

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 February 2003 (27.02.2003)

PCT

(10) International Publication Number
WO 03/016481 A2

(51) International Patent Classification⁷: **C12N**

(21) International Application Number: PCT/US02/26031

(22) International Filing Date: 14 August 2002 (14.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/312,300 14 August 2001 (14.08.2001) US

(71) Applicant: **LEXICON GENETICS INCORPORATED**
[US/US]; 8800 Technology Forest Place, The Woodlands,
TX 77381 (US).

(72) Inventors: **YU, Xuanchuan**; 6 Danville Crossing, Conroe,
TX 77385 (US). **XIE, Qiongshu**; 1912 Carroll #1, Hous-
ton, TX 77030 (US). **HU, Yi**; 333 Holly Creek Ct. #203,
The Woodlands, TX 77381 (US).

(74) Agent: **ISHIMOTO, Lance, K.**; Lexicon Genetics Incor-
porated, 8800 Technology Forest Place, 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, VC,
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, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK,
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 "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: NOVEL HUMAN COLLAGEN 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 03/016481 A2

NOVEL HUMAN COLLAGEN PROTEINS AND
POLYNUCLEOTIDES ENCODING THE SAME

The present application claims the benefit of U.S.
5 Provisional Application Number 60/312,300 which was filed on
August 14, 2001 and is herein incorporated by reference in its
entirety.

1. INTRODUCTION

The present invention relates to the discovery,
10 identification, and characterization of novel human
polynucleotides encoding proteins that share sequence
similarity with animal collagen proteins. The invention
encompasses the described polynucleotides, host cell
expression systems, the encoded proteins, fusion proteins,
15 polypeptides and peptides, antibodies to the encoded proteins
and peptides and genetically engineered animals that either
lack or overexpress the disclosed genes, antagonists and
agonists of the proteins, and other compounds that modulate
the expression or activity of the proteins encoded by the
20 disclosed genes, 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 Collagens are a family of proteins that are among the
most abundant proteins in the body. Biosynthetically produced
collagens find medical utility in prosthetic and cosmetic
applications.

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 animal collagens, including, but not limited to, the human collagen alpha 2 (VIII) chain. As such, the novel polynucleotides encode novel collagen proteins having homologues and orthologs across a range of phyla and species.

The novel human polynucleotides described herein encode alternative open reading frames (ORFs) encoding proteins of 717 and 703 amino acids in length (see 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 thereof, that compete with native NHP, 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 cells ("ES cells") 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:1-4 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:1-4 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.

5 Additionally, the unique NHP sequences described in SEQ ID NOS:1-4 are useful for the identification of protein coding sequence and mapping a unique gene to a particular chromosome. These sequences identify actual, biologically verified, and therefore relevant, exon splice junctions as opposed to those that may have been bioinformatically predicted from genomic
10 sequence alone. The sequences of the present invention are also useful as additional DNA markers for restriction fragment length polymorphism (RFLP) analysis, as is utilized in, *intra alia*, population and forensic biology.

Further, the present invention also relates to processes
15 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 product, or cells expressing the same. Such compounds can be used as therapeutic agents for the treatment of any of a wide
20 variety of symptoms associated with biological disorders or imbalances.

4. DESCRIPTION OF THE SEQUENCE LISTING AND FIGURES

The Sequence Listing provides the nucleic acid and amino
25 acid sequences of novel human ORFs that encode the described novel human collagen proteins. SEQ ID NO:1 is a nucleic acid sequence that encodes the amino acid sequence of SEQ ID NO:2. Nucleic acid SEQ ID NO:3 is a transcript that begins at an internal ATG, located at position 43, of SEQ ID NO:1. SEQ ID
30 NO:3 is therefore fully contained within SEQ ID NO:1. SEQ ID NO:4 is the amino acid sequence encoded by SEQ ID NO:3.

5. DETAILED DESCRIPTION OF THE INVENTION

The NHP described for the first time herein, is a novel protein that is expressed in, *inter alia*, human cell lines, and human pituitary, lymph node, fetal kidney, and osteocarcinoma cells. The described sequences were compiled from human genomic sequence and cDNAs made from human fetal lung and lymph node mRNAs (Edge Biosystems, Gaithersburg, MD). 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 gene, including the specifically described NHPs, and related NHP products; (b) nucleotides that encode one or more portions of a NHP corresponding to a NHP functional domain(s), 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 NHP 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 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: (a) 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 NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel F.M. *et al*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., NY, at p. 2.10.3) and encodes a functionally equivalent expression product. Additionally contemplated are any nucleotide sequences that hybridize to the complement of the DNA sequence that encode and express an amino acid sequence presented in the Sequence Listing under moderately stringent conditions, *e.g.*, washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel *et al*, 1989, *supra*), yet still encode 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 herein incorporated by reference). The invention also includes degenerate nucleic acid variants of the disclosed NHP polynucleotide sequence.

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 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 gene nucleotide sequences. Such hybridization conditions may be highly stringent or less highly stringent, as described above. 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, 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
5 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
10 gene expression patterns (particularly using a micro array or high-throughput "chip" format). Additionally, a series of the described 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
15 sequence first disclosed in at least a portion of one or more of the sequences of SEQ ID NOS: 1 and 3 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
20 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
25 array comprises an oligonucleotide or polynucleotide sequence first disclosed in at least one of the sequences of SEQ ID NOS: 1 and 3, or an amino acid sequence encoded thereby. Methods for attaching biopolymers to, or synthesizing biopolymers on, solid support matrices, and conducting binding
30 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.

35 Addressable arrays comprising sequences first disclosed in SEQ ID NOS:1-4 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 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 and 3.

For example, a series of the described 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:1-4 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:1-4 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 drugs intended target. 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:1-4 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:1-4 *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:1-4 can be used to identify mutations associated with a particular disease and also as a diagnostic or prognostic assay.

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 the SEQ ID NOS: 1 and 3. 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.

5 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
10 NHP gene antisense molecules, useful, for example, in NHP gene regulation and/or as antisense primers in amplification reactions of NHP gene 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
15 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 which is selected from the group including, but not limited
20 to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine,
25 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,
30 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-
35 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.

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 phosphoramidate, a phosphordiamidate,
10 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
15 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'-O-methylribonucleotide (Inoue *et al*, 1987, Nucl. Acids Res. 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue *et al*,
20 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
25 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
30 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
35 sequences are derived. For guidance regarding such conditions see, for example, Sambrook *et al*, 1989, Molecular Cloning, A

Laboratory Manual (and periodic updates thereof), Cold Spring Harbor Press, NY; and Ausubel et al, 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, NY.

5 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,
10 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
15 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.

20 In another 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
25 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
30 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 and their descendants). Actual base sequence information can be used for identification as an accurate
35 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 total RNA, mRNA, and/or cDNA obtained by reverse transcription of mRNA prepared from human or non-human cell lines or tissue known or suspected to express 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, or suspected, to express a NHP gene, such as, for example, testis tissue). 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 or suspected to be expressed 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 or known to carry a mutant NHP allele (e.g., a person manifesting a NHP-associated phenotype such as, for example, paralysis or palsy, nerve damage or degeneration, an inflammatory disorder, vision disorders, etc.), or a cDNA library can be constructed using RNA from a tissue known, or suspected, to express 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, or suspected, to express a mutant NHP allele in an individual suspected of 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, E. and Lane, eds., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor, NY.

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 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 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),
5 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
10 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 NHP or NHP peptides, NHP fusion proteins, NHP nucleotide sequences, antibodies, antagonists and agonists can
15 be useful for the detection of mutant NHPs or inappropriately expressed NHPs for the diagnosis of disease. The NHP or NHP peptides, NHP fusion proteins, NHP nucleotide sequences, host cell expression systems, antibodies, antagonists, agonists and genetically engineered cells and animals can be used for
20 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 NHP in the body. The use of engineered host cells and/or animals may offer an advantage in that such
25 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.
30 For example, soluble derivatives such as a mature NHP, or NHP peptides/domains corresponding to the NHP, 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),
35 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 soluble NHP, or a NHP-IgFc fusion protein or an anti-idiotypic antibody (or its Fab) that mimics the NHP could activate or effectively
5 antagonize the endogenous NHP receptor. Soluble NHP can also be modified by proteolytic cleavage to active peptide products (e.g., any novel peptide sequence initiating at any one of the amino acids presented in the Sequence Listing and ending at any downstream amino acid). Such products or peptides can be
10 further subject to modification such as the construction of NHP fusion proteins and/or can be derivatized by being combined with pharmaceutically acceptable agents such as, but not limited to, polyethylene glycol (PEG).

Nucleotide constructs encoding such NHP products can be
15 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 a functional NHP, mutant NHPs,
20 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.

25 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
30 acid sequence of the described NHP are presented in the Sequence Listing. The NHP nucleotide sequences were obtained from cDNAs obtained using probes and/or primers generated from human genomic sequence.

Expression analysis has provided evidence that the
35 described NHPs are expressed in a limited number of tissues, and that the NHPs share significant similarity with human

collagens. Given the physiological importance of collagen proteins, they have been subject to intense scrutiny as exemplified and discussed in U.S. Patent Nos. 5,925,736 and 5,807,581 herein incorporated by reference in their entirety which additionally describe a variety of uses and applications applicable to the described NHPs.

The gene encoding the described NHPs is apparently encoded by several exons dispersed on human chromosome 1 (see GENBANK accession no. AL138787). Accordingly, the described NHPs can be used to map the coding regions of the corresponding human genomic locus (*i.e.*, chromosome mapping), and to identify exon splice junctions.

Several polymorphisms were identified in the disclosed sequences including: a G/A polymorphism at the nucleotide position represented by, for example, position 274 of SEQ ID NO:1 (which can result in a glu or lys at the region corresponding to amino acid (aa) position 92 of, for example, SEQ ID NO:2); a C/A polymorphism at nucleotide position 424 (which can result in a pro or thr at aa position 142); a C/T polymorphism at nucleotide position 732 (both of which can result in leu, therefore it is a silent mutation at aa position 244; a G/A polymorphism at the nucleotide position represented by, for example, position 787 of SEQ ID NO:1 (which can result in a gly or arg at the region corresponding to aa position 263 of, for example, SEQ ID NO:2), and a G/A polymorphism at nucleotide position 1090 (which can result in a glu or lys at corresponding aa position 364). These polymorphisms, even the silent one, are particularly useful in the fields of forensic science and population biology, as markers identifying a particular individual and their descendants.

The described novel human polynucleotide sequences can be used, among other things, 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, P.C. and Wagner, T.E., 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 the NHP transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, i.e., mosaic animals or somatic cell transgenic animals. The transgene may be integrated as a single transgene or in concatamers, e.g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell-type by following, for example, the teaching of Lakso *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
5 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).

10 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
15 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
20 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 which include, but are not
25 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.

30 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 the sequence is expressed. Examples include, but are not
35 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
5 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.

10 5.2 NHP 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,
15 as reagents in diagnostic assays, for the identification of other cellular gene products related to a NHP, as reagents in assays for screening for compounds that can be as pharmaceutical reagents useful in the therapeutic treatment of mental, biological, or medical disorders and disease.

20 The Sequence Listing discloses the amino acid sequence encoded by the described NHP-encoding polynucleotides. The NHPs have initiator methionines in DNA sequence contexts consistent with eucaryotic translation initiation sites and signal-like sequences indicating that the NHPs can be secreted
25 or membrane associated.

The NHP amino acid sequences of the invention include the amino acid sequences presented in the Sequence Listing as well as analogues and derivatives thereof. Further, corresponding NHP homologues from other species are encompassed by the
30 invention. In fact, any NHP product encoded by the NHP nucleotide sequences described above 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
35 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, NY, 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 NHP 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, or 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 above, but which 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 a soluble or secreted molecule, the 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*.

5 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
10 the structural and functional characteristics of the 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
15 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;
20 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
25 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.,
30 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
35 quantity of such a protein is to be produced for the generation of pharmaceutical compositions of or containing

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
5 (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 & Inouye, 1985, Nucleic Acids Res. 13:3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem.
10 264:5503-5509); 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
15 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.

20 In an 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 non-essential regions (for example
25 the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of NHP coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the
30 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).

35 In mammalian host cells, a number of viral-based expression systems can 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 gene
5 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 & Shenk, 1984,
10 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
15 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,
20 must be provided. Furthermore, the initiation codon must 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
25 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
30 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
35 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 correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper 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 which stably express the NHP sequences described above can be engineered. Rather than using expression vectors which 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 are 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 which express the NHP product. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the 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⁻, hgp⁻ 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 (Colbere-Garapin *et al*, 1981, J. Mol. Biol. 150:1); and 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. For example, a system described by Janknecht *et al* 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 Ni²⁺-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, NY, 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 which 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 which accompany it in its natural state. Thus, a polypeptide which 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, or 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 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 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, 5 such antibodies can be used in conjunction gene therapy to, for example, evaluate the normal and/or engineered NHP-expressing cells prior to their introduction into the patient. Such antibodies may additionally be used as a method for the inhibition of abnormal NHP activity. Thus, such antibodies 10 may, therefore, be utilized as part of treatment methods.

For the production of antibodies, various host animals may be immunized by injection with the NHP, an NHP peptide (e.g., one corresponding to a functional domain of an NHP), truncated NHP polypeptides (NHP in which one or more domains 15 have been deleted), functional equivalents of the NHP or mutated variant of the 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 20 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 25 (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 30 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 which provides for the production of antibody 35 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; Cote *et al*, 1983, Proc. Natl. Acad. Sci. USA 80:2026-2030), and the EBV-hybridoma technique
5 (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, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. Production
10 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-
15 608; 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 (see U.S. Patent Nos. 6,075,181 and 5,877,397 both of which are herein
20 incorporated by reference in their entirety). 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.
25 Patent Nos. 5,877,397; 6,075,181 and 6,150,584 and their respective disclosures which are herein incorporated by reference in their entirety.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778; Bird, 1988,
30 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
35 region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include, but are not limited to: the $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ 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-idiotypic antibodies that "mimic" a given NHP, using techniques well-known to those skilled in the art. (See, *e.g.*, Greenspan & Bona, 1993, *FASEB J* 7(5):437-444; and Nisonoff, 1991, *J. Immunol.* 147(8):2429-2438). For example antibodies which bind to a NHP domain and competitively inhibit the binding of 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 signaling pathway.

Additionally given the high degree of relatedness of mammalian NHPs, the presently described knock-out mice (having never seen NHP, and thus never been tolerized to NHP) have a unique utility, as they can be advantageously applied to the generation of antibodies against the disclosed mammalian NHP (*i.e.*, 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 polynucleotide encoding an amino acid sequence drawn from the group consisting of SEQ ID NO:2
5 and SEQ ID NO:4.
2. An isolated cDNA molecule comprising a nucleotide sequence that:
 - (a) encodes the amino acid sequence shown in SEQ ID
10 NO: 2; and
 - (b) hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO: 1 or the complement thereof.
- 15 3. An isolated recombinant expression vector comprising a nucleotide sequence encoding the amino acid sequence shown in SEQ ID NO: 2.
- 20 4. A substantially isolated protein comprising the amino acid sequence shown in SEQ ID NO:2.
- 25 5. An isolated recombinant expression vector comprising a nucleotide sequence encoding the amino acid sequence shown in SEQ ID NO: 4.
6. A cell comprising the expression vector of Claim 3.
7. A cell comprising the expression vector of
30 Claim 5.

SEQUENCE LISTING

<110> LEXICON GENETICS INCORPORATED

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

<130> LEX-0363-PCT

<150> US 60/312,300

<151> 2001-08-14

<160> 4

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 2154

<212> DNA

<213> homo sapiens

<400> 1

```

atgactttct ctgtttcttct ttgcagcacg tccacggacg ccatgctggg gactctgaca 60
cccctgtctt cgctgctgct gctgctactg gtgctgggtg tggggtgtgg gccgcgggcg 120
tcctctgggtg gcggggcccg tggggcggcg ggctatgcc cagtgaagta catccagccc 180
atgcagaaaag gacctgtggg accgcccttc cgtgagggca aaggccagta cctggaaatg 240
cctctaccgc tgctgccgat ggacctgaag ggagagcccg gccccctgg gaagcccg 300
cctcggggtc cccctggccc ccttgcttcc ccaggaaaac caggcatggg aaagccagga 360
ctccatgggc agcctggccc tgctggggccc cctggcttct cccggatggg caaggctggg 420
ccccaggggc tccctggcaa ggtcggggcca ccagggcagc cggggcttcg gggggagcca 480
ggaatacagag gggaccaggg cctccgggga ccccaggac cccctggcct cccgggcccc 540
tcaggcatta ctatccctgg aaaaccaggt gcccaagggg tgccagggcc cccaggattc 600
cagggggaac cagggcccca gggggagcct ggccccccag gtgatcgagg cctcaagggg 660
gataatggag tgggccagcc cgggctgcct ggggccccag ggcagggggg tgcccccg 720
cccccgggcc tccctgggtc agctggctta ggcaaacctg gtttgatgg gcttctctgg 780
gccccaggag acaagggtga gtctgggcct cctggagttc caggccccag gggggagcca 840
ggagctgtgg gccaaaagg acctcctgga gtagacggtg tgggagtccc aggggcagca 900
gggttgccag gaccacaggg cccatcaggg gccaaagggg agccaggac cggggcccc 960
cctgggctga taggccccac tggctatggg atgccaggac tgccaggccc caagggggac 1020
aggggccag ctggggtccc aggactcttg ggggacaggg gtgagccagg ggaggatggg 1080
gagccagggg agcagggccc acaggggtctt gggggtcccc ctggacttcc tgggtctgca 1140
gggcttctct gcagacgtgg gccccctggg cctaaggggtg aggcagggcc tggaggacct 1200
ccaggagtgc ctggcattcg aggtgaccag gggcctagtg gcctggctgg gaaaccaggg 1260
gtcccaggtg agaggggact tcctggggcc catggacccc ctggaccaac tgggccaag 1320
ggtgagccgg gtttcacggg tcgccctgga ggaccagggg tggcaggagc cctggggcag 1380
aaaggtgact tggggctccc tgggcagcct ggcctgaggg gtccctcagg aatcccagga 1440
ctccagggtc cagctggccc tattggggccc caaggcctgc cgggcctgaa ggggggaacca 1500
ggcctgccag ggccccctgg agagggggaga gcaggggaac ctggcacggc tggggccacg 1560
gggccccccag gggtccctgg ctcccctgga atcacggggc ctccggggcc tcccggggcc 1620
ccgggacccc ctgggtgccc tggggccttc gatgagactg gcatcgaggg cttgcacctg 1680
cccaacggcg gtgtggaggg tgccgtgctg ggcaaggggg gcaagccaca gtttgggctg 1740
ggcgagctgt ctgcccagc cacaccggcc ttcactgcgg tgctcacctc gcccttcccc 1800
gcctcgggca tgcccgtgaa atttgaccgg actctctaca atggccacag cggctacaac 1860
ccagccactg gcatcttcac ctgccctgtg ggcggcgtct actactttgc ttaccatgtg 1920
cacgtcaagg gcaccaacgt gtgggtggcc ctgtacaaga acaacgtgcc ggccacctat 1980
acctacgatg agtacaagaa gggctacctg gaccaggcat ctggtggggc cgtgctccag 2040
ctgcccccca acgaccaggt ctgggtgcag atgccgtcgg accaggccaa cggcctctac 2100
tccacggagt acatccactc ctcttttca ggattcttgc tctgccccac ataa 2154

```

<210> 2

<211> 717

<212> PRT

<213> homo sapiens

<400> 2

```

Met Thr Phe Ser Val Leu Leu Cys Ser Thr Ser Thr Asp Ala Met Leu
1           5           10           15

```

Gly Thr Leu Thr Pro Leu Ser Ser Leu Leu Leu Leu Leu Leu Val Leu
 20 25 30
 Val Leu Gly Cys Gly Pro Arg Ala Ser Ser Gly Gly Gly Ala Gly Gly
 35 40 45
 Ala Ala Gly Tyr Ala Pro Val Lys Tyr Ile Gln Pro Met Gln Lys Gly
 50 55 60
 Pro Val Gly Pro Pro Phe Arg Glu Gly Lys Gly Gln Tyr Leu Glu Met
 65 70 75 80
 Pro Leu Pro Leu Leu Pro Met Asp Leu Lys Gly Glu Pro Gly Pro Pro
 85 90 95
 Gly Lys Pro Gly Pro Arg Gly Pro Pro Gly Pro Pro Gly Phe Pro Gly
 100 105 110
 Lys Pro Gly Met Gly Lys Pro Gly Leu His Gly Gln Pro Gly Pro Ala
 115 120 125
 Gly Pro Pro Gly Phe Ser Arg Met Gly Lys Ala Gly Pro Pro Gly Leu
 130 135 140
 Pro Gly Lys Val Gly Pro Pro Gly Gln Pro Gly Leu Arg Gly Glu Pro
 145 150 155 160
 Gly Ile Arg Gly Asp Gln Gly Leu Arg Gly Pro Pro Gly Pro Pro Gly
 165 170 175
 Leu Pro Gly Pro Ser Gly Ile Thr Ile Pro Gly Lys Pro Gly Ala Gln
 180 185 190
 Gly Val Pro Gly Pro Pro Gly Phe Gln Gly Glu Pro Gly Pro Gln Gly
 195 200 205
 Glu Pro Gly Pro Pro Gly Asp Arg Gly Leu Lys Gly Asp Asn Gly Val
 210 215 220
 Gly Gln Pro Gly Leu Pro Gly Ala Pro Gly Gln Gly Gly Ala Pro Gly
 225 230 235 240
 Pro Pro Gly Leu Pro Gly Pro Ala Gly Leu Gly Lys Pro Gly Leu Asp
 245 250 255
 Gly Leu Pro Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly Pro Pro Gly
 260 265 270
 Val Pro Gly Pro Arg Gly Glu Pro Gly Ala Val Gly Pro Lys Gly Pro
 275 280 285
 Pro Gly Val Asp Gly Val Gly Val Pro Gly Ala Ala Gly Leu Pro Gly
 290 295 300
 Pro Gln Gly Pro Ser Gly Ala Lys Gly Glu Pro Gly Thr Arg Gly Pro
 305 310 315 320
 Pro Gly Leu Ile Gly Pro Thr Gly Tyr Gly Met Pro Gly Leu Pro Gly
 325 330 335
 Pro Lys Gly Asp Arg Gly Pro Ala Gly Val Pro Gly Leu Leu Gly Asp
 340 345 350
 Arg Gly Glu Pro Gly Glu Asp Gly Glu Pro Gly Glu Gln Gly Pro Gln
 355 360 365
 Gly Leu Gly Gly Pro Pro Gly Leu Pro Gly Ser Ala Gly Leu Pro Gly
 370 375 380
 Arg Arg Gly Pro Pro Gly Pro Lys Gly Glu Ala Gly Pro Gly Gly Pro
 385 390 395 400
 Pro Gly Val Pro Gly Ile Arg Gly Asp Gln Gly Pro Ser Gly Leu Ala
 405 410 415
 Gly Lys Pro Gly Val Pro Gly Glu Arg Gly Leu Pro Gly Ala His Gly
 420 425 430
 Pro Pro Gly Pro Thr Gly Pro Lys Gly Glu Pro Gly Phe Thr Gly Arg
 435 440 445
 Pro Gly Gly Pro Gly Val Ala Gly Ala Leu Gly Gln Lys Gly Asp Leu
 450 455 460
 Gly Leu Pro Gly Gln Pro Gly Leu Arg Gly Pro Ser Gly Ile Pro Gly
 465 470 475 480
 Leu Gln Gly Pro Ala Gly Pro Ile Gly Pro Gln Gly Leu Pro Gly Leu
 485 490 495
 Lys Gly Glu Pro Gly Leu Pro Gly Pro Pro Gly Glu Gly Arg Ala Gly
 500 505 510
 Glu Pro Gly Thr Ala Gly Pro Thr Gly Pro Pro Gly Val Pro Gly Ser
 515 520 525
 Pro Gly Ile Thr Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro
 530 535 540
 Gly Ala Pro Gly Ala Phe Asp Glu Thr Gly Ile Ala Gly Leu His Leu

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

<210> 3
 <211> 2112
 <212> DNA
 <213> homo sapiens

<400> 3
 atgctgggga ctctgacacc cctgtcttcg ctgctgctgc tgctactggt gctgggtgctg 60
 ggggtgtgggc cgcgggcgct ctctggtggc ggggcccgtg gggcggcggg ctatgcccc 120
 gtgaagtaca tccagcccat gcagaaagga cctgtgggac cgcccttccg tgagggcaaa 180
 ggccagtacc tggaaatgcc tctaccgctg ctgccgatgg acctgaaggg agagcccggc 240
 ccccttgagg agcccgggcc tcggggtccc cctggcccc ctggcttccc aggaaaacca 300
 ggcatgggaa agccaggact ccatgggcag cctggccctg ctgggcccc tggtttctcc 360
 cggatgggca aggtggtcc cccagggctc cctggcaagg tcgggccacc agggcagccg 420
 gggcttcggg gggagccagg aatacgaggg gaccaggcc tccggggacc cccaggaccc 480
 cctggcctcc cgggcccctc aggcattact atccctggaa aaccagggtg ccaaggggtg 540
 ccaggggccc caggattcca ggggaacca gggccccagg gggagcctgg gccccagggt 600
 gatcgaggcc tcaaggggga taatggagtg ggccagccc ggctgcctgg ggccccaggg 660
 caggggggtg cccccggccc ccccgccctc cctggtccag ctggcttagg caaacctggt 720
 ttggatgggc ttcttggggc cccaggagac aaggggtgagt ctgggcctcc tggagttcca 780
 ggccccaggg gggagccagg agctgtgggc ccaaaaggac ctctggagt agacggtgtg 840
 ggagtcccag gggcagcagg gttgccagga ccacagggcc catcaggggc caaaggggag 900
 ccagggaccc ggggcccccc tgggctgata ggccccactg gctatgggat gccaggactg 960
 ccaggcccca agggggacag gggcccagct ggggtcccag gactcttggg ggacaggggt 1020
 gagccagggg aggatgggga gccaggggag caggggccac aggggtcttg ggggtcccc 1080
 ggacttccct ggtctgcagg gcttctggc agacgtgggc cccctggggc taagggtag 1140
 gcagggcctg gaggaccccc aggagtgcct ggcatcag gtgaccaggg gcctagtggc 1200
 ctggctggga aaccaggggt cccagggtgag aggggacttc ctggggcca tggaccccct 1260
 ggaccaactg ggcccaaggg tagccgggt ttcacgggtc gccctggagg accaggggtg 1320
 gcaggagccc tggggcagaa aggtgacttg gggctccctg ggcagcctgg cctgaggggt 1380
 ccctcaggaa tcccaggact ccagggtcca gctggcccta ttgggcccc aggcctgccg 1440
 ggctgaagg ggaaccagg cctgccaggg cccctgggag aggggagagc aggggaacct 1500
 ggcacggctg ggcccacggg gccccaggg gtccctggct cccctggaat cacgggccct 1560
 ccggggcctc ccgggcccc gggaccccct ggtgccctg gggccttcga tgagactggc 1620
 atcgaggct tgcacctgcc caacggcggt gtggagggtg ccgtgctggg caaggggggc 1680
 aagccacagt ttgggctggg cgagctgtct gccatgcca caccggcctt cactgcggtg 1740
 ctacactcgc ccttcccgc ctcgggcatg cccgtgaaat ttgaccggac tctctacaat 1800
 ggccacagcg gctacaacct agccactggc atcttcacct gccctgtggg cggcgtctac 1860
 tactttgctt accatgtgca cgtcaagggt accaacgtgt ggggtggcct gtacaagaac 1920
 aacgtgccg ccacctatac ctacgatgag tacaagaagg gctacctgga ccaggcatct 1980
 ggtggggccg tgctccagct gcggcccaac gaccagggtc ggggtgcagat gccgtcggag 2040
 caggccaacg gcctctactc cacggagtac atccactcct ccttttcagg attcttgcct 2100
 tgccccacat aa 2112

<210> 4

<211> 703
 <212> PRT
 <213> homo sapiens

<400> 4

```

Met Leu Gly Thr Leu Thr Pro Leu Ser Ser Leu Leu Leu Leu Leu Leu
 1      5      10      15
Val Leu Val Leu Gly Cys Gly Pro Arg Ala Ser Ser Gly Gly Ala
 20      25      30
Gly Gly Ala Ala Gly Tyr Ala Pro Val Lys Tyr Ile Gln Pro Met Gln
 35      40      45
Lys Gly Pro Val Gly Pro Pro Phe Arg Glu Gly Lys Gly Gln Tyr Leu
 50      55      60
Glu Met Pro Leu Pro Leu Leu Pro Met Asp Leu Lys Gly Glu Pro Gly
 65      70      75      80
Pro Pro Gly Lys Pro Gly Pro Arg Gly Pro Pro Gly Pro Pro Gly Phe
 85      90      95
Pro Gly Lys Pro Gly Met Gly Lys Pro Gly Leu His Gly Gln Pro Gly
 100     105     110
Pro Ala Gly Pro Pro Gly Phe Ser Arg Met Gly Lys Ala Gly Pro Pro
 115     120     125
Gly Leu Pro Gly Lys Val Gly Pro Pro Gly Gln Pro Gly Leu Arg Gly
 130     135     140
Glu Pro Gly Ile Arg Gly Asp Gln Gly Leu Arg Gly Pro Pro Gly Pro
 145     150     155     160
Pro Gly Leu Pro Gly Pro Ser Gly Ile Thr Ile Pro Gly Lys Pro Gly
 165     170     175
Ala Gln Gly Val Pro Gly Pro Pro Gly Phe Gln Gly Glu Pro Gly Pro
 180     185     190
Gln Gly Glu Pro Gly Pro Pro Gly Asp Arg Gly Leu Lys Gly Asp Asn
 195     200     205
Gly Val Gly Gln Pro Gly Leu Pro Gly Ala Pro Gly Gln Gly Gly Ala
 210     215     220
Pro Gly Pro Pro Gly Leu Pro Gly Pro Ala Gly Leu Gly Lys Pro Gly
 225     230     235     240
Leu Asp Gly Leu Pro Gly Ala Pro Gly Asp Lys Gly Glu Ser Gly Pro
 245     250     255
Pro Gly Val Pro Gly Pro Arg Gly Glu Pro Gly Ala Val Gly Pro Lys
 260     265     270
Gly Pro Pro Gly Val Asp Gly Val Gly Val Pro Gly Ala Ala Gly Leu
 275     280     285
Pro Gly Pro Gln Gly Pro Ser Gly Ala Lys Gly Glu Pro Gly Thr Arg
 290     295     300
Gly Pro Pro Gly Leu Ile Gly Pro Thr Gly Tyr Gly Met Pro Gly Leu
 305     310     315     320
Pro Gly Pro Lys Gly Asp Arg Gly Pro Ala Gly Val Pro Gly Leu Leu
 325     330     335
Gly Asp Arg Gly Glu Pro Gly Glu Asp Gly Glu Pro Gly Glu Gln Gly
 340     345     350
Pro Gln Gly Leu Gly Gly Pro Pro Gly Leu Pro Gly Ser Ala Gly Leu
 355     360     365
Pro Gly Arg Arg Gly Pro Pro Gly Pro Lys Gly Glu Ala Gly Pro Gly
 370     375     380
Gly Pro Pro Gly Val Pro Gly Ile Arg Gly Asp Gln Gly Pro Ser Gly
 385     390     395     400
Leu Ala Gly Lys Pro Gly Val Pro Gly Glu Arg Gly Leu Pro Gly Ala
 405     410     415
His Gly Pro Pro Gly Pro Thr Gly Pro Lys Gly Glu Pro Gly Phe Thr
 420     425     430
Gly Arg Pro Gly Gly Pro Gly Val Ala Gly Ala Leu Gly Gln Lys Gly
 435     440     445
Asp Leu Gly Leu Pro Gly Gln Pro Gly Leu Arg Gly Pro Ser Gly Ile
 450     455     460
Pro Gly Leu Gln Gly Pro Ala Gly Pro Ile Gly Pro Gln Gly Leu Pro
 465     470     475     480
Gly Leu Lys Gly Glu Pro Gly Leu Pro Gly Pro Pro Gly Glu Gly Arg
 485     490     495

```

Ala	Gly	Glu	Pro	Gly	Thr	Ala	Gly	Pro	Thr	Gly	Pro	Pro	Gly	Val	Pro
			500					505					510		
Gly	Ser	Pro	Gly	Ile	Thr	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly
		515					520					525			
Pro	Pro	Gly	Ala	Pro	Gly	Ala	Phe	Asp	Glu	Thr	Gly	Ile	Ala	Gly	Leu
		530				535					540				
His	Leu	Pro	Asn	Gly	Gly	Val	Glu	Gly	Ala	Val	Leu	Gly	Lys	Gly	Gly
545				550						555					560
Lys	Pro	Gln	Phe	Gly	Leu	Gly	Glu	Leu	Ser	Ala	His	Ala	Thr	Pro	Ala
				565					570					575	
Phe	Thr	Ala	Val	Leu	Thr	Ser	Pro	Phe	Pro	Ala	Ser	Gly	Met	Pro	Val
			580					585					590		
Lys	Phe	Asp	Arg	Thr	Leu	Tyr	Asn	Gly	His	Ser	Gly	Tyr	Asn	Pro	Ala
		595					600					605			
Thr	Gly	Ile	Phe	Thr	Cys	Pro	Val	Gly	Gly	Val	Tyr	Tyr	Phe	Ala	Tyr
		610				615					620				
His	Val	His	Val	Lys	Gly	Thr	Asn	Val	Trp	Val	Ala	Leu	Tyr	Lys	Asn
625				630						635					640
Asn	Val	Pro	Ala	Thr	Tyr	Thr	Tyr	Asp	Glu	Tyr	Lys	Lys	Gly	Tyr	Leu
				645					650					655	
Asp	Gln	Ala	Ser	Gly	Gly	Ala	Val	Leu	Gln	Leu	Arg	Pro	Asn	Asp	Gln
			660					665					670		
Val	Trp	Val	Gln	Met	Pro	Ser	Asp	Gln	Ala	Asn	Gly	Leu	Tyr	Ser	Thr
		675					680					685			
Glu	Tyr	Ile	His	Ser	Ser	Phe	Ser	Gly	Phe	Leu	Leu	Cys	Pro	Thr	
		690				695					700				