggtttg acttttttatg atgggatttt ttttttctgg 240
gtatgtaatc tatttttttt ttaaactgga aagcattttt gtcagtgtga atgagggtca 300
atagtgcagc cagtggtgac atttttcttt attttgcaaa atgcttttaa aaccaaaggc 360
tgctctagtt gatggacagt atcagtcttg atctaaattg taggacactt tttcatgtaa 420
cataacattt ggggattggg tttatttagt gtaatgaaga taatttgata taaaaatatt 480
ttgtgtatat atatattttt actttgtttt ctaaattgct gtttgcagta acagtaagcg 540
caaagcaaaa tatataagtt atgactgtat gatcagatga agtatgagtt cttttggttt 600
gcatccttaa atagttagag atctctgata aaaactttgg aatctttgca aaacaataca 660
aaaatgccaa aatgtgagca tgtcaatgaa aactaaagac aaatacttca ctctttttca 720
tactattata agttattctg gtattaaata tgttaataaa agtgtttttg ttttgacata 780
tttcagttaa atgaatgaat gctggttgta ttttatttga atgagtcatg attcatgttt 840
gccatctttt taaaaaaatc agcaaatttc ttctatgtta taaattatag atgacaaggc 900
aatataggac aactattcac atgatttttt ttaataccaa aggttggaag attttataat 960
taacatgtca agaagacttt atagtaagca catccttggt aatatctcca attgcaatga 1020
ctttttaatt tatttttttct tttgctgctt taacattttc tggatattaa aatcccccca 1080
gtcctttaaa agaatcttga acaatgctga gccggcagct gaaaatctaa ctcataattt 1140
atgttgtaga gaaatagaat tacctctatt ctttgttttg ccatatgtaa tcattttaat 1200
aaaattaata actgccagga gttcttgaca gatttaaaat aaaagttaat ttctagaaaa 1260
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaggcggc c 1301

<210> 111
<211> 1117
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (49)
<223> a, c, g or t

<220>
<221> unsure
<222> (60)
<223> a, c, g or t

<400> 111
atggaaagct ggagtgaagg ctgagtggga atgtgcctcc tcgcctggna tagcaacaan 60
acagcatgct gcagtggtgt acatgggatg ggagaagtgg caggggaggc cacgtcttgc 120
aaggcctgcc aagcctcagt aaggactgtg gctttactgt gagaattggg agccattgaa 180

gggtttggag cagaggagtg tcatgatctg acgtaggtat gaaaatgatg ggtttggctg 240
ttgcactgag aatgagctgt aatgggtcag gggagaagtg gaacatggct tggcaggtta 300
ttgtgctaat ctgggggggg gtgatagttg ctgggaccat ggggacagca aagatggtaa 360
gattggattc tctatatatt ttgaaggaga gccagtaaga ttggctgaca gcttggttgg 420
cagacagagg aactgaggac agatatttgt tctgagcaaa tgaaaggaag aagttactat 480
caactgagat gaggaaaact atgtgtgagg caggttttgg gggcaagagc aagagtttac 540
ttactcactt gggcttattg agtttgagtt gtctcttaga catccaagtg gaggtgaccc 600
acaagtctgg agtttggaga acgaaatctg tatctgggct ggaggtgtag atttgggagt 660
tgtgagcata cacgtgacat ctaatgcctt gaacctggat gtcataaccc agggagtgag 720
ggcaggtgac gagggctgac ctgtggtgta tccctaccta aaggggtcag tctgatgaga 780
aggaaccagc agaaaagaga agaagggatc agtaggtagg aagaaaacgt agagtatagc 840
ttcgggaagc cacgaggaaa ggggatctca agaagggaag ggagttatta ccagtgtcac 900
ctctgctgtg ttaggcttag taacatggag gtcagcattt cccttgatga gagtttttgg 960
ggccaacgtc tacttgggtt tgagaaagaa gcaagaagag gaattagaga cattacatac 1020
agtctttctg gaagttttgc tgcaaaggac agcaaagaat ttgtaagtaa caggtcttcg 1080
ttataaaaat ttagataaca ttaagaacat aaagaag 1117

<210> 112
<211> 1129
<212> DNA
<213> Homo sapiens

<400> 112
tggaaagctg gagtgaaggc tgagtgggaa tgtgcctcct cgcctgggat agcaacaaga 60
cagcatgctg cagtggtgta catgggatgg gagaagtggc aggggaggcc acgtcttgca 120
aggcctgcca agcctcagta aggactgtgg ctttactgtg agaattggga gccattgaag 180
ggtttggagc agaggagtgt catgatctga cgtaggtatg aaaatgatgg gtttggctgt 240
tgcactgaga atgagctgta atgggtcagg ggagaagtgg aacatggctt ggcaggttat 300
tgtgctaatc tgggggggggg tgatagttgc tgggaccatg gggacagcaa agatggtaag 360
attggattct ctatatattt tgaaggagag ccagtaagat tggctgacag cttggttggc 420
agacagagga actgaggaca gatatttgtt ctgagcaaat gaaaggaaga agttactatc 480
aactgagatg aggaaaacta tgtgtgaggc aggttttggg ggcaagagca agagtttact 540
tactcacttg ggcttattga gtttgagttg tctcttagac atccaagtgg aggtgaccca 600
caagtctgga gtttggagaa cgaaatctgt atctgggctg gaggtgtaga tttgggagtt 660
gtgagcatac acgtgacatc taatgccttg aacctggatg tcataaccca gggagtgagg 720
gcaggtgacg agggctgacc tgtggtgtat ccctacctaa aggggtcagt ctgatgagaa 780
ggaaccagca gaaaagagaa gaagggatca gtaggtagga agaaaacgta gagtatagct 840
tcgggaagcc acgaggaaag gggatctcaa gaagggaagg gagttattac cagtgtcacc 900
tctgctgtgt taggcttagt aacatggagg tcagcatttc ccttgatgag agtttttggg 960
gccaacgtct acttgggttt gagaaagaag caagaagagg aattagagac attacataca 1020
gtctttctgg aagttttgct gcaaaggaca gcaaagaatt tgtaagtaac aggtcttcgt 1080
tataaaaatt tagataacat taagaacata aagaagggct gggcgcggt 1129

<210> 113
<211> 229
<212> DNA
<213> Homo sapiens

<400> 113
atgagttcac actgttatct ccaactctgt tacacagatt gttctagctt cttcctctta 60
cttatctgta aattcctact ccaatggtaa aaaagtggct tcccatgatc ttccatccat 120
ttacttaact gcacaactgc tgtataggtg tataacacta acagaattgt taacctgtac 180
catatggaac aggacattat caattagagt acagtgcata tatatagta 229

<210> 114
<211> 262

<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (111)
<223> a, c, g or t

<400> 114
gaattccaaa ttccttttca agcttttttat atatttgatt ttctttgaag tgtgtacagg 60
aatattattg tgttcccttg cagatgtcat attttcttgc tttttcatgt ntttgtattc 120
ctacattgat atctgtgcat ctggtggaat tctcacctct tccaatttta tggagtggct 180
ttctaagaaa aagatttttt ctgtagttgt gacctatagt gttggttggg tagggtgctt 240
tggtattggt tctggatgca tg 262

<210> 115
<211> 274
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (176)..(205)
<223> a, c, g or t

<220>
<221> unsure
<222> (207)
<223> a, c, g or t

<220>
<221> unsure
<222> (216)
<223> a, c, g or t

<220>
<221> unsure
<222> (226)
<223> a, c, g or t

<220>
<221> unsure
<222> (259)
<223> a, c, g or t

<400> 115
tgtgtttcga gatttatcaa cattacacaa aagttattgg ctctttatat taattagttt 60
atacaaacag ccctcaagtc tcagtggctt aatgttgctg tccatatcat agttgatttt 120
ggggtagata gccatcttcc atcttaaagc tgcgccatct gaatttattt ttattnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnantta tttttntgag atggancctc gctgcactcc 240
agctgggcag cagagtgana acctgtctca aacc 274

<210> 116
<211> 148
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (113)
<223> a, c, g or t

<220>
<221> unsure
<222> (131)
<223> a, c, g or t

<400> 116
gataaaacag agtttgtcag ttttagtctc ttatgtaatg gaacagaaat acgatgagcc 60
cagtgggaga aagcaggagg agttcttgtc cgtctctact tatacttttt gtnttttta 120
agttactaaa nattttgac actgattt 148

<210> 117
<211> 145
<212> DNA
<213> Homo sapiens

<400> 117
atataaacca tgtgtttata atgtttcaaa catgctttaa attttcttca ctagtctata 60
tttgcacctt tattagtatt attcatgaag caaaattaag gtctagaaaa aaaaagactt 120
gaagttcagt taaaaagctt ggtca 145

<210> 118
<211> 479
<212> DNA
<213> Homo sapiens

<400> 118
aatccaccaa atataaacag ccatccgtca ctgcactcat gcctccctct gtttactttc 60
atactaaggg tacaaaaatt ccaagtctct tttgaactgt attttgtatg ccaatttcat 120
gcttattttt cctttatcag agagagttaa ggtggacgag catgcccttt ttgtcatatc 180
agcctgaaaa tgttaaaaag ctaggtggag acagattagt tgtttcattt ttgtttaaca 240
aggtatttat acttttagct taatttcatt aagaggaaca tcaggcattg caatcagtat 300
taatcagggg ctcaaataca gactatctgg gtgaccttga ctaagcatca aggaggtagc 360
ctttatttcc ccttaaaatt agtttaacat ctctgttcca ttattcagat ctacacaaac 420
aaggcttcct caacagctat ctatttttac tcgcgtcttt ttttaaaact aaaactaac 479

<210> 119
<211> 2561
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (30)
<223> a, c, g or t

<400> 119
aacccgtcca ggtgccccaa gaggctcgtn aatatggacg gacccatgag gccacgatcg 60
gcctccctcg ttgactttca gtttggagtt gtcgccacag agacgattga agacgccctg 120
cttcacttgg cccagcagaa tgagcaagca gtgagggagg cttcggggcg gctgggccgc 180
ttcagggagc cccagatcag tttgtttttc tcctgtctga acaatggtgt ctggagaaat 240
ctgtgagcta ccaggctgta gaaatcctag aaaggtttat ggtaaaacag gcagagaaca 300

tctgcaggca agccacaatc cagccaagag ataataagag agagtctcag aattggaggg 360
ctctgaaaca gcagcttgtc aacaagttta ctctccgtct tgtgtcatgt gttcagctgc 420
ccagcaaact ttccttccga aacaaaataa tcagcaacat tacagtcttg aatttcctcc 480
aggctctagg ctatctacac actaaagaag aactgctgga atcagagctt gatgttttga 540
agtccttgaa cttccgaatt aatctgccca ctcccctggc atatgtggag acgctcctag 600
aggttttagg atacaatggc tgtttggttc cagccatgag gctgcatgca acctgcctga 660
cactgctcga cctggtctat cttctgcatg aacccatata tgagagcctg ttgagggctt 720
caattgagaa ctccactccc agtcagctgc aaggggaaaa gtttacttca gtgaaggaag 780
acttcatgct gttggcagta ggaatcattg cagcaagtgc tttcatccaa aaccatgagt 840
gttggagcca ggttgtgggg catttgcaga gcatcactgg tattgccttg gcaagcattg 900
ctgagttctc ttatgcaatc ctgactcacg gagtgggagc caacactccg gggagacagc 960
agtctattcc tccccacctg gcagccagag ctctgaagac tgttgcttcc tctaacacat 1020
gagggaggct gaatccacca aatataaaca gccatccgtc actgcactca tgcctccctc 1080
tgtttacttt catactaagg gtacaaaaat tccaagtctc ttttgaactg tattttgtat 1140
gccaatttca tgcttatttt tcctttatca gagagagtta aggtggacga gcatgccctt 1200
tttgtcatat cagcctgaaa atgttaaaaa gctaggtgga gacagattag ttgtttcatt 1260
tttgtttaac aaggtattta tacttttagc ttaatttcat taagaggaac atcaggcatt 1320
gcaatcagta ttaatcaggg gctcaaatac agactatctg ggtgaccttg actaagcatc 1380
aaggaggtag cctttatttc cccttaaaat tagtttaaca tctctgttcc attattcaga 1440
tctacacaaa caaggcttcc tcaacagcta tctattttta ctagagtctt tttttaaaac 1500
taaaactaac tctaaagaag tttcaacaga atttccacat acctgcattc attagaactt 1560
gattctccca gaatacaaag tactctattt taaagaaaaa cccaacagtg caccctggg 1620
cagttttcag actgcagcaa atcttttatt acaaataatt aaatctctcc ataatgtctc 1680
aaacagtatc aaacaccatt tcatatctct aacacagagc agagtcggca ttcagtataa 1740
gaaccaagtg aaaagtgtta aatttcaagc atctgatcac atcacatggt gaccaggtaa 1800
agcttagatg tcatttccc acattatcca actgtgcatc tcaaacatat cctcatctca 1860
gtaaagacaa aagtttctat ttcatattgt taagtgcagg aagttgagag agataaaaat 1920
ccagtgaaaa cacatcaatc tcaattcaac tcagttaaaa aaaagaaaag caaatttaaa 1980
ttagtttttt tcagagaaga aagggaaagg agtccatggg gttaagaatc aaaactgacc 2040
agggctggca actatagatg gcatgttgta gctctggaaa gtatctgtca catgatattt 2100
taaaataaag tggcttttgt ggatttttc ttttttggt attgtaaaca tgtactgttt 2160
aatattaccc gaatttaatt taaaacatgt ttgcaaacaa aacaaaatta aaagccttta 2220
aggcaaacct ccccctaagg aaaaaaagtc atttgttata aaattgtgag gacacccaag 2280
caagacccca cttaagattc gtcagcatga aactttgaaa gtagccttgt tcgactggaa 2340
ttcctccaga attaaactgg gttcatgatg gaataaagaa cccgacaact gcctcctggt 2400
gcttttcaat acttgccttt ctgaccatcc atcgtctgaa atctcagacc catcttattg 2460
gccagagctg gagcaagcaa actagtactg gcccgacaga ataattctct gtcttccacg 2520
gaacatgagc tagcgacaag actgaagtaa agatgtgccc c 2561

<210> 120
<211> 215
<212> DNA
<213> Homo sapiens

<400> 120
atttgtgact actgctaggt gattctagga tccagacgcc agtttctgtt ctgtcagagg 60
taggtattac cttccaggct gtgaagtgaa tagagtccct tttgggaata acgattcttg 120
ttgctccctg gagaaagaat acaatttcta ggaagtcctc tgtgctactt ctgcatgcgg 180
ttgtgttctt ttaattttca tattttgtca gctta 215

<210> 121
<211> 753
<212> DNA
<213> Homo sapiens

<400> 121
ggggcagaac cctttggttt taaaggtaga gaaaagaatc tctaacaata gggatggtgg 60

ggcttttctt tttcttttta aggagattac cttgttgcaa ggaaacaata aatttcttta 120
aatgggggat taccccaaaa aaaaaaaaaa ggtatttgtg actactgcta ggtgattcta 180
ggatccagac gccagtttct gttctgtcag aggtaggtat taccttccag gctgtgaagt 240
gaatagagtc ccttttggga ataacgattc ttgttgctcc ctggagaaag aatacaattt 300
ctaggaagtc ctctgtgcta cttctgcatg cggttgtgtt cttttaattt tcatattttg 360
tcagcttatt taaaaaaaaa atctcttcca tgctttagaa ggagaaagga aaacaatgta 420
tgtaccttga gtgtatacta attaaaggtt tacttatgta tgtttgttta atgttagaga 480
tacgtatctc taacattacc tggtacctca actcatgcta aacattctaa tttgagcaaa 540
gctaaatcat gggtcttagt ctgtttagca aaatctccac aagatgaaat ttagcttact 600
accctaaact gttcaatgtt atgggtccaa tttagacaca acaaaaatat ttgatagatc 660
ctcagacatt agattaatgc acttacttac ataccaagct catgtttgtg cttagacagt 720
tatgcataat gttagcagag acgtgtacac tct 753

<210> 122
<211> 248
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (120)
<223> a, c, g or t

<220>
<221> unsure
<222> (122)
<223> a, c, g or t

<400> 122
gaattcgaaa tttttctttc agcctgggat ttcaagtgta ctacagcatc ctcttcacct 60
atacaagcgt ctgttgtttt aaaaggaaac tggtaagtag aattacaaat gtacaaaatn 120
cagatngtta aaaacagagt tatagaattt gcaggttaga gtacatcttc ttagaagact 180
gtattgcaga aatgttaatt gtaacattat tgcaaaggca aaaagtatta gaatcagtct 240
agatgtcc 248

<210> 123
<211> 241
<212> DNA
<213> Homo sapiens

<400> 123
caatttcaga aatgccaaaa tgcaaaaata atatacctct tgggactgag gaagtacaat 60
acgtcctcac attatcccag tgagctgggt actgaggctt ccagagttta agttgctgca 120
caaaacctgg aagaatagaa tcatgatcca aatgcatgac tgatatttca atccagaaac 180
tgccaaattt tttttaccat aaaggtctgg atagtaaaca tgtttgactc cgtggccaga 240
t 241

<210> 124
<211> 82
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (31)
<223> a, c, g or t

<400> 124
aagaaatgta tgtaattcct tcttaataag nttctgttcc tgattcaagc accagagata 60
gaagaaagag aggtttgtac tt 82

<210> 125
<211> 56
<212> DNA
<213> Homo sapiens

<400> 125
ggtctcactg ctgcactcca gtgctgcctc ttaataaaaa aaaaatggca cctgcc 56

<210> 126
<211> 357
<212> DNA
<213> Homo sapiens

<400> 126
ctcgagccga tatggatcat ccttctttcc agaaggaaac ttggcaggct gcataagcct 60
tattcccatt gatgcgccgc ctttggggcc tgcgggaatc attcacctgt tcttgtgatg 120
tgctggggca ctaggaccca cctgggttgt ctcagtcatt ggcgcaggca caccaagcag 180
acgtatccac cctgacacct ccctgtcctc cccgcctcta ccagcacagg caagtgccat 240
gccccccttcc ctcttatttc ccagttccac tgggaggtta ttattcactc ctctcaggat 300
ctgcctggag gtgggtggag tttaggggcc ttccgtagga ctccccggtg tctaata 357

<210> 127
<211> 260
<212> DNA
<213> Homo sapiens

<400> 127
cggctcgagc aaattacatc cagaaaagcc ttcccttgga gaaaaaaaat aacatatttt 60
taaaagcccc catcactacg tatgcattag tatccactac tagtcttttt ctccttccta 120
acatcgtaca atctgtttct ctgcctcact caatggtaaa gtttataaag gttattcctt 180
tgcctgtatt gttcaccgat ccatcctact gtatctaaaa cagtgccttg cacagtgatg 240
atatataaaa gttacttagt 260

<210> 128
<211> 162
<212> DNA
<213> Homo sapiens

<400> 128
ttgaagaata atttaataat ttgtatagga aggtatactt gtaaaataca gccagtcaaa 60
aatatctttt ccccattccc cttgatgcgt ccattacata catgttcacc ctgagattga 120
tcttccatca gcattgaaat gaacattgaa agtaataaga tg 162

<210> 129
<211> 98
<212> DNA
<213> Homo sapiens

<400> 129

gtgatctctt catatcacgt aaccaaaaaa ttcatacctg accttgagtt ttcccaggac 60
gggattctgt gacataaacc cttccatgct tctaccttt 98

<210> 130
<211> 1218
<212> DNA
<213> Homo sapiens

<400> 130
aagaagaaga cggtgacgat gaagaggaac ctgaacccca tcttcaatga gtccttcgcc 60
ttcgatatcc ccacggagaa gctgagggag acgaccatca tcatcactgt catggacaag 120
gacaagctca gccgcaatga cgtcatcggc aaggtagggg cgaggcaggt ggtgtggtgc 180
acctgctggc accagtgagg ctccgttcct taaaagaagc agcgggggta gtggacgggc 240
acaggctgga caccaggaga gattgggatc agtctcctct cctctcaggc ctctgtttct 300
tcatccaaaa ataaggggat gagatgaccc cctccctagg gttcctctct gagtgtcctc 360
ctatgggtaa gtcagctcag gacagactgc cagctgagca gggagaaaag gagtgcaaga 420
gaggaatggc ctggatctca aaagaaaggt ttactggaca ggatgatgac gttggagcag 480
tattgcccgg ggtcccatct gcacagggtg agctgcgtgg ggcagtgttg ggagctgcag 540
tcagactcca aagacctcgt ttcagttccc actccatctg ccgataaaca accatatcac 600
cttgggcaag tggctcggcc tctccaagct ttagtgttct catctttaaa atggagcaac 660
taatactgtc cctgtagaac tttacagggt ggcgtaagaa ttgcttgagg aggcagaggt 720
tgtagtgagc caagattgca agactgcact ccagcctgcg caacagagaa aactctgcct 780
gaaaaaaaaa aaaaaaaaaa aagtgatctc ttcatatcac gtaaccaaaa aattcatacc 840
tgaccttgag ttttcccagg acgggattct gtgacataaa cccttccatg cttctacctt 900
aaaaatggtc tactcctggt gtctatttca tgggttatag tggtagagag catgggctcg 960
ggagtcagat tggcctgggg tccaaatcct aagcgtaact atttaacgct gtaacctctc 1020
tgagcctcat tttctcctac tcaaaacgag agtgggatgg gtcctaccct tacaatggtg 1080
ctgagaggat aagtaagatc acacaggtga gctggttaag cataggactg ggaacatatg 1140
tagctagtat gataataaag gggcatgtat ataccattca ggcaggacca gttggacccg 1200
gagtctgtga ggaaccag 1218

<210> 131
<211> 905
<212> DNA
<213> Homo sapiens

<400> 131
atttgaggaa gccagggact ctctggaaga ggaaacaggg aggcaccata atgggaatag 60
gaaaaacaca gaaagcagga ctgtggtaat atcaaaggca gaccctcacc gtacatacct 120
cactactcag ctaggatcta actttggaaa aaaatgaact atacaggctg ccttagatat 180
aagaagcata atcaaaatag tttatataaa tgggagggaa ggaactatat tttagtggct 240
gtggggaatc aggaaaaaca aaccagatac aaggctacca tccacaatta ttaagaaaat 300
agaattttag aatctgggag tgcacgagaa gtggcagttg agctgaagcg aaggtggcac 360
aatagtgcca tagcttgcca agagaatata tatgaaggct ccctaactgg ccatttggag 420
ctgactgtag cctgcctggg aaagcaagta cactcaagag ctgttcttgt ttgaaagtgg 480
agagtttgcc tagagagggt agatgttgtc ctgagaatgt tcgtggcatg cagtgtcagt 540
caagattact gcaagagtta aaaagaatac tgttaatagc tgtcctgaaa tttagacaat 600
ttaaacgtga cactggatga ctgatattgt tcaaagatta cagtaagtac tagtaatcca 660
ggaataaaaa gacctaagag ttattctcgt gaaaaatgga tgacatggag tgtataaata 720
atgggagcaa atgcagcaag atacagggaa tcttggaaca tttaaagtct aatacattac 780
aaagggagtt caatactgat aaatgtagaa tataagtttt atggggtttt atcatatgct 840
gtaagtttgt gtggatcata aaatatgagg aataagaaag aaggggagtt gaaagacaag 900
tctgt 905

<210> 132
<211> 351

<212> DNA
<213> Homo sapiens

<400> 132
catcctttca tattttaatg aattttcaga tactctacca ttggtttctt tggcttcagc 60
ttttgtatgt agggcactat tgtttctctt catgtctgtg taacctatat agaaagcatt 120
aggccataac tttcttaact gggtcatgtt tcttgttaat ttaccaattg tatccagatt 180
ctcaactgga gctaggatat acttagggcc taagttttct aactgtggac ctgagctggg 240
acccacagtg ttctctgttc cattttcatc ttcatgattt tgtggaaaga aacctttata 300
agtcagtaac ctaaaagccc caagttagcc atcctgctgc ctgggaggct g 351

<210> 133
<211> 477
<212> DNA
<213> Homo sapiens

<400> 133
tttagattct aacccagtac tgcttcaaca aaatgacctg aaggaaggtc ttgttgtata 60
tcatcttcta tagtctacca caatggtgct cttacttagt tcatattttc atttagttgt 120
tttccccatc ctttcatatt ttaatgaatt ttcagatact ctaccattgg tttctttggc 180
ttcagctttt gtatgtaggg cactattgtt tctcttcatg tctgtgtaac ctatatagaa 240
agcattaggc cataactttc ttaactgggt catgtttctt gttaatttac caattgtatc 300
cagattctca actggagcta ggatatactt agggcctaag ttttctaact gtggacctga 360
gctgggaccc acagtgttct ctgttccatt ttcatcttca tgattttgtg gaaagaaacc 420
tttataagtc agtaacctaa aagccccaag ttagccatcc tgctgcctgg gaggctg 477

<210> 134
<211> 126
<212> DNA
<213> Homo sapiens

<400> 134
agcgttggac atttagatag tttctgccaa cccaacctgt ctggacattg acacactggc 60
cagattcccg tgtttggacg ttttggatcc gaccaggcca ctgaagttgt cctgaacaat 120
ctcgtg 126

<210> 135
<211> 140
<212> DNA
<213> Homo sapiens

<400> 135
caccgcgtct ggccagcgtt ggacatttag atagtttctg ccaacccaac ctgtctggac 60
attgacacac tggccagatt cccgtgtttg gacgttttgg atccgaccag gccactgaag 120
ttgtcctgaa caatctcgtg 140

<210> 136
<211> 160
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (14)
<223> a, c, g or t

<220>
<221> unsure
<222> (22)
<223> a, c, g or t

<220>
<221> unsure
<222> (32)
<223> a, c, g or t

<220>
<221> unsure
<222> (111)
<223> a, c, g or t

<220>
<221> unsure
<222> (136)
<223> a, c, g or t

<220>
<221> unsure
<222> (139)
<223> a, c, g or t

<400> 136
ggaggtgctc aganacacac anacacacac anacattcat tctcactcat ttaccaatcc 60
agaaacaaag agtttttagg ttttgattaa caccttagcg ttaacaatgc natataaaca 120
cagagaaatg ctgaangtnt cccagaagaa caaaactctt 160

<210> 137
<211> 336
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (14)
<223> a, c, g or t

<220>
<221> unsure
<222> (22)
<223> a, c, g or t

<220>
<221> unsure
<222> (32)
<223> a, c, g or t

<400> 137
ggaggtgctc aganacacac anacacacac anacattcat tctcactcat ttaccaatcc 60
agaaacaaag agtttttagg ttttgattaa caccttagcg ttaacaatgc aatataaaca 120
cacagaaatg ctgaaggtat cccagaagaa caaaactctt atatgctttg atatatatat 180
atcctatata tttcagacac tacaatgtgg aaatggcatg tatgtgtgtg tgtatttggc 240
taaaaaatta tactgccaaa attactgatt ataaatactt gactacactg attgatggga 300
caaaatgatt aaagtatttt cagggatctt attcca 336

<210> 138
<211> 297
<212> DNA
<213> Homo sapiens

<400> 138
ggccaatgtt ttgagacttt tctttaagta aataagggaa tgtgtagtgt ggggacttgg 60
gttccagggt gctcactggg aactgcacag ggatccaccc tcccacatgg aaatgcacca 120
ctgattgggg actgaggtca agttcatacc ctcccaaact gtacttggac cagagaaggt 180
tgtcagatat ttggttttcc aagaatttct cctcatataa atggcagcag atctggaaac 240
agaatggatg cttattgctt atcaacaagt ttgaaaccaa gatttgtcaa agatgaa 297

<210> 139
<211> 441
<212> DNA
<213> Homo sapiens

<400> 139
catgtttgtt tttacatttt ttaaaggaaa atttcaagcc tgtacaaaag cagagaggat 60
gataaagtga acacctgtgt agccactacc cagcttcaac agtggcccac ttgcaaacac 120
tcttgccctt tctacagccc acccctggct atgtcaaact ggacaccagg caccaggcat 180
catacctttc caccacttca ggagtgtgga atgtggggcc agtcagcccc gatgttgtgg 240
agtctcagcc aaaagggata gaaagagccc cctctacaca tgctctcacc tgttgacaca 300
cccagagaag gggtatcatg tgctgcagcc ccagtctcat ttccaagaga gcatctcacc 360
agctcagctt gggtcacacg ctctcctcga atccagccaa ccctggttat gagagagtgg 420
ggtcgcaccg tacaggagag t 441

<210> 140
<211> 675
<212> DNA
<213> Homo sapiens

<400> 140
atcatttggc tcagaaattc tacttgtagg gattcatctt aaggaaattt tcacaaatgt 60
acacaaagat taattacaaa gatgtttagc ttcaatgttg tttggaaact aacataacgt 120
gcaatagagg cctggtgaac tgagtactgg gaccttcaca ctgcacaggg gagcaagaac 180
ctcttctctt gcctctgctt tgtgggattt tccaaaatcc tcctggcttc ttttcccttc 240
ccttccttct tcccagagca ggggccagct ttacctcagt tggtaggtca cacacattgg 300
gtggagatgg tcaaaggtcc tctgcttgga caagttgctc cctgtcttca ttcctctggc 360
caaggccagt gttaggtggg gactggtgag ctgtctggcc ccacagggct aagctttcct 420
ttaggtgggt ttatcccact tctcaacagt ctaacaaaga acagattcca caagcccact 480
tcgtctcttg ctgcaactat gccttccgat tctgccttca ttagaaataa gggtgatatt 540
ttggttctcc acctgctatt tctcgatttg ggtggggaag atgggtgcta cttactgaac 600
tctcccaaag accttcaaga acactgagca gaataccatc tagccataaa aactgatgta 660
gaaaaatact tatga 675

<210> 141
<211> 686
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (442)

<223> a, c, g or t

<220>
<221> unsure
<222> (635)
<223> a, c, g or t

<220>
<221> unsure
<222> (647)
<223> a, c, g or t

<220>
<221> unsure
<222> (649)
<223> a, c, g or t

<220>
<221> unsure
<222> (651)
<223> a, c, g or t

<220>
<221> unsure
<222> (653)
<223> a, c, g or t

<220>
<221> unsure
<222> (655)
<223> a, c, g or t

<220>
<221> unsure
<222> (657)
<223> a, c, g or t

<220>
<221> unsure
<222> (661)
<223> a, c, g or t

<220>
<221> unsure
<222> (664)
<223> a, c, g or t

<220>
<221> unsure
<222> (666)
<223> a, c, g or t

<220>
<221> unsure
<222> (676)
<223> a, c, g or t

<220>
<221> unsure

<222> (682)
<223> a, c, g or t

<400> 141
aagcaaattt gatagaccat agcaaaaaga catgttacat ttgattattc tcttatttga 60
aaagtacgct tttctacatt ttcctaggta accctgtttc agaaccatgg gccctctgga 120
agttaaaatc actctggagg tctcactggt tgctctgaca gcttcttcct cccgataacc 180
ggctttcctc atctcaaggc acattccaaa ctgccgcccg gtggttgaaa cagggaataa 240
accaaagaaa aaatactgtt ctcttccttt ctctaggaaa ttatgcttgt ggcattttct 300
ccctctgttt ccattaccct aggataatct ctctttcttc tgcacatcag tactcatgca 360
gaagacaagg gttgtaatct ttgtcccccc tctcacctgc ctcttgccta gtcacactca 420
ctaaatctct caagccatgt cnttatatag tttgctttaa aaaaataccc cctaagcaca 480
gtaatgcatt ttgttattaa ggaatagaaa tgcaaatttt gagtgtgaaa atttgatcta 540
gtaaacaaaa atcaacctct aagacccttt taacctaaaa ttgttttcag agtcttcatg 600
tcttttcata tgttcatccc tttctttttt ttttngtttt ttttttngng ngngngngtg 660
ngtntngttt tttttngttt tntttg 686

<210> 142
<211> 845
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (636)
<223> a, c, g or t

<220>
<221> unsure
<222> (667)
<223> a, c, g or t

<220>
<221> unsure
<222> (822)
<223> a, c, g or t

<400> 142
aaagcaaatt tgatagacca tagcaaaaag acatgttaca tttgattatt ctcttatttg 60
aaaagtacgc ttttctacat tttcctaggt aaccctgttt cagaaccatg ggccctctgg 120
aagttaaaat cactctggag gtctcactgg ttgctctgac agcttcttcc tcccgataac 180
cggctttcct catctcaagg cacattccaa actgccgccc ggtggttgaa acagggaata 240
aaccaaagaa aaaatactgt tctcttcctt tctctaggaa attatgcttg tggcattttc 300
tccctctgtt tccattaccc taggataatc tctctttctt ctgcacatca gtactcatgc 360
agaagacaag ggttgtaatc tttgtccccc ttctcacctg cctcttgcct agtcacactc 420
actaaatctc tcaagccatg tcttcatata gtttgcttta aaaaaatacc ccctaagcac 480
agtaatgcat tttgttatta aggaatagaa atgcaaattt tgagtgtgaa aatttgatct 540
agtaaacaaa aatcaacctc taagacccct ttaacctaaa attgttttca gagtcttcat 600
gtcttttcat atgttcatcc ctttcttttt ttttngtttt ttttttttgt gtgtgtgtgt 660
gtttttngtt tttttttgtt ttgtttgaga catagtctca ctctgtcacc caggctggag 720
tgcaatggca caatctcagc tcactgcaac ctccacctcc caggaggatt gcttgaggcc 780
aggagttcga gaccagcctg agcaacatag tgaggcccca tntttacaga agtttttat 840
aaatt 845

<210> 143
<211> 25
<212> PRT

<213> Homo sapiens

<400> 143
Met Val Gln Asp Ala Ser Met Ser Met Lys Phe His Gly Phe Ile Phe
1 5 10 15

Lys Glu Arg Lys Glu Thr Gly Ile Tyr
20 25

<210> 144
<211> 66
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (2)

<220>
<221> UNSURE
<222> (10)

<220>
<221> UNSURE
<222> (13)

<220>
<221> UNSURE
<222> (20)

<400> 144
Met Xaa Phe His Cys Arg Phe Tyr Ile Xaa Asn Leu Xaa Phe Ser Ser
1 5 10 15

Leu Asn Phe Xaa Ser Thr Lys Asp Leu Gln Pro Tyr Cys His Trp Arg
20 25 30

Arg Ile Cys Ser Ser Ser Leu Lys Phe Leu Gly Cys Ser Ser Leu Trp
35 40 45

Gln Trp Gln Tyr Arg Glu Ser Phe Lys Val Leu Phe Ser Asp Val Phe
50 55 60

Pro Ser
65

<210> 145
<211> 55
<212> PRT
<213> Homo sapiens

<400> 145
Met Thr Leu Lys Leu Leu Phe Ile Leu Gly Lys Gly Glu Gln Thr Arg
1 5 10 15

Gly Cys Asp Gln Glu Ala Thr Ser Asp His Arg His Leu Gly Ile Ser
20 25 30

```
Arg Gly Val Gln Arg Ile Leu Gln Asn Phe Phe Gly Leu Trp Leu Val
         35                  40                  45

His Ser Val Pro Ile Asn Leu
     50                  55


<210> 146
<211> 118
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (10)

<220>
<221> UNSURE
<222> (111)

<400> 146
Met Ala Ser Phe Ser Arg Pro Ala Ser Xaa Leu Cys Val Pro Thr Thr
  1               5                  10                  15

His Thr Arg Leu Gln Cys Ala Gly Val Gly Gly Gly Ala Trp Ala Gly
             20                  25                  30

Cys Arg Met Glu Lys Ser Trp Phe Ser Arg Asp Ala Arg Asp Leu Lys
         35                  40                  45

Arg Glu Arg Leu Ser Gln Ser Trp Glu Glu Ser Lys Cys Phe Cys Pro
     50                  55                  60

Phe Tyr Lys Arg Cys Phe Ser Lys Ala Phe Thr Thr His Val Leu His
 65                  70                  75                  80

Phe Pro Ser Ala Lys Gly Pro His Ser Phe Thr Met Ala Pro Ser Glu
                 85                  90                  95

Gly Cys Cys Pro Arg Ser Leu Cys Pro Asn Ser Cys Thr Lys Xaa Pro
            100                 105                 110

Pro Leu Phe Val Leu Gln
        115


<210> 147
<211> 60
<212> PRT
<213> Homo sapiens

<400> 147
Met Arg Leu Ala Ser Ile His Arg Pro Pro His Thr Gln Pro Ser Thr
  1               5                  10                  15

Ala Gly Glu Ser Asn Thr Gly Val Arg Lys Pro Gly Tyr Leu Pro Ser
             20                  25                  30
```

Val Arg Thr Asn Leu Thr Asp Arg Glu Lys Leu Tyr Phe Ile Gln Leu
35 40 45

Lys Thr Pro Ile Phe Tyr Ile Leu Lys Phe Leu Asn
50 55 60

<210> 148
<211> 35
<212> PRT
<213> Homo sapiens

<400> 148
Met Leu Lys Ala Ser Asn Leu Phe Arg Lys Ser Thr Gly His Arg Ser
1 5 10 15

Cys Cys Gly Leu Ser Phe Leu Pro Arg His Leu Leu Asn Leu Gly Lys
20 25 30

Ile Asn Phe
35

<210> 149
<211> 46
<212> PRT
<213> Homo sapiens

<400> 149
Met Pro Gly Ile Gln Val Thr Val Asn Thr Leu Trp Ala Phe Cys Asn
1 5 10 15

Cys Asp Leu Asp Gln Lys Lys Thr Lys Glu Gly Ile Asn Met Lys Leu
20 25 30

Tyr Ile Leu Leu Leu Leu Leu Cys Thr Cys Leu Arg Phe Leu
35 40 45

<210> 150
<211> 85
<212> PRT
<213> Homo sapiens

<400> 150
Met Arg Asn Ser His His Leu Val Gly Glu Gly Gly Cys Thr Val Thr
1 5 10 15

Val Gly Leu Ser Leu Leu Ala Arg Phe Val Gln Lys Glu Tyr Leu Pro
20 25 30

Thr Ala Thr Phe Ser Gln Thr Gly Thr Arg Ser Ala Phe Leu Ile Phe
35 40 45

Ile Leu Leu Cys Val Asn Leu Leu His Leu Val Tyr His Leu Glu Arg
50 55 60

Asp Gly Gln Glu Arg Pro Ala Ala Gly Glu Asn Leu Cys Phe Ile Val
65 70 75 80

Gln Gln Leu Lys Val
85

<210> 151
<211> 56
<212> PRT
<213> Homo sapiens

<400> 151
Met Cys Phe Leu Ser Thr Cys Arg Arg Lys Gln Ser Leu Arg Ser Leu
1 5 10 15

Ser Phe Met Ala Pro His Lys Lys Ala Glu Ser Arg Ser Glu Glu Leu
20 25 30

Glu Ile Leu Gln Ser Gly Ser Ser Pro Tyr Leu Ser Ala Leu Lys Gly
35 40 45

Arg Arg Gly Arg Gly Met Gly Trp
50 55

<210> 152
<211> 91
<212> PRT
<213> Homo sapiens

<400> 152
Phe Phe Phe Phe Glu Met Leu Ser Leu Cys Arg Pro Gly Trp Ser Ala
1 5 10 15

Ala Ala Pro Cys Gln Leu Thr Ala Ala Ser Thr Tyr Trp Val Lys Arg
20 25 30

Phe Ser Cys Leu Arg Leu Pro Ser Ser Trp Asp Tyr Arg Arg Ala Pro
35 40 45

Gln His Pro Ala Asn Ser Phe Cys Ile Phe Ser Arg Asp Arg Ala Leu
50 55 60

Pro Cys Trp Arg Leu Val Ser Asn Ser Ala Pro Gln Val Ile Arg Leu
65 70 75 80

Pro Gln Pro Pro Lys Val Met Arg Leu Gln Ala
85 90

<210> 153
<211> 84
<212> PRT
<213> Homo sapiens

<400> 153
Met Leu Ala Thr Val Tyr Ala Asn Ala Lys Lys Gly Phe Phe Ile Tyr
1 5 10 15

Ser Cys Thr Glu Ile Cys Tyr Thr Phe Leu Ala Ser Phe Gln Glu Gln

20 25 30

Lys Phe Lys Asp Thr Gln Thr Leu Leu Ala Leu Asn Glu Phe Gln Leu
35 40 45

His Ile Leu Cys Ser Gln Glu Lys Arg Tyr Leu Ser Tyr Ile Leu Phe
50 55 60

Leu Ser Lys Arg Gln Asn Ile His Gln Trp Leu Tyr Arg Ile Leu Met
65 70 75 80

Val Leu Leu Ser

<210> 154
<211> 100
<212> PRT
<213> Homo sapiens

<400> 154
Met Phe Ser Phe Ser Met Pro Leu Asn Thr Leu Pro Ala Ala Met Gln
1 5 10 15

Arg Ala Ile His Gly Lys Arg Leu Leu Tyr Ile Asp Pro Cys Phe Trp
20 25 30

Cys Phe Asp Leu Leu Leu Cys Ile Glu Leu Ile Cys Pro Ser Ser His
35 40 45

Trp Cys Pro Pro Pro Pro Pro Asn Pro Ser Pro Leu Pro Ser Ser Phe
50 55 60

Phe Ser Ser Leu Leu Leu Cys Ser Leu Asn Cys Ile Pro Thr Pro Ser
65 70 75 80

Asp Phe Ser Leu Pro Lys Lys Ala Glu Glu Glu Arg Met Arg Glu Tyr
85 90 95

Val Leu Gly Arg
100

<210> 155
<211> 37
<212> PRT
<213> Homo sapiens

<400> 155
Met Pro Gly Ile Gly Gln Gly Pro Ile Gly Tyr Thr Glu Met Thr Asp
1 5 10 15

Thr Ala Phe Ser Phe Ser Glu Ser His Arg Ile Glu Glu Thr Ile Gln
20 25 30

Ala Glu Ser Thr Ile
35

<210> 156
<211> 35
<212> PRT
<213> Homo sapiens

<400> 156
Met Leu Asn Thr Cys Cys Cys Gly Ala Pro Gln Trp Gly His Val Ser
1 5 10 15

Ser Leu Arg Ser Trp Pro Arg Arg Ala Ala Val Thr Arg Ser Gln Arg
20 25 30

Val Gln Ala
35

<210> 157
<211> 67
<212> PRT
<213> Homo sapiens

<400> 157
Met Lys Ala Leu Pro Lys Ile Ser Pro Thr Pro Asn Phe Pro Leu Pro
1 5 10 15

Pro Thr Phe Pro Thr Ser Ser Thr Thr Leu Phe Gly Ala Thr Ala Gly
20 25 30

Pro Glu Gly Thr Lys Cys Gly Phe Pro Ser Leu Cys Pro Ser Gln Pro
35 40 45

Pro Glu Tyr Ile Cys Ala Trp Gly Ile Ser His Arg Asn Ser Gly Ala
50 55 60

Pro Pro Ala
65

<210> 158
<211> 144
<212> PRT
<213> Homo sapiens

<400> 158
Met Ser Thr Ala Lys Leu Thr Pro Gln Lys Arg Pro Leu Ser Glu His
1 5 10 15

Pro Arg Leu Arg Ser Ile Ser Pro Thr Val Met Pro Gly Leu Arg Ala
20 25 30

Ala Cys Leu Leu Val Ala Phe Leu Glu Asp Leu Leu Leu Val His Leu
35 40 45

Pro Leu Arg Ser Thr Val Pro Cys Leu His Gly Arg Ala Leu Pro Ala
50 55 60

Gly Met Gln Ala His Ser Ala Leu Gly Leu Asp Thr Thr Gly Arg Ser
65 70 75 80

Met Ala Asp Ser Thr His Gly Pro Gly Arg Glu Pro Trp Lys Leu Tyr
85 90 95

Thr Asp Gly Glu Leu Ser His Ser Thr Cys Ala Phe Ala Gln His Asn
100 105 110

Ala Tyr Tyr Lys Pro Thr Cys Thr Ser Phe Gln Leu Val Ala Phe Tyr
115 120 125

Cys Cys Cys Leu Lys Leu Gln Ser Phe Lys Gly Asn Leu Leu Lys Arg
130 135 140

<210> 159
<211> 17
<212> PRT
<213> Homo sapiens

<400> 159
Met Val Val Thr Met Val Leu Ser Ser Gly Ser Pro Pro Thr Gly Gly
1 5 10 15

Tyr

<210> 160
<211> 59
<212> PRT
<213> Homo sapiens

<400> 160
Met Gln Leu Ile Ala Pro Lys Thr Asp His Gly Gln Gly Lys Gly Arg
1 5 10 15

Lys Ile Asn Glu Lys Ile Cys Glu Phe Cys Phe Cys Ala Gly Phe Phe
20 25 30

Leu Lys Thr Asn Tyr Leu Leu Ala Asp Leu Gly Ala Leu Pro Gly Ser
35 40 45

Gln Ala Phe Pro Gly Asp Ala Leu Ser Gly Gly
50 55

<210> 161
<211> 35
<212> PRT
<213> Homo sapiens

<400> 161
Met Tyr Leu Leu His Arg Arg Ala Cys Arg Val Lys Ser Ser Arg Ser
1 5 10 15

Thr Cys Gly Lys Leu Asn Trp Asp Ser Thr Val Val Ile Ser Gly His
20 25 30

Ser Gly His
35

<210> 162
<211> 108
<212> PRT
<213> Homo sapiens

<400> 162
Met Pro Ala Ile Leu Ser Val Ser Ala Glu Pro His Leu Pro Pro Gly
1 5 10 15

Pro Leu Gly Ala Pro Glu Leu Cys Pro His Ser Leu Ser Leu Lys Val
20 25 30

Arg Pro Leu Leu Leu Pro Ala Leu Gly Arg Ile Arg Ala Gly Ser Glu
35 40 45

Ser Cys Glu Gln Val Ala Pro Gly Ala Trp Val Trp Thr Pro Arg Ile
50 55 60

Phe Arg Asp Pro Lys Ser Cys Arg Val Cys Gly Thr Arg Gln Glu Leu
65 70 75 80

Thr Ser Leu Cys Phe Cys Pro Ser Leu Leu Ser Leu Arg Thr Leu Gln
85 90 95

Leu Ala Ser Ala Arg Cys Leu Thr Ala Leu Trp Asn
100 105

<210> 163
<211> 53
<212> PRT
<213> Homo sapiens

<400> 163
Met Gly Ile His Phe Thr Ser Leu Thr Leu Cys Met Phe Val Ile Phe
1 5 10 15

His Lys Thr Lys Phe Cys Lys Val Val Tyr Leu Gly Leu His Cys Ile
20 25 30

Ser Thr Phe Phe Asn Ser Leu Ser Ala Arg Gly Ser Leu Gln Leu Ser
35 40 45

Lys Val Lys Phe Lys
50

<210> 164
<211> 54
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE

<222> (26)

<220>
<221> UNSURE
<222> (31)

<400> 164
Met Phe Lys Asn Ile Tyr Phe Val Leu Leu Tyr Cys Gln Thr Val Phe
1 5 10 15

Tyr Lys Ile Leu Ile Met Ser Ser Phe Xaa Val His Thr Ser Xaa Thr
20 25 30

Val Leu Pro Val Gln Val Gln Phe Pro Ile Ser Leu His Ser Ile Asp
35 40 45

Ile Ser Ser Gly Cys Pro
50

<210> 165
<211> 53
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (17)..(27)

<400> 165
Met Thr Asp Tyr Met Lys Leu Glu Lys Met Leu Ser Asp Lys Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Leu Lys Asn Phe Trp
20 25 30

Leu Val Ile Lys Glu Tyr Phe Leu Ile Ser Lys Asn Ile Leu Leu Thr
35 40 45

Ser Ile Asn Arg Lys
50

<210> 166
<211> 38
<212> PRT
<213> Homo sapiens

<400> 166
Met Lys Leu Ile His Arg Gly Arg Thr Thr Cys Leu Val Trp Tyr Gly
1 5 10 15

Asp Trp Asn Ser Cys Ser Pro Thr Arg Leu His Val Gly Val Lys Ser
20 25 30

Phe Lys Lys Tyr Cys Cys
35

<210> 167
<211> 28
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (28)

<400> 167
Met Phe Arg Phe Lys Leu Phe Tyr Ser Val Pro Phe Phe Gln Pro Glu
1 5 10 15

Glu Leu Ser Leu Val Phe Pro Val Asn Arg Lys Xaa
20 25

<210> 168
<211> 96
<212> PRT
<213> Homo sapiens

<400> 168
Phe Phe Phe Leu Thr Asp Ser Thr Leu Ser Pro Arg Leu Glu Cys Ser
1 5 10 15

Gly Ala Ile Ser Ala Tyr Cys Asn Leu His Phe Pro Gly Ser Ser Asn
20 25 30

Ser Pro Ala Ser Ala Ser Arg Ile Ala Gly Thr Thr Gly Lys Arg His
35 40 45

His Ala Gln Leu Ile Phe Val Phe Ala Val Glu Thr Gly Phe His His
50 55 60

Val Gly Gln Asp Gly Leu Asp Leu Leu Thr Ser Asp Leu Pro Ala Ser
65 70 75 80

Ala Ser Gln Ser Ala Glu Ile Thr Gly Met Asn His His Ala Trp Pro
85 90 95

<210> 169
<211> 19
<212> PRT
<213> Homo sapiens

<400> 169
Met Ala Thr Gln Lys Thr Ala Ser Gly Thr Ser Tyr Met Phe Pro Arg
1 5 10 15

Ala Ala Arg

<210> 170

<211> 33
<212> PRT
<213> Homo sapiens

<400> 170
Met Tyr Thr Val Leu Glu Ile Lys Thr Glu Lys Asn Phe Arg Cys Leu
1 5 10 15

Phe Ile His Leu Gln Ile Ile His Leu Leu His Ile Asn Met Asn Ile
20 25 30

Asn

<210> 171
<211> 40
<212> PRT
<213> Homo sapiens

<400> 171
Met Pro Phe Pro Leu Ile Ile Ile Phe Phe Leu Gln Asn Lys Gly Gln
1 5 10 15

Pro Leu Phe Pro Leu Lys Tyr Phe Leu Arg Leu Leu Val His Pro Ser
20 25 30

Leu Cys Pro Leu Phe Pro Leu Leu
35 40

<210> 172
<211> 113
<212> PRT
<213> Homo sapiens

<400> 172
Met Ala Phe Glu Arg Gly Gly Ile Pro Ala Gly Glu Leu Leu Leu Ala
1 5 10 15

Ser Phe Leu Gly Ser Arg Leu Arg Ile Phe Leu Thr Ser Lys Glu Lys
20 25 30

Tyr Pro Leu Ser Thr Glu Glu Ser Leu Leu Glu Leu Phe Leu Asn Thr
35 40 45

Gln Phe Asp Pro Ala Leu Arg Gly Phe Ser Thr Thr Leu Asn Ile Leu
50 55 60

Gly Glu Ser Cys Tyr Phe Gly Leu Met Ala Ala His Leu Glu Met Glu
65 70 75 80

Tyr Cys Leu Gly Thr Arg Gly Gly Glu Val Gly Leu Lys Gln His Tyr
85 90 95

His Leu Phe Pro Thr Ser Gly Val Lys Ile Leu Arg Ala Ala Lys Tyr
100 105 110

Asn

<210> 173
<211> 388
<212> PRT
<213> Homo sapiens

<400> 173
Met Thr Ala Ser Val Leu Leu His Pro Arg Trp Ile Glu Pro Thr Val
1 5 10 15

Met Phe Leu Tyr Asp Asn Gly Gly Gly Leu Val Ala Asp Glu Leu Asn
20 25 30

Lys Asn Met Glu Gly Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
35 40 45

Ala Ala Ala Gly Ala Gly Gly Gly Gly Phe Pro His Pro Ala Ala Ala
50 55 60

Ala Ala Gly Gly Asn Phe Ser Val Ala Ala Ala Ala Ala Ala Ala Ala
65 70 75 80

Ala Ala Ala Ala Asn Gln Cys Arg Asn Leu Met Ala His Pro Ala Pro
85 90 95

Leu Ala Pro Gly Ala Ala Ser Ala Tyr Ser Ser Ala Pro Gly Glu Ala
100 105 110

Pro Pro Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
115 120 125

Ala Ala Ala Ala Ala Ser Ser Ser Gly Gly Pro Gly Pro Ala Gly Pro
130 135 140

Ala Gly Ala Glu Ala Ala Lys Gln Cys Ser Pro Cys Ser Ala Ala Ala
145 150 155 160

Gln Ser Ser Ser Gly Pro Ala Ala Leu Pro Tyr Gly Tyr Phe Gly Ser
165 170 175

Gly Tyr Tyr Pro Cys Ala Arg Met Gly Pro His Pro Asn Ala Ile Lys
180 185 190

Ser Cys Ala Gln Pro Ala Ser Ala Ala Ala Ala Ala Ala Phe Ala Asp
195 200 205

Lys Tyr Met Asp Thr Ala Gly Pro Ala Ala Glu Glu Phe Ser Ser Arg
210 215 220

Ala Lys Glu Phe Ala Phe Tyr His Gln Gly Tyr Ala Ala Gly Pro Tyr
225 230 235 240

His His His Gln Pro Met Pro Gly Tyr Leu Asp Met Pro Val Val Pro
245 250 255

Gly Leu Gly Gly Pro Gly Glu Ser Arg His Glu Pro Leu Gly Leu Pro
260 265 270

```
Met Glu Ser Tyr Gln Pro Trp Ala Leu Pro Asn Gly Trp Asn Gly Gln
        275                 280                 285

Met Tyr Cys Pro Lys Glu Gln Ala Gln Pro Pro His Leu Trp Lys Ser
    290                 295                 300

Thr Leu Pro Asp Val Val Ser His Pro Ser Asp Ala Ser Ser Tyr Arg
305                 310                 315                 320

Arg Gly Arg Lys Lys Arg Val Pro Tyr Thr Lys Val Gln Leu Lys Glu
                325                 330                 335

Leu Glu Arg Glu Tyr Ala Thr Asn Lys Phe Ile Thr Lys Asp Lys Arg
            340                 345                 350

Arg Arg Ile Ser Ala Thr Thr Asn Leu Ser Glu Arg Gln Val Thr Ile
        355                 360                 365

Trp Phe Gln Asn Arg Arg Val Lys Glu Lys Lys Val Ile Asn Lys Leu
    370                 375                 380

Lys Thr Thr Ser
385
```

<210> 174
<211> 31
<212> PRT
<213> Homo sapiens

<400> 174

```
Met Asn Val Leu Leu Leu Ala Lys Phe Cys Phe Ser Ser Lys Ala Gln
  1               5                  10                  15

Phe Asn Ile Leu Val Val Arg Lys Asp Phe Phe Asp Pro Lys Lys
             20                  25                  30
```

<210> 175
<211> 60
<212> PRT
<213> Homo sapiens

<400> 175

```
Met Ser Ala Ser Thr Arg Tyr Lys Ser Ala Phe Ser Gln Pro Ser Leu
  1               5                  10                  15

Leu Gly Ala Glu Val Pro Glu Leu Leu Ser Gln Leu Ser Ala Gln Leu
             20                  25                  30

Gly Glu Gln Pro His Leu Pro Gly Leu Gly Ser Asn Ala Pro Gly Gly
         35                  40                  45

Ser Gly Glu Pro Phe Arg Ala Pro Asp Glu Gly Arg
     50                  55                  60
```

<210> 176

<211> 35
<212> PRT
<213> Homo sapiens

<400> 176
Met Ala Leu Gln His Tyr Phe Lys Ile Ala Leu Leu Asn Ser Phe Ile
1 5 10 15

Lys Lys Phe Gln Gln Met Thr Cys Gly Pro Val Asp Glu Lys Lys Phe
20 25 30

Cys Asn Tyr
35

<210> 177
<211> 730
<212> PRT
<213> Homo sapiens

<400> 177
Met Ala Asp Glu Asp Leu Ile Phe Arg Leu Glu Gly Val Asp Gly Gly
1 5 10 15

Gln Ser Pro Arg Ala Gly His Asp Gly Asp Ser Asp Gly Asp Ser Asp
20 25 30

Asp Glu Glu Gly Tyr Phe Ile Cys Pro Ile Thr Asp Asp Pro Ser Ser
35 40 45

Asn Gln Asn Val Asn Ser Lys Val Asn Lys Tyr Tyr Ser Asn Leu Thr
50 55 60

Lys Ser Glu Arg Tyr Ser Ser Ser Gly Ser Pro Ala Asn Ser Phe His
65 70 75 80

Phe Lys Glu Ala Trp Lys His Ala Ile Gln Lys Ala Lys His Met Pro
85 90 95

Asp Pro Trp Ala Glu Phe His Leu Glu Asp Ile Ala Thr Glu Arg Ala
100 105 110

Thr Arg His Arg Tyr Asn Ala Val Thr Gly Glu Trp Leu Asp Asp Glu
115 120 125

Val Leu Ile Lys Met Ala Ser Gln Pro Phe Gly Arg Gly Ala Met Arg
130 135 140

Glu Cys Phe Arg Thr Lys Lys Leu Ser Asn Phe Leu His Ala Gln Gln
145 150 155 160

Trp Lys Gly Ala Ser Asn Tyr Val Ala Lys Arg Tyr Ile Glu Pro Val
165 170 175

Asp Arg Asp Val Tyr Phe Glu Asp Val Arg Leu Gln Met Glu Ala Lys
180 185 190

Leu Trp Gly Glu Glu Tyr Asn Arg His Lys Pro Pro Lys Gln Val Asp
195 200 205

```
Ile Met Gln Met Cys Ile Ile Glu Leu Lys Asp Arg Pro Gly Lys Pro
    210                 215                 220

Leu Phe His Leu Glu His Tyr Ile Glu Gly Lys Tyr Ile Lys Tyr Asn
225                 230                 235                 240

Ser Asn Ser Gly Phe Val Arg Asp Asp Asn Ile Arg Leu Thr Pro Gln
                245                 250                 255

Ala Phe Ser His Phe Thr Phe Glu Arg Ser Gly His Gln Leu Ile Val
            260                 265                 270

Val Asp Ile Gln Gly Val Gly Asp Leu Tyr Thr Asp Pro Gln Ile His
        275                 280                 285

Thr Glu Thr Gly Thr Asp Phe Gly Asp Gly Asn Leu Gly Val Arg Gly
    290                 295                 300

Met Ala Leu Phe Phe Tyr Ser His Ala Cys Asn Arg Ile Cys Glu Ser
305                 310                 315                 320

Met Gly Leu Ala Pro Phe Asp Leu Ser Pro Arg Glu Arg Asp Ala Val
                325                 330                 335

Asn Gln Asn Thr Lys Leu Leu Gln Ser Ala Lys Thr Ile Leu Arg Gly
            340                 345                 350

Thr Glu Glu Lys Cys Gly Ser Pro Arg Val Arg Thr Leu Ser Gly Ser
        355                 360                 365

Arg Pro Pro Leu Leu Arg Pro Leu Ser Glu Asn Ser Gly Asp Glu Asn
    370                 375                 380

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

Thr Pro His Ser Gln Lys Leu Asp His Leu His Trp Pro Val Phe Ser
                405                 410                 415

Asp Leu Asp Asn Met Ala Ser Arg Asp His Asp His Leu Asp Asn His
            420                 425                 430

Arg Glu Ser Glu Asn Ser Gly Asp Ser Gly Tyr Pro Ser Glu Lys Arg
        435                 440                 445

Gly Glu Leu Asp Asp Pro Glu Pro Arg Glu His Gly His Ser Tyr Ser
    450                 455                 460

Asn Arg Lys Tyr Glu Ser Asp Glu Asp Ser Leu Gly Ser Ser Gly Arg
465                 470                 475                 480

Val Cys Val Glu Lys Trp Asn Leu Leu Asn Ser Ser Arg Leu His Leu
                485                 490                 495

Pro Arg Ala Ser Ala Val Ala Leu Glu Val Gln Arg Leu Asn Ala Leu
            500                 505                 510

Asp Leu Glu Lys Lys Ile Gly Lys Ser Ile Leu Gly Lys Val His Leu
```

```
        515                 520                 525

Ala Met Val Arg Tyr His Glu Gly Gly Arg Phe Cys Glu Lys Gly Glu
    530                 535                 540

Glu Trp Asp Gln Glu Ser Ala Val Phe His Leu Glu His Ala Ala Asn
545                 550                 555                 560

Leu Gly Glu Leu Glu Ala Ile Val Gly Leu Gly Leu Met Tyr Ser Gln
                565                 570                 575

Leu Pro His His Ile Leu Ala Asp Val Ser Leu Lys Glu Thr Glu Glu
            580                 585                 590

Asn Lys Thr Lys Gly Phe Asp Tyr Leu Leu Lys Ala Ala Glu Ala Gly
        595                 600                 605

Asp Arg Gln Ser Met Ile Leu Val Ala Arg Ala Phe Asp Ser Gly Gln
    610                 615                 620

Asn Leu Ser Pro Asp Arg Cys Gln Asp Trp Leu Glu Ala Leu His Trp
625                 630                 635                 640

Tyr Asn Thr Ala Leu Glu Met Thr Asp Cys Asp Glu Gly Gly Glu Tyr
                645                 650                 655

Asp Gly Met Gln Asp Glu Pro Arg Tyr Met Met Leu Ala Arg Glu Ala
            660                 665                 670

Glu Met Leu Phe Thr Gly Gly Tyr Gly Leu Glu Lys Asp Pro Gln Arg
        675                 680                 685

Ser Gly Asp Leu Tyr Thr Gln Ala Ala Glu Ala Ala Met Glu Ala Met
    690                 695                 700

Lys Gly Arg Leu Ala Asn Gln Tyr Tyr Gln Lys Ala Glu Glu Ala Trp
705                 710                 715                 720

Ala Gln Met Glu Glu Ala Gln Met Glu Glu
                725                 730


<210> 178
<211> 14
<212> PRT
<213> Homo sapiens

<400> 178
Met Cys Leu Ala Phe His Asp Ser Leu Ala Thr Leu Lys Met
  1               5                  10


<210> 179
<211> 97
<212> PRT
<213> Homo sapiens

<400> 179
Met Gly Asp Cys Phe Arg Ser Ala Gln Arg Asp Thr Leu Glu Ile Glu
```

```
  1               5                  10                  15

Tyr Phe Asn Leu Lys Lys Gln Gln His Leu Leu Val Ala Gly Ser Leu
             20                  25                  30

His Phe Trp Ser Pro Ala Val Val Trp Ser His Gln Ala Ser Ala Glu
         35                  40                  45

Trp Ala Tyr Ala Gln Gln Leu Val Gly Val Gly Ala Val Pro Ala Gly
     50                  55                  60

Leu Asn Met Asn Gln Ser Val Gln Asp Ala His Leu Gln Asp Ser Leu
 65                  70                  75                  80

Ala Ala Arg Thr Pro Cys Pro Leu Pro Val Val Val Ala Gly Ala Leu
                 85                  90                  95

Glu
```

<210> 180
<211> 48
<212> PRT
<213> Homo sapiens

<400> 180

```
Met Arg Tyr Leu Arg Lys Met Ser Ser Lys Gln Leu Thr Ile Gln Thr
  1               5                  10                  15

Trp Ser Ser Gly Asp Leu Asn Val Glu Val Asp Ile Gly Glu Ser Val
             20                  25                  30

Ala Leu Ser Glu Lys Lys Ala Cys Ser Leu Glu Gly Val Gly Ser Gly
         35                  40                  45
```

<210> 181
<211> 85
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (11)..(12)

<220>
<221> UNSURE
<222> (30)

<400> 181

```
Met Ser Arg Asp Ala Gly Gly Ser Lys Ala Xaa Xaa Leu Ser Thr His
  1               5                  10                  15

Trp Glu Asn Ala Leu Gln Gly Pro Gln Thr Gly Arg Thr Xaa Leu Val
             20                  25                  30
```

```
Glu Gly Thr Gly Ala Leu Asp Cys Pro Pro Trp Ala Gln Met Glu Thr
         35                  40                  45

Arg Gln Asp Gln Thr Gly Asn Leu Ser Leu Asp Lys Ser Leu Lys Val
     50                  55                  60

Thr Arg Ser Lys Leu Ile Ile Tyr Arg Gly Gly Lys Lys Ala Lys Gln
 65                  70                  75                  80

Val Asn Ser His Val
                 85
```

<210> 182
<211> 11
<212> PRT
<213> Homo sapiens

<400> 182

```
Met Ile Leu Phe Lys Cys Phe Met Arg Phe His
  1               5                  10
```

<210> 183
<211> 56
<212> PRT
<213> Homo sapiens

<400> 183

```
Met Glu Lys Thr Asp Gly Glu Asp Cys Leu Ser Leu Gly Arg Cys Ile
  1               5                  10                  15

Val Arg Ile Met Glu Gly His Asp Ile Leu Glu Arg Thr Val Leu Lys
             20                  25                  30

Trp Leu Leu Asp Arg Phe Lys Leu Tyr Arg Glu Thr Ile Lys Pro Ser
         35                  40                  45

Gly Gly Lys Glu Gln Val Tyr Asn
     50                  55
```

<210> 184
<211> 77
<212> PRT
<213> Homo sapiens

<400> 184

```
Val Thr Gln Ala Gly Val Gln Trp Phe Asn Thr Ser Ser Leu Gln Pro
  1               5                  10                  15

Pro Pro Pro Lys Pro Lys Arg Ser Ser His Leu Ser Pro Pro Ser Ser
             20                  25                  30

Trp Asp Tyr Lys Cys Ala Pro Pro His Pro Ala Lys Phe Val Ile Phe
         35                  40                  45

Gly Arg Asp Glu Val Ser Ser Cys Cys Pro Ala Trp Ser Arg Thr Pro
```

50 55 60

Glu Leu Lys Gln Tyr Ala His Leu Ser Leu Pro Asn Cys
65 70 75

<210> 185
<211> 77
<212> PRT
<213> Homo sapiens

<400> 185
Met Val Asn Ser Arg Gly Arg Asp Arg Lys Gly Gly Leu Leu Arg Glu
1 5 10 15

Ala Arg Pro Glu Ala Ala Ser Pro His Gln Cys His Val Gln Gly Leu
20 25 30

Ser His Ser Ser Gln Arg Gly Lys Phe Gln Ser Asn Pro Ala Ser Gly
35 40 45

Leu Tyr Trp Thr Leu Glu Lys Lys Arg Leu Ser Phe Tyr Arg Glu Thr
50 55 60

Gln Glu Pro Thr Ser Asp Tyr Ser Leu Ala Lys Gly Phe
65 70 75

<210> 186
<211> 245
<212> PRT
<213> Homo sapiens

<400> 186
Ala Met Glu Ser Lys Leu Leu Ile Gly Gly Arg Asn Ile Met Asp His
1 5 10 15

Thr Asn Glu Gln Gln Lys Met Leu Glu Leu Lys Arg Gln Glu Ile Ala
20 25 30

Glu Gln Lys Arg Arg Glu Arg Glu Met Gln Gln Glu Met Met Leu Arg
35 40 45

Asp Glu Glu Thr Met Glu Leu Arg Gly Thr Tyr Thr Ser Leu Gln Gln
50 55 60

Glu Val Glu Val Lys Thr Lys Lys Leu Lys Lys Leu Tyr Ala Lys Leu
65 70 75 80

Gln Ala Val Lys Ala Glu Ile Gln Asp Gln His Asp Glu Tyr Ile Arg
85 90 95

Val Arg Gln Asp Leu Glu Glu Ala Gln Asn Glu Gln Thr Arg Glu Leu
100 105 110

Lys Leu Lys Tyr Leu Ile Ile Glu Asn Phe Ile Pro Pro Glu Glu Lys
115 120 125

Asn Lys Ile Met Asn Arg Leu Phe Leu Asp Cys Glu Glu Glu Gln Trp

```
    130                 135                 140

Lys Phe Gln Pro Leu Val Pro Ala Gly Val Ser Ser Ser Gln Met Lys
145                 150                 155                 160

Lys Arg Pro Thr Ser Ala Val Gly Tyr Lys Arg Pro Ile Ser Gln Tyr
                165                 170                 175

Ala Arg Val Ala Met Ala Met Gly Ser His Pro Arg Tyr Arg Ala Val
            180                 185                 190

Phe Glu Met Glu Phe Ser His Asp Gln Glu Gln Asp Pro Arg Ala Leu
        195                 200                 205

His Ile Glu Arg Leu Met Arg Leu Asp Ser Phe Leu Glu Arg Pro Ser
    210                 215                 220

Thr Ser Lys Val Arg Lys Ser Arg Ser Cys Ser Ser Ser Gln Met Lys
225                 230                 235                 240

Lys Arg Pro Thr Ser
                245
```

<210> 187
<211> 14
<212> PRT
<213> Homo sapiens

<400> 187
Met Leu Ile Phe Lys Asn Gly Lys Met Leu Phe Asn Leu Lys
1 5 10

<210> 188
<211> 44
<212> PRT
<213> Homo sapiens

<400> 188
Met His Lys Ile Ile Asn Ser Asn Gly Ile Thr Thr Thr Leu Pro Asn
1 5 10 15

Pro Pro Glu Tyr Lys Ser Pro Met Met Ile Leu Ser Phe His Arg Ile
20 25 30

Leu Leu Glu Gly His Leu Asn Thr Phe Ser Ser Glu
35 40

<210> 189
<211> 21
<212> PRT
<213> Homo sapiens

<400> 189
Met Ser Gln Arg Gln Thr Gly Ile Ile Asp Phe Ala Val Val Leu Ser
1 5 10 15

Ser Ile Asn Ser Ile
20

<210> 190
<211> 23
<212> PRT
<213> Homo sapiens

<400> 190
Met Glu Lys Tyr Leu Leu Gly Ser Leu Leu Leu Phe Ala Arg Asn Arg
1 5 10 15

Gly Lys Gly Cys Phe Ser Ile
20

<210> 191
<211> 67
<212> PRT
<213> Homo sapiens

<400> 191
Met Thr Glu Ala Glu Ser Ala Ser Phe Leu Gln Ala Gly Arg Pro Glu
1 5 10 15

Thr Asp Lys Tyr Ile Asn Asn Gln Gly Cys Arg Leu Ser Cys Val Cys
20 25 30

Pro Phe Leu Ser Ala Glu Pro Thr Asn Gln Ile Ser Tyr Ser Ser Ser
35 40 45

Pro Gly Val Ile Glu Gln Gln Gln Tyr Tyr Val Asn Gly Ser Ser Phe
50 55 60

Gln Met Thr
65

<210> 192
<211> 55
<212> PRT
<213> Homo sapiens

<400> 192
Met Arg Gly Ser Gly Met Val Arg Gly Asp Pro Leu Glu Arg Gly Lys
1 5 10 15

Arg Pro Gln Glu Gly Leu Pro Pro Pro Leu Thr Glu Met Ala Leu Val
20 25 30

Glu Thr Phe Gly Gly Leu Glu Pro Leu Asp Ser Pro Cys Ser Asn Ser
35 40 45

His Thr Leu Leu Ser Glu Thr
50 55

<210> 193

<211> 70
<212> PRT
<213> Homo sapiens

<400> 193
Met Ala Pro Ser Gly Val Gln Trp Pro Gln Val Arg Gln Val Cys Ser
1 5 10 15

Gly Ser Arg Ala Gly Thr Pro His Leu His Pro Gly Thr Glu Leu Arg
20 25 30

Pro Trp Ala Lys Ala Gly Leu Pro Val Tyr Gln Gln Pro Gln Thr Thr
35 40 45

Ser Thr Cys Val Ala Gly Ala Val Ile Ala Ala Asp Ile Leu Ser Ser
50 55 60

Thr Ser Xaa Glu Thr Gly
65 70

<210> 194
<211> 67
<212> PRT
<213> Homo sapiens

<400> 194
Met Lys Val Lys Phe Ala Lys Ser Met Ser Phe Leu Val Ser Gly Phe
1 5 10 15

Glu Asp Asn Asp Phe Tyr Phe Arg Cys Val Leu Gly Pro Ala Ala Ser
20 25 30

Phe Tyr Ser Cys Leu Lys Cys Phe Ile Leu Gly Lys Leu Phe Asp Leu
35 40 45

Pro Gln Ser Lys Leu Lys Asn Leu Lys Val Cys Leu Lys Ala Thr Ile
50 55 60

Glu Lys Ile
65

<210> 195
<211> 39
<212> PRT
<213> Homo sapiens

<400> 195
Met Arg Arg Thr Phe Met Phe Ser Glu Tyr Ile Phe Lys Ser Arg Tyr
1 5 10 15

Leu Gly Ile Leu Cys Pro Phe Phe Phe Pro Leu Lys Leu Ile Thr Asn
20 25 30

His Met Arg Ser Asn Leu Ile
35

<210> 196
<211> 41
<212> PRT
<213> Homo sapiens

<400> 196
Met Ala Phe Glu Ile Tyr Ser Ile Thr Thr Leu Leu Cys Leu Ala Phe
1 5 10 15

Leu Gln Cys Gln Leu Gln Val Asp Glu Ser Lys Val Asn Gly Thr Glu
20 25 30

Lys Thr Ser Ser Arg Ser Gly Arg Gly
35 40

<210> 197
<211> 33
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (11)

<220>
<221> UNSURE
<222> (14)

<220>
<221> UNSURE
<222> (20)

<220>
<221> UNSURE
<222> (27)

<400> 197
Met Gln His Ser Ala Val Thr Val Gln Leu Xaa Tyr Ile Xaa Thr Xaa
1 5 10 15

Gln Ser Lys Xaa Lys Lys Lys Ile Gln Gly Xaa Ala Phe Asn Arg Asn
20 25 30

Leu

<210> 198
<211> 49
<212> PRT
<213> Homo sapiens

<400> 198
Met Pro Ser Leu Leu Ser Ser Asn Lys Thr Lys Leu Lys Asn Asn Ile
1 5 10 15

Ile Thr Ile Ile Val Thr Thr Lys Ile Val Pro His Lys Tyr Pro Ser
20 25 30

Thr His Gln Gly Ile Arg Arg Phe Arg Leu Lys Thr Ile Gln Arg Gln
35 40 45

Arg

<210> 199
<211> 82
<212> PRT
<213> Homo sapiens

<400> 199
Val Ser Pro Arg Leu Glu Cys Arg Gly Met Ile Ser Ala His Arg Lys
1 5 10 15

Leu His Leu Leu Gly Lys Gln Phe Ser Cys Leu Ser Leu Leu Ser Ser
20 25 30

Trp Asp Tyr Arg His Pro Pro Pro His Gln Leu Thr Leu Val Ser Ser
35 40 45

Val Glu Thr Gly Leu His Arg Val Gly Gln Ala Ser Leu Lys Leu Leu
50 55 60

Thr Ser Ser Asp Pro His Trp Asp Tyr Arg Arg Gln Ser Pro Arg Pro
65 70 75 80

Pro Ser

<210> 200
<211> 90
<212> PRT
<213> Homo sapiens

<400> 200
Met Gly Arg Lys Phe Ile Cys Phe Ala Leu Pro Ile Leu Tyr Gln Cys
1 5 10 15

Phe Pro Lys Cys Ile Pro Ser Val Leu Glu Gln Pro Gly Leu Leu Leu
20 25 30

Gly Thr Ser Pro Leu Pro Gln Pro Met Gly Asn His Thr Trp Ser Pro
35 40 45

Arg Asp Cys Ile Phe Ile Ser His Thr Thr Gln Gln Ser Val Asn Arg
50 55 60

Ser Tyr Ile Tyr Asp Ser Ser Phe Glu Met Ser Ser Ser Val Val Leu
65 70 75 80

Leu His Leu Ser Leu Thr Ser Ala Thr Ser
85 90

<210> 201

<211> 47
<212> PRT
<213> Homo sapiens

<400> 201
Met Glu Leu Pro Ser Lys Ala Ser Lys Lys Thr Ile Val Ser Phe Phe
1 5 10 15

Tyr Glu Glu Lys Asn Phe Leu His Leu Ser His Val Asn Leu Ser Pro
20 25 30

Ser Val Val Leu Pro Tyr Arg Pro Cys Asp Ser Arg Ala Phe Arg
35 40 45

<210> 202
<211> 33
<212> PRT
<213> Homo sapiens

<400> 202
Met Asn Thr Arg Met Leu Ser Ser Thr Ser Val Ala Pro Phe Val Ala
1 5 10 15

Thr Leu Tyr Val Ser His Cys Tyr Tyr Cys Phe Thr Gln Ser Met Thr
20 25 30

Val

<210> 203
<211> 26
<212> PRT
<213> Homo sapiens

<400> 203
Met Glu Leu Arg Ser Pro Ile Ile Leu Met Tyr Leu Ser Ile Gly Lys
1 5 10 15

His Leu Lys Asn His Lys Arg Thr Gln Leu
20 25

<210> 204
<211> 46
<212> PRT
<213> Homo sapiens

<400> 204
Met His Phe Ser His Ser Cys Arg Tyr Gly Gly Asp Gln Leu Phe Ile
1 5 10 15

Pro Pro Arg Val Thr Pro Ile Pro Phe Glu Val Leu Pro Tyr Gly Ile
20 25 30

Ser Leu Phe Ile Arg Cys Ser Asn Ser Tyr Arg Ser Leu Leu
35 40 45

<210> 205
<211> 49
<212> PRT
<213> Homo sapiens

<400> 205
Phe Phe Phe Phe Phe Cys Ile Phe Phe Val Glu Thr Gly Phe His His
1 5 10 15

Val Ala Gln Ala Gly Leu Lys Leu Leu Gly Ser Ser Asp Leu Pro Thr
20 25 30

Ser Ala Ser Gln Ser Pro Gly Ile Thr Gly Val Thr Thr Cys Val Ala
35 40 45

Gln

<210> 206
<211> 20
<212> PRT
<213> Homo sapiens

<400> 206
Met Leu Leu Phe Leu Tyr Lys Leu Tyr Pro Pro Gly Pro Leu Val Val
1 5 10 15

Phe Phe Gln Glu
20

<210> 207
<211> 36
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (25)..(33)

<400> 207
Met Asp Ile Leu Gln Trp Thr Ser Leu Cys Ala Arg Asn Leu Phe Ile
1 5 10 15

Leu Leu Leu Lys Asn Lys His Ser Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30

Xaa Ile Ile Asn
35

<210> 208
<211> 20
<212> PRT
<213> Homo sapiens

<400> 208

Met Lys Asn Asn Asn Asn Arg Phe Ile Ser Phe Arg Arg Glu Ala Ser
1 5 10 15

Lys Tyr Phe Leu
20

<210> 209
<211> 31
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (23)

<400> 209
Met Pro Ile Phe Lys Asp Tyr Leu Tyr Met Arg Asp Phe Ser Phe Asn
1 5 10 15

Tyr Thr Ala Pro Ser Gly Xaa Val Phe Leu Tyr Ile Phe Leu Gln
20 25 30

<210> 210
<211> 37
<212> PRT
<213> Homo sapiens

<400> 210
Met Arg Phe Cys Phe Glu Ser Ser Gln Cys Val Glu Ile Gln Leu Leu
1 5 10 15

Leu His Gln Asn Tyr Phe His Leu Cys Thr Thr Trp Leu Lys Thr Thr
20 25 30

Asp Arg Gln Glu Ser
35

<210> 211
<211> 21
<212> PRT
<213> Homo sapiens

<400> 211
Met Ser Glu Asn Phe Ile Ile Trp Ile Leu Cys Gly Met Phe Leu Leu
1 5 10 15

Pro Val Ala Phe Phe
20

<210> 212
<211> 57
<212> PRT
<213> Homo sapiens

<220>

<221> UNSURE
<222> (28)..(45)

<400> 212
Met Ile Leu Thr Asp Ser Leu Asp Leu Thr Gly Asp Ala Pro Val Val
1 5 10 15

Lys Thr His Phe Pro Gln Gly Gln Gln His Tyr Xaa Xaa Xaa Xaa Xaa
20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Gln Gln
35 40 45

Ile Lys Ala Gly Ser Gln Ser Ser Ile
50 55

<210> 213
<211> 105
<212> PRT
<213> Homo sapiens

<400> 213
Phe Phe Phe Phe Leu Arg Tyr Gly Leu Thr Arg Ser Thr Arg Leu Glu
1 5 10 15

Cys Ser Gly Thr Ile Met Ala His Cys Ser Leu Asp Leu Pro Gly Ser
20 25 30

Ser Asp Cys Pro Ala Leu Thr Ser Ala Val Ala Gly Thr Lys Asp Val
35 40 45

Cys Ala His Ser Gln Leu Ile Phe Ala Asn Tyr Phe Leu Val Glu Met
50 55 60

Gly Ser Pro Tyr Val Ile Glu Ala Gly Ile Glu Phe Leu Ala Ala Ser
65 70 75 80

Ser Pro Pro Ile Leu Ala Ser Gln Ser Ala Gly Leu Lys Gly Met Ser
85 90 95

His His Ile Trp Leu Asn Phe Phe Leu
100 105

<210> 214
<211> 33
<212> PRT
<213> Homo sapiens

<400> 214
Met Glu Gly Pro Ser Leu Thr Pro Thr Arg Lys Val Arg Gly Gly Asn
1 5 10 15

Thr Ser Ser Phe Leu Lys Gly Gln Asp Gly Cys Phe Ser Thr Ala Ala
20 25 30

Thr

<210> 215
<211> 79
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (3)

<220>
<221> UNSURE
<222> (5)

<220>
<221> UNSURE
<222> (12)

<220>
<221> UNSURE
<222> (17)

<400> 215
Met Arg Xaa Ala Xaa Val Pro Pro Val Leu Asp Xaa Gln Leu Phe Arg
1 5 10 15

Xaa Ser Glu Ile Tyr Leu Arg Asp Ser Leu Ala Phe Tyr Phe Ser Thr
20 25 30

Ala Asn Ala Asp Arg Gln Ser Gly Gly Phe Ser Gln Cys Ser His Leu
35 40 45

Leu Pro Asn Cys Tyr Arg Asp Arg His Ile Leu Leu Pro Ala Lys Met
50 55 60

Ala Cys Leu Cys Asp Ser Leu Phe Gly Phe Ile Ser Pro Thr Ser
65 70 75

<210> 216
<211> 17
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (2)

<400> 216
Met Xaa Tyr Asn Gly Tyr Ala Ala Ala Ile Tyr Asn Leu Thr His Thr
1 5 10 15

Leu

<210> 217
<211> 20

<212> PRT
<213> Homo sapiens

<400> 217
Met Thr Ile Gln Val Ala His Ser Leu Val Asn Ile Trp Cys Ser Thr
1 5 10 15

Val Ala His Ala
20

<210> 218
<211> 50
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (21)

<220>
<221> UNSURE
<222> (29)

<400> 218
Met Ser Leu Tyr Ile Ile Ser Phe Asn Ala His Asn His Ile Lys Gly
1 5 10 15

Ser Leu Thr Tyr Xaa Thr Asp Glu Lys Thr Gly Ser Xaa Ile Phe Pro
20 25 30

Ser Ser Pro Ser His Ala Ser Xaa Trp Gln Asp Gln Tyr Leu Tyr Thr
35 40 45

Asp Xaa
50

<210> 219
<211> 40
<212> PRT
<213> Homo sapiens

<400> 219
Met Asn Leu Asn Leu Lys Thr Thr Gln Thr His Phe Lys Tyr Phe Glu
1 5 10 15

Val Ile Pro Gln Leu Thr Val Phe Leu Val Phe Asn Ile Val Thr Gln
20 25 30

Ser Phe Ser Lys Ile Thr Leu Glu
35 40

<210> 220
<211> 39
<212> PRT
<213> Homo sapiens

<400> 220
Met Lys Ser Gln His Asp Arg Val Met Arg Ser Met Asp Asn Lys Leu
1 5 10 15

Trp Gln Trp Val Arg Gly Arg Glu Ile Arg Arg Leu Ile Val Lys His
20 25 30

Leu Ser Phe Ile Asn Ile Tyr
35

<210> 221
<211> 92
<212> PRT
<213> Homo sapiens

<400> 221
Phe Phe Phe Phe Phe Glu Thr Glu Phe Leu Leu Val Thr Arg Leu Glu
1 5 10 15

Cys Ser Gly Ala Ile Lys Thr His Cys Asn Leu Arg Leu Leu Gly Pro
20 25 30

Ser Asp Ser Leu Ala Ser Ala Ser Ala Val Ala Trp Thr Thr Gly Thr
35 40 45

His His His Ile Gln Leu Ile Ser Val Phe Leu Val Glu Thr Gly Phe
50 55 60

His His Phe Gly Gln Gly Asp Ser Asn Ser Ala Pro Gln Val Ile His
65 70 75 80

Pro Pro Ala Pro Pro Lys Val Leu Arg Leu Gln Ala
85 90

<210> 222
<211> 55
<212> PRT
<213> Homo sapiens

<400> 222
Met Cys Cys Ser Pro Leu Ile Gln Ile Ser Arg Val Glu Cys Val His
1 5 10 15

Gln Phe Pro Thr Leu Ser Ser Thr Thr Ser Pro Gly Gln Leu Gln Cys
20 25 30

Gly Arg Ser Ile Phe Thr Lys Glu Ile Lys Cys Tyr Lys Ser Tyr His
35 40 45

His Thr Thr Gly Leu Gly Ser
50 55

<210> 223
<211> 55
<212> PRT
<213> Homo sapiens

<400> 223
Met Glu Pro Glu Phe Asn Ala Ala Ser Val Val Ala Leu Gln Ser Met
1 5 10 15

Leu Ser Thr Ser Ala Ala Tyr His Phe Cys Ile Ser His Met Ser Cys
20 25 30

Gly Phe Gly Ser Ala Asn Ser Tyr Val Ile Ser His Ser Ser Ser Leu
35 40 45

Arg Glu Ile Thr Ala Gly Gln
50 55

<210> 224
<211> 87
<212> PRT
<213> Homo sapiens

<400> 224
Met Ser Arg Arg Leu Tyr Ser Lys His Ile Leu Gly Asn Ile Ser Asn
1 5 10 15

Cys Asn Asp Phe Leu Ile Tyr Phe Phe Phe Cys Cys Phe Asn Ile Phe
20 25 30

Trp Ile Leu Lys Ser Pro Gln Ser Phe Lys Arg Ile Leu Asn Asn Ala
35 40 45

Glu Pro Ala Ala Glu Asn Leu Thr His Asn Leu Cys Cys Arg Glu Ile
50 55 60

Glu Leu Pro Leu Phe Phe Val Leu Pro Tyr Val Ile Ile Leu Ile Lys
65 70 75 80

Leu Ile Thr Ala Arg Ser Ser
85

<210> 225
<211> 113
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (94)

<220>
<221> UNSURE
<222> (98)

<400> 225
Met Val Pro Ala Thr Ile Thr Pro Pro Gln Ile Ser Thr Ile Thr Cys
1 5 10 15

Gln Ala Met Phe His Phe Ser Pro Asp Pro Leu Gln Leu Ile Leu Ser
20 25 30

```
Ala Thr Ala Lys Pro Ile Ile Phe Ile Pro Thr Ser Asp His Asp Thr
         35                  40                  45

Pro Leu Leu Gln Thr Leu Gln Trp Leu Pro Ile Leu Thr Val Lys Pro
     50                  55                  60

Gln Ser Leu Leu Arg Leu Gly Arg Pro Cys Lys Thr Trp Pro Pro Leu
 65                  70                  75                  80

Pro Leu Leu Pro Ser His Val His His Cys Ser Met Leu Xaa Cys Cys
                 85                  90                  95

Tyr Xaa Arg Arg Gly Gly Thr Phe Pro Leu Ser Leu His Ser Ser Phe
            100                 105                 110

Pro
```

<210> 226
<211> 55
<212> PRT
<213> Homo sapiens

<400> 226

```
Met Ser Cys Ser Ile Trp Tyr Arg Leu Thr Ile Leu Leu Val Leu Tyr
  1               5                  10                  15

Thr Tyr Thr Ala Val Val Gln Leu Ser Lys Trp Met Glu Asp His Gly
             20                  25                  30

Lys Pro Leu Phe Tyr His Trp Ser Arg Asn Leu Gln Ile Ser Lys Arg
         35                  40                  45

Lys Lys Leu Glu Gln Ser Val
     50                  55
```

<210> 227
<211> 52
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (2)

<400> 227

```
Met Xaa Leu Tyr Ser Tyr Ile Asp Ile Cys Ala Ser Gly Gly Ile Leu
  1               5                  10                  15

Thr Ser Ser Asn Phe Met Glu Trp Leu Ser Lys Lys Lys Ile Phe Ser
             20                  25                  30

Val Val Val Thr Tyr Ser Val Gly Trp Val Gly Cys Phe Gly Ile Gly
         35                  40                  45

Ser Gly Cys Met
```

50

<210> 228
<211> 33
<212> PRT
<213> Homo sapiens

<400> 228
Met Ala Ile Tyr Pro Lys Ile Asn Tyr Asp Met Asp Ser Asn Ile Lys
1 5 10 15

Pro Leu Arg Leu Glu Gly Cys Leu Tyr Lys Leu Ile Asn Ile Lys Ser
20 25 30

Gln

<210> 229
<211> 31
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (26)

<400> 229
Met Ser Pro Val Gly Glu Ser Arg Arg Ser Ser Cys Pro Ser Leu Leu
1 5 10 15

Ile Leu Phe Val Phe Phe Lys Leu Leu Xaa Ile Phe Asp Thr Asp
20 25 30

<210> 230
<211> 33
<212> PRT
<213> Homo sapiens

<400> 230
Met Phe Gln Thr Cys Phe Lys Phe Ser Ser Leu Val Tyr Ile Cys Thr
1 5 10 15

Phe Ile Ser Ile Ile His Glu Ala Lys Leu Arg Ser Arg Lys Lys Lys
20 25 30

Thr

<210> 231
<211> 51
<212> PRT
<213> Homo sapiens

<400> 231
Met Pro Ile Ser Cys Leu Phe Phe Leu Tyr Gln Arg Glu Leu Arg Trp

```
  1               5                  10                  15

Thr Ser Met Pro Phe Leu Ser Tyr Gln Pro Glu Asn Val Lys Lys Leu
             20                  25                  30

Gly Gly Asp Arg Leu Val Val Ser Phe Leu Phe Asn Lys Val Phe Ile
         35                  40                  45

Leu Leu Ala
     50
```

<210> 232
<211> 330
<212> PRT
<213> Homo sapiens

<400> 232

```
Asn Met Asp Gly Pro Met Arg Pro Arg Ser Ala Ser Leu Val Asp Phe
  1               5                  10                  15

Gln Phe Gly Val Val Ala Thr Glu Thr Ile Glu Asp Ala Leu Leu His
             20                  25                  30

Leu Ala Gln Gln Asn Glu Gln Ala Val Arg Glu Ala Ser Gly Arg Leu
         35                  40                  45

Gly Arg Phe Arg Glu Pro Gln Ile Gln Phe Val Phe Leu Leu Ser Glu
     50                  55                  60

Gln Trp Cys Leu Glu Lys Ser Val Ser Tyr Gln Ala Val Glu Ile Leu
 65                  70                  75                  80

Glu Arg Phe Met Val Lys Gln Ala Glu Asn Ile Cys Arg Gln Ala Thr
                 85                  90                  95

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

Lys Gln Gln Leu Val Asn Lys Phe Thr Leu Arg Leu Val Ser Cys Val
        115                 120                 125

Gln Leu Pro Ser Lys Leu Ser Phe Arg Asn Lys Ile Ile Ser Asn Ile
    130                 135                 140

Thr Val Leu Asn Phe Leu Gln Ala Leu Gly Tyr Leu His Thr Lys Glu
145                 150                 155                 160

Glu Leu Leu Glu Ser Glu Leu Asp Val Leu Lys Ser Leu Asn Phe Arg
                165                 170                 175

Ile Asn Leu Pro Thr Pro Leu Ala Tyr Val Glu Thr Leu Leu Glu Val
            180                 185                 190

Leu Gly Tyr Asn Gly Cys Leu Val Pro Ala Met Arg Leu His Ala Thr
        195                 200                 205

Cys Leu Thr Leu Leu Asp Leu Val Tyr Leu Leu His Glu Pro Ile Tyr
    210                 215                 220
```

```
Glu Ser Leu Leu Arg Ala Ser Ile Glu Asn Ser Thr Pro Ser Gln Leu
225                 230                 235                 240

Gln Gly Glu Lys Phe Thr Ser Val Lys Glu Asp Phe Met Leu Leu Ala
                245                 250                 255

Val Gly Ile Ile Ala Ala Ser Ala Phe Ile Gln Asn His Glu Cys Trp
            260                 265                 270

Ser Gln Val Val Gly His Leu Gln Ser Ile Thr Gly Ile Ala Leu Ala
        275                 280                 285

Ser Ile Ala Glu Phe Ser Tyr Ala Ile Leu Thr His Gly Val Gly Ala
    290                 295                 300

Asn Thr Pro Gly Arg Gln Gln Ser Ile Pro Pro His Leu Ala Ala Arg
305                 310                 315                 320

Ala Leu Lys Thr Val Ala Ser Ser Asn Thr
                325                 330


<210> 233
<211> 17
<212> PRT
<213> Homo sapiens

<400> 233
Met Lys Ile Lys Arg Thr Gln Pro His Ala Glu Val Ala Gln Arg Thr
  1               5                  10                  15

Ser



<210> 234
<211> 34
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (28)

<220>
<221> UNSURE
<222> (30)

<400> 234
Met Leu Gln Leu Thr Phe Leu Gln Tyr Ser Leu Leu Arg Arg Cys Thr
  1               5                  10                  15

Leu Thr Cys Lys Phe Tyr Asn Ser Val Phe Asn Xaa Leu Xaa Phe Val
            20                  25                  30

His Leu
```

<210> 235
<211> 52
<212> PRT
<213> Homo sapiens

<400> 235
Met His Leu Asp His Asp Ser Ile Leu Pro Gly Phe Val Gln Gln Leu
1 5 10 15

Lys Leu Trp Lys Pro Gln Tyr Pro Ala His Trp Asp Asn Val Arg Thr
20 25 30

Tyr Cys Thr Ser Ser Val Pro Arg Gly Ile Leu Phe Leu His Phe Gly
35 40 45

Ile Ser Glu Ile
50

<210> 236
<211> 45
<212> PRT
<213> Homo sapiens

<400> 236
Met Ala Leu Ala Cys Ala Gly Arg Gly Gly Glu Asp Arg Glu Val Ser
1 5 10 15

Gly Trp Ile Arg Leu Leu Gly Val Pro Ala Pro Met Thr Glu Thr Thr
20 25 30

Gln Val Gly Pro Ser Ala Pro Ala His His Lys Asn Arg
35 40 45

<210> 237
<211> 36
<212> PRT
<213> Homo sapiens

<400> 237
Met Leu Gly Arg Arg Lys Arg Leu Val Val Asp Thr Asn Ala Tyr Val
1 5 10 15

Val Met Gly Ala Phe Lys Asn Met Leu Phe Phe Phe Ser Lys Gly Arg
20 25 30

Leu Phe Trp Met
35

<210> 238
<211> 48
<212> PRT
<213> Homo sapiens

<400> 238
Met Phe Ile Ser Met Leu Met Glu Asp Gln Ser Gln Gly Glu His Val

```
  1               5                  10                  15

Cys Asn Gly Arg Ile Lys Gly Asn Gly Glu Lys Ile Phe Leu Thr Gly
             20                  25                  30

Cys Ile Leu Gln Val Tyr Leu Pro Ile Gln Ile Ile Lys Leu Phe Phe
         35                  40                  45
```

<210> 239
<211> 25
<212> PRT
<213> Homo sapiens

<400> 239

```
Met Glu Gly Phe Met Ser Gln Asn Pro Val Leu Gly Lys Leu Lys Val
  1               5                  10                  15

Arg Tyr Glu Phe Phe Gly Tyr Val Ile
             20                  25
```

<210> 240
<211> 52
<212> PRT
<213> Homo sapiens

<400> 240

```
Lys Lys Lys Thr Val Thr Met Lys Arg Asn Leu Asn Pro Ile Phe Asn
  1               5                  10                  15

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

Ile Ile Ile Thr Val Met Asp Lys Asp Lys Leu Ser Arg Asn Asp Val
         35                  40                  45

Ile Gly Lys Val
     50
```

<210> 241
<211> 84
<212> PRT
<213> Homo sapiens

<400> 241

```
Met Pro Arg Thr Phe Ser Gly Gln His Leu Pro Ser Leu Gly Lys Leu
  1               5                  10                  15

Ser Thr Phe Lys Gln Glu Gln Leu Leu Ser Val Leu Ala Phe Pro Gly
             20                  25                  30

Arg Leu Gln Ser Ala Pro Asn Gly Gln Leu Gly Ser Leu His Ile Tyr
         35                  40                  45
```

Ser Leu Gly Lys Leu Trp His Tyr Cys Ala Thr Phe Ala Ser Ala Gln
50 55 60

Leu Pro Leu Leu Val His Ser Gln Ile Leu Lys Phe Tyr Phe Leu Asn
65 70 75 80

Asn Cys Gly Trp

<210> 242
<211> 49
<212> PRT
<213> Homo sapiens

<400> 242
Met Thr Gln Leu Arg Lys Leu Trp Pro Asn Ala Phe Tyr Ile Gly Tyr
1 5 10 15

Thr Asp Met Lys Arg Asn Asn Ser Ala Leu His Thr Lys Ala Glu Ala
20 25 30

Lys Glu Thr Asn Gly Arg Val Ser Glu Asn Ser Leu Lys Tyr Glu Arg
35 40 45

Met

<210> 243
<211> 11
<212> PRT
<213> Homo sapiens

<400> 243
Met Ser Arg Gln Val Gly Leu Ala Glu Thr Ile
1 5 10

<210> 244
<211> 18
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (10)

<400> 244
Met Xaa Tyr Lys His Arg Glu Met Leu Xaa Val Ser Gln Lys Asn Lys
1 5 10 15

Thr Leu

<210> 245
<211> 47
<212> PRT

<213> Homo sapiens

<400> 245
Met Arg Arg Asn Ser Trp Lys Thr Lys Tyr Leu Thr Thr Phe Ser Gly
1 5 10 15

Pro Ser Thr Val Trp Glu Gly Met Asn Leu Thr Ser Val Pro Asn Gln
20 25 30

Trp Cys Ile Ser Met Trp Glu Gly Gly Ser Leu Cys Ser Ser Gln
35 40 45

<210> 246
<211> 54
<212> PRT
<213> Homo sapiens

<400> 246
Met Leu Ser Pro Val Asp Thr Pro Arg Glu Gly Val Ser Cys Ala Ala
1 5 10 15

Ala Pro Val Ser Phe Pro Arg Glu His Leu Thr Ser Ser Ala Trp Val
20 25 30

Thr Arg Ser Pro Arg Ile Gln Pro Thr Leu Val Met Arg Glu Trp Gly
35 40 45

Arg Thr Val Gln Glu Ser
50

<210> 247
<211> 49
<212> PRT
<213> Homo sapiens

<400> 247
Met Lys Ala Glu Ser Glu Gly Ile Val Ala Ala Arg Asp Glu Val Gly
1 5 10 15

Leu Trp Asn Leu Phe Phe Val Arg Leu Leu Arg Ser Gly Ile Asn Pro
20 25 30

Pro Lys Gly Lys Leu Ser Pro Val Gly Pro Asp Ser Ser Pro Val Pro
35 40 45

Thr

<210> 248
<211> 68
<212> PRT
<213> Homo sapiens

<400> 248
Met Cys Leu Glu Met Arg Lys Ala Gly Tyr Arg Glu Glu Glu Ala Val
1 5 10 15

```
Arg Ala Thr Ser Glu Thr Ser Arg Val Ile Leu Thr Ser Arg Gly Pro
            20                  25                  30

Met Val Leu Lys Gln Gly Tyr Leu Gly Lys Cys Arg Lys Ala Tyr Phe
        35                  40                  45

Ser Asn Lys Arg Ile Ile Lys Cys Asn Met Ser Phe Cys Tyr Gly Leu
    50                  55                  60

Ser Asn Leu Leu
65
```

<210> 249
<211> 62
<212> PRT
<213> Homo sapiens

<400> 249

```
Ile Val Phe Arg Val Phe Met Ser Phe His Met Phe Ile Pro Phe Phe
  1               5                  10                  15

Phe Phe Ala Phe Phe Phe Phe Val Cys Val Cys Val Phe Ala Phe Phe
            20                  25                  30

Leu Phe Cys Leu Arg His Ser Leu Thr Leu Ser Pro Arg Leu Glu Cys
        35                  40                  45

Asn Gly Thr Ile Ser Ala His Cys Asn Leu His Leu Pro Gly
    50                  55                  60
```