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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2002 (27.12.2002)

PCT

(10) International Publication Number
WO 02/102986 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/19106

(22) International Filing Date: 14 June 2002 (14.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/298,241 14 June 2001 (14.06.2001) US

(71) Applicant: LEXICON GENETICS INCORPORATED [US/US]; 4000 Research Forest Drive, The Woodlands, TX 77381 (US).

(72) Inventors: HU, Yi; 333 Holly Creek Ct. #203, The Woodlands, TX 77381 (US). NEPOMNICHY, Boris; 905 Cypress Station Dr. #1515, Houston, TX 77909 (US). TURNER, C., Alexander, Jr.; 67 Winter Wheat Place, The Woodlands, TX 77381 (US). MATHUR, Brian; 1071 Thorne Avenue, Wooster, OH 44691 (US). FRIDDLE, Carl, Johan; 127 S. Goldenvine Circle, The Woodlands, TX 77382 (US).

(74) Agents: ISHIMOTO, Lance, K. et al.; LEXICON GENETICS INCORPORATED, 4000 Research Forest Drive, The Woodlands, TX 77381 (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, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, 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 02/102986 A2

(54) Title: NOVEL HUMAN TRANSPORTER PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

(57) Abstract: Novel human polynucleotide and polypeptide sequences are disclosed that can be used in therapeutic, diagnostic, and pharmacogenomic applications.

NOVEL HUMAN TRANSPORTER PROTEINS AND POLYNUCLEOTIDES ENCODING THE SAME

The present application claims the benefit of U.S. Provisional Application Number 60/298,241, which was filed on June 14, 2001, and is herein incorporated by reference in its entirety.

1. INTRODUCTION

The present invention relates to the discovery, identification, and characterization of novel human polynucleotides encoding proteins that share sequence similarity with mammalian transporter proteins. The invention encompasses the described polynucleotides, host cell expression systems, the encoded proteins, fusion proteins, polypeptides and peptides, antibodies to the encoded proteins and peptides, and genetically engineered animals that either lack or overexpress the disclosed polynucleotides, antagonists and agonists of the proteins, and other compounds that modulate the expression or activity of the proteins encoded by the disclosed polynucleotides, which can be used for diagnosis, drug screening, clinical trial monitoring, the treatment of diseases and disorders, and cosmetic or nutriceutical applications.

2. BACKGROUND OF THE INVENTION

Transporter proteins are integral membrane proteins that mediate or facilitate the passage of materials across the lipid bilayer. Given that the transport of materials across the membrane can play an important physiological role, transporter proteins are good drug targets. Additionally, one of the mechanisms of drug resistance involves diseased cells using cellular transporter systems to export chemotherapeutic agents from the cell. Such mechanisms are particularly relevant to cells manifesting resistance to a multiplicity of drugs.

3. SUMMARY OF THE INVENTION

The present invention relates to the discovery, identification, and characterization of nucleotides that encode novel human proteins, and the corresponding amino acid sequences of these proteins. The novel human proteins (NHPs) described for the first time herein share structural similarity with mammalian ATP-binding cassette (ABC) transporters, organic ion transporters/symporters, and sodium-glucose cotransporters.

The novel human nucleic acid sequences described herein encode alternative proteins/open reading frames (ORFs) of 1205 and 1207 amino acids in length (ABC transporter, SEQ ID NOS:3 and 4, respectively), and 681, 674, 745 and 738 amino acids in length (sodium/glucose-like cotransporter, SEQ ID NOS:7, 9 11 and 13, respectively).

The invention also encompasses agonists and antagonists of the described NHPs, including small molecules, large molecules, mutant NHPs, or portions thereof, that compete with native NHPs, peptides, and antibodies, as well as nucleotide sequences that can be used to inhibit the expression of the described NHPs (*e.g.*, antisense and ribozyme molecules, and open reading frame or regulatory sequence replacement constructs) or to enhance the expression of the described NHPs (*e.g.*, expression constructs that place the described polynucleotide under the control of a strong promoter system), and transgenic animals that express a NHP sequence, or "knock-outs" (which can be conditional) that do not express a functional NHP. Knock-out mice can be produced in several ways, one of which involves the use of mouse embryonic stem cell ("ES cell") lines that contain gene trap mutations in a murine homolog of at least one of the described NHPs. When the unique NHP sequences described in SEQ ID NOS:1-13 are "knocked-out" they provide a method of identifying phenotypic expression of the particular gene, as well as a method of assigning function to previously unknown genes.

In addition, animals in which the unique NHP sequences described in SEQ ID NOS:1-13 are "knocked-out" provide an unique source in which to elicit antibodies to homologous and orthologous proteins, which would have been previously viewed by the immune system as "self" and therefore would have failed to elicit significant antibody responses. To these ends, gene trapped knockout ES cells have been generated in murine homologs of certain of the described NHPs.

Additionally, the unique NHP sequences described in SEQ ID NOS:1-13 are useful for the identification of protein coding sequences, and mapping an unique gene to a particular chromosome. These sequences identify biologically verified exon splice junctions, as opposed to splice junctions that may have been bioinformatically predicted from genomic sequence alone. The sequences of the present invention are also useful as additional DNA markers for restriction fragment length polymorphism (RFLP) analysis, and in forensic biology, particularly given the presence of nucleotide polymorphisms within the described sequences.

Further, the present invention also relates to processes for identifying compounds that modulate, *i.e.*, act as agonists or antagonists of, NHP expression and/or NHP activity that utilize purified preparations of the described NHPs and/or NHP products, or cells expressing the same. Such compounds can be used as therapeutic agents for the treatment of any of a wide variety of symptoms associated with biological disorders or imbalances.

4. DESCRIPTION OF THE SEQUENCE LISTING AND FIGURES

The Sequence Listing provides the sequences of the described NHP ORFs that encode the described NHP amino acid sequences. SEQ ID NO:5 describes a polynucleotide encoding a NHP ORF along with regions of flanking sequence.

5. DETAILED DESCRIPTION OF THE INVENTION

The NHPs described for the first time herein are novel proteins that can be expressed in, *inter alia,* human cell lines, bone marrow, and osteocarcinoma cells (SEQ ID NOS:1-5), or lymph node, kidney, fetal liver, liver, testis, thyroid, adrenal gland, small intestine, uterus, bladder, hypothalamus, fetal kidney, and fetal lung cells (SEQ ID NOS:6-13).

The present invention encompasses the nucleotides presented in the Sequence Listing, host cells expressing such nucleotides, the expression products of such nucleotides, and: (a) nucleotides that encode mammalian homologs of the described polynucleotides, including the specifically described NHPs, and the NHP products; (b) nucleotides that encode one or more portions of the NHPs that correspond to functional domains, and the polypeptide products specified by such nucleotide sequences, including, but not limited to, the novel regions of any active domain(s); (c) isolated nucleotides that encode mutant versions, engineered or naturally occurring, of the described NHPs in which all or a part of at least one domain is deleted or altered, and the polypeptide products specified by such nucleotide sequences, including, but not limited to, soluble proteins and peptides in which all or a portion of the signal (or one or more hydrophobic transmembrane) sequence is deleted; (d) nucleotides that encode chimeric fusion proteins containing all or a portion of a coding region of a NHP, or one of its domains (*e.g.*, a receptor or ligand binding domain, accessory protein/self-association domain, *etc.*) fused to another peptide or polypeptide; or (e) therapeutic or diagnostic derivatives of the described polynucleotides, such as oligonucleotides, antisense polynucleotides, ribozymes, dsRNA, or gene therapy constructs comprising a sequence first disclosed in the Sequence Listing.

As discussed above, the present invention includes the human DNA sequences presented in the Sequence Listing (and vectors comprising the same), and additionally contemplates any nucleotide sequence encoding a contiguous NHP open reading frame (ORF) that hybridizes to a complement of a DNA sequence presented in the Sequence Listing under highly stringent conditions, *e.g.*, hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1x SSC/0.1% SDS at 68°C (Ausubel *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., N.Y., at p. 2.10.3) and encodes a functionally equivalent expression product. Additionally contemplated are any nucleotide sequences that hybridize to the complement of a DNA sequence that encodes and expresses an amino acid sequence presented in the Sequence Listing under moderately stringent conditions, *e.g.*, washing in 0.2x SSC/0.1% SDS at 42°C (Ausubel *et al.*, 1989, *supra*), yet still encodes a functionally equivalent NHP product. Functional equivalents of a NHP include naturally occurring NHPs present in other species, and mutant NHPs, whether naturally occurring or engineered (by site directed mutagenesis, gene shuffling, directed evolution as described in, for example, U.S. Patent No. 5,837,458). The invention also includes degenerate nucleic acid variants of the disclosed NHP polynucleotide sequences.

Additionally contemplated are polynucleotides encoding NHP ORFs, or their functional equivalents, encoded by polynucleotide sequences that are about 99, 95, 90, or about 85 percent similar or identical to corresponding regions of the nucleotide sequences of the Sequence Listing (as measured by BLAST sequence comparison analysis using, for example, the GCG sequence analysis package, as described herein, using standard default settings).

The invention also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are

therefore the complements of, the described NHP nucleotide sequences. Such hybridization conditions may be highly stringent or less highly stringent, as described herein. In instances where the nucleic acid molecules are deoxyoligonucleotides ("DNA oligos"), such molecules are generally about 16 to about 100 bases long, or about 20 to about 80 bases long, or about 34 to about 45 bases long, or any variation or combination of sizes represented therein that incorporate a contiguous region of sequence first disclosed in the Sequence Listing. Such oligonucleotides can be used in conjunction with the polymerase chain reaction (PCR) to screen libraries, isolate clones, and prepare cloning and sequencing templates, *etc.*

Alternatively, such NHP oligonucleotides can be used as hybridization probes for screening libraries, and assessing gene expression patterns (particularly using a microarray or high-throughput "chip" format). Additionally, a series of NHP oligonucleotide sequences, or the complements thereof, can be used to represent all or a portion of the described NHP sequences. An oligonucleotide or polynucleotide sequence first disclosed in at least a portion of one or more of the sequences of SEQ ID NOS:1-13 can be used as a hybridization probe in conjunction with a solid support matrix/substrate (resins, beads, membranes, plastics, polymers, metal or metallized substrates, crystalline or polycrystalline substrates, *etc.*). Of particular note are spatially addressable arrays (*i.e.*, gene chips, microtiter plates, *etc.*) of oligonucleotides and polynucleotides, or corresponding oligopeptides and polypeptides, wherein at least one of the biopolymers present on the spatially addressable array comprises an oligonucleotide or polynucleotide sequence first disclosed in at least one of the sequences of SEQ ID NOS:1-13, or an amino acid sequence encoded thereby. Methods for attaching biopolymers to, or synthesizing biopolymers on, solid support matrices, and conducting binding studies thereon,

are disclosed in, *inter alia*, U.S. Patent Nos. 5,700,637, 5,556,752, 5,744,305, 4,631,211, 5,445,934, 5,252,743, 4,713,326, 5,424,186, and 4,689,405, the disclosures of which are herein incorporated by reference in their entirety.

Addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-13 can be used to identify and characterize the temporal and tissue specific expression of a gene. These addressable arrays incorporate oligonucleotide sequences of sufficient length to confer the required specificity, yet be within the limitations of the production technology. The length of these probes is usually within a range of between about 8 to about 2000 nucleotides. Preferably the probes consist of 60 nucleotides, and more preferably 25 nucleotides, from the sequences first disclosed in SEQ ID NOS:1-13.

For example, a series of NHP oligonucleotide sequences, or the complements thereof, can be used in chip format to represent all or a portion of the described sequences. The oligonucleotides, typically between about 16 to about 40 (or any whole number within the stated range) nucleotides in length, can partially overlap each other, and/or the sequence may be represented using oligonucleotides that do not overlap. Accordingly, the described polynucleotide sequences shall typically comprise at least about two or three distinct oligonucleotide sequences of at least about 8 nucleotides in length that are each first disclosed in the described Sequence Listing. Such oligonucleotide sequences can begin at any nucleotide present within a sequence in the Sequence Listing, and proceed in either a sense (5'-to-3') orientation vis-a-vis the described sequence or in an antisense orientation.

Microarray-based analysis allows the discovery of broad patterns of genetic activity, providing new understanding of gene functions, and generating novel and unexpected insight into transcriptional processes and

biological mechanisms. The use of addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-13 provides detailed information about transcriptional changes involved in a specific pathway, potentially leading to the identification of novel components, or gene functions that manifest themselves as novel phenotypes.

Probes consisting of sequences first disclosed in SEQ ID NOS:1-13 can also be used in the identification, selection, and validation of novel molecular targets for drug discovery. The use of these unique sequences permits the direct confirmation of drug targets, and recognition of drug dependent changes in gene expression that are modulated through pathways distinct from the intended target of the drug. These unique sequences therefore also have utility in defining and monitoring both drug action and toxicity.

As an example of utility, the sequences first disclosed in SEQ ID NOS:1-13 can be utilized in microarrays, or other assay formats, to screen collections of genetic material from patients who have a particular medical condition. These investigations can also be carried out using the sequences first disclosed in SEQ ID NOS:1-13 *in silico*, and by comparing previously collected genetic databases and the disclosed sequences using computer software known to those in the art.

Thus the sequences first disclosed in SEQ ID NOS:1-13 can be used to identify mutations associated with a particular disease, and also in diagnostic or prognostic assays.

Although the presently described sequences have been specifically described using nucleotide sequence, it should be appreciated that each of the sequences can uniquely be described using any of a wide variety of additional structural attributes, or combinations thereof. For example, a given sequence can be described by the net composition of the nucleotides present within a given

region of the sequence, in conjunction with the presence of one or more specific oligonucleotide sequence(s) first disclosed in SEQ ID NOS:1-13. Alternatively, a restriction map specifying the relative positions of restriction endonuclease digestion sites, or various palindromic or other specific oligonucleotide sequences, can be used to structurally describe a given sequence. Such restriction maps, which are typically generated by widely available computer programs (*e.g.*, the University of Wisconsin GCG sequence analysis package, SEQUENCHER 3.0, Gene Codes Corp., Ann Arbor, MI, *etc.*), can optionally be used in conjunction with one or more discrete nucleotide sequence(s) present in the sequence that can be described by the relative position of the sequence relative to one or more additional sequence(s) or one or more restriction sites present in the disclosed sequence.

For oligonucleotide probes, highly stringent conditions may refer, *e.g.*, to washing in 6x SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligos), 48°C (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base oligos). These nucleic acid molecules may encode or act as NHP antisense molecules, useful, for example, in NHP gene regulation and/or as antisense primers in amplification reactions of NHP nucleic acid sequences. With respect to NHP gene regulation, such techniques can be used to regulate biological functions. Further, such sequences may be used as part of ribozyme and/or triple helix sequences that are also useful for NHP gene regulation.

Inhibitory antisense or double stranded oligonucleotides can additionally comprise at least one modified base moiety that is selected from the group including, but not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine,

5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine.

The antisense oligonucleotide can also comprise at least one modified sugar moiety selected from the group including, but not limited to, arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense oligonucleotide will comprise at least one modified phosphate backbone selected from the group including, but not limited to, a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the antisense oligonucleotide is an α-anomeric oligonucleotide. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier *et al.*, 1987, Nucl. Acids Res. *15*:6625-6641). The oligonucleotide is a 2'-0-methylribonucleotide (Inoue *et al.*, 1987, Nucl. Acids Res. *15*:6131-6148), or a chimeric RNA-DNA analogue (Inoue *et al.*, 1987, FEBS Lett. *215*:327-330). Alternatively, double

stranded RNA can be used to disrupt the expression and function of a targeted NHP.

Oligonucleotides of the invention can be synthesized by standard methods known in the art, *e.g.*, by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, *etc.*). As examples, phosphorothioate oligonucleotides can be synthesized by the method of Stein *et al.* (1988, Nucl. Acids Res. *16*:3209), and methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin *et al.*, 1988, Proc. Natl. Acad. Sci. USA *85*:7448-7451), *etc.*

Low stringency conditions are well-known to those of skill in the art, and will vary predictably depending on the specific organisms from which the library and the labeled sequences are derived. For guidance regarding such conditions, see, for example, Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (and periodic updates thereof), and Ausubel *et al.*, 1989, *supra*.

Alternatively, suitably labeled NHP nucleotide probes can be used to screen a human genomic library using appropriately stringent conditions or by PCR. The identification and characterization of human genomic clones is helpful for identifying polymorphisms (including, but not limited to, nucleotide repeats, microsatellite alleles, single nucleotide polymorphisms, or coding single nucleotide polymorphisms), determining the genomic structure of a given locus/allele, and designing diagnostic tests. For example, sequences derived from regions adjacent to the intron/exon boundaries of the human gene can be used to design primers for use in amplification assays to detect mutations within the exons, introns, splice sites (*e.g.*, splice acceptor and/or donor sites), *etc.*, that can be used in diagnostics and pharmacogenomics.

For example, the present sequences can be used in restriction fragment length polymorphism (RFLP) analysis to identify specific individuals. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification (as generally described in U.S. Patent No. 5,272,057, incorporated herein by reference). In addition, the sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.*, another DNA sequence that is unique to a particular individual). Actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments.

Further, a NHP gene homolog can be isolated from nucleic acid from an organism of interest by performing PCR using two degenerate or "wobble" oligonucleotide primer pools designed on the basis of amino acid sequences within the NHP products disclosed herein. The template for the reaction may be genomic DNA, or total RNA, mRNA, and/or cDNA obtained by reverse transcription of mRNA prepared from human or non-human cell lines or tissue known to express, or suspected of expressing, an allele of a NHP gene.

The PCR product can be subcloned and sequenced to ensure that the amplified sequences represent the sequence of the desired NHP gene. The PCR fragment can then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment can be labeled and used to screen a cDNA library, such as a bacteriophage cDNA library. Alternatively, the labeled fragment can be used to isolate genomic clones via the screening of a genomic library.

PCR technology can also be used to isolate full length cDNA sequences. For example, RNA can be isolated, following standard procedures, from an appropriate cellular or tissue source (*i.e.*, one known to express, or suspected of expressing, a NHP gene). A reverse transcription (RT) reaction can be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" using a standard terminal transferase reaction, the hybrid may be digested with RNase H, and second strand synthesis may then be primed with a complementary primer. Thus, cDNA sequences upstream of the amplified fragment can be isolated. For a review of cloning strategies that can be used, see, *e.g.*, Sambrook *et al.*, 1989, *supra*.

A cDNA encoding a mutant NHP sequence can be isolated, for example, by using PCR. In this case, the first cDNA strand may be synthesized by hybridizing an oligo-dT oligonucleotide to mRNA isolated from tissue known to express, or suspected of expressing, a NHP, in an individual putatively carrying a mutant NHP allele, and by extending the new strand with reverse transcriptase. The second strand of the cDNA is then synthesized using an oligonucleotide that hybridizes specifically to the 5' end of the normal sequence. Using these two primers, the product is then amplified via PCR, optionally cloned into a suitable vector, and subjected to DNA sequence analysis through methods well-known to those of skill in the art. By comparing the DNA sequence of the mutant NHP allele to that of a corresponding normal NHP allele, the mutation(s) responsible for the loss or alteration of function of the mutant NHP gene product can be ascertained.

Alternatively, a genomic library can be constructed using DNA obtained from an individual suspected of carrying, or known to carry, a mutant NHP allele (*e.g.*, a person manifesting a NHP-associated phenotype such as, for

example, obesity, high blood pressure, connective tissue disorders, infertility, *etc.*), or a cDNA library can be constructed using RNA from a tissue known to express, or suspected of expressing, a mutant NHP allele. A normal NHP gene, or any suitable fragment thereof, can then be labeled and used as a probe to identify the corresponding mutant NHP allele in such libraries. Clones containing mutant NHP sequences can then be purified and subjected to sequence analysis according to methods well-known to those skilled in the art.

Additionally, an expression library can be constructed utilizing cDNA synthesized from, for example, RNA isolated from a tissue known to express, or suspected of expressing, a mutant NHP allele in an individual suspected of carrying, or known to carry, such a mutant allele. In this manner, gene products made by the putatively mutant tissue can be expressed and screened using standard antibody screening techniques in conjunction with antibodies raised against a normal NHP product, as described below (for screening techniques, see, for example, Harlow and Lane, eds., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor, N.Y.).

Additionally, screening can be accomplished by screening with labeled NHP fusion proteins, such as, for example, alkaline phosphatase-NHP or NHP-alkaline phosphatase fusion proteins. In cases where a NHP mutation results in an expression product with altered function (*e.g.*, as a result of a missense or a frameshift mutation), polyclonal antibodies to a NHP are likely to cross-react with a corresponding mutant NHP expression product. Library clones detected via their reaction with such labeled antibodies can be purified and subjected to sequence analysis according to methods well-known in the art.

The invention also encompasses: (a) DNA vectors that contain any of the foregoing NHP coding sequences and/or

their complements (*i.e.*, antisense); (b) DNA expression vectors that contain any of the foregoing NHP coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences (for example, baculovirus as described in U.S. Patent No. 5,869,336, herein incorporated by reference); (c) genetically engineered host cells that contain any of the foregoing NHP coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences in the host cell; and (d) genetically engineered host cells that express an endogenous NHP sequence under the control of an exogenously introduced regulatory element (*i.e.*, gene activation). As used herein, regulatory elements include, but are not limited to, inducible and non-inducible promoters, enhancers, operators, and other elements known to those skilled in the art that drive and regulate expression. Such regulatory elements include, but are not limited to, the cytomegalovirus (hCMV) immediate early gene, regulatable, viral elements (particularly retroviral LTR promoters), the early or late promoters of SV40 or adenovirus, the *lac* system, the *trp* system, the *TAC* system, the *TRC* system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase (PGK), the promoters of acid phosphatase, and the promoters of the yeast α-mating factors.

The present invention also encompasses antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists and agonists of a NHP, as well as compounds or nucleotide constructs that inhibit expression of a NHP sequence (transcription factor inhibitors, antisense and ribozyme molecules, or open reading frame sequence or regulatory sequence replacement constructs), or promote the expression of a NHP (*e.g.*, expression constructs in which NHP coding sequences are operatively associated with

expression control elements such as promoters, promoter/enhancers, *etc.*).

The NHPs or NHP peptides, NHP fusion proteins, NHP nucleotide sequences, antibodies, antagonists and agonists can be useful for the detection of mutant NHPs, or inappropriately expressed NHPs, for the diagnosis of disease. The NHP proteins or peptides, NHP fusion proteins, NHP nucleotide sequences, host cell expression systems, antibodies, antagonists, agonists and genetically engineered cells and animals can be used for screening for drugs (or high throughput screening of combinatorial libraries) effective in the treatment of the symptomatic or phenotypic manifestations of perturbing the normal function of a NHP in the body. The use of engineered host cells and/or animals may offer an advantage in that such systems allow not only for the identification of compounds that bind to the endogenous receptor for a NHP, but can also identify compounds that trigger NHP-mediated activities or pathways.

Finally, the NHP products can be used as therapeutics. For example, soluble derivatives such as NHP peptides/domains corresponding to NHPs, NHP fusion protein products (especially NHP-Ig fusion proteins, *i.e.*, fusions of a NHP, or a domain of a NHP, to an IgFc), NHP antibodies and anti-idiotypic antibodies (including Fab fragments), antagonists or agonists (including compounds that modulate or act on downstream targets in a NHP-mediated pathway) can be used to directly treat diseases or disorders. For instance, the administration of an effective amount of a soluble NHP, a NHP-IgFc fusion protein, or an anti-idiotypic antibody (or its Fab) that mimics the NHP, could activate or effectively antagonize an endogenous NHP receptor. Nucleotide constructs encoding such NHP products can be used to genetically engineer host cells to express such products *in vivo*; these genetically engineered cells function as "bioreactors" in the body delivering a

continuous supply of a NHP, a NHP peptide, or a NHP fusion protein to the body. Nucleotide constructs encoding functional NHPs, mutant NHPs, as well as antisense and ribozyme molecules can also be used in "gene therapy" approaches for the modulation of NHP expression. Thus, the invention also encompasses pharmaceutical formulations and methods for treating biological disorders.

Various aspects of the invention are described in greater detail in the subsections below.

5.1 THE NHP SEQUENCES

The cDNA sequences and the corresponding deduced amino acid sequences of the described NHPs are presented in the Sequence Listing. The NHP nucleotides were obtained from clustered human genomic sequences, and human cDNAs made from bone marrow and trachea mRNA (SEQ ID NOS:1-5), while SEQ ID NOS:6-13 were generated using cDNAs generated from human lymph node, thyroid, adrenal gland, uterus, and small intestine mRNAs (Edge Biosystems, Gaithersburg, MD, Clontech, Palo Alto, CA).

A number of polymorphisms were identified during the sequencing of the NHPs, including: a T/C polymorphism at the nucleotide position represented by, for example, position 462 of SEQ ID NO:1 (or position 468 of SEQ ID NO:2), both of which result in a leu at the region corresponding to amino acid (aa) position 154 of, for example, SEQ ID NO:3 (or position 156 of SEQ ID NO:4); a G/A polymorphism at the nucleotide position represented by, for example, position 123 of SEQ ID NO:6 (and the corresponding location in SEQ ID NOS:8. 10 and 12), both of which result in a val at the region corresponding to aa position 41 of, for example, SEQ ID NO:7 (and the corresponding location in SEQ ID NOS:9, 11 and 13); a G/A polymorphism at the nucleotide position represented by, for example, position 370 of SEQ ID NO:6 (and the corresponding location in SEQ ID NOS:8. 10 and 12), which can result in a

val or ile at the region corresponding to aa position 124 of, for example, SEQ ID NO:7 (and the corresponding location in SEQ ID NOS:9, 11 and 13); and a G/A polymorphism at the nucleotide position represented by, for example, position 454 of SEQ ID NO:6 (and the corresponding location in SEQ ID NOS:8. 10 and 12), which can result in a val or met at the region corresponding to aa position 152 of, for example, SEQ ID NO:7 (and the corresponding location in SEQ ID NOS:9, 11 and 13). As these polymorphisms are coding single nucleotide polymorphisms (SNPs), they are particularly useful in forensic analysis.

SEQ ID NOS:1-5 describe sequences that are similar to, *inter alia*, mammalian ABC transporter proteins, and are apparently encoded on human chromosome 7 (see GenBank Accession Number AC073424). SEQ ID NOS:6-13 describe sequences that are similar to, *inter alia*, mammalian sodium symporter proteins, and are apparently encoded on either human chromosome 1 or 4 (see GenBank Accession Numbers AL359959 and AC055887). Accordingly, the described sequences are useful for mapping and/or defining the corresponding coding regions of the human genome and identifying exon splice junctions.

An additional application of the described novel human polynucleotide sequences is their use in the molecular mutagenesis/evolution of proteins that are at least partially encoded by the described novel sequences using, for example, polynucleotide shuffling or related methodologies. Such approaches are described in U.S. Patent Nos. 5,830,721 and 5,837,458, which are herein incorporated by reference in their entirety.

NHP gene products can also be expressed in transgenic animals. Animals of any species, including, but not limited to, worms, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, birds, goats, and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees, may be used to generate NHP transgenic animals.

Any technique known in the art may be used to introduce a NHP transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Hoppe and Wagner, 1989, U.S. Patent No. 4,873,191); retrovirus-mediated gene transfer into germ lines (Van der Putten *et al.*, 1985, Proc. Natl. Acad. Sci. USA *82*:6148-6152); gene targeting in embryonic stem cells (Thompson *et al.*, 1989, Cell *56*:313-321); electroporation of embryos (Lo, 1983, Mol Cell. Biol. *3*:1803-1814); and sperm-mediated gene transfer (Lavitrano *et al.*, 1989, Cell *57*:717-723); *etc.* For a review of such techniques, see Gordon, 1989, Transgenic Animals, Intl. Rev. Cytol. *115*:171-229, which is incorporated by reference herein in its entirety.

The present invention provides for transgenic animals that carry a NHP transgene in all their cells, as well as animals that carry a transgene in some, but not all their cells, *i.e.*, mosaic animals or somatic cell transgenic animals. A transgene may be integrated as a single transgene, or in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. A transgene may also be selectively introduced into and activated in a particular cell-type by following, for example, the teaching of Lasko *et al.*, 1992, Proc. Natl. Acad. Sci. USA *89*:6232-6236. 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 a NHP transgene be integrated into the chromosomal site of the endogenous NHP gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous NHP 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 NHP gene (*i.e.*, "knockout" animals).

The transgene can also be selectively introduced into a particular cell-type, thus inactivating the endogenous NHP gene in only that cell-type, by following, for example, the teaching of Gu *et al.*, 1994, Science *265*:103-106. 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 NHP gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether 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 that include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR. Samples of NHP gene-expressing tissue may also be evaluated immunocytochemically using antibodies specific for the NHP transgene product.

The present invention also provides for "knock-in" animals. Knock-in animals are those in which a polynucleotide sequence (*i.e.*, a gene or a cDNA) that the animal does not naturally have in its genome is inserted in such a way that it is expressed. Examples include, but are not limited to, a human gene or cDNA used to replace its murine ortholog in the mouse, a murine cDNA used to replace the murine gene in the mouse, and a human gene or cDNA or murine cDNA that is tagged with a reporter construct used to replace the murine ortholog or gene in the mouse. Such replacements can occur at the locus of the murine ortholog or gene, or at another specific site. Such knock-in animals are useful for the *in vivo* study, testing and validation of, *intra alia*, human drug targets, as well as

for compounds that are directed at the same, and therapeutic proteins.

5.2 NHPS AND NHP POLYPEPTIDES

NHPs, NHP polypeptides, NHP peptide fragments, mutated, truncated, or deleted forms of the NHPs, and/or NHP fusion proteins can be prepared for a variety of uses. These uses include, but are not limited to, the generation of antibodies, as reagents in diagnostic assays, for the identification of other cellular gene products related to a NHP, and as reagents in assays for screening for compounds that can be used as pharmaceutical reagents useful in the therapeutic treatment of mental, biological, or medical disorders and diseases. Given the similarity information and expression data, the described NHPs can be targeted (by drugs, oligos, antibodies, *etc.*) in order to treat disease, or to therapeutically augment the efficacy of, for example, chemotherapeutic agents used in the treatment of breast or prostate cancer.

The Sequence Listing discloses the amino acid sequences encoded by the described NHP polynucleotides. The NHPs typically display initiator methionines in DNA sequence contexts consistent with a translation initiation site. SEQ ID NOS:3 and 4 display signal type sequences similar to those often found on membrane proteins; however, all of the described proteins display multiple transmembrane hydrophobic domains typical of membrane associated proteins.

The NHP amino acid sequences of the invention include the amino acid sequence presented in the Sequence Listing, as well as analogues and derivatives thereof. Further, corresponding NHP homologues from other species are encompassed by the invention. In fact, any NHP protein encoded by the NHP nucleotide sequences described herein are within the scope of the invention, as are any novel polynucleotide sequences encoding all or any novel portion

of an amino acid sequence presented in the Sequence Listing. The degenerate nature of the genetic code is well-known, and, accordingly, each amino acid presented in the Sequence Listing is generically representative of the well-known nucleic acid "triplet" codon, or in many cases codons, that can encode the amino acid. As such, as contemplated herein, the amino acid sequences presented in the Sequence Listing, when taken together with the genetic code (see, for example, Table 4-1 at page 109 of "Molecular Cell Biology", 1986, J. Darnell *et al.*, eds., Scientific American Books, New York, N.Y., herein incorporated by reference), are generically representative of all the various permutations and combinations of nucleic acid sequences that can encode such amino acid sequences.

The invention also encompasses proteins that are functionally equivalent to the NHPs encoded by the presently described nucleotide sequences, as judged by any of a number of criteria, including, but not limited to, the ability to bind and cleave a substrate of a NHP, the ability to effect an identical or complementary downstream pathway, or a change in cellular metabolism (*e.g.*, proteolytic activity, ion flux, tyrosine phosphorylation, *etc.*). Such functionally equivalent NHP proteins include, but are not limited to, additions or substitutions of amino acid residues within the amino acid sequence encoded by the NHP nucleotide sequences described herein, but that result in a silent change, thus producing a functionally equivalent expression product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino

acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

A variety of host-expression vector systems can be used to express the NHP nucleotide sequences of the invention. Where, as in the present instance, the NHP peptide or polypeptide is thought to be from a membrane protein, the hydrophobic regions of the protein can be excised, and the resulting soluble peptide or polypeptide can be recovered from the culture media. Such expression systems also encompass engineered host cells that express a NHP, or functional equivalent, *in situ*. Purification or enrichment of a NHP from such expression systems can be accomplished using appropriate detergents and lipid micelles and methods well-known to those skilled in the art. However, such engineered host cells themselves may be used in situations where it is important not only to retain the structural and functional characteristics of a NHP, but to assess biological activity, *e.g.*, in certain drug screening assays.

The expression systems that may be used for purposes of the invention include, but are not limited to, microorganisms such as bacteria (*e.g.*, *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing NHP nucleotide sequences; yeast (*e.g.*, *Saccharomyces*, *Pichia*) transformed with recombinant yeast expression vectors containing NHP nucleotide sequences; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing NHP nucleotide sequences; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing NHP nucleotide sequences; or mammalian cell systems (*e.g.*, COS, CHO, BHK, 293, 3T3) harboring recombinant expression

constructs containing NHP nucleotide sequences and promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the NHP product being expressed. For example, when a large quantity of such a protein is to be produced for the generation of pharmaceutical compositions of or containing a NHP, or for raising antibodies to a NHP, vectors that direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited to, the *E. coli* expression vector pUR278 (Ruther *et al.*, 1983, EMBO J. *2*:1791), in which a NHP coding sequence may be ligated individually into the vector in-frame with the *lacZ* coding region so that a fusion protein is produced; pIN vectors (Inouye and Inouye, 1985, Nucleic Acids Res. *13*:3101-3109; Van Heeke and Schuster, 1989, J. Biol. Chem. *264*:5503-5509); and the like. pGEX vectors (Pharmacia or American Type Culture Collection) can also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads, followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target expression product can be released from the GST moiety.

In an exemplary insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign polynucleotide sequences. The virus grows in *Spodoptera frugiperda* cells. A NHP coding sequence can be cloned individually into a non-essential region (for example the polyhedrin gene) of the virus and placed under

control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of a NHP coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted sequence is expressed (*e.g.*, see Smith *et al.*, 1983, J. Virol. *46*:584; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the NHP nucleotide sequence of interest may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric sequence may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing a NHP product in infected hosts (*e.g.*, see Logan and Shenk, 1984, Proc. Natl. Acad. Sci. USA *81*:3655-3659). Specific initiation signals may also be required for efficient translation of inserted NHP nucleotide sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire NHP gene or cDNA, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of a NHP coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, may be provided. Furthermore, the initiation codon should be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of

origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, *etc.* (see Bitter *et al.*, 1987, Methods in Enzymol. *153*:516-544).

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the expression product in the specific fashion desired. Such modifications (*e.g.*, glycosylation) and processing (*e.g.*, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and expression products. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for the desired processing of the primary transcript, glycosylation, and phosphorylation of the expression product may be used. Such mammalian host cells include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and in particular, human cell lines.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express the NHP sequences described herein can be engineered. Rather than using expression vectors that contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (*e.g.*, promoter, enhancer sequences, transcription terminators, polyadenylation sites, *etc.*), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection

and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci, which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines that express a NHP product. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of a NHP product.

A number of selection systems may be used, including, but not limited to, the herpes simplex virus thymidine kinase (Wigler *et al.*, 1977, Cell *11*:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska and Szybalski, 1962, Proc. Natl. Acad. Sci. USA *48*:2026), and adenine phosphoribosyltransferase (Lowy *et al.*, 1980, Cell *22*:817) genes, which can be employed in tk^-, $hgprt^-$ or $aprt^-$ cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler *et al.*, 1980, Proc. Natl. Acad. Sci. USA 77:3567; O'Hare *et al.*, 1981, Proc. Natl. Acad. Sci. USA *78*:1527); gpt, which confers resistance to mycophenolic acid (Mulligan and Berg, 1981, Proc. Natl. Acad. Sci. USA *78*:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin *et al.*, 1981, J. Mol. Biol. *150*:1); and hygro, which confers resistance to hygromycin (Santerre *et al.*, 1984, Gene *30*:147).

Alternatively, any fusion protein can be readily purified by utilizing an antibody specific for the fusion protein being expressed. Another exemplary system allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht *et al.*, 1991, Proc. Natl. Acad. Sci. USA *88*:8972-8976). In this system, the sequence of interest is subcloned into a vaccinia recombination plasmid such that the sequence's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells

infected with recombinant vaccinia virus are loaded onto Ni^{2+}·nitriloacetic acid-agarose columns, and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

Also encompassed by the present invention are fusion proteins that direct a NHP to a target organ and/or facilitate transport across the membrane into the cytosol. Conjugation of NHPs to antibody molecules or their Fab fragments could be used to target cells bearing a particular epitope. Attaching an appropriate signal sequence to a NHP would also transport a NHP to a desired location within the cell. Alternatively targeting of a NHP or its nucleic acid sequence might be achieved using liposome or lipid complex based delivery systems. Such technologies are described in "Liposomes: A Practical Approach", New, R.R.C., ed., Oxford University Press, N.Y., and in U.S. Patent Nos. 4,594,595, 5,459,127, 5,948,767 and 6,110,490 and their respective disclosures, which are herein incorporated by reference in their entirety. Additionally embodied are novel protein constructs engineered in such a way that they facilitate transport of NHPs to a target site or desired organ, where they cross the cell membrane and/or the nucleus where the NHPs can exert their functional activity. This goal may be achieved by coupling of a NHP to a cytokine or other ligand that provides targeting specificity, and/or to a protein transducing domain (see generally U.S. Provisional Patent Application Ser. Nos. 60/111,701 and 60/056,713, both of which are herein incorporated by reference, for examples of such transducing sequences), to facilitate passage across cellular membranes, and can optionally be engineered to include nuclear localization signals.

Additionally contemplated are oligopeptides that are modeled on an amino acid sequence first described in the Sequence Listing. Such NHP oligopeptides are generally between about 10 to about 100 amino acids long, or between

about 16 to about 80 amino acids long, or between about 20 to about 35 amino acids long, or any variation or combination of sizes represented therein that incorporate a contiguous region of sequence first disclosed in the Sequence Listing. Such NHP oligopeptides can be of any length disclosed within the above ranges, and can initiate at any amino acid position represented in the Sequence Listing.

The invention also contemplates "substantially isolated" or "substantially pure" proteins or polypeptides. By a "substantially isolated" or "substantially pure" protein or polypeptide is meant a protein or polypeptide that has been separated from at least some of those components that naturally accompany it. Typically, the protein or polypeptide is substantially isolated or pure when it is at least 60%, by weight, free from the proteins and other naturally-occurring organic molecules with which it is naturally associated *in vivo*. Preferably, the purity of the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight. A substantially isolated or pure protein or polypeptide may be obtained, for example, by extraction from a natural source, by expression of a recombinant nucleic acid encoding the protein or polypeptide, or by chemically synthesizing the protein or polypeptide.

Purity can be measured by any appropriate method, *e.g.*, column chromatography such as immunoaffinity chromatography using an antibody specific for the protein or polypeptide, polyacrylamide gel electrophoresis, or HPLC analysis. A protein or polypeptide is substantially free of naturally associated components when it is separated from at least some of those contaminants that accompany it in its natural state. Thus, a polypeptide that is chemically synthesized or produced in a cellular system different from the cell from which it naturally originates will be, by definition, substantially free from its

naturally associated components. Accordingly, substantially isolated or pure proteins or polypeptides include eukaryotic proteins synthesized in *E. coli*, other prokaryotes, or any other organism in which they do not naturally occur.

5.3 ANTIBODIES TO NHP PRODUCTS

Antibodies that specifically recognize one or more epitopes of a NHP, epitopes of conserved variants of a NHP, or peptide fragments of a NHP, are also encompassed by the invention. Such antibodies include, but are not limited to, polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above.

The antibodies of the invention may be used, for example, in the detection of a NHP in a biological sample and may, therefore, be utilized as part of a diagnostic or prognostic technique whereby patients may be tested for abnormal amounts of a NHP. Such antibodies may also be utilized in conjunction with, for example, compound screening schemes for the evaluation of the effect of test compounds on expression and/or activity of a NHP expression product. Additionally, such antibodies can be used in conjunction with gene therapy to, for example, evaluate normal and/or engineered NHP-expressing cells prior to their introduction into a patient. Such antibodies may additionally be used in methods for the inhibition of abnormal NHP activity. Thus, such antibodies may be utilized as a part of treatment methods.

For the production of antibodies, various host animals may be immunized by injection with a NHP, a NHP peptide (*e.g.*, one corresponding to a functional domain of a NHP), a truncated NHP polypeptide (a NHP in which one or more

domains have been deleted), functional equivalents of a NHP or mutated variants of a NHP. Such host animals may include, but are not limited to, pigs, rabbits, mice, goats, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including, but not limited to, Freund's adjuvant (complete and incomplete), mineral salts such as aluminum hydroxide or aluminum phosphate, chitosan, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Alternatively, the immune response could be enhanced by combination and/or coupling with molecules such as keyhole limpet hemocyanin, tetanus toxoid, diphtheria toxoid, ovalbumin, cholera toxin, or fragments thereof. Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of the immunized animals.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, can be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler and Milstein, (1975, Nature *256*:495-497; and U.S. Patent No. 4,376,110), the human B-cell hybridoma technique (Kosbor *et al.*, 1983, Immunology Today *4*:72; Cole *et al.*, 1983, Proc. Natl. Acad. Sci. USA *80*:2026-2030), and the EBV-hybridoma technique (Cole *et al.*, 1985, Monoclonal Antibodies And Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class, including IgG, IgM, IgE, IgA, and IgD, and any subclass thereof. The hybridomas producing the mAbs of this invention may be cultivated *in vitro* or *in vivo*. Production of high titers of mAbs *in vivo* makes this the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison *et al.*, 1984, Proc. Natl. Acad. Sci. USA *81*:6851-6855; Neuberger *et al.*, 1984, Nature, *312*:604-608; Takeda *et al.*, 1985, Nature, *314*:452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. Such technologies are described in U.S. Patent Nos. 6,114,598, 6,075,181 and 5,877,397 and their respective disclosures, which are herein incorporated by reference in their entirety. Also encompassed by the present invention is the use of fully humanized monoclonal antibodies, as described in U.S. Patent No. 6,150,584 and respective disclosures, which are herein incorporated by reference in their entirety.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778; Bird, 1988, Science *242*:423-426; Huston *et al.*, 1988, Proc. Natl. Acad. Sci. USA *85*:5879-5883; and Ward *et al.*, 1989, Nature *341*:544-546) can be adapted to produce single chain antibodies against NHP expression products. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments that recognize specific epitopes may be generated by known techniques. For example, such fragments include, but are not limited to: $F(ab')_2$ fragments, which can be produced by pepsin digestion of an antibody molecule; and Fab fragments, which can be generated by reducing the disulfide bridges of $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed (Huse *et al.*, 1989, Science, *246*:1275-1281) to

allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Antibodies to a NHP can, in turn, be utilized to generate anti-idiotype antibodies that "mimic" a given NHP, using techniques well-known to those skilled in the art (see, *e.g.*, Greenspan and Bona, 1993, FASEB J. 7:437-444; and Nissinoff, 1991, J. Immunol. *147*:2429-2438). For example, antibodies that bind to a NHP domain and competitively inhibit the binding of a NHP to its cognate receptor can be used to generate anti-idiotypes that "mimic" the NHP and, therefore, bind and activate or neutralize a receptor. Such anti-idiotypic antibodies, or Fab fragments of such anti-idiotypes, can be used in therapeutic regimens involving a NHP-mediated pathway.

Additionally, given the high degree of relatedness of mammalian NHPs, the presently described knock-out mice (having never seen a NHP, and thus never been tolerized to a NHP) have an unique utility, as they can be advantageously applied to the generation of antibodies against the disclosed mammalian NHPs (*i.e.*, a NHP will be immunogenic in NHP knock-out animals).

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended as single illustrations of individual aspects of the invention, and functionally equivalent methods and components are within the scope of the invention. Indeed, various modifications of the invention, in addition to those shown and described herein, will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims. All cited publications, patents, and patent applications are herein incorporated by reference in their entirety.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:2.

2. An isolated nucleic acid molecule comprising a nucleotide sequence that:

(a) encodes the amino acid sequence shown in SEQ ID NO:3 or SEQ ID NO:4; and

(b) hybridizes under highly stringent conditions to the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:2, or the complement thereof.

3. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes the amino acid sequence shown in SEQ ID NO:3 or SEQ ID NO:4.

4. A recombinant expression vector comprising the isolated nucleic acid molecule of claim 1.

5. A host cell comprising the recombinant expression vector of claim 4.

6. A substantially isolated protein having the activity of the protein shown in SEQ ID NOS:3 or 4, which is encoded by a nucleotide sequence that hybridizes to the complement of SEQ ID NO:1 or SEQ ID NO:2 under highly stringent conditions.

7. An isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10 or SEQ ID NO:12.

8. An isolated nucleic acid molecule comprising a nucleotide sequence that:

(c) encodes the amino acid sequence of SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13; and

(d) hybridizes under highly stringent conditions to the nucleotide sequence of SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10 or SEQ ID NO:12, or the complement thereof.

9. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes the amino acid sequence shown in SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13.

10. A recombinant expression vector comprising the isolated nucleic acid molecule of claim 7.

11. A host cell comprising the recombinant expression vector of claim 10.

12. The host cell of Claim 11, wherein said cell is procaryotic.

13. The host cell of Claim 11, wherein said cell is eucaryotic.

14. A substantially isolated protein having the activity of the protein shown in SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13, which is encoded by a nucleotide sequence that hybridizes to the complement of SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10 or SEQ ID NO:12 under highly stringent conditions.

SEQUENCE LISTING

<110> LEXICON GENETICS INCORPORATED

<120> Novel Human Transporter Proteins and
Polynucleotides Encoding the Same

<130> LEX-0358-PCT

<150> US 60/298,241
<151> 2001-06-14

<160> 13

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 3618
<212> DNA
<213> homo sapiens

<400> 1
atgggttgca cattttacc cttttatgtc attgtatata ttttttgct aagtgttgtt 60
gagatttgtg aagttttcca gcagactgtg aagccctcag aagccatgga gatgctgcag 120
aaagtgaaga tgatggtcgt acgtgtgctc accatcgttg cagaaaaccc ttcctggacc 180
aaggacattt tgtgtgctac tctgagttgc aagcaaaatg ggataaggca tctcatttta 240
tctgctatac aaggggtcac tttggcgcag gaccacttcc aggaaattga aaagatatgg 300
tcctcgccga atcagctaaa ttgtgaaagt cttagcaaga atctttctag caccttggag 360
agcttcaaga gcagcttgga aaatgccact ggccaggact gcacaagcca gccgaggctg 420
gagacggtgc agcagcactt gtacatgttg gccaaaagcc tygaggaaac ttggtcatca 480
gggaatccca tcatgacttt tctcagcaat ttcacagtaa ctgaggatgt aaaaataaaa 540
gatttgatga agaatatcac caagttgact gaggagcttc gctcttccat ccaaatctcg 600
aatgagacta tccatagcat tctagaagca aatatttccc actccaaggt tctcttcagt 660
gccctcaccg tagctctgtc tggaaagtgt gatcaggaaa tccttcatct cctgctgaca 720
tttcccaaag gggaaaaatc ttggatcgca gcggaggaac tctgtagcct gccagggtca 780
aaagtgtatt ctctgattgt gttgctgagt cgaaacttgg atgtgcgagc tttcatttac 840
aagactctga tgccttctga agcaaatggc ttgctcaact ccttgctgga tatagtttcc 900
agcctcagcg ccttgcttgc caaagcccag cacgtctttg agtatcttcc tgagtttctt 960
cacacattta aaatcactgc cttgctagaa accctggact ttcaacaggt ttcacaaaat 1020
gtccaggcca gaagttcagc ttttggttct ttccagtttg tgatgaagat ggtttgcaag 1080
gaccaagcat cattccttag cgattctaat atgtttatta atttgcccag agttaaggaa 1140
ctcttggaag atgacaaaga aaaattcaac attcctgaag attcaacacc gttttgcttg 1200
aagctttatc aggaaattct acaattgcca aatggtgctt tggtgtggac cttcctaaaa 1260
cccatattgc atggaaaaat actatacaca ccaaacactc cagaaattaa caaggtcatt 1320
caaaaggcta attacacctt ttatattgtg gacaaactaa aaactttatc agaaacactg 1380
ctggaaatgt ccagcctttt ccagagaagt ggaagtggcc agatgttcaa ccagctgcag 1440
gaggccctga gaaacaaatt tgtaagaaac tttgtagaaa accagttgca cattgatgta 1500
gacaaactta ctgaaaaact ccagacatac ggagggctgc tggatgagat gtttaaccat 1560
gcaggcgctg gacgcttccg tttcttgggc agcatcttgg tcaatctctc ttcctgcgtg 1620
gcactgaacc gtttccaggc tctgcagtct gtcgacatcc tggagactaa agcacatgaa 1680
ctcttgcagc agaacagctt cttggccagt atcattttca gcaattcctt attcgacaag 1740
aacttcagat cagagtctgt caaactgcca ccccatgtct catacacaat ccggaccaat 1800
gtgttataca gcgtgcgaac agatgtggta aaaaaccctt cttggaagtt ccaccctcag 1860
aatctaccag ctgatgggtt caaatataac tacgtctttg ccccactgca agacatgatc 1920
gaaagagcca tcattttggt gcagactggg caggaagccc tggaaccagc agcacagact 1980
caggcggccc cttacccctg ccataccagc gacctattcc tgaacaacgt tggtttcttt 2040
tttccactga taatgatgct gacgtggatg gtgtctgtgg ccagcatggt cagaaagttg 2100
gtgtatgagc aggagataca gatagaagag tatatgcgga tgatgggagt gcatccagtg 2160
atccatttcc tggcctggtt cctggagaac atggctgtgt tgaccataag cagtgctact 2220
ctggccatcg ttctgaaaac aagtggcatc tttgcacaca gcaatacctt tattgttttc 2280
ctctttctct tggattttgg gatgtcagtc gtcatgctga gctacctctt gagtgcattt 2340
ttcagccaag ctaatacagc ggccctttgt accagcctgg tgtacatgat cagctttctg 2400

ccctacatag ttctattggt tctacataac caattaagtt ttgttaatca gacatttctg 2460
tgccttcttt cgacaaccgc ctttggacaa ggggtatttt ttattacatt cctggaagga 2520
caagagacag ggattcaatg gaataatatg taccaggctc tggaacaagg gggcatgaca 2580
tttggctggg tttgctggat gattcttttt gattcaagcc tttatttttt gtgtggatgg 2640
tacttgagca acttgattcc tggaacattt ggtttacgga aaccatggta tttccccttt 2700
actgcctcat attggaagag tgtgggtttc ttggtggaga aaaggcaata ctttctaagt 2760
tctagtctgt tcttcttcaa tgagaacttt gacaataaag ggtcatcact gcaaaacagg 2820
gaaggagagc ttgaaggaag tgccccggga gtcaccctgg tgtctgtgac caaggaatat 2880
gagggccaca aggctgtggt ccaagacctc agcctgacct tctacagaga ccaaatcacc 2940
gccctgctgg ggacaaacgg tgccgggaaa accactatca tatccatgtt gacggggctc 3000
caccctccca cttctggaac catcatcatc aatggcaaga acctacagac agacctgtcg 3060
agggtcagaa tggagcttgg tgtgtgtccg cagcaggaca tcctgttgga caacctcacc 3120
gtccgggaac atttgctgct ctttgcttcc ataaaggcgc ctcagtggac caagaaggag 3180
ctgcatcagc aagtcaatca aactcttcag gatgtggact taactcagca tcagcacaaa 3240
cagacccgag ctctgtctgg aggcctgaag aggaagctct cccttggcat tgctttcatg 3300
ggcatgtcga ggaccgtggt tctggatgag cccaccagtg gggtggaccc ttgctcccgg 3360
catagcctgt gggacattct gctcaagtac cgagaaggta ggcactgggc ctcattctgc 3420
cttctcttcc cacaatattg tgttgcagga aatgcattgc tactgtacag tagaatcaag 3480
ttgtatccca gtgaggctac attatccttt tcagaaaaat ataaattttt aaaagcactt 3540
atagggatat attcgttaga taacatctct atagtgctta gaattgctta ctttgtgttt 3600
gacctttaa ctcaataa 3618

<210> 2
<211> 3624
<212> DNA
<213> homo sapiens

<400> 2
atgcacatgg gttgcacatt tttacccttt tatgtcattg tatatatttt tttgctaagt 60
gttgttgaga tttgtgaagt tttccagcag actgtgaagc cctcagaagc catggagatg 120
ctgcagaaag tgaagatgat ggtcgtacgt gtgctcacca tcgttgcaga aaacccttcc 180
tggaccaagg acattttgtg tgctactctg agttgcaagc aaaatgggat aaggcatctc 240
attttatctg ctatacaagg ggtcactttg gcgcaggacc acttccagga aattgaaaag 300
atatggtcct cgccgaatca gctaaattgt gaaagtctta gcaagaatct ttctagcacc 360
ttggagagct tcaagagcag cttggaaaat gccactggcc aggactgcac aagccagccg 420
aggctggaga cggtgcagca gcacttgtac atgttggcca aaagcctyga ggaaacttgg 480
tcatcaggga atcccatcat gacttttctc agcaatttca cagtaactga ggatgtaaaa 540
ataaaagatt tgatgaagaa tatcaccaag ttgactgagg agcttcgctc ttccatccaa 600
atctcgaatg agactatcca tagcattcta gaagcaaata tttcccactc caaggttctc 660
ttcagtgccc tcaccgtagc tctgtctgga aagtgtgatc aggaaatcct tcatctcctg 720
ctgacatttc ccaaagggga aaaatcttgg atcgcagcgg aggaactctg tagcctgcca 780
gggtcaaaag tgtattctct gattgtgttg ctgagtcgaa acttggatgt gcgagctttc 840
atttacaaga ctctgatgcc ttctgaagca aatggcttgc tcaactcctt gctggatata 900
gtttccagcc tcagcgcctt gcttgccaaa gcccagcacg tctttgagta tcttcctgag 960
tttcttcaca catttaaaat cactgccttg ctagaaaccc tggactttca acaggtttca 1020
caaaatgtcc aggccagaag ttcagctttt ggttctttcc agtttgtgat gaagatggtt 1080
tgcaaggacc aagcatcatt ccttagcgat tctaatatgt ttattaattt gcccagagtt 1140
aaggaactct tggaagatga caaagaaaaa ttcaacattc ctgaagattc aacaccgttt 1200
tgcttgaagc tttatcagga aattctacaa ttgccaaatg gtgctttggt gtggaccttc 1260
ctaaaaccca tattgcatgg aaaaatacta tacacaccaa acactccaga aattaacaag 1320
gtcattcaaa aggctaatta caccttttat attgtggaca aactaaaaac tttatcagaa 1380
acactgctgg aaatgtccag ccttttccag agaagtggaa gtggccagat gttcaaccag 1440
ctgcaggagg ccctgagaaa caaatttgta agaaactttg tagaaaacca gttgcacatt 1500
gatgtagaca aacttactga aaaactccag acatacggag ggctgctgga tgagatgttt 1560
aaccatgcag gcgctggacg cttccgtttc ttgggcagca tcttggtcaa tctctcttcc 1620
tgcgtggcac tgaaccgttt ccaggctctg cagtctgtcg acatcctgga gactaaagca 1680
catgaactct tgcagcagaa cagcttcttg gccagtatca ttttcagcaa ttccttattc 1740
gacaagaact tcagatcaga gtctgtcaaa ctgccacccc atgtctcata cacaatccgg 1800
accaatgtgt tatacagcgt gcgaacagat gtggtaaaaa acccttcttg gaagttccac 1860
cctcagaatc taccagctga tgggttcaaa tataactacg tctttgcccc actgcaagac 1920
atgatcgaaa gagccatcat tttggtgcag actgggcagg aagccctgga accagcagca 1980
cagactcagg cggcccctta cccctgccat accagcgacc tattcctgaa caacgttggt 2040
ttctttttttc cactgataat gatgctgacg tggatggtgt ctgtggccag catggtcaga 2100

```
aagttggtgt atgagcagga gatacagata gaagagtata tgcggatgat gggagtgcat 2160
ccagtgatcc atttcctggc ctggttcctg gagaacatgg ctgtgttgac cataagcagt 2220
gctactctgg ccatcgttct gaaaacaagt ggcatctttg cacacagcaa tacctttatt 2280
gttttcctct ttctcttgga ttttgggatg tcagtcgtca tgctgagcta cctcttgagt 2340
gcatttttca gccaagctaa tacagcggcc ctttgtacca gcctggtgta catgatcagc 2400
tttctgccct acatagttct attggttcta cataaccaat taagttttgt taatcagaca 2460
tttctgtgcc ttctttcgac aaccgccttt ggacaagggg tatttttttat tacattcctg 2520
gaaggacaag agacagggat tcaatggaat aatatgtacc aggctctgga acaaggggggc 2580
atgacatttg gctgggtttg ctggatgatt ctttttgatt caagccttta ttttttgtgt 2640
ggatggtact tgagcaactt gattcctgga acatttggtt tacggaaacc atggtatttc 2700
ccctttactg cctcatattg gaagagtgtg ggtttcttgg tggagaaaag gcaatacttt 2760
ctaagttcta gtctgttctt cttcaatgag aactttgaca ataaagggtc atcactgcaa 2820
aacagggaag gagagcttga aggaagtgcc ccgggagtca ccctggtgtc tgtgaccaag 2880
gaatatgagg gccacaaggc tgtggtccaa gacctcagcc tgaccttcta cagagaccaa 2940
atcaccgccc tgctggggac aaacggtgcc gggaaaacca ctatcatatc catgttgacg 3000
gggctccacc ctcccacttc tggaaccatc atcatcaatg gcaagaacct acagacagac 3060
ctgtcgaggg tcagaatgga gcttggtgtg tgtccgcagc aggacatcct gttggacaac 3120
ctcaccgtcc gggaacattt gctgctcttt gcttccataa aggcgcctca gtggaccaag 3180
aaggagctgc atcagcaagt caatcaaact cttcaggatg tggacttaac tcagcatcag 3240
cacaaacaga cccgagctct gtctggaggc ctgaagagga agctctccct tggcattgct 3300
ttcatgggca tgtcgaggac cgtggttctg gatgagccca ccagtggggt ggacccttgc 3360
tcccggcata gcctgtggga cattctgctc aagtaccgag aaggtaggca ctgggcctca 3420
ttctgccttc tcttcccaca atattgtgtt gcaggaaatg cattgctact gtacagtaga 3480
atcaagttgt atcccagtga ggctacatta tccttttcag aaaaatataa atttttaaaa 3540
gcacttatag ggatatattc gttagataac atctctatag tgcttagaat tgcttacttt 3600
gtgtttgacc ttttaactca ataa                                       3624
```

<210> 3
<211> 1205
<212> PRT
<213> homo sapiens

<400> 3

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

```
Phe Pro Lys Gly Glu Lys Ser Trp Ile Ala Ala Glu Glu Leu Cys Ser
                245                 250                 255
Leu Pro Gly Ser Lys Val Tyr Ser Leu Ile Val Leu Leu Ser Arg Asn
            260                 265                 270
Leu Asp Val Arg Ala Phe Ile Tyr Lys Thr Leu Met Pro Ser Glu Ala
        275                 280                 285
Asn Gly Leu Leu Asn Ser Leu Leu Asp Ile Val Ser Ser Leu Ser Ala
    290                 295                 300
Leu Leu Ala Lys Ala Gln His Val Phe Glu Tyr Leu Pro Glu Phe Leu
305                 310                 315                 320
His Thr Phe Lys Ile Thr Ala Leu Leu Glu Thr Leu Asp Phe Gln Gln
                325                 330                 335
Val Ser Gln Asn Val Gln Ala Arg Ser Ser Ala Phe Gly Ser Phe Gln
            340                 345                 350
Phe Val Met Lys Met Val Cys Lys Asp Gln Ala Ser Phe Leu Ser Asp
        355                 360                 365
Ser Asn Met Phe Ile Asn Leu Pro Arg Val Lys Glu Leu Leu Glu Asp
    370                 375                 380
Asp Lys Glu Lys Phe Asn Ile Pro Glu Asp Ser Thr Pro Phe Cys Leu
385                 390                 395                 400
Lys Leu Tyr Gln Glu Ile Leu Gln Leu Pro Asn Gly Ala Leu Val Trp
                405                 410                 415
Thr Phe Leu Lys Pro Ile Leu His Gly Lys Ile Leu Tyr Thr Pro Asn
            420                 425                 430
Thr Pro Glu Ile Asn Lys Val Ile Gln Lys Ala Asn Tyr Thr Phe Tyr
        435                 440                 445
Ile Val Asp Lys Leu Lys Thr Leu Ser Glu Thr Leu Leu Glu Met Ser
    450                 455                 460
Ser Leu Phe Gln Arg Ser Gly Ser Gly Gln Met Phe Asn Gln Leu Gln
465                 470                 475                 480
Glu Ala Leu Arg Asn Lys Phe Val Arg Asn Phe Val Glu Asn Gln Leu
                485                 490                 495
His Ile Asp Val Asp Lys Leu Thr Glu Lys Leu Gln Thr Tyr Gly Gly
            500                 505                 510
Leu Leu Asp Glu Met Phe Asn His Ala Gly Ala Gly Arg Phe Arg Phe
        515                 520                 525
Leu Gly Ser Ile Leu Val Asn Leu Ser Ser Cys Val Ala Leu Asn Arg
    530                 535                 540
Phe Gln Ala Leu Gln Ser Val Asp Ile Leu Glu Thr Lys Ala His Glu
545                 550                 555                 560
Leu Leu Gln Gln Asn Ser Phe Leu Ala Ser Ile Ile Phe Ser Asn Ser
                565                 570                 575
Leu Phe Asp Lys Asn Phe Arg Ser Glu Ser Val Lys Leu Pro Pro His
            580                 585                 590
Val Ser Tyr Thr Ile Arg Thr Asn Val Leu Tyr Ser Val Arg Thr Asp
        595                 600                 605
Val Val Lys Asn Pro Ser Trp Lys Phe His Pro Gln Asn Leu Pro Ala
    610                 615                 620
Asp Gly Phe Lys Tyr Asn Tyr Val Phe Ala Pro Leu Gln Asp Met Ile
625                 630                 635                 640
Glu Arg Ala Ile Ile Leu Val Gln Thr Gly Gln Glu Ala Leu Glu Pro
                645                 650                 655
Ala Ala Gln Thr Gln Ala Ala Pro Tyr Pro Cys His Thr Ser Asp Leu
            660                 665                 670
Phe Leu Asn Asn Val Gly Phe Phe Phe Pro Leu Ile Met Met Leu Thr
        675                 680                 685
Trp Met Val Ser Val Ala Ser Met Val Arg Lys Leu Val Tyr Glu Gln
    690                 695                 700
Glu Ile Gln Ile Glu Glu Tyr Met Arg Met Met Gly Val His Pro Val
705                 710                 715                 720
Ile His Phe Leu Ala Trp Phe Leu Glu Asn Met Ala Val Leu Thr Ile
                725                 730                 735
Ser Ser Ala Thr Leu Ala Ile Val Leu Lys Thr Ser Gly Ile Phe Ala
```

```
            740                 745                 750
His Ser Asn Thr Phe Ile Val Phe Leu Phe Leu Leu Asp Phe Gly Met
        755                 760                 765
Ser Val Val Met Leu Ser Tyr Leu Leu Ser Ala Phe Phe Ser Gln Ala
    770                 775                 780
Asn Thr Ala Ala Leu Cys Thr Ser Leu Val Tyr Met Ile Ser Phe Leu
785                 790                 795                 800
Pro Tyr Ile Val Leu Leu Val Leu His Asn Gln Leu Ser Phe Val Asn
                805                 810                 815
Gln Thr Phe Leu Cys Leu Leu Ser Thr Thr Ala Phe Gly Gln Gly Val
            820                 825                 830
Phe Phe Ile Thr Phe Leu Glu Gly Gln Glu Thr Gly Ile Gln Trp Asn
        835                 840                 845
Asn Met Tyr Gln Ala Leu Glu Gln Gly Gly Met Thr Phe Gly Trp Val
    850                 855                 860
Cys Trp Met Ile Leu Phe Asp Ser Ser Leu Tyr Phe Leu Cys Gly Trp
865                 870                 875                 880
Tyr Leu Ser Asn Leu Ile Pro Gly Thr Phe Gly Leu Arg Lys Pro Trp
                885                 890                 895
Tyr Phe Pro Phe Thr Ala Ser Tyr Trp Lys Ser Val Gly Phe Leu Val
            900                 905                 910
Glu Lys Arg Gln Tyr Phe Leu Ser Ser Ser Leu Phe Phe Phe Asn Glu
        915                 920                 925
Asn Phe Asp Asn Lys Gly Ser Ser Leu Gln Asn Arg Glu Gly Glu Leu
    930                 935                 940
Glu Gly Ser Ala Pro Gly Val Thr Leu Val Ser Val Thr Lys Glu Tyr
945                 950                 955                 960
Glu Gly His Lys Ala Val Val Gln Asp Leu Ser Leu Thr Phe Tyr Arg
                965                 970                 975
Asp Gln Ile Thr Ala Leu Leu Gly Thr Asn Gly Ala Gly Lys Thr Thr
            980                 985                 990
Ile Ile Ser Met Leu Thr Gly Leu His Pro Pro Thr Ser Gly Thr Ile
        995                 1000                1005
Ile Ile Asn Gly Lys Asn Leu Gln Thr Asp Leu Ser Arg Val Arg Met
    1010                1015                1020
Glu Leu Gly Val Cys Pro Gln Gln Asp Ile Leu Leu Asp Asn Leu Thr
1025                1030                1035                1040
Val Arg Glu His Leu Leu Leu Phe Ala Ser Ile Lys Ala Pro Gln Trp
                1045                1050                1055
Thr Lys Lys Glu Leu His Gln Gln Val Asn Gln Thr Leu Gln Asp Val
            1060                1065                1070
Asp Leu Thr Gln His Gln His Lys Gln Thr Arg Ala Leu Ser Gly Gly
        1075                1080                1085
Leu Lys Arg Lys Leu Ser Leu Gly Ile Ala Phe Met Gly Met Ser Arg
    1090                1095                1100
Thr Val Val Leu Asp Glu Pro Thr Ser Gly Val Asp Pro Cys Ser Arg
1105                1110                1115                1120
His Ser Leu Trp Asp Ile Leu Leu Lys Tyr Arg Glu Gly Arg His Trp
                1125                1130                1135
Ala Ser Phe Cys Leu Leu Phe Pro Gln Tyr Cys Val Ala Gly Asn Ala
            1140                1145                1150
Leu Leu Leu Tyr Ser Arg Ile Lys Leu Tyr Pro Ser Glu Ala Thr Leu
        1155                1160                1165
Ser Phe Ser Glu Lys Tyr Lys Phe Leu Lys Ala Leu Ile Gly Ile Tyr
    1170                1175                1180
Ser Leu Asp Asn Ile Ser Ile Val Leu Arg Ile Ala Tyr Phe Val Phe
1185                1190                1195                1200
Asp Leu Leu Thr Gln
                1205
```

<210> 4
<211> 1207

<212> PRT
<213> homo sapiens

<400> 4

```
Met His Met Gly Cys Thr Phe Leu Pro Phe Tyr Val Ile Val Tyr Ile
 1               5                  10                  15
Phe Leu Leu Ser Val Val Glu Ile Cys Glu Val Phe Gln Gln Thr Val
            20                  25                  30
Lys Pro Ser Glu Ala Met Glu Met Leu Gln Lys Val Lys Met Met Val
        35                  40                  45
Val Arg Val Leu Thr Ile Val Ala Glu Asn Pro Ser Trp Thr Lys Asp
    50                  55                  60
Ile Leu Cys Ala Thr Leu Ser Cys Lys Gln Asn Gly Ile Arg His Leu
65                  70                  75                  80
Ile Leu Ser Ala Ile Gln Gly Val Thr Leu Ala Gln Asp His Phe Gln
                85                  90                  95
Glu Ile Glu Lys Ile Trp Ser Ser Pro Asn Gln Leu Asn Cys Glu Ser
            100                 105                 110
Leu Ser Lys Asn Leu Ser Ser Thr Leu Glu Ser Phe Lys Ser Ser Leu
        115                 120                 125
Glu Asn Ala Thr Gly Gln Asp Cys Thr Ser Gln Pro Arg Leu Glu Thr
    130                 135                 140
Val Gln Gln His Leu Tyr Met Leu Ala Lys Ser Leu Glu Glu Thr Trp
145                 150                 155                 160
Ser Ser Gly Asn Pro Ile Met Thr Phe Leu Ser Asn Phe Thr Val Thr
                165                 170                 175
Glu Asp Val Lys Ile Lys Asp Leu Met Lys Asn Ile Thr Lys Leu Thr
            180                 185                 190
Glu Glu Leu Arg Ser Ser Ile Gln Ile Ser Asn Glu Thr Ile His Ser
        195                 200                 205
Ile Leu Glu Ala Asn Ile Ser His Ser Lys Val Leu Phe Ser Ala Leu
    210                 215                 220
Thr Val Ala Leu Ser Gly Lys Cys Asp Gln Glu Ile Leu His Leu Leu
225                 230                 235                 240
Leu Thr Phe Pro Lys Gly Glu Lys Ser Trp Ile Ala Ala Glu Glu Leu
                245                 250                 255
Cys Ser Leu Pro Gly Ser Lys Val Tyr Ser Leu Ile Val Leu Leu Ser
            260                 265                 270
Arg Asn Leu Asp Val Arg Ala Phe Ile Tyr Lys Thr Leu Met Pro Ser
        275                 280                 285
Glu Ala Asn Gly Leu Leu Asn Ser Leu Leu Asp Ile Val Ser Ser Leu
    290                 295                 300
Ser Ala Leu Leu Ala Lys Ala Gln His Val Phe Glu Tyr Leu Pro Glu
305                 310                 315                 320
Phe Leu His Thr Phe Lys Ile Thr Ala Leu Leu Glu Thr Leu Asp Phe
                325                 330                 335
Gln Gln Val Ser Gln Asn Val Gln Ala Arg Ser Ser Ala Phe Gly Ser
            340                 345                 350
Phe Gln Phe Val Met Lys Met Val Cys Lys Asp Gln Ala Ser Phe Leu
        355                 360                 365
Ser Asp Ser Asn Met Phe Ile Asn Leu Pro Arg Val Lys Glu Leu Leu
    370                 375                 380
Glu Asp Asp Lys Glu Lys Phe Asn Ile Pro Glu Asp Ser Thr Pro Phe
385                 390                 395                 400
Cys Leu Lys Leu Tyr Gln Glu Ile Leu Gln Leu Pro Asn Gly Ala Leu
                405                 410                 415
Val Trp Thr Phe Leu Lys Pro Ile Leu His Gly Lys Ile Leu Tyr Thr
            420                 425                 430
Pro Asn Thr Pro Glu Ile Asn Lys Val Ile Gln Lys Ala Asn Tyr Thr
        435                 440                 445
Phe Tyr Ile Val Asp Lys Leu Lys Thr Leu Ser Glu Thr Leu Leu Glu
    450                 455                 460
Met Ser Ser Leu Phe Gln Arg Ser Gly Ser Gly Gln Met Phe Asn Gln
```

```
465                 470                 475                 480
Leu Gln Glu Ala Leu Arg Asn Lys Phe Val Arg Asn Phe Val Glu Asn
                485                 490                 495
Gln Leu His Ile Asp Val Asp Lys Leu Thr Glu Lys Leu Gln Thr Tyr
            500                 505                 510
Gly Gly Leu Leu Asp Glu Met Phe Asn His Ala Gly Ala Gly Arg Phe
        515                 520                 525
Arg Phe Leu Gly Ser Ile Leu Val Asn Leu Ser Ser Cys Val Ala Leu
    530                 535                 540
Asn Arg Phe Gln Ala Leu Gln Ser Val Asp Ile Leu Glu Thr Lys Ala
545                 550                 555                 560
His Glu Leu Leu Gln Gln Asn Ser Phe Leu Ala Ser Ile Ile Phe Ser
                565                 570                 575
Asn Ser Leu Phe Asp Lys Asn Phe Arg Ser Glu Ser Val Lys Leu Pro
            580                 585                 590
Pro His Val Ser Tyr Thr Ile Arg Thr Asn Val Leu Tyr Ser Val Arg
        595                 600                 605
Thr Asp Val Val Lys Asn Pro Ser Trp Lys Phe His Pro Gln Asn Leu
    610                 615                 620
Pro Ala Asp Gly Phe Lys Tyr Asn Tyr Val Phe Ala Pro Leu Gln Asp
625                 630                 635                 640
Met Ile Glu Arg Ala Ile Ile Leu Val Gln Thr Gly Gln Glu Ala Leu
                645                 650                 655
Glu Pro Ala Ala Gln Thr Gln Ala Ala Pro Tyr Pro Cys His Thr Ser
            660                 665                 670
Asp Leu Phe Leu Asn Asn Val Gly Phe Phe Phe Pro Leu Ile Met Met
        675                 680                 685
Leu Thr Trp Met Val Ser Val Ala Ser Met Val Arg Lys Leu Val Tyr
    690                 695                 700
Glu Gln Glu Ile Gln Ile Glu Glu Tyr Met Arg Met Met Gly Val His
705                 710                 715                 720
Pro Val Ile His Phe Leu Ala Trp Phe Leu Glu Asn Met Ala Val Leu
                725                 730                 735
Thr Ile Ser Ser Ala Thr Leu Ala Ile Val Leu Lys Thr Ser Gly Ile
            740                 745                 750
Phe Ala His Ser Asn Thr Phe Ile Val Phe Leu Phe Leu Leu Asp Phe
        755                 760                 765
Gly Met Ser Val Val Met Leu Ser Tyr Leu Leu Ser Ala Phe Phe Ser
    770                 775                 780
Gln Ala Asn Thr Ala Ala Leu Cys Thr Ser Leu Val Tyr Met Ile Ser
785                 790                 795                 800
Phe Leu Pro Tyr Ile Val Leu Leu Val Leu His Asn Gln Leu Ser Phe
                805                 810                 815
Val Asn Gln Thr Phe Leu Cys Leu Leu Ser Thr Thr Ala Phe Gly Gln
            820                 825                 830
Gly Val Phe Phe Ile Thr Phe Leu Glu Gly Gln Glu Thr Gly Ile Gln
        835                 840                 845
Trp Asn Asn Met Tyr Gln Ala Leu Glu Gln Gly Gly Met Thr Phe Gly
    850                 855                 860
Trp Val Cys Trp Met Ile Leu Phe Asp Ser Ser Leu Tyr Phe Leu Cys
865                 870                 875                 880
Gly Trp Tyr Leu Ser Asn Leu Ile Pro Gly Thr Phe Gly Leu Arg Lys
                885                 890                 895
Pro Trp Tyr Phe Pro Phe Thr Ala Ser Tyr Trp Lys Ser Val Gly Phe
            900                 905                 910
Leu Val Glu Lys Arg Gln Tyr Phe Leu Ser Ser Ser Leu Phe Phe Phe
        915                 920                 925
Asn Glu Asn Phe Asp Asn Lys Gly Ser Ser Leu Gln Asn Arg Glu Gly
    930                 935                 940
Glu Leu Glu Gly Ser Ala Pro Gly Val Thr Leu Val Ser Val Thr Lys
945                 950                 955                 960
Glu Tyr Glu Gly His Lys Ala Val Val Gln Asp Leu Ser Leu Thr Phe
                965                 970                 975
```

```
Tyr Arg Asp Gln Ile Thr Ala Leu Leu Gly Thr Asn Gly Ala Gly Lys
            980                 985                 990
Thr Thr Ile Ile Ser Met Leu Thr Gly Leu His Pro Pro Thr Ser Gly
        995                 1000                1005
Thr Ile Ile Ile Asn Gly Lys Asn Leu Gln Thr Asp Leu Ser Arg Val
    1010                1015                1020
Arg Met Glu Leu Gly Val Cys Pro Gln Gln Asp Ile Leu Leu Asp Asn
1025                1030                1035                1040
Leu Thr Val Arg Glu His Leu Leu Leu Phe Ala Ser Ile Lys Ala Pro
                1045                1050                1055
Gln Trp Thr Lys Lys Glu Leu His Gln Gln Val Asn Gln Thr Leu Gln
            1060                1065                1070
Asp Val Asp Leu Thr Gln His Gln His Lys Gln Thr Arg Ala Leu Ser
        1075                1080                1085
Gly Gly Leu Lys Arg Lys Leu Ser Leu Gly Ile Ala Phe Met Gly Met
    1090                1095                1100
Ser Arg Thr Val Val Leu Asp Glu Pro Thr Ser Gly Val Asp Pro Cys
1105                1110                1115                1120
Ser Arg His Ser Leu Trp Asp Ile Leu Leu Lys Tyr Arg Glu Gly Arg
                1125                1130                1135
His Trp Ala Ser Phe Cys Leu Leu Phe Pro Gln Tyr Cys Val Ala Gly
            1140                1145                1150
Asn Ala Leu Leu Leu Tyr Ser Arg Ile Lys Leu Tyr Pro Ser Glu Ala
        1155                1160                1165
Thr Leu Ser Phe Ser Glu Lys Tyr Lys Phe Leu Lys Ala Leu Ile Gly
    1170                1175                1180
Ile Tyr Ser Leu Asp Asn Ile Ser Ile Val Leu Arg Ile Ala Tyr Phe
1185                1190                1195                1200
Val Phe Asp Leu Leu Thr Gln
                1205
```

<210> 5
<211> 4165
<212> DNA
<213> homo sapiens

<400> 5

```
tggagaccta aagttttcta aaggtccaga atgtgttatc tgtgttttct tatgttccta 60
tgaagaaaat atgattatgc agggagggag gatggttctt acatgtgtgt tataacttaa 120
cctacacagt agagatgcac atgggttgca cattttacc cttttatgtc attgtatata 180
ttttttgct aagtgttgtt gagatttgtg aagttttcca gcagactgtg aagccctcag 240
aagccatgga gatgctgcag aaagtgaaga tgatggtcgt acgtgtgctc accatcgttg 300
cagaaaaccc ttcctggacc aaggacattt tgtgtgctac tctgagttgc aagcaaaatg 360
ggataaggca tctcatttta tctgctatac aaggggtcac tttggcgcag gaccacttcc 420
aggaaattga aaagatatgg tcctcgccga atcagctaaa ttgtgaaagt cttagcaaga 480
atctttctag caccttggag agcttcaaga gcagcttgga aaatgccact ggccaggact 540
gcacaagcca gccgaggctg gagacggtgc agcagcactt gtacatgttg gccaaaagcc 600
tygaggaaac ttggtcatca gggaatccca tcatgacttt tctcagcaat ttcacagtaa 660
ctgaggatgt aaaaataaaa gatttgatga agaatatcac caagttgact gaggagcttc 720
gctcttccat ccaaatctcg aatgagacta tccatagcat tctagaagca aatatttccc 780
actccaaggt tctcttcagt gccctcaccg tagctctgtc tggaaagtgt gatcaggaaa 840
tccttcatct cctgctgaca tttcccaaag gggaaaaatc ttggatcgca gcggaggaac 900
tctgtagcct gccagggtca aaagtgtatt ctctgattgt gttgctgagt cgaaacttgg 960
atgtgcgagc tttcatttac aagactctga tgccttctga agcaaatggc ttgctcaact 1020
ccttgctgga tatagtttcc agcctcagcg ccttgcttgc caaagcccag cacgtctttg 1080
agtatcttcc tgagtttctt cacacattta aaatcactgc cttgctagaa accctggact 1140
ttcaacaggt ttcacaaaat gtccaggcca gaagttcagc ttttggttct ttccagtttg 1200
tgatgaagat ggtttgcaag gaccaagcat cattccttag cgattctaat atgtttatta 1260
atttgcccag agttaaggaa ctcttggaag atgacaaaga aaaattcaac attcctgaag 1320
attcaacacc gttttgcttg aagctttatc aggaaattct acaattgcca aatggtgctt 1380
tggtgtggac cttcctaaaa cccatattgc atggaaaaat actatacaca ccaaacactc 1440
cagaaattaa caaggtcatt caaaaggcta attacacctt ttatattgtg gacaaactaa 1500
```

aaactttatc agaaacactg ctggaaatgt ccagcctttt ccagagaagt ggaagtggcc 1560
agatgttcaa ccagctgcag gaggccctga gaaacaaatt tgtaagaaac tttgtagaaa 1620
accagttgca cattgatgta gacaaactta ctgaaaaact ccagacatac ggagggctgc 1680
tggatgagat gtttaaccat gcaggcgctg gacgcttccg tttcttgggc agcatcttgg 1740
tcaatctctc ttcctgcgtg gcactgaacc gtttccaggc tctgcagtct gtcgacatcc 1800
tggagactaa agcacatgaa ctcttgcagc agaacagctt cttggccagt atcattttca 1860
gcaattcctt attcgacaag aacttcagat cagagtctgt caaactgcca ccccatgtct 1920
catacacaat ccggaccaat gtgttataca gcgtgcgaac agatgtggta aaaaaccctt 1980
cttggaagtt ccaccctcag aatctaccag ctgatgggtt caaatataac tacgtctttg 2040
ccccactgca agacatgatc gaaagagcca tcattttggt gcagactggg caggaagccc 2100
tggaaccagc agcacagact caggcggccc cttacccctg ccataccagc gacctattcc 2160
tgaacaacgt tggtttcttt tttccactga taatgatgct gacgtggatg gtgtctgtgg 2220
ccagcatggt cagaaagttg gtgtatgagc aggagataca gatagaagag tatatgcgga 2280
tgatgggagt gcatccagtg atccatttcc tggcctggtt cctggagaac atggctgtgt 2340
tgaccataag cagtgctact ctggccatcg ttctgaaaac aagtggcatc tttgcacaca 2400
gcaatacctt tattgttttc ctctttctct tggattttgg gatgtcagtc gtcatgctga 2460
gctacctctt gagtgcattt ttcagccaag ctaatacagc ggccctttgt accagcctgg 2520
tgtacatgat cagctttctg ccctacatag ttctattggt tctacataac caattaagtt 2580
ttgttaatca gacatttctg tgccttcttt cgacaaccgc ctttggacaa ggggtatttt 2640
ttattacatt cctggaagga caagagacag ggattcaatg gaataatatg taccaggctc 2700
tggaacaagg gggcatgaca tttggctggg tttgctggat gattcttttt gattcaagcc 2760
tttatttttt gtgtggatgg tacttgagca acttgattcc tggaacattt ggtttacgga 2820
aaccatggta tttccccttt actgcctcat attggaagag tgtgggtttc ttggtggaga 2880
aaaggcaata ctttctaagt tctagtctgt tcttcttcaa tgagaacttt gacaataaag 2940
ggtcatcact gcaaaacagg gaaggagagc ttgaaggaag tgccccggga gtcaccctgg 3000
tgtctgtgac caaggaatat gagggccaca aggctgtggt ccaagacctc agcctgacct 3060
tctacagaga ccaaatcacc gccctgctgg ggacaaacgg tgccgggaaa accactatca 3120
tatccatgtt gacggggctc caccctccca cttctggaac catcatcatc aatggcaaga 3180
acctacagac agacctgtcg agggtcagaa tggagcttgg tgtgtgtccg cagcaggaca 3240
tcctgttgga caacctcacc gtccgggaac atttgctgct ctttgcttcc ataaaggcgc 3300
ctcagtggac caagaaggag ctgcatcagc aagtcaatca aactcttcag gatgtggact 3360
taactcagca tcagcacaaa cagacccgag ctctgtctgg aggcctgaag aggaagctct 3420
cccttggcat tgctttcatg ggcatgtcga ggaccgtggt tctggatgag cccaccagtg 3480
gggtggaccc ttgctcccgg catagcctgt gggacattct gctcaagtac cgagaaggta 3540
ggcactgggc ctcattctgc cttctcttcc cacaatattg tgttgcagga aatgcattgc 3600
tactgtacag tagaatcaag ttgtatccca gtgaggctac attatccttt tcagaaaaat 3660
ataaattttt aaaagcactt atagggatat attcgttaga taacatctct atagtgctta 3720
gaattgctta ctttgtgttt gacctttaa ctcaataaca gcaatgacat ctatgtacat 3780
tatacattat catacatgat ttcaaggaaa attgtcttct tctggaagca tagtttctta 3840
gaagaggcat cccagatcat aggacaagcc tcccttgtct cagatgaaga aatgaaggct 3900
cagagagacg ggcatgtgat ttacttgtag ctacagagaa agtttcctga actgagggtg 3960
gatgttgaac ctcttgtcca tgtttctcac atctattatt gtttctttcc aatttaggac 4020
atttgatggg cagttactaa tttccaactt ctgattcttt ctgcaatcct gacagctagg 4080
aagcattgtt ctatgtattt tctgtgagaa tactcccttt tggaaagaaa cattgcaaca 4140
gtaaaacaca tcttggtgct ggtaa 4165

<210> 6
<211> 2046
<212> DNA
<213> homo sapiens

<400> 6
atgagcaagg agctggcagc aatggggcct ggagcttcag gggacggggt caggactgag 60
acagctccac acatagcact ggactccaga gttggtctgc acgcctacga catcagcgtg 120
gtrgtcatct actttgtctt cgtcattgct gtggggatct ggtcgtccat ccgtgcaagt 180
cgagggacca ttggcggcta tttcctggcc gggaggtcca tgagctggtg gccaattgga 240
gcatctctga tgtccagcaa tgtgggcagt ggcttgttca tcggcctggc tgggacaggg 300
gctgccggag gccttgccgt aggtggcttc gagtggaacg caacctggct gctcctggcc 360
cttggctggr tcttcgtccc tgtgtacatc gcagcaggtg tggtcacaat gccgcagtat 420
ctgaagaagc gatttggggg ccagaggatc cagrtgtaca tgtctgtcct gtctctcatc 480
ctctacatct tcaccaagat ctcgactgac atcttctctg gagccctctt catccagatg 540
gcattgggct ggaacctgta cctctccaca gggatcctgc tggtggtgac tgccgtctac 600
accattgcag gtggcctcat ggccgtgatc tacacagatg ctctgcagac ggtgatcatg 660

gtagggggag ccctggtcct catgtttctg ggctttcagg acgtgggctg gtacccaggc 720
ctggagcagc ggtacaggca ggccatccct aatgtcacag tccccaacac cacctgtcac 780
ctcccacggc ccgatgcttt ccacatgctt cgggaccctg tgagcgggga catcccttgg 840
ccaggtctca ttttcgggct cacagtgctg gccacctggt gttggtgcac agaccaggtc 900
attgtgcagc ggtctctctc ggccaagagt ctgtctcatg ccaagggagg ctccgtgctg 960
gggggctacc tgaagatcct ccccatgttc ttcatcgtca tgcctggcat gatcagccgg 1020
gccctgttcc cagacgaggt gggctgcgtg gaccctgatg tctgccaaag aatctgtggg 1080
gcccgagtgg gatgttccaa cattgcctac cctaagttgg tcatggccct catgcctgtt 1140
ggtctgcggg ggctgatgat tgccgtgatc atggccgctc tcatgagctc actcacctcc 1200
atcttcaaca gcagcagcac cctgttcacc attgatgtgt ggcagcgctt ccgcaggaag 1260
tcaacagagc aggagctgat ggtggtgggc agagtgtttg tggtgttcct ggttgtcatc 1320
agcatcctct ggatccccat catccaaagc tccaacagtg ggcagctctt cgactacatc 1380
caggctgtca ccagttacct ggccccaccc atcaccgctc tcttcctgct ggccatcttc 1440
tgcaagaggg tcacagagcc cggagctttc tggggcctcg tgtttggcct gggagtgggg 1500
cttctgcgta tgatcctgga gttctcatac ccagcgccag cctgtgggga ggtggaccgg 1560
aggccagcag tgctgaagga cttccactac ctgtactttg caatcctcct ctgcgggctc 1620
actgccatcg tcattgtcat tgtcagcctc tgtacaactc ccatccctga ggaacagctc 1680
acacgcctca catggtggac tcggaactgc cccctctctg agctggagaa ggaggcccac 1740
gagagcacac cggagatatc cgagaggcca gccggggagt gccctgcagg aggtggagcg 1800
gcagagaact cgagcctggg ccaggagcag cctgaagccc caagcaggtc ctggggaaag 1860
ttgctctgga gctggttctg tgggctctct ggaacaccgg agcaggccct gagcccagca 1920
gagaaggctg cgctagaaca gaagctgaca agcattgagg aggagccact ctggagacat 1980
gtctgcaaca tcaatgctgt ccttttgctg gccatcaaca tcttcctctg gggctatttt 2040
gcgtga 2046

<210> 7
<211> 681
<212> PRT
<213> homo sapiens

<220>
<221> VARIANT
<222> 124, 152
<223> Xaa = Any Amino Acid

<400> 7
Met Ser Lys Glu Leu Ala Ala Met Gly Pro Gly Ala Ser Gly Asp Gly
1 5 10 15
Val Arg Thr Glu Thr Ala Pro His Ile Ala Leu Asp Ser Arg Val Gly
20 25 30
Leu His Ala Tyr Asp Ile Ser Val Val Val Ile Tyr Phe Val Phe Val
35 40 45
Ile Ala Val Gly Ile Trp Ser Ser Ile Arg Ala Ser Arg Gly Thr Ile
50 55 60
Gly Gly Tyr Phe Leu Ala Gly Arg Ser Met Ser Trp Trp Pro Ile Gly
65 70 75 80
Ala Ser Leu Met Ser Ser Asn Val Gly Ser Gly Leu Phe Ile Gly Leu
85 90 95
Ala Gly Thr Gly Ala Ala Gly Gly Leu Ala Val Gly Gly Phe Glu Trp
100 105 110
Asn Ala Thr Trp Leu Leu Leu Ala Leu Gly Trp Xaa Phe Val Pro Val
115 120 125
Tyr Ile Ala Ala Gly Val Val Thr Met Pro Gln Tyr Leu Lys Lys Arg
130 135 140
Phe Gly Gly Gln Arg Ile Gln Xaa Tyr Met Ser Val Leu Ser Leu Ile
145 150 155 160
Leu Tyr Ile Phe Thr Lys Ile Ser Thr Asp Ile Phe Ser Gly Ala Leu
165 170 175
Phe Ile Gln Met Ala Leu Gly Trp Asn Leu Tyr Leu Ser Thr Gly Ile
180 185 190
Leu Leu Val Val Thr Ala Val Tyr Thr Ile Ala Gly Gly Leu Met Ala
195 200 205
Val Ile Tyr Thr Asp Ala Leu Gln Thr Val Ile Met Val Gly Gly Ala

```
    210                 215                 220
Leu Val Leu Met Phe Leu Gly Phe Gln Asp Val Gly Trp Tyr Pro Gly
225                 230                 235                 240
Leu Glu Gln Arg Tyr Arg Gln Ala Ile Pro Asn Val Thr Val Pro Asn
                245                 250                 255
Thr Thr Cys His Leu Pro Arg Pro Asp Ala Phe His Met Leu Arg Asp
            260                 265                 270
Pro Val Ser Gly Asp Ile Pro Trp Pro Gly Leu Ile Phe Gly Leu Thr
        275                 280                 285
Val Leu Ala Thr Trp Cys Trp Cys Thr Asp Gln Val Ile Val Gln Arg
    290                 295                 300
Ser Leu Ser Ala Lys Ser Leu Ser His Ala Lys Gly Gly Ser Val Leu
305                 310                 315                 320
Gly Gly Tyr Leu Lys Ile Leu Pro Met Phe Phe Ile Val Met Pro Gly
                325                 330                 335
Met Ile Ser Arg Ala Leu Phe Pro Asp Glu Val Gly Cys Val Asp Pro
            340                 345                 350
Asp Val Cys Gln Arg Ile Cys Gly Ala Arg Val Gly Cys Ser Asn Ile
        355                 360                 365
Ala Tyr Pro Lys Leu Val Met Ala Leu Met Pro Val Gly Leu Arg Gly
    370                 375                 380
Leu Met Ile Ala Val Ile Met Ala Ala Leu Met Ser Ser Leu Thr Ser
385                 390                 395                 400
Ile Phe Asn Ser Ser Ser Thr Leu Phe Thr Ile Asp Val Trp Gln Arg
                405                 410                 415
Phe Arg Arg Lys Ser Thr Glu Gln Glu Leu Met Val Val Gly Arg Val
            420                 425                 430
Phe Val Val Phe Leu Val Val Ile Ser Ile Leu Trp Ile Pro Ile Ile
        435                 440                 445
Gln Ser Ser Asn Ser Gly Gln Leu Phe Asp Tyr Ile Gln Ala Val Thr
    450                 455                 460
Ser Tyr Leu Ala Pro Pro Ile Thr Ala Leu Phe Leu Leu Ala Ile Phe
465                 470                 475                 480
Cys Lys Arg Val Thr Glu Pro Gly Ala Phe Trp Gly Leu Val Phe Gly
                485                 490                 495
Leu Gly Val Gly Leu Leu Arg Met Ile Leu Glu Phe Ser Tyr Pro Ala
            500                 505                 510
Pro Ala Cys Gly Glu Val Asp Arg Arg Pro Ala Val Leu Lys Asp Phe
        515                 520                 525
His Tyr Leu Tyr Phe Ala Ile Leu Leu Cys Gly Leu Thr Ala Ile Val
    530                 535                 540
Ile Val Ile Val Ser Leu Cys Thr Thr Pro Ile Pro Glu Glu Gln Leu
545                 550                 555                 560
Thr Arg Leu Thr Trp Trp Thr Arg Asn Cys Pro Leu Ser Glu Leu Glu
                565                 570                 575
Lys Glu Ala His Glu Ser Thr Pro Glu Ile Ser Glu Arg Pro Ala Gly
            580                 585                 590
Glu Cys Pro Ala Gly Gly Gly Ala Ala Glu Asn Ser Ser Leu Gly Gln
        595                 600                 605
Glu Gln Pro Glu Ala Pro Ser Arg Ser Trp Gly Lys Leu Leu Trp Ser
    610                 615                 620
Trp Phe Cys Gly Leu Ser Gly Thr Pro Glu Gln Ala Leu Ser Pro Ala
625                 630                 635                 640
Glu Lys Ala Ala Leu Glu Gln Lys Leu Thr Ser Ile Glu Glu Glu Pro
                645                 650                 655
Leu Trp Arg His Val Cys Asn Ile Asn Ala Val Leu Leu Leu Ala Ile
            660                 665                 670
Asn Ile Phe Leu Trp Gly Tyr Phe Ala
        675                 680
```

<210> 8
<211> 2025

<212> DNA
<213> homo sapiens

<400> 8
atggggcctg gagcttcagg ggacggggtc aggactgaga cagctccaca catagcactg 60
gactccagag ttggtctgca cgcctacgac atcagcgtgg tggtcatcta ctttgtcttc 120
gtcattgctg tggggatctg gtcgtccatc cgtgcaagtc gagggaccat tggcggctat 180
ttcctggccg ggaggtccat gagctggtgg ccaattggag catctctgat gtccagcaat 240
gtgggcagtg gcttgttcat cggcctggct gggacagggg ctgccggagg ccttgccgta 300
ggtggcttcg agtggaacgc aacctggctg ctcctggccc ttggctgggt cttcgtccct 360
gtgtacatcg cagcaggtgt ggtcacaatg ccgcagtatc tgaagaagcg atttgggggc 420
cagaggatcc aggtgtacat gtctgtcctg tctctcatcc tctacatctt caccaagatc 480
tcgactgaca tcttctctgg agccctcttc atccagatgg cattgggctg gaacctgtac 540
ctctccacag ggatcctgct ggtggtgact gccgtctaca ccattgcagg tggcctcatg 600
gccgtgatct acacagatgc tctgcagacg gtgatcatgg taggggagc cctggtcctc 660
atgtttctgg gctttcagga cgtgggctgg tacccaggcc tggagcagcg gtacaggcag 720
gccatcccta atgtcacagt ccccaacacc acctgtcacc tcccacggcc cgatgctttc 780
cacatgcttc gggaccctgt gagcggggac atcccttggc caggtctcat tttcgggctc 840
acagtgctgg ccacctggtg ttggtgcaca gaccaggtca ttgtgcagcg gtctctctcg 900
gccaagagtc tgtctcatgc caagggaggc tccgtgctgg ggggctacct gaagatcctc 960
cccatgttct tcatcgtcat gcctggcatg atcagccggg ccctgttccc agacgaggtg 1020
ggctgcgtgg accctgatgt ctgccaaaga atctgtgggg cccgagtggg atgttccaac 1080
attgcctacc ctaagttggt catggccctc atgcctgttg gtctgcgggg gctgatgatt 1140
gccgtgatca tggccgctct catgagctca ctcacctcca tcttcaacag cagcagcacc 1200
ctgttcacca ttgatgtgtg gcagcgcttc cgcaggaagt caacagagca ggagctgatg 1260
gtggtgggca gagtgtttgt ggtgttcctg gttgtcatca gcatcctctg gatccccatc 1320
atccaaagct ccaacagtgg gcagctcttc gactacatcc aggctgtcac cagttacctg 1380
gccccaccca tcaccgctct cttcctgctg gccatcttct gcaagagggt cacagagccc 1440
ggagctttct ggggcctcgt gtttggcctg ggagtggggc ttctgcgtat gatcctggag 1500
ttctcatacc cagcgccagc ctgtggggag gtggaccgga ggccagcagt gctgaaggac 1560
ttccactacc tgtactttgc aatcctcctc tgcgggctca ctgccatcgt cattgtcatt 1620
gtcagcctct gtacaactcc catccctgag gaacagctca cacgcctcac atggtggact 1680
cggaactgcc ccctctctga gctggagaag gaggcccacg agagcacacc ggagatatcc 1740
gagaggccag ccggggagtg ccctgcagga ggtggagcgg cagagaactc gagcctgggc 1800
caggagcagc ctgaagcccc aagcaggtcc tggggaaagt tgctctggag ctggttctgt 1860
gggctctctg gaacaccgga gcaggccctg agcccagcag agaaggctgc gctagaacag 1920
aagctgacaa gcattgagga ggagccactc tggagacatg tctgcaacat caatgctgtc 1980
cttttgctgg ccatcaacat cttcctctgg ggctattttg cgtga 2025

<210> 9
<211> 674
<212> PRT
<213> homo sapiens

<400> 9
Met Gly Pro Gly Ala Ser Gly Asp Gly Val Arg Thr Glu Thr Ala Pro
 1   5   10   15
His Ile Ala Leu Asp Ser Arg Val Gly Leu His Ala Tyr Asp Ile Ser
            20   25   30
Val Val Val Ile Tyr Phe Val Phe Val Ile Ala Val Gly Ile Trp Ser
        35   40   45
Ser Ile Arg Ala Ser Arg Gly Thr Ile Gly Gly Tyr Phe Leu Ala Gly
    50   55   60
Arg Ser Met Ser Trp Trp Pro Ile Gly Ala Ser Leu Met Ser Ser Asn
65   70   75   80
Val Gly Ser Gly Leu Phe Ile Gly Leu Ala Gly Thr Gly Ala Ala Gly
                85   90   95
Gly Leu Ala Val Gly Gly Phe Glu Trp Asn Ala Thr Trp Leu Leu Leu
            100   105   110
Ala Leu Gly Trp Val Phe Val Pro Val Tyr Ile Ala Ala Gly Val Val
        115   120   125
Thr Met Pro Gln Tyr Leu Lys Lys Arg Phe Gly Gly Gln Arg Ile Gln
    130   135   140

```
Val Tyr Met Ser Val Leu Ser Leu Ile Leu Tyr Ile Phe Thr Lys Ile
145                 150                 155                 160
Ser Thr Asp Ile Phe Ser Gly Ala Leu Phe Ile Gln Met Ala Leu Gly
                165                 170                 175
Trp Asn Leu Tyr Leu Ser Thr Gly Ile Leu Leu Val Val Thr Ala Val
            180                 185                 190
Tyr Thr Ile Ala Gly Gly Leu Met Ala Val Ile Tyr Thr Asp Ala Leu
        195                 200                 205
Gln Thr Val Ile Met Val Gly Gly Ala Leu Val Leu Met Phe Leu Gly
    210                 215                 220
Phe Gln Asp Val Gly Trp Tyr Pro Gly Leu Glu Gln Arg Tyr Arg Gln
225                 230                 235                 240
Ala Ile Pro Asn Val Thr Val Pro Asn Thr Thr Cys His Leu Pro Arg
                245                 250                 255
Pro Asp Ala Phe His Met Leu Arg Asp Pro Val Ser Gly Asp Ile Pro
            260                 265                 270
Trp Pro Gly Leu Ile Phe Gly Leu Thr Val Leu Ala Thr Trp Cys Trp
        275                 280                 285
Cys Thr Asp Gln Val Ile Val Gln Arg Ser Leu Ser Ala Lys Ser Leu
    290                 295                 300
Ser His Ala Lys Gly Gly Ser Val Leu Gly Gly Tyr Leu Lys Ile Leu
305                 310                 315                 320
Pro Met Phe Phe Ile Val Met Pro Gly Met Ile Ser Arg Ala Leu Phe
                325                 330                 335
Pro Asp Glu Val Gly Cys Val Asp Pro Asp Val Cys Gln Arg Ile Cys
            340                 345                 350
Gly Ala Arg Val Gly Cys Ser Asn Ile Ala Tyr Pro Lys Leu Val Met
        355                 360                 365
Ala Leu Met Pro Val Gly Leu Arg Gly Leu Met Ile Ala Val Ile Met
    370                 375                 380
Ala Ala Leu Met Ser Ser Leu Thr Ser Ile Phe Asn Ser Ser Ser Thr
385                 390                 395                 400
Leu Phe Thr Ile Asp Val Trp Gln Arg Phe Arg Arg Lys Ser Thr Glu
                405                 410                 415
Gln Glu Leu Met Val Val Gly Arg Val Phe Val Val Phe Leu Val Val
            420                 425                 430
Ile Ser Ile Leu Trp Ile Pro Ile Ile Gln Ser Ser Asn Ser Gly Gln
        435                 440                 445
Leu Phe Asp Tyr Ile Gln Ala Val Thr Ser Tyr Leu Ala Pro Pro Ile
    450                 455                 460
Thr Ala Leu Phe Leu Leu Ala Ile Phe Cys Lys Arg Val Thr Glu Pro
465                 470                 475                 480
Gly Ala Phe Trp Gly Leu Val Phe Gly Leu Gly Val Gly Leu Leu Arg
                485                 490                 495
Met Ile Leu Glu Phe Ser Tyr Pro Ala Pro Ala Cys Gly Glu Val Asp
            500                 505                 510
Arg Arg Pro Ala Val Leu Lys Asp Phe His Tyr Leu Tyr Phe Ala Ile
        515                 520                 525
Leu Leu Cys Gly Leu Thr Ala Ile Val Ile Val Ile Val Ser Leu Cys
    530                 535                 540
Thr Thr Pro Ile Pro Glu Glu Gln Leu Thr Arg Leu Thr Trp Trp Thr
545                 550                 555                 560
Arg Asn Cys Pro Leu Ser Glu Leu Glu Lys Glu Ala His Glu Ser Thr
                565                 570                 575
Pro Glu Ile Ser Glu Arg Pro Ala Gly Glu Cys Pro Ala Gly Gly Gly
            580                 585                 590
Ala Ala Glu Asn Ser Ser Leu Gly Gln Glu Gln Pro Glu Ala Pro Ser
        595                 600                 605
Arg Ser Trp Gly Lys Leu Leu Trp Ser Trp Phe Cys Gly Leu Ser Gly
    610                 615                 620
Thr Pro Glu Gln Ala Leu Ser Pro Ala Glu Lys Ala Ala Leu Glu Gln
625                 630                 635                 640
Lys Leu Thr Ser Ile Glu Glu Glu Pro Leu Trp Arg His Val Cys Asn
```

```
                645                 650                 655
Ile Asn Ala Val Leu Leu Leu Ala Ile Asn Ile Phe Leu Trp Gly Tyr
            660                 665                 670
Phe Ala
```

<210> 10
<211> 2238
<212> DNA
<213> homo sapiens

<400> 10
atgagcaagg agctggcagc aatggggcct ggagcttcag gggacggggt caggactgag 60
acagctccac acatagcact ggactccaga gttggtctgc acgcctacga catcagcgtg 120
gtggtcatct actttgtctt cgtcattgct gtggggatct ggtcgtccat ccgtgcaagt 180
cgagggacca ttggcggcta tttcctggcc gggaggtcca tgagctggtg gccaattgga 240
gcatctctga tgtccagcaa tgtgggcagt ggcttgttca tcggcctggc tgggacaggg 300
gctgccggag gccttgccgt aggtggcttc gagtggaaca tgaggaaatc aaggtctgga 360
ggagacagag ggatccatcc aaggtcacac gggaggactg gggtcaggtc ccaggtctct 420
tatttctctg ttcgggggcc tcccacagca cagcactgcc tctgggtggg aagccgcccc 480
tctgtctaca tccaggacct ggataccttc ttcttctccc cactctccca ggcaacctgg 540
ctgctcctgg cccttggctg ggtcttcgtc cctgtgtaca tcgcagcagg tgtggtcaca 600
atgccgcagt atctgaagaa gcgatttggg ggccagagga tccaggtgta catgtctgtc 660
ctgtctctca tcctctacat cttcaccaag atctcgactg acatcttctc tggagccctc 720
ttcatccaga tggcattggg ctggaacctg tacctctcca cagggatcct gctggtggtg 780
actgccgtct acaccattgc aggtggcctc atggccgtga tctacacaga tgctctgcag 840
acggtgatca tggtaggggg agccctggtc ctcatgtttc tgggctttca ggacgtgggc 900
tggtacccag gcctggagca gcggtacagg caggccatcc ctaatgtcac agtccccaac 960
accacctgtc acctcccacg gcccgatgct ttccacatgc ttcgggaccc tgtgagyggg 1020
gacatccctt ggccaggtct cattttcggg ctcacagtgc tggccacctg gtgttggtgc 1080
acagaccagg tcattgtgca gcggtctctc tcggccaaga gtctgtctca tgccaaggga 1140
ggctccgtgc tgggggggcta cctgaagatc ctccccatgt tcttcatcgt catgcctggc 1200
atgatcagcc gggccctgtt cccagacgag gtgggctgcg tggaccctga tgtctgccaa 1260
agaatctgtg gggcccgagt gggatgttcc aacattgcct accctaagtt ggtcatggcc 1320
ctcatgcctg ttggtctgcg ggggctgatg attgccgtga tcatggccgc tctcatgagc 1380
tcactcacct ccatcttcaa cagcagcagc accctgttca ccattgatgt gtggcagcgc 1440
ttccgcagga agtcaacaga gcaggagctg atggtggtgg gcagagtgtt tgtggtgttc 1500
ctggttgtca tcagcatcct ctggatcccc atcatccaaa gctccaacag tgggcagctc 1560
ttcgactaca tccaggctgt caccagttac ctggccccac ccatcaccgc tctcttcctg 1620
ctggccatct tctgcaagag ggtcacagag cccggagctt tctggggcct cgtgtttggc 1680
ctgggagtgg ggcttctgcg tatgatcctg gagttctcat acccagcgcc agcctgtggg 1740
gaggtggacc ggaggccagc agtgctgaag gacttccact acctgtactt tgcaatcctc 1800
ctctgcgggc tcactgccat cgtcattgtc attgtcagcc tctgtacaac tcccatccct 1860
gaggaacagc tcacacgcct cacatggtgg actcggaact gccccctctc tgagctggag 1920
aaggaggccc acgagagcac accggagata tccgagaggc cagccgggga gtgccctgca 1980
ggaggtggag cggcagagaa ctcgagcctg ggccaggagc agcctgaagc cccaagcagg 2040
tcctggggaa agttgctctg gagctggttc tgtgggctct ctggaacacc ggagcaggcc 2100
ctgagcccag cagagaaggc tgcgctagaa cagaagctga caagcattga ggaggagcca 2160
ctctggagac atgtctgcaa catcaatgct gtccttttgc tggccatcaa catcttcctc 2220
tggggctatt ttgcgtga 2238

<210> 11
<211> 745
<212> PRT
<213> homo sapiens

<400> 11
```
Met Ser Lys Glu Leu Ala Ala Met Gly Pro Gly Ala Ser Gly Asp Gly
 1               5                  10                  15
Val Arg Thr Glu Thr Ala Pro His Ile Ala Leu Asp Ser Arg Val Gly
            20                  25                  30
Leu His Ala Tyr Asp Ile Ser Val Val Val Ile Tyr Phe Val Phe Val
```

```
        35                  40                  45
Ile Ala Val Gly Ile Trp Ser Ser Ile Arg Ala Ser Arg Gly Thr Ile
    50                  55                  60
Gly Gly Tyr Phe Leu Ala Gly Arg Ser Met Ser Trp Trp Pro Ile Gly
65                  70                  75                  80
Ala Ser Leu Met Ser Ser Asn Val Gly Ser Gly Leu Phe Ile Gly Leu
                85                  90                  95
Ala Gly Thr Gly Ala Ala Gly Gly Leu Ala Val Gly Gly Phe Glu Trp
            100                 105                 110
Asn Met Arg Lys Ser Arg Ser Gly Gly Asp Arg Gly Ile His Pro Arg
        115                 120                 125
Ser His Gly Arg Thr Gly Val Arg Ser Gln Val Ser Tyr Phe Ser Val
    130                 135                 140
Arg Gly Pro Pro Thr Ala Gln His Cys Leu Trp Val Gly Ser Arg Pro
145                 150                 155                 160
Ser Val Tyr Ile Gln Asp Leu Asp Thr Phe Phe Phe Ser Pro Leu Ser
                165                 170                 175
Gln Ala Thr Trp Leu Leu Leu Ala Leu Gly Trp Val Phe Val Pro Val
            180                 185                 190
Tyr Ile Ala Ala Gly Val Val Thr Met Pro Gln Tyr Leu Lys Lys Arg
        195                 200                 205
Phe Gly Gly Gln Arg Ile Gln Val Tyr Met Ser Val Leu Ser Leu Ile
    210                 215                 220
Leu Tyr Ile Phe Thr Lys Ile Ser Thr Asp Ile Phe Ser Gly Ala Leu
225                 230                 235                 240
Phe Ile Gln Met Ala Leu Gly Trp Asn Leu Tyr Leu Ser Thr Gly Ile
                245                 250                 255
Leu Leu Val Val Thr Ala Val Tyr Thr Ile Ala Gly Gly Leu Met Ala
            260                 265                 270
Val Ile Tyr Thr Asp Ala Leu Gln Thr Val Ile Met Val Gly Gly Ala
        275                 280                 285
Leu Val Leu Met Phe Leu Gly Phe Gln Asp Val Gly Trp Tyr Pro Gly
    290                 295                 300
Leu Glu Gln Arg Tyr Arg Gln Ala Ile Pro Asn Val Thr Val Pro Asn
305                 310                 315                 320
Thr Thr Cys His Leu Pro Arg Pro Asp Ala Phe His Met Leu Arg Asp
                325                 330                 335
Pro Val Ser Gly Asp Ile Pro Trp Pro Gly Leu Ile Phe Gly Leu Thr
            340                 345                 350
Val Leu Ala Thr Trp Cys Trp Cys Thr Asp Gln Val Ile Val Gln Arg
        355                 360                 365
Ser Leu Ser Ala Lys Ser Leu Ser His Ala Lys Gly Gly Ser Val Leu
    370                 375                 380
Gly Gly Tyr Leu Lys Ile Leu Pro Met Phe Phe Ile Val Met Pro Gly
385                 390                 395                 400
Met Ile Ser Arg Ala Leu Phe Pro Asp Glu Val Gly Cys Val Asp Pro
                405                 410                 415
Asp Val Cys Gln Arg Ile Cys Gly Ala Arg Val Gly Cys Ser Asn Ile
            420                 425                 430
Ala Tyr Pro Lys Leu Val Met Ala Leu Met Pro Val Gly Leu Arg Gly
        435                 440                 445
Leu Met Ile Ala Val Ile Met Ala Ala Leu Met Ser Ser Leu Thr Ser
    450                 455                 460
Ile Phe Asn Ser Ser Ser Thr Leu Phe Thr Ile Asp Val Trp Gln Arg
465                 470                 475                 480
Phe Arg Arg Lys Ser Thr Glu Gln Glu Leu Met Val Val Gly Arg Val
                485                 490                 495
Phe Val Val Phe Leu Val Val Ile Ser Ile Leu Trp Ile Pro Ile Ile
            500                 505                 510
Gln Ser Ser Asn Ser Gly Gln Leu Phe Asp Tyr Ile Gln Ala Val Thr
        515                 520                 525
Ser Tyr Leu Ala Pro Pro Ile Thr Ala Leu Phe Leu Leu Ala Ile Phe
    530                 535                 540
```

```
Cys Lys Arg Val Thr Glu Pro Gly Ala Phe Trp Gly Leu Val Phe Gly
545                 550                 555                 560
Leu Gly Val Gly Leu Leu Arg Met Ile Leu Glu Phe Ser Tyr Pro Ala
                565                 570                 575
Pro Ala Cys Gly Glu Val Asp Arg Arg Pro Ala Val Leu Lys Asp Phe
            580                 585                 590
His Tyr Leu Tyr Phe Ala Ile Leu Leu Cys Gly Leu Thr Ala Ile Val
        595                 600                 605
Ile Val Ile Val Ser Leu Cys Thr Thr Pro Ile Pro Glu Glu Gln Leu
    610                 615                 620
Thr Arg Leu Thr Trp Trp Thr Arg Asn Cys Pro Leu Ser Glu Leu Glu
625                 630                 635                 640
Lys Glu Ala His Glu Ser Thr Pro Glu Ile Ser Glu Arg Pro Ala Gly
                645                 650                 655
Glu Cys Pro Ala Gly Gly Gly Ala Ala Glu Asn Ser Ser Leu Gly Gln
            660                 665                 670
Glu Gln Pro Glu Ala Pro Ser Arg Ser Trp Gly Lys Leu Leu Trp Ser
        675                 680                 685
Trp Phe Cys Gly Leu Ser Gly Thr Pro Glu Gln Ala Leu Ser Pro Ala
    690                 695                 700
Glu Lys Ala Ala Leu Glu Gln Lys Leu Thr Ser Ile Glu Glu Glu Pro
705                 710                 715                 720
Leu Trp Arg His Val Cys Asn Ile Asn Ala Val Leu Leu Leu Ala Ile
                725                 730                 735
Asn Ile Phe Leu Trp Gly Tyr Phe Ala
            740                 745
```

<210> 12
<211> 2217
<212> DNA
<213> homo sapiens

<400> 12

```
atggggcctg gagcttcagg ggacggggtc aggactgaga cagctccaca catagcactg 60
gactccagag ttggtctgca cgcctacgac atcagcgtgg tggtcatcta ctttgtcttc 120
gtcattgctg tggggatctg gtcgtccatc cgtgcaagtc gagggaccat tggcggctat 180
ttcctggccg ggaggtccat gagctggtgg ccaattggag catctctgat gtccagcaat 240
gtgggcagtg gcttgttcat cggcctggct gggacagggg ctgccggagg ccttgccgta 300
ggtggcttcg agtggaacat gaggaaatca aggtctggag gagacagagg gatccatcca 360
aggtcacacg ggaggactgg ggtcaggtcc caggtctctt atttctctgt tcgggggcct 420
cccacagcac agcactgcct ctgggtggga agccgcccct ctgtctacat ccaggacctg 480
gataccttct tcttctcccc actctcccag gcaacctggc tgctcctggc ccttggctgg 540
gtcttcgtcc ctgtgtacat cgcagcaggt gtggtcacaa tgccgcagta tctgaagaag 600
cgatttgggg gccagaggat ccaggtgtac atgtctgtcc tgtctctcat cctctacatc 660
ttcaccaaga tctcgactga catcttctct ggagccctct tcatccagat ggcattgggc 720
tggaacctgt acctctccac agggatcctg ctggtggtga ctgccgtcta caccattgca 780
ggtggcctca tggccgtgat ctacacagat gctctgcaga cggtgatcat ggtaggggga 840
gccctggtcc tcatgtttct gggctttcag gacgtgggct ggtacccagg cctggagcag 900
cggtacaggc aggccatccc taatgtcaca gtccccaaca ccacctgtca cctcccacgg 960
cccgatgctt tccacatgct tcgggaccct gtgagygggg acatcccttg gccaggtctc 1020
attttcgggc tcacagtgct ggccacctgg tgttggtgca cagaccaggt cattgtgcag 1080
cggtctctct cggccaagag tctgtctcat gccaagggag gctccgtgct gggggggctac 1140
ctgaagatcc tccccatgtt cttcatcgtc atgcctggca tgatcagccg ggccctgttc 1200
ccagacgagg tgggctgcgt ggaccctgat gtctgccaaa gaatctgtgg ggcccgagtg 1260
ggatgttcca acattgccta ccctaagttg gtcatggccc tcatgcctgt tggtctgcgg 1320
gggctgatga ttgccgtgat catggccgct ctcatgagct cactcacctc catcttcaac 1380
agcagcagca ccctgttcac cattgatgtg tggcagcgct tccgcaggaa gtcaacagag 1440
caggagctga tggtggtggg cagagtgttt gtggtgttcc tggttgtcat cagcatcctc 1500
tggatcccca tcatccaaag ctccaacagt gggcagctct tcgactacat ccaggctgtc 1560
accagttacc tggccccacc catcaccgct ctcttcctgc tggccatctt ctgcaagagg 1620
gtcacagagc ccggagcttt ctggggcctc gtgtttggcc tgggagtggg gcttctgcgt 1680
atgatcctgg agttctcata cccagcgcca gcctgtgggg aggtggaccg gaggccagca 1740
```

```
gtgctgaagg acttccacta cctgtacttt gcaatcctcc tctgcgggct cactgccatc 1800
gtcattgtca ttgtcagcct ctgtacaact cccatccctg aggaacagct cacacgcctc 1860
acatggtgga ctcggaactg ccccctctct gagctggaga aggaggccca cgagagcaca 1920
ccggagatat ccgagaggcc agccggggag tgccctgcag gaggtggagc ggcagagaac 1980
tcgagcctgg gccaggagca gcctgaagcc ccaagcaggt cctggggaaa gttgctctgg 2040
agctggttct gtgggctctc tggaacaccg gagcaggccc tgagcccagc agagaaggct 2100
gcgctagaac agaagctgac aagcattgag gaggagccac tctggagaca tgtctgcaac 2160
atcaatgctg tccttttgct ggccatcaac atcttcctct ggggctattt tgcgtga    2217

<210> 13
<211> 738
<212> PRT
<213> homo sapiens

<400> 13
Met Gly Pro Gly Ala Ser Gly Asp Gly Val Arg Thr Glu Thr Ala Pro
 1               5                  10                  15
His Ile Ala Leu Asp Ser Arg Val Gly Leu His Ala Tyr Asp Ile Ser
            20                  25                  30
Val Val Val Ile Tyr Phe Val Phe Val Ile Ala Val Gly Ile Trp Ser
        35                  40                  45
Ser Ile Arg Ala Ser Arg Gly Thr Ile Gly Gly Tyr Phe Leu Ala Gly
    50                  55                  60
Arg Ser Met Ser Trp Trp Pro Ile Gly Ala Ser Leu Met Ser Ser Asn
65                  70                  75                  80
Val Gly Ser Gly Leu Phe Ile Gly Leu Ala Gly Thr Gly Ala Ala Gly
                85                  90                  95
Gly Leu Ala Val Gly Gly Phe Glu Trp Asn Met Arg Lys Ser Arg Ser
            100                 105                 110
Gly Gly Asp Arg Gly Ile His Pro Arg Ser His Gly Arg Thr Gly Val
        115                 120                 125
Arg Ser Gln Val Ser Tyr Phe Ser Val Arg Gly Pro Pro Thr Ala Gln
    130                 135                 140
His Cys Leu Trp Val Gly Ser Arg Pro Ser Val Tyr Ile Gln Asp Leu
145                 150                 155                 160
Asp Thr Phe Phe Phe Ser Pro Leu Ser Gln Ala Thr Trp Leu Leu Leu
                165                 170                 175
Ala Leu Gly Trp Val Phe Val Pro Val Tyr Ile Ala Ala Gly Val Val
            180                 185                 190
Thr Met Pro Gln Tyr Leu Lys Lys Arg Phe Gly Gly Gln Arg Ile Gln
        195                 200                 205
Val Tyr Met Ser Val Leu Ser Leu Ile Leu Tyr Ile Phe Thr Lys Ile
    210                 215                 220
Ser Thr Asp Ile Phe Ser Gly Ala Leu Phe Ile Gln Met Ala Leu Gly
225                 230                 235                 240
Trp Asn Leu Tyr Leu Ser Thr Gly Ile Leu Leu Val Val Thr Ala Val
                245                 250                 255
Tyr Thr Ile Ala Gly Gly Leu Met Ala Val Ile Tyr Thr Asp Ala Leu
            260                 265                 270
Gln Thr Val Ile Met Val Gly Gly Ala Leu Val Leu Met Phe Leu Gly
        275                 280                 285
Phe Gln Asp Val Gly Trp Tyr Pro Gly Leu Glu Gln Arg Tyr Arg Gln
    290                 295                 300
Ala Ile Pro Asn Val Thr Val Pro Asn Thr Thr Cys His Leu Pro Arg
305                 310                 315                 320
Pro Asp Ala Phe His Met Leu Arg Asp Pro Val Ser Gly Asp Ile Pro
                325                 330                 335
Trp Pro Gly Leu Ile Phe Gly Leu Thr Val Leu Ala Thr Trp Cys Trp
            340                 345                 350
Cys Thr Asp Gln Val Ile Val Gln Arg Ser Leu Ser Ala Lys Ser Leu
        355                 360                 365
Ser His Ala Lys Gly Gly Ser Val Leu Gly Gly Tyr Leu Lys Ile Leu
    370                 375                 380
```

```
Pro Met Phe Phe Ile Val Met Pro Gly Met Ile Ser Arg Ala Leu Phe
385                 390                 395                 400
Pro Asp Glu Val Gly Cys Val Asp Pro Asp Val Cys Gln Arg Ile Cys
                405                 410                 415
Gly Ala Arg Val Gly Cys Ser Asn Ile Ala Tyr Pro Lys Leu Val Met
            420                 425                 430
Ala Leu Met Pro Val Gly Leu Arg Gly Leu Met Ile Ala Val Ile Met
        435                 440                 445
Ala Ala Leu Met Ser Ser Leu Thr Ser Ile Phe Asn Ser Ser Ser Thr
    450                 455                 460
Leu Phe Thr Ile Asp Val Trp Gln Arg Phe Arg Arg Lys Ser Thr Glu
465                 470                 475                 480
Gln Glu Leu Met Val Val Gly Arg Val Phe Val Val Phe Leu Val Val
                485                 490                 495
Ile Ser Ile Leu Trp Ile Pro Ile Ile Gln Ser Ser Asn Ser Gly Gln
            500                 505                 510
Leu Phe Asp Tyr Ile Gln Ala Val Thr Ser Tyr Leu Ala Pro Pro Ile
        515                 520                 525
Thr Ala Leu Phe Leu Leu Ala Ile Phe Cys Lys Arg Val Thr Glu Pro
    530                 535                 540
Gly Ala Phe Trp Gly Leu Val Phe Gly Leu Gly Val Gly Leu Leu Arg
545                 550                 555                 560
Met Ile Leu Glu Phe Ser Tyr Pro Ala Pro Ala Cys Gly Glu Val Asp
                565                 570                 575
Arg Arg Pro Ala Val Leu Lys Asp Phe His Tyr Leu Tyr Phe Ala Ile
            580                 585                 590
Leu Leu Cys Gly Leu Thr Ala Ile Val Ile Val Ile Val Ser Leu Cys
        595                 600                 605
Thr Thr Pro Ile Pro Glu Glu Gln Leu Thr Arg Leu Thr Trp Trp Thr
    610                 615                 620
Arg Asn Cys Pro Leu Ser Glu Leu Glu Lys Glu Ala His Glu Ser Thr
625                 630                 635                 640
Pro Glu Ile Ser Glu Arg Pro Ala Gly Glu Cys Pro Ala Gly Gly Gly
                645                 650                 655
Ala Ala Glu Asn Ser Ser Leu Gly Gln Glu Gln Pro Glu Ala Pro Ser
            660                 665                 670
Arg Ser Trp Gly Lys Leu Leu Trp Ser Trp Phe Cys Gly Leu Ser Gly
        675                 680                 685
Thr Pro Glu Gln Ala Leu Ser Pro Ala Glu Lys Ala Ala Leu Glu Gln
    690                 695                 700
Lys Leu Thr Ser Ile Glu Glu Glu Pro Leu Trp Arg His Val Cys Asn
705                 710                 715                 720
Ile Asn Ala Val Leu Leu Leu Ala Ile Asn Ile Phe Leu Trp Gly Tyr
                725                 730                 735
Phe Ala
```