

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number  
**WO 02/097095 A1**

(51) International Patent Classification<sup>7</sup>: **C12N 15/12**,  
C07K 14/435

(21) International Application Number: PCT/US02/16078

(22) International Filing Date: 21 May 2002 (21.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/293,710 25 May 2001 (25.05.2001) US

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

(72) Inventors: **WALKE, D., Wade**; 7507 Danehill Drive,  
Spring, TX 77389 (US). **FRIDDLE, Carl, Johan**; 127  
S. Goldenvine Circle, The Woodlands, TX 77382 (US).  
**SCOVILLE, John**; 8222 Kingsbrook Dr. #543, Houston,  
TX 77024 (US). **TURNER, C., Alexander, Jr.**; 67 Winter  
Wheat Place, The Woodlands, TX 77381 (US).

(74) Agents: **ISHIMOTO, Lance, K.** et al.; LEXICON GE-  
NETICS 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:**

- with international search report
- before the expiration of the time limit for amending the  
claims and to be republished in the event of receipt of  
amendments

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

(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.



**WO 02/097095 A1**

NOVEL HUMAN TRANSPORTER PROTEINS AND  
POLYNUCLEOTIDES ENCODING THE SAME

The present application claims the benefit of U.S. Provisional Application Number 60/293,710, which was  
5 filed on May 25, 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  
10 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  
15 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  
20 disclosed polynucleotides, which can be used for diagnosis, drug screening, clinical trial monitoring, the treatment of diseases and disorders, and cosmetic or nutraceutical applications.

2. BACKGROUND OF THE INVENTION

25 Transporter proteins are integral membrane proteins that mediate or facilitate the passage of materials across lipid bilayers. Given that the transport of materials across the membrane can play an important physiological role, transporter proteins are good drug  
30 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, 5 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 organic ion 10 transporters and sugar and cationic ion transporters. Organic transporters have been implicated in the transport and secretion of drugs, and drug resistance.

The novel human nucleic acid sequences described herein encode alternative proteins/open reading frames 15 (ORFs) of 547 and 476 amino acids in length (see, respectively, SEQ ID NOS:2 and 4).

The invention also encompasses agonists and antagonists of the described NHPs, including small molecules, large molecules, mutant NHPs, or portions 20 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 25 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 30 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. 35 When the unique NHP sequences described in SEQ ID NOS:1-

5 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:1-5 are "knocked-out" provide a 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.

Additionally, the unique NHP sequences described in SEQ ID NOS:1-5 are useful for the identification of protein coding sequences, and mapping a 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 NHP ORFs encoding the described NHP amino acid

sequences. SEQ ID NO:5 describes a NHP ORF and flanking sequences.

## 5. DETAILED DESCRIPTION OF THE INVENTION

5 The NHPs described for the first time herein are novel proteins that may be expressed in, *inter alia*, human cell lines, fetal brain, brain, pituitary, cerebellum, spinal cord, thymus, spleen, lymph node, bone marrow, trachea, lung, kidney, fetal liver,  
10 prostate, testis, thyroid, adrenal gland, pancreas, salivary gland, stomach, small intestine, colon, skeletal muscle, heart, uterus, placenta, mammary gland, adipose, skin, bladder, cervix, pericardium, hypothalamus, ovary, fetal kidney, fetal lung, gall  
15 bladder, tongue, aorta, 6-, 9-, and 12-week old embryos, adenocarcinoma, osteosarcoma, and embryonic carcinoma cells.

The present invention encompasses the nucleotides presented in the Sequence Listing, host cells expressing  
20 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  
25 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  
30 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  
35 of a 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<sub>4</sub>, 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 Nos. 5,837,458 and 5,723,323 both of which are herein incorporated by

reference in their entirety). The invention also includes degenerate nucleic acid variants of the disclosed NHP polynucleotide sequences.

Additionally contemplated are polynucleotides  
5 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  
10 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  
15 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  
20 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  
25 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  
30 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  
35 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:1-5 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:1-5, 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:1-5 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:1-5.

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.

5 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  
10 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  
15 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  
20 NOS:1-5 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.

25 Probes consisting of sequences first disclosed in SEQ ID NOS:1-5 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  
30 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.

35 As an example of utility, the sequences first disclosed in SEQ ID NOS:1-5 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  
5 in SEQ ID NOS:1-5 *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:1-5 can be used to identify mutations associated with a  
10 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  
15 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  
20 presence of one or more specific oligonucleotide sequence(s) first disclosed in SEQ ID NOS:1-5.

Alternatively, a restriction map specifying the relative positions of restriction endonuclease digestion sites, or various palindromic or other specific oligonucleotide  
25 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,  
30 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  
35 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  
5 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  
10 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  
15 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,  
20 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine,  
25 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,  
30 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-  
35 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.

5 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  
10 phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the antisense oligonucleotide is an  $\alpha$ -anomeric oligonucleotide. An  $\alpha$ -  
15 anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gautier *et al.*, 1987, Nucl. Acids Res. 15:6625-6641). The oligonucleotide is a 2'-O-methylribonucleotide (Inoue  
20 *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.

25 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  
30 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),  
35 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  
5 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  
10 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,  
15 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  
20 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  
25 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  
30 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  
35 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  
5 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  
10 sequences within the NHP products disclosed herein. The template for the reaction may be total RNA, mRNA, genomic DNA 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  
15 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  
20 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.

25 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  
30 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  
35 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*.

5        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  
10    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.  
15    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  
20    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  
25    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  
30    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  
35    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  
5 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  
10 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  
15 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  
20 phosphatase, and the promoters of the yeast  $\alpha$ -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  
25 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  
30 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  
35 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  
5 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  
10 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.

15 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  
20 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  
25 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 the endogenous NHP receptor. Nucleotide constructs encoding such NHP products can be  
30 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  
35 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.

5        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  
10 amino acid sequences of the described NHPs are presented in the Sequence Listing. The NHP nucleotides were obtained from clustered human ESTs, and cDNAs made from pituitary, bone marrow, brain, fetal kidney, adrenal gland, and mammary gland mRNAs (Edge Biosystems,  
15 Gaithersburg, MD).

Similar transporter sequences, as well as uses and applications that are germane to the described NHPs, are described in U.S. Patent Nos. 5,198,344, 5,866,699, and 6,080,842, which are herein incorporated by reference in  
20 their entirety.

A number of polymorphisms were detected during the sequencing of the NHP sequences, including: a C/T polymorphism at the region represented by nucleotide position number 288 of, for example, SEQ ID NO:1, both  
25 of which result in a ser being present at corresponding amino acid (aa) position 96 of, for example, SEQ ID NO:2; an A/C polymorphism at the region represented by nucleotide position number 485 of, for example, SEQ ID NO:1, or position number 272 of, for example, SEQ ID  
30 NO:3, which can result in a tyr or ser being present at corresponding aa position 162 of, for example, SEQ ID NO:2, or aa position 91 of, for example, SEQ ID NO:4; and a C/G polymorphism at the region represented by nucleotide position number 615 of, for example, SEQ ID  
35 NO:1, or position number 402 of, for example, SEQ ID NO:3, both of which result in a leu being present at

corresponding aa position 205 of, for example, SEQ ID NO:2, or aa position 134 of, for example, SEQ ID NO:4. The present invention contemplates sequences comprising any and all combinations and permutations of the above  
5 polymorphisms. As these polymorphisms are coding single nucleotide polymorphisms (SNPs), they are particularly useful in forensic analysis.

The disclosed NHPs are apparently encoded on human chromosome 1 (see GENBANK ACCESSION NO. AL365318).  
10 Accordingly, the described sequences are useful for mapping the coding regions of the human genome and biologically verifying mRNA splice junctions.

An additional application of the described novel human polynucleotide sequences is their use in the  
15 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  
20 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  
25 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  
30 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  
35 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  
5 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  
10 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  
15 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.

20 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  
25 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).

30 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  
35 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  
5 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  
10 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  
15 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  
20 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  
25 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  
30 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

35 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 cancer.

15       The Sequence Listing discloses the amino acid sequences encoded by the described NHP polynucleotides. The NHPs display initiator methionines in DNA sequence contexts consistent with a translation initiation site.

20       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  
5 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  
10 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, phosphorylation,  
15 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,  
20 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  
25 (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)  
30 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  
35 invention. Where, as in the present instance, the NHP  
peptide or polypeptide is thought to be 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  
5 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  
10 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.

15 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  
20 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  
25 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  
30 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  
35 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  
5 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  
10 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  
15 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  
20 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,  
25 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  
30 (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  
35 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  
5 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,  
10 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  
15 (Lowy *et al.*, 1980, *Cell* 22:817) genes, which can be employed in tk<sup>-</sup>, hgp<sup>r</sup>t<sup>-</sup> or ap<sup>r</sup>t<sup>-</sup> 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,  
20 *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  
25 (Colberre-Garapin *et al.*, 1981, *J. Mol. Biol.* 150:1); and hyg<sup>r</sup>, 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  
30 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  
35 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  $\text{Ni}^{2+}$ -nitriloacetic acid-agarose columns, and  
5 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  
10 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  
15 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.  
20 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  
25 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  
30 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  
35 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  
5 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  
10 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  
15 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  
20 antibodies, Fab fragments, F(ab')<sub>2</sub> 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  
25 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  
30 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  
35 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  
5 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.  
10 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  
15 (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  
20 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.  
25 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  
30 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.  
35 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.

10 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.

20 25 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.

30 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

35

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  
5 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  $F(ab')_2$   
10 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  
15 generate anti-idiotypic 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  
20 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  
25 be used in therapeutic regimens involving a NHP-mediated pathway.

Additionally given the high degree of relatedness of mammalian NHPs, NHP knock-out mice (having never seen a NHP, and thus never been tolerized to a NHP) have a  
30 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  
5 ID NO:3.
2. An isolated nucleic acid molecule comprising a nucleotide sequence that:
  - (a) encodes the amino acid sequence shown in  
10 SEQ ID NO:2; and
  - (b) hybridizes under highly stringent conditions to the nucleotide sequence of SEQ ID NO:1 or the complement thereof.
- 15 3. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes the amino acid sequence shown in SEQ ID NO:2.
- 20 4. An isolated nucleic acid molecule comprising a nucleotide sequence that encodes the amino acid sequence shown in SEQ ID NO:4.
- 25 5. A substantially isolated protein having the activity of the protein shown in SEQ ID NO:2 or SEQ ID NO:4, which is encoded by a nucleotide sequence that hybridizes to SEQ ID NO:1 or SEQ ID NO:3 under highly stringent conditions.

## SEQUENCE LISTING

&lt;110&gt; LEXICON GENETICS INCORPORATED

&lt;120&gt; Novel Human Transporter Proteins and Polynucleotides Encoding the Same

&lt;130&gt; LEX-0346-PCT

&lt;150&gt; US 60/293,710

&lt;151&gt; 2001-05-25

&lt;160&gt; 5

&lt;170&gt; FastSEQ for Windows Version 4.0

&lt;210&gt; 1

&lt;211&gt; 1644

&lt;212&gt; DNA

&lt;213&gt; homo sapiens

&lt;400&gt; 1

```

atggaggtgg aggagggcgtt ccaggcgggtg ggggagatgg gcatctacca gatgtacttg 60
tgcttcctgc tggccgtgct gctgcagctc tacgtggcca cggaggccat cctcattgca 120
ctgggtgggg ccacgccatc ctaccactgg gacctggcag agtcctgcc aaatcagagc 180
cacggtaacc agtcagctgg tgaagaccag gcctttgggg actggctcct gacagccaac 240
ggcagtgaga tccataagca cgtgcatttc agcagcagct tcacctccat cgcctcggag 300
tggtttttaa ttgccaacag atcctacaaa gtcagtgcag caagctcttt tttcttcagt 360
gggtgtatttg ttggagttat ctcttttggg cagctttcag atcgcttcgg aaggaaaaaa 420
gtctatctca cagggttttg tcttgacatc ttatttgcaa ttgcaaattg attttcccc 480
tcatatgagt tctttgcagt aactcgcttc ctgggtgggca tgatgaatgg agggatgtcg 540
ctgggtggcct ttgtcttgct taatgaatgt gtgggcaccg cctactgggc acttgcagga 600
tcgattggcg gcctcttctt tgcagttggc attgccaat atgccctgtt aggatacttc 660
atccgctcct ggaggaccct agccattctg gttaacctgc agggaaacgt ggtctttctc 720
ttatctttat tcattcctga atcacctcgt tggttatact cccagggtcg actgagttag 780
gctgaagagg cgctgtacct cattgccaa aggaaccgca aactcaagtg cacgttctca 840
ctaacacacc cagccaacag gagctgcagg gagactggaa gtttcctgga tctctttcgt 900
taccgggtcc tgtaggaca cactttgatc ctgatgttca tctggtttgt gtgcagcttg 960
gtgtattatg gcctaactct gagtgcgggt gatctagggt gaagtattta tgccaacctg 1020
gccctgtctg gcctcataga gattccatct taccctctct gtatctactt gattaaccaa 1080
aaatgggttg gtcggaagcg aacattatca gcatttctgt gcctaggagg actggcttgt 1140
cttattgtaa tgtttcttcc agaaaagaaa gacacaggtg tgtttgcagt ggtgaacagc 1200
cattccttgt ccttgctggg gaagctgacc atcagtgtct cctttaacat tgtttatatc 1260
tacacctctg agctttaccc tacagtcac aggaatgttg ggcttggaac ttgttccatg 1320
ttctcccagag ttggtgggat tattgtctcc ttcacccct cactgaaata tgtgcaatgg 1380
tctttaccat tcattgtctt cggagccacg ggtctgacct ccggcctcct gagtttgtta 1440
ttgccggaga cccttaacag tccgctgcta gaaacattct ccgaccttca ggtgtattcg 1500
tatcgcaggc tgggagaaga agcattatct ttacaggctt tggaccccca acagtgtgtg 1560
gacaaggaga gctctttagg gagtgagagt gaggaagagg aagaatttta tgatgcagat 1620
gaagagactc agatgatcaa gtga                                     1644

```

&lt;210&gt; 2

&lt;211&gt; 547

&lt;212&gt; PRT

&lt;213&gt; homo sapiens

&lt;400&gt; 2

```

Met Glu Val Glu Glu Ala Phe Gln Ala Val Gly Glu Met Gly Ile Tyr
  1                      5                      10                      15
Gln Met Tyr Leu Cys Phe Leu Leu Ala Val Leu Leu Gln Leu Tyr Val
                20                25                30
Ala Thr Glu Ala Ile Leu Ile Ala Leu Val Gly Ala Thr Pro Ser Tyr

```

2 / 6

Met Ile Lys  
545

<210> 3  
<211> 1431  
<212> DNA  
<213> homo sapiens

<400> 3  
atggagggtgg aggaggcggtt ccaggcggtg ggggagatgg gcatctacca gatgtacttg 60  
tgcttctctgc tggccgtgct gctgcagtgg tttttaattg ccaacagatc ctacaaagtc 120  
agtgcagcaa gctctttttt cttcagtggt gtatttggtg gagttatctc ttttggtcag 180  
ctttcagatc gcttcggaag gaaaaaagtc tatctcacag gttttgctct tgacatctta 240  
tttgcaattg caaatggatt ttccccctca tatgagttct ttgcagtaac tcgcttcctg 300  
gtgggcatga tgaatggagg gatgtcgctg gtggcctttg tcttgcttaa tgaatgtgtg 360  
ggcaccgcct actgggcact tgcaggatcg attggcggcc tcttctttgc agttggcatt 420  
gccaatatg ccctgttagg atacttcacg cgctcctgga ggaccctagc cattctggtt 480  
aacctgcagg gaacgggtgg ttttctctta tctttattca ttctgaatc acctcgttgg 540  
ttatactccc agggctcgact gagtgaggct gaagaggcgc tgtacctcat tgccaagagg 600  
aaccgcaaac tcaagtgcac gttctcacta acacaccag ccaacaggag ctgcagggag 660  
actggaagtt tcctggatct ctttcgttac cgggtcctgt taggacacac tttgatcctg 720  
atgttcatct ggtttgtgtg cagcttggtg tattatggcc taactctgag tgcgggtgat 780  
ctagggtggaa gtatttatgc caacctggcc ctgtctggcc tcatagagat tccatcttac 840  
cctctctgta tctacttgat taacaaaaaa tggtttggtc ggaagcgaa attatcagca 900  
tttctgtgcc taggaggact ggcttgtctt attgtaatgt ttcttccaga aaagaaaagac 960  
acaggtgtgt ttgcagtggt gaacagccat tccttgtcct tgctggggaa gctgaccatc 1020  
agtgtgcct ttaacattgt ttatatctac acctctgagc tttaccctac agtcatcagg 1080  
aatgttgggc ttggaacttg ttccatgttc tcccagattg gtgggattat tgctcccttc 1140  
atccccctac tgaaatatgt gcaatggtct ttaccattca ttgtcttcgg agccacgggt 1200  
ctgacctccg gcctcctgag tttgttattg cgggagacc ttaacagtcc gctgctagaa 1260  
acattctccg accttcaggt gtattcgtat cgcaggctgg gagaagaagc attatcttta 1320  
caggcttttg accccaaca gtgtgtggac aaggagagct ctttagggag tgagagtga 1380  
gaagaggaag aattttatga tgcagatgaa gagactcaga tgatcaagtg a 1431

<210> 4  
<211> 476  
<212> PRT  
<213> homo sapiens

<400> 4  
Met Glu Val Glu Glu Ala Phe Gln Ala Val Gly Glu Met Gly Ile Tyr  
1 5 10 15  
Gln Met Tyr Leu Cys Phe Leu Leu Ala Val Leu Leu Gln Trp Phe Leu  
20 25 30  
Ile Ala Asn Arg Ser Tyr Lys Val Ser Ala Ala Ser Ser Phe Phe Phe  
35 40 45  
Ser Gly Val Phe Val Gly Val Ile Ser Phe Gly Gln Leu Ser Asp Arg  
50 55 60  
Phe Gly Arg Lys Lys Val Tyr Leu Thr Gly Phe Ala Leu Asp Ile Leu  
65 70 75 80  
Phe Ala Ile Ala Asn Gly Phe Ser Pro Ser Tyr Glu Phe Phe Ala Val  
85 90 95  
Thr Arg Phe Leu Val Gly Met Met Asn Gly Gly Met Ser Leu Val Ala  
100 105 110  
Phe Val Leu Leu Asn Glu Cys Val Gly Thr Ala Tyr Trp Ala Leu Ala  
115 120 125  
Gly Ser Ile Gly Gly Leu Phe Phe Ala Val Gly Ile Ala Gln Tyr Ala  
130 135 140  
Leu Leu Gly Tyr Phe Ile Arg Ser Trp Arg Thr Leu Ala Ile Leu Val  
145 150 155 160  
Asn Leu Gln Gly Thr Val Val Phe Leu Leu Ser Leu Phe Ile Pro Glu  
165 170 175



Ser Pro Arg Trp Leu Tyr Ser Gln Gly Arg Leu Ser Glu Ala Glu Glu  
 180 185 190  
 Ala Leu Tyr Leu Ile Ala Lys Arg Asn Arg Lys Leu Lys Cys Thr Phe  
 195 200 205  
 Ser Leu Thr His Pro Ala Asn Arg Ser Cys Arg Glu Thr Gly Ser Phe  
 210 215 220  
 Leu Asp Leu Phe Arg Tyr Arg Val Leu Leu Gly His Thr Leu Ile Leu  
 225 230 235 240  
 Met Phe Ile Trp Phe Val Cys Ser Leu Val Tyr Tyr Gly Leu Thr Leu  
 245 250 255  
 Ser Ala Gly Asp Leu Gly Gly Ser Ile Tyr Ala Asn Leu Ala Leu Ser  
 260 265 270  
 Gly Leu Ile Glu Ile Pro Ser Tyr Pro Leu Cys Ile Tyr Leu Ile Asn  
 275 280 285  
 Gln Lys Trp Phe Gly Arg Lys Arg Thr Leu Ser Ala Phe Leu Cys Leu  
 290 295 300  
 Gly Gly Leu Ala Cys Leu Ile Val Met Phe Leu Pro Glu Lys Lys Asp  
 305 310 315 320  
 Thr Gly Val Phe Ala Val Val Asn Ser His Ser Leu Ser Leu Leu Gly  
 325 330 335  
 Lys Leu Thr Ile Ser Ala Ala Phe Asn Ile Val Tyr Ile Tyr Thr Ser  
 340 345 350  
 Glu Leu Tyr Pro Thr Val Ile Arg Asn Val Gly Leu Gly Thr Cys Ser  
 355 360 365  
 Met Phe Ser Arg Val Gly Gly Ile Ile Ala Pro Phe Ile Pro Ser Leu  
 370 375 380  
 Lys Tyr Val Gln Trp Ser Leu Pro Phe Ile Val Phe Gly Ala Thr Gly  
 385 390 395 400  
 Leu Thr Ser Gly Leu Leu Ser Leu Leu Leu Pro Glu Thr Leu Asn Ser  
 405 410 415  
 Pro Leu Leu Glu Thr Phe Ser Asp Leu Gln Val Tyr Ser Tyr Arg Arg  
 420 425 430  
 Leu Gly Glu Glu Ala Leu Ser Leu Gln Ala Leu Asp Pro Gln Gln Cys  
 435 440 445  
 Val Asp Lys Glu Ser Ser Leu Gly Ser Glu Ser Glu Glu Glu Glu  
 450 455 460  
 Phe Tyr Asp Ala Asp Glu Glu Thr Gln Met Ile Lys  
 465 470 475

&lt;210&gt; 5

&lt;211&gt; 5166

&lt;212&gt; DNA

&lt;213&gt; homo sapiens

&lt;400&gt; 5

ttctctgaag gctctagggc tccccagcc cccgcattaa actggtttgg gcaaacgac 60  
 aggcattgct gagagccgcc aggcgcgcg ccagctgcgg gtcccctgta ggcagagccc 120  
 tgggtcacac gcggacgcgg tcccgatgtt tcttctcagg ccctcagccc ggcccacttc 180  
 ctctttcctc actttaaggc gcagacagcg gtccctctca aaacaaaacg tacacaaaaa 240  
 caccatTTTT tttttaaaca ggaagtccag aggcccggtt ccagggcatc cagccccgga 300  
 tctcgaacgc cctcctcgcc cctcgccccg cgtagcccgg ctgcgctccc cgctccccgc 360  
 ccggcgtcca cctctcgagc ctgacagccc cgccgcccgc gcaggcgccc cgccccggcc 420  
 cggcccgccc cgccccgcc gccagcgctt ccattccccgc cccggcggtt ccaagccggt 480  
 gccgggcgcc caggggttgc cgcgctgggc gggagggcag cgcctgagag ggcggtgggg 540  
 tggcggggtt cctgcgcgcg gcccgccatg gaggtggagg aggcgttcca ggcggtgggg 600  
 gagatgggca tctaccagat gtacttgtgc ttcctgctgg ccgtgctgct gcagctctac 660  
 gtggccacgg aggccatcct cattgcactg gttggggcca cgccatccta cactggggac 720  
 ctggcagagc tcctgcaaaa tcagagccac ggtaaccagt cagctggtga agaccaggcc 780  
 tttggggact ggctcctgac agccaacggc agtgagatcc ataagcacgt gcatttcagc 840  
 agcagcttca cctccatcgc ctccggagtgg tttttaattg ccaacagatc ctacaaagtc 900  
 agtgcagcaa gctctTTTT cttcagtggt gtatttgttg gagttatctc ttttggtcag 960  
 ctttcagatc gcttcggaag gaaaaaagtc tatctcacag gttttgctct tgacatctta 1020

tttgcaattg	caaattggatt	ttccccctca	tatgagttct	ttgcagtaac	tcgcttcctg	1080
gtgggcatga	tgaatggagg	gatgtcgtcg	gtggcctttg	tcttgcttaa	tgaaatgtgtg	1140
ggcaccgcct	actgggcact	tgcaggatcg	attggcggcc	tcttctttgc	agttggcatt	1200
gccaatatg	ccctggttag	atacttcatc	cgctcctgga	ggaccctagc	cattctgggt	1260
aacctgcagg	gaacgggtgt	ctttctctta	cttttattca	ttcctgaatc	acctcgttgg	1320
ttatactccc	agggctcgact	gagtgaggct	gaagaggcgc	tgtacctcat	tgccaagagg	1380
aaccgcaaac	tcaagtgcac	gttctcacta	acacaccag	ccaacaggag	ctgcaggagg	1440
actggaagtt	tcctggatct	ctttcgttac	cgggtcctgt	taggacacac	tttgatcctg	1500
atgttcatct	ggttttgtgtg	cagcttggtg	tattatggcc	taactctgag	tgcggttgat	1560
ctaggtggaa	gtattttatgc	caacctggcc	ctgtctggcc	tcatagagat	tccatcttac	1620
cctctctgta	tctacttgat	taacccaaaa	tggtttggtc	ggaagcgaac	attatcagca	1680
tttctgtgcc	taggaggact	ggcttgtctt	attgtaatgt	ttcttcagca	aaagaaagac	1740
acaggtgtgt	ttgcagtggt	gaacagccat	tccttgtcct	tgctggggaa	gctgaccatc	1800
agtgtgcct	ttaacattgt	ttatatctac	acctctgagc	ttaccctac	agtcacagg	1860
aatgttgggc	ttggaacttg	ttccatgttc	tcccagattg	gtgggattat	tgctcccctc	1920
atccccctac	tgaaatatgt	gcaatggctc	ttaccattca	ttgtcttcgg	agccacgggt	1980
ctgacctccg	gcctcctgag	tttgttattg	cggagagccc	ttaacagtcc	gctgctagaa	2040
acattctccg	accttcagggt	gtattcgtat	cgcaggctgg	gagaagaagc	attatcttta	2100
caggcttttg	acccccaaaca	gtgtgtggac	aaggagagct	ctttaggagg	tgagagttag	2160
gaagaggaag	aatttttatga	tgcagatgaa	gagactcaga	tgatcaagtg	aagagcccca	2220
gattccccct	aagaagcaaa	ggatcgtctt	ttatgcctct	ggctaaggca	ggttcttcca	2280
tgactcctaa	gagagttgaa	aaaatagagg	cttggcttga	atgtacatag	atggtacctg	2340
gcatggactg	atgttttttag	gcacagaagt	tggagaagag	atttcatgaa	agacaacatc	2400
actgcattga	gagaatagtt	gttaattttg	ttagaattta	agttctactc	agaatcataa	2460
catctggcag	aacagcccaa	acctcacattc	caaagtggta	ggctcatttg	tttctagaga	2520
tttcatcatg	tcgctttttcc	ttcatcatga	tctaaataaa	ggcagatatg	taaaatttct	2580
caccattttg	gtggggtaag	ataagctatt	attaagattt	aatccttata	ccatgttggg	2640
catttgcccc	tatcagttgc	tcctcaggaa	tcttctggta	caggttaaca	tcagcatttt	2700
cattttgtat	ccagggaaaa	gcacccagag	gtcatctgtg	tgtcccgaga	ccctccagct	2760
ttttcttagc	tgatgaaata	tgagtccctca	gcttgggtcc	cagcctgctg	attgacttgg	2820
gctgctgggtg	ccttgagtcc	cacagatgat	tcattaggaa	aagccagatg	taccaaagcg	2880
gtttactcag	agtcaggggt	gtagctctgg	ctgcctgtca	gctccccttg	atactataat	2940
gtatgatttc	ttccttttccc	actaatatgc	acatccagaa	aaatttccat	ctgagattct	3000
agtacttcaa	aatcatgcac	agtaaatgag	aaagctttaa	gtagagggca	gttaaaccgt	3060
gacatgttga	gcacctggag	gaaaaaaaag	gtgcagtttt	taataagaga	gaaaatgaaa	3120
ttatctttga	taaatttttg	tttgttttgc	tttcagcatt	gtgccatgag	ggatttggac	3180
aatattttaag	aacttcttgt	cctagatcag	ccccaatctg	tttaatcaaa	atggaagggt	3240
cagtaatttc	atgggaaacc	ttgggttttc	attaagtgtc	accaactttc	aagtgaatct	3300
tgtatttgat	ttcctaaaaat	catgtcttga	aaacatgttt	tctcatgaaa	cttgaatact	3360
atctcaaata	ggaatataaa	cctggagtca	acaagcttag	gcagcattga	tttaggtcac	3420
tttcccagtg	aggaaaattt	ctgtgttttc	agaattttca	tttctactaa	cctcttggag	3480
aaaaagaaat	tgaatatagag	gtaaatagaa	gacgtcactg	tggctgcttc	tgggaagtgc	3540
ggaagcatca	cccccaattgg	ctccaaatac	tgtcatgttt	tcttgcacac	tgacttcttg	3600
tttccactgt	atcagtatgt	acctttgtaa	ttgttatttt	tatgtctttt	atgcccttga	3660
ttattagttg	ggctcttcat	aaacagaggc	catctctact	actgtttatt	tttccctgct	3720
gtgcccagaa	cattggcgta	gacacagtaa	gaacctagta	aatattactg	tttctagcca	3780
tcagggagat	tgtggaactc	ctcccagtat	aattttttaca	aactccaagc	aaatctgacc	3840
caaactccca	aattgtcaag	tcctgcttaa	ctttctcttg	aaaatagacc	ccttctcaac	3900
atcagaatag	gaagagagga	agaacttaca	aatagactta	aaagttattc	ttaaatgggt	3960
gttgggcatt	taaaacagtg	aactaacata	tataataatt	ttgattagtt	ggagctttct	4020
ttgtatttatg	agagtaatat	atctcattac	agaaaaattg	gaaactataa	atttagaaac	4080
gtatcaccca	tacgtccaac	atcgaaagaa	aaccagtgtt	atgactttgt	tccatttgaa	4140
gactaattgg	gagtccatct	ctctattggc	actgggttcg	attgcctctg	gctaataagag	4200
ttcaattagt	tctatccctg	ggtttccctt	cttagctatg	gggtggaaga	taggaggggg	4260
agatctacaa	tttgaatatg	tgttacttaa	taaggctagg	ctggccatca	gttgcttatt	4320
tcagatgtgt	cactaaattt	tccttctaga	tggtccttga	gcaaaactta	ataattactg	4380
ttttttattt	ctactgcctt	tataaaatca	aaattttctc	cttttgataa	aaactgttga	4440
atactattga	tgtagagaat	gtgtatatgt	gtataatttgc	attgattaaa	ttattggaaa	4500
acttttcat	gacaggtaat	gtattttggg	tgtgcggatg	gtttgtaaaa	agtatcctgt	4560
tttttttaatt	ctattcaaaa	ttaaaaattt	tttttcttaa	aataaagcaa	gaaaaataaa	4620
tgtaagatat	ataagaaacc	ccagagtttc	caggggtgaa	gttcacatta	gcagcttgtc	4680
caaaatatatt	acacctcaca	aatatttttg	aagaaaccat	gttgaaattt	tctgattttg	4740
tgagttacaa	ccacagccca	gtaaaaccaa	tacttgttat	ttggctactt	tcaagcaatg	4800

```
cagatattct ttagtattta aaaatgagca aataaacaaa atgaggcaaa taaatgaaga 4860
aaatgagttt ctgagacact gtttaggtta ccataaatca gacagatatt tttatggtgg 4920
gcaaattgat gaagatactt ttagtaggtg atttatattc ataacaaata catattgaaa 4980
taattgtcac tactgtatca gaattaaaaa ccagcttcac attaggactt ctccatacaa 5040
atgctttatc attggttcct aaattggaaa aatatctata aatgattaca tttttaagct 5100
aaataaaatg cagatatttt ggcttcttgt gaatgtaata aagaatcata aaacataaaa 5160
aaaaaa
```

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US02/16078

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 15/12; C07K 14/435

US CL : 536/23.5; 530/350

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.5; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

GENCORE

search terms: SEQ ID NO: 1 &amp; 2

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- A	Database GENBANK/EMBL/DBJ, AN Q9NSH5. ANSORGE, W. et al. 'Hypothetical protein'. April 2000 .	5 -- 1-3



Further documents are listed in the continuation of Box C.



See patent family annex.

"	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

02 OCTOBER 2002

Date of mailing of the international search report

Authorized officer

ROBERT C. HAYES, PH.D.

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US02/16078

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
1-3 & 5 (in-part)

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US02/16078

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-3 and 5, drawn to nucleic acids encoding SEQ ID NO:2, and the protein of SEQ ID NO:2.

Group II, claim(s) 1 and 4-5, drawn to nucleic acids encoding SEQ ID NO:4, and the protein of SEQ ID NO:4.

The inventions listed as Groups I-II do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Group I is directed toward nucleic acids encoding SEQ ID NO:2, which is the first product and method of using the product. The special technical feature is the nucleic acid encoding SEQ ID NO:2. Group II is drawn to structurally different products, which do not require each other for their practice and do not share the same or a corresponding technical feature. Note that PCT Rule 13 does not provide for multiple products or methods within a single application. Since the special technical feature of the Group I invention is not present in the Group II claims, and the special technical feature of the Group II invention is not present in the Group I claims, unity of invention is lacking.