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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 October 2002 (31.10.2002)

PCT

(10) International Publication Number
WO 02/086091 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/13164

(22) International Filing Date: 25 April 2002 (25.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/286,314 25 April 2001 (25.04.2001) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 5th Floor, 1111 Franklin Street, Oakland, CA 94607-5200 (US).

(72) Inventors: **CHIEN, Kenneth, R.**; 7460 Pepita Way, La Jolla, CA 92037 (US). **HOSHIJIMA, Masahiko**; 3373 Lebon Drive, #203, San Diego, CA 92122 (US).

(74) Agent: **MCCLAIN, James, W.**; Brown Martin Haller & McClain, 1660 Union Street, San Diego, CA 92101-2926 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

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

WO 02/086091 A2

(54) Title: METHOD TO TREAT HEMOPHILIA BY HEPATIC GENE TRANSFER OF FACTOR VIII/IX WITH VESICLE VECTOR

(57) Abstract: Hemophilia is one of the most common genetic disorders. Standard therapies include transfusions with plasma products to provide clotting factors. The invention is a non-viral vesicle vector and method for the treatment of hemophilia. The vesicle vector contains the hepatitis B envelope protein to target the vesicle ot the liver for delivery of an expression construct containing the coding sequence for factor VIII or IX driven by an appropriate promoter of factor VIII or IX protein.

METHOD TO TREAT HEMOPHILIA BY HEPATIC GENE TRANSFER OF FACTOR VIII/IX WITH VESICLE VECTOR

CROSS-REFERENCES TO RELATED APPLICATIONS

This application claims the benefit of priority of United States provisional application Serial Number 60/286,314 filed April 25,2001 which is incorporated herein by reference in its entirety.

SEQUENCE LISTING

A sequence listing is submitted herewith under 35 C.F.R. §1.821 and is incorporated herein by reference.

BACKGROUND OF THE INVENTION

Hemophilia is one of the most common genetic disorders. Hemophilia A caused by deficiency of Factor VIII occurs in about 1 in 5000 male births, while hemophilia B caused by a defect in Factor IX is around 1 in 30,000 male births. The prevalence is very general in all populations studied. Hemophilia has long been treated with clotting factor concentrates, but the aim of this therapy is to control bleeding and requires lifelong repetitive intravenous infusions. Because of the increasing awareness of the risk of plasma derived products, the importance of the development of new and effective treatments is increased.

Gene therapy approaches have been developed for the treatment of hemophilia. Hemophilia is a particularly attractive model for developing a gene transfer approach for the treatment of disease. The proteins are well characterized, the genes are cloned and available, and there are large and small animal models of the disease. Moreover, there is no essential requirement for tissue specific delivery of the gene product and as protein function is regulated by activation of the protein; therefore, expression levels of the protein need not be tightly regulated. Additionally, only a low level of protein expression is required for phenotypic correction of the disease. The major hurdle of treatment of hemophilia by gene therapy is

that the expression of the gene product must be sustained throughout the life of the individual; therefore, effective therapy would likely require re-administration of the gene therapy vector.

Clinical trials for the treatment of hemophilia using retroviral and adeno-associated viral (AAV) vectors are ongoing. Adenoviral and lentiviral vectors have been used experimentally. However, the problem with all of these viral vectors is that they have a limited capacity for nucleic acid and have been shown to elicit an immune response. The use of DNA or RNA with or without synthetic liposomes results in low efficiency gene transfer. Non-viral methods achieve only short term, non-targeted gene expression.

A novel, liver-specific vesicle vector expressing modified surface proteins of the hepatitis B virus was recently described by Yamada et al (2001a). The vesicles containing the hepatitis B membrane proteins are generated by the methods well known to those skilled in the art (Kuroda et al,1992, and Yamada et al., 2001b, incorporated herein by reference). Briefly, a modified hepatitis B envelope (env) L protein, containing the pre-S1 +pre-S2 + S peptides, can be effectively generated in yeast by fusing the coding sequence for the chicken lysozyme signal sequence in frame to the beginning of the coding sequence for the modified env L protein (SEQ ID 1). The signal sequence directs the insertion of the proteins into the endoplasmic reticulum during translation. Protein rich vesicles bud from the endoplasmic reticulum and accumulate in the cytoplasm of the yeast cell. The vesicles are composed of lipid bilayers derived from the ER and the modified env L proteins as the major protein component. Particles formed by this method are very stable and can be easily purified through repetitive cesium chloride and sucrose gradients by methods well known to those skilled in the art.

Plasmid DNA can be incorporated into the env L containing particles by electroporation (Yamada et al. 2001a). Such DNA containing particles were demonstrated to facilitate entry of the DNA specifically into liver cells both in culture and upon systemic administration to nude mice in which human

hepatoma cells were transplanted. Yamada et al. (2001a) suggested that such a vesicle vector could be used for tissue specific delivery of nucleic acid and other compounds to the liver.

SUMMARY OF THE INVENTION

The invention is a non-viral vesicle vector for the treatment of hemophilia comprising a lipid bilayer containing a modified hepatitis B env L protein such that recognition of the S-peptide by the immune system is attenuated or abrogated, but the liver targeting signals are still exposed on the surface of the vesicle, and an expression construct for the expression of Factor VIII or IX for the treatment of hemophilia A or B, respectively. The expression construct may be single or double stranded DNA containing any of a number of promoters including, but not limited to general (e.g. cytomegalovirus, Rous sarcoma virus) and tissue specific (e.g. alpha fetoprotein, globulin, albumin, α1-microglobulin) promoters. The construct may contain additional regulatory elements including, but not limited to enhancers, introns, poly A sequences, RNA targeting sequences. Sequences to promote replication of the plasmid including SV40 origin of replication can be included. Inverted terminal repeat (ITR) sequences from AAV can be included in the construct to promote expression cassette stability or to enhance integration into the host DNA with the AAV Rep protein. In lieu of ITR sequences, eukaryotic DNA transposon/transposases systems can be used to promote integration.

The invention is a method for the treatment of hemophilia by administration of the non-viral vesicle vector of the invention. The vesicle vector containing the nucleic acid construct with the appropriate coding sequence is administered intravenously or intraarterially. The individual is monitored for expression of the gene product of interest by detection of the protein or mRNA or by phenotypic recovery.

DETAILED DESCRIPTION AND PREFERRED EMBODIMENT

Hemophilia is one of the most common genetic disorders and is a result of a mutation or deletion in any of the clotting factors, most commonly Factor VIII or IX. Treatment requires the lifelong replacement of clotting factors which requires repetitive intravenous infusions and exposes patients to the dangers associated with plasma derived products.

Hemophilia is amenable to treatment with gene therapy for a number of reasons. First, the genes involved are cloned and available. Second, the proteins are well characterized and their activation is regulated by cleavage of the protein rather than at the transcriptional or translational level; therefore, the expression level does not need to be tightly regulated. Third, low levels of protein expression have been demonstrated to be sufficient for phenotypic recovery. Fourth, although the liver is the physiological site of production of most of the Factor VIII and IX, the site of production of the protein within the body is relatively unimportant. Fifth, a number of animal models are available for analysis of various therapies. However, to date no effective gene transfer vectors or methods for the treatment of hemophilia have been developed.

The invention is a vesicle vector for the treatment of hemophilia comprising a natural lipid vesicle preferably produced in yeast or insect cells, such as Sf9 cells, containing modified hepatitis B env L protein integrated into the membrane and an expression construct inside the vesicle for the expression of Factor VIII or IX. The vesicles are prepared by the vaccine production method of Kuroda (1992) further refined by Yamada (2001b). Briefly, the hepatitis B env L protein is composed of three regions: the 108- or 119-residue pre-S1 region involved in the direct interaction with hepatocytes, the 55-residue pre-S2 region associated with the polymerized albumin-mediated interaction and the major 226-residue S-protein region. Attempts to produce L protein in various eukaryotic cells had been unsuccessful, probably due to the presence of the N-terminus of

the pre-S1 peptide. The coding sequence of the N-terminus of the L protein was replaced by a chicken lysosome signal sequence to direct the translocation of the N-terminus through the endoplasmic reticulum (ER). The chimeric sequence was inserted into a yeast (*S. cerevisiae*) expression vector and inserted into yeast using a standard transformation protocol. The chimeric L-protein was produced in abundance, up to 42% of the total yeast protein, and was properly inserted into the membrane. Vesicles budded off of the ER to form 23 nm spherical and filamentous particles containing the protein in the membrane of the vesicles. The yeast cells were disrupted with glass beads to release the vesicles. Vesicles were purified by serial rounds of discontinuous cesium and sucrose gradients. Production and purification of vesicles from insect cells would be performed in a similar method. A crude membrane fraction could be prepared as with the yeast cells, by homogenization and differential centrifugation. The fraction can be loaded onto cesium or sucrose gradients as with the yeast extract for purification of vesicles. The methods are amenable to inexpensive, large scale production of vesicles which is necessary for gene transfer. Vesicles are stable for long term storage at a low temperature but are unstable upon repeated freeze-thaw cycles.

The vesicle vectors can be used for the delivery of any nucleic acid construct, single- or double-stranded DNA or RNA, or gene product to the liver. In a preferred embodiment of the invention, the nucleic acid is a double stranded DNA plasmid. The construct minimally contains the coding sequence for human Factor VIII (SEQ ID 2) or IX (SEQ ID 3) for the treatment of hemophilia A or B respectively and a promoter to allow for transcription of the hemophilia gene. The construct may optionally contain additional regulatory and enhancer elements to modulate gene expression, intron and poly-A sequences to promote RNA processing and gene expression, RNA targeting sequences, AAV-ITR or eukaryotic transposon

sequences to promote stabilization of expression cassettes and integration into the host genome and viral origin of replication sequences to promote amplification of the plasmid in host cells. Such sequences are well known to those skilled in the art. The number of elements that can be inserted into the nucleic acid construct as the size is not limited by the requirements of a viral genome as is the case with many gene transfer protocols.

Any of a number of promoter sequences are known to be functional in liver cells. These include both non-tissue specific promoters such as CMV, RSV, ubiquitin, chicken β-actin and elongation factor (EF)-1α; and tissue specific promoters such as alpha-fetoprotein, globulin, α1-microglobulin and albumin.

AAV-ITR sequences may be incorporated into the construct flanking all of the coding and regulatory sequences, other than any origins of replication. The AAV-ITR sequences have been demonstrated to increase the stability of transferred constructs in gene therapy protocols. Alternatively, the AAV-ITR sequences may enhance integration into the human genome at a specific site with the cooperation of the AAV-Rep protein, which may be supplied by incorporation into the vesicles with the nucleic acid construct or by expression cassettes packaged into the same vesicle.

Eukaryotic transposon sequences can be incorporated into the construct flanking all of the coding sequences and regulatory elements, similar to the AAV-ITR sequences. Transposase to promote integration may be expressed from the same expression cassette or from a separate expression cassette packaged into the same vesicle.

Special considerations may be taken when expressing Factor VIII. Studies have demonstrated that human Factor VIII contains a sequence (nulceotides 1741 to 1771 in SEQ ID 2) that decreases heterologous expression of proteins (Fallaux et al., 1996). The sequence is AT-rich and has been demonstrated to bind a nuclear factor and repress expression of

a reporter construct in cells. Deletion or random mutation of the sequence results in a non-functional Factor VIII. However, silent mutations that result in no change in the amino acid sequence of the gene product can be introduced into the coding sequence by methods well known to those skilled in the art to enhance expression of Factor VIII.

In a preferred embodiment, the nucleic acid construct of the invention is introduced into the vesicles by electroporation. The nucleic acid construct is mixed thoroughly with the vesicles, brought to a final volume in water and transferred to an electroporation cuvette. Voltage and resistance vary widely depending on the size (gap length) of the cuvette and the volume of material in the cuvette. Such parameters can be readily modified by methods well known to those skilled in the art to result in maximum transfer of nucleic acid into vesicles with minimum destruction of vesicles.

Alternatively the nucleic acid may be introduced into the vesicle by fusion with nucleic acid containing liposomes by methods well known to those skilled in the art (Dzau et al, 1996). The construct of the invention is encapsulated into liposomes prepared by vortexing. Liposomes may be composed of known phospholipids and membrane components (e.g. phosphatidyl-choline, cholesterol) or of commercially available proprietary mixtures of membrane components (e.g. Lipofectamine from Gibco-BRL). Nucleic acid encapsulated in liposomes will fuse with the yeast or insect cell derived vesicles upon incubation at 37°C for 10-30 minutes.

Alternatively, factor VIII or IX protein may be incorporated into the vesicle vector of the invention. Factor VIII (SEQ ID 4) and IX (SEQ ID 5) protein may be produced using any of a number of methods well known to those skilled in the art. A solution containing a high concentration of protein may be mixed with purified vesicles and subjected to osmotic shock or sonication to promote incorporation of the protein into the vesicles. Protein may also be incorporated into artificial membranes by

vortexing or sonication. The artificial membranes containing the protein can be fused with the hepatitis B vesicles.

The nucleic acid or protein containing non-viral vesicle vectors of the invention are administered to the individual intravenously or intraarterially. To increase delivery, the vesicle vector can be administered directly into the hepatic or portal artery. The individual is monitored on regular intervals for the presence of factor VIII or IX or for phenotypic recovery. The amount of the non-viral vesicle to be administered would depend on the strength of the promoter, integration sequences, number of plasmids per vesicle and a number of other considerations well know to those skilled in the art. As methods for monitoring the state of health of individuals are well known, an effective dose can be readily determined.

Although an exemplary embodiment of the invention has been described above by way of example only, it will be understood by those skilled in the field that modifications may be made to the disclosed embodiment without departing from the scope of the invention, which is defined by the appended claims.

REFERENCES

Fallaux, F.J. et al (1996) The human clotting factor VIII cDNA contains an autonomously replicating sequence consensus- and matrix attachment region-like sequence that binds a nuclear factor, represses heterologous gene expression, and mediates the transcriptional effects of sodium butyrate. *Mol. Cell. Biol.* **16**:4264-4272.

Kuroda, S. et al (1992) Hepatitis B virus envelope L protein particles. *J. Biol. Chem.* **267**:1953-1961.

Yamada, T. et al (2001a) A new pinpoint gene delivery system using genetically engineered hepatitis B virus envelope L particles. *Molecular Biology and New Therapeutic Strategies: Cancer Research in the 21st*

Century. 5th Joint Conference of the American Association for Cancer Research and the Japanese Cancer Association. Hawaii, USA, February 12-16, 2001.

Yamada. T. et al (2001b) Physiochemical and immunological characterization of hepatitis B virus envelope particles exclusively consisting of the entire L (pre-S1 + pre- S2 + S) protein. *Vaccine* **19**:3154-3163.

WE CLAIM:

CLAIMS

1. A non-viral vesicle vector comprising:

a vesicular membrane with hepatitis B envelope (env) protein exposed on the vesicle surface and

a nucleic acid expression construct comprising a complete factor VIII or factor IX coding sequence and a promoter sequence functional in liver cells.

2. The vesicle vector of claim 1, wherein the env protein contains mutations to reduce antigenicity.

3. The vesicle vector of claim 1, wherein the expression construct is DNA.

4. The vesicle vector of claim 1, wherein the expression construct is double stranded plasmid DNA.

5. The vesicle vector of claim 1, wherein the expression construct is RNA.

6. The vesicle vector of claim 1, wherein the promoter is a non-tissue specific promoter.

7. The vesicle vector of claim 6, wherein the non-tissue specific promoter is selected from the group consisting of cytomegalovirus promoter, Rous sarcoma virus promoter, ubiquitin promoter, chicken β-actin promoter and elongation factor 1α promoter.

8. The vesicle vector of claim 1, wherein the promoter is a liver specific promoter.

9. The vesicle vector of claim 8, wherein the liver specific promoter is selected from the group consisting of alpha-fetoprotein promoter, globulin promoter, α1-microglobulin and albumin.

10. The vesicle vector of claim 1, wherein the expression construct comprises inverted terminal repeat sequences from adeno-associated virus (AAV-ITR).

11. The vesicle vector of claim 1, wherein the expression construct comprises eukaryotic transposon and transposase sequences.

12. The vesicle vector of claim 1, wherein the expression construct comprises the coding sequence of factor VIII.

13. The vesicle vector of claim 12, wherein the factor VIII comprises silent mutations to enhance expression.

14. The vesicle vector of claim 1, wherein the expression construct comprises the coding sequence of factor IX.

15. A non-viral vesicle vector comprising:

a vesicular membrane with hepatitis B envelope (env) protein exposed on the vesicle surface and

a protein comprising a complete factor VIII or factor IX.

16. The vesicle vector of claim 15, wherein the env protein contains mutations to reduce antigenicity.

17. A method for treatment of hemophilia comprising:

administration into circulation of an individual with hemophilia a non-viral vesicle vector comprising a vesicular membrane with hepatitis B env protein exposed on the vesicle surface and

a nucleic acid expression construct comprising a complete factor VIII or IX coding sequence and a promoter sequence functional in liver cells and

monitoring the individual for amelioration of disease.

18. The method of claim 17, wherein administration into circulation comprises intravenous administration.

19. The method of claim 17, wherein administration into circulation comprises administration into a hepatic or portal artery

SEQUENCE LISTING

<110> Chien, Kenneth R
Hoshijima, Masahiko

<120> Method to treat hemophilia by hepatic gene transfer of Factor VIII/IX with vesicle vector

<130> 6627-PA1170

<150> 60/286,314
<151> 2001-04-25

<160> 5

<170> PatentIn version 3.1

<210> 1
<211> 9029
<212> DNA
<213> Homo sapiens

<400> 1
gcttagtgct gagcacatcc agtgggtaaa gttccttaaa atgctctgca aagaaattgg 60

gacttttcat taaatcagaa attttacttt tttcccctcc tgggagctaa agatatttta 120

gagaagaatt aacctttgc ttctccagtt gaacatttgt agcaataagt catgcaaata 180

gagctctcca cctgcttctt tctgtgcctt ttgcgattct gctttagtgc caccagaaga 240

tactacctgg gtgcagtgga actgtcatgg gactatatgc aaagtgatct cggtgagctg 300

cctgtggacg caagatttcc tcctagagtg ccaaaatctt ttccattcaa cacctcagtc 360

gtgtacaaaa agactctgtt tgtagaattc acggatcacc ttttcaacat cgctaagcca 420

aggccaccct ggatgggtct gctaggtcct accatccagg ctgaggttta tgatacagtg 480

gtcattacac ttaagaacat ggcttcccat cctgtcagtc ttcatgctgt tggtgtatcc 540

tactggaaag cttctgaggg agctgaatat gatgatcaga ccagtcaaag ggagaaagaa 600

gatgataaag tcttccctgg tggaagccat acatatgtct ggcaggtcct gaaagagaat 660

ggtccaatgg cctctgaccc actgtgcctt acctactcat atctttctca tgtggacctg 720

gtaaaagact tgaattcagg cctcattgga gccctactag tatgtagaga agggagtctg 780

gccaaggaaa agacacagac cttgcacaaa tttatactac tttttgctgt atttgatgaa 840

gggaaaagtt ggcactcaga aacaaagaac tccttgatgc aggataggga tgctgcatct 900

gctcgggcct ggcctaaaat gcacacagtc aatggttatg taaacaggtc tctgccaggt 960

ctgattggat gccacaggaa atcagtctat tggcatgtga ttggaatggg caccactcct 1020

gaagtgcact caatattcct cgaaggtcac acatttcttg tgaggaacca tcgccaggcg 1080

tccttggaaa tctcgccaat aactttcctt actgctcaaa cactcttgat ggaccttgga 1140

cagtttctac tgttttgtca tatctcttcc caccaacatg atggcatgga agcttatgtc 1200

aaagtagaca gctgtccaga ggaaccccaa ctacgaatga aaaataatga agaagcggaa 1260

gactatgatg atgatcttac tgattctgaa atggatgtgg tcaggtttga tgatgacaac 1320

tctccttcct ttatccaaat tcgctcagtt gccaagaagc atcctaaaac ttgggtacat 1380

tacattgctg ctgaagagga ggactgggac tatgctccct tagtcctcgc ccccgatgac 1440

agaagttata aaagtcaata tttgaacaat ggccctcagc ggattggtag gaagtacaaa 1500

aaagtccgat ttatggcata cacagatgaa acctttaaga ctcgtgaagc tattcagcat 1560

gaatcaggaa tcttgggacc tttactttat ggggaagttg gagacacact gttgattata 1620

tttaagaatc aagcaagcag accatataac atctaccctc acggaatcac tgatgtccgt 1680

cctttgtatt caaggagatt accaaaaggt gtaaaacatt tgaaggattt tccaattctg 1740

ccaggagaaa tattcaaata taaatggaca gtgactgtag aagatgggcc aactaaatca 1800

gatcctcggt gcctgacccg ctattactct agtttcgtta atatggagag agatctagct 1860

tcaggactca ttggccctct cctcatctgc tacaaagaat ctgtagatca aagaggaaac 1920

cagataatgt cagacaagag gaatgtcatc ctgttttctg tatttgatga gaaccgaagc 1980

tggtacctca cagagaatat acaacgcttt ctccccaatc cagctggagt gcagcttgag 2040

gatccagagt tccaagcctc caacatcatg cacagcatca atggctatgt ttttgatagt 2100

ttgcagttgt cagtttgttt gcatgaggtg gcatactggt acattctaag cattggagca 2160

cagactgact tcctttctgt cttcttctct ggatatacct tcaaacacaa aatggtctat 2220

gaagacacac tcaccctatt cccattctca ggagaaactg tcttcatgtc gatggaaaac 2280

ccaggtctat ggattctggg gtgccacaac tcagactttc ggaacagagg catgaccgcc 2340

ttactgaagg tttctagttg tgacaagaac actggtgatt attacgagga cagttatgaa 2400

gatatttcag catacttgct gagtaaaaac aatgccattg aaccaagaag cttctcccag 2460

aattcaagac accctagcac taggcaaaag caatttaatg ccaccacaat tccagaaaat 2520

gacatagaga agactgaccc ttggtttgca cacagaacac ctatgcctaa aatacaaaat 2580

gtctcctcta gtgatttgtt gatgctcttg cgacagagtc ctactccaca tgggctatcc 2640

ttatctgatc tccaagaagc caaatatgag acttttttctg atgatccatc acctggagca 2700

atagacagta ataacagcct gtctgaaatg acacacttca ggccacagct ccatcacagt 2760

ggggacatgg tatttacccc tgagtcaggc ctccaattaa gattaaatga gaaactgggg 2820

acaactgcag caacagagtt gaagaaactt gatttcaaag tttctagtac atcaaataat 2880

ctgatttcaa caattccatc agacaatttg gcagcaggta ctgataatac aagttcctta 2940

ggacccccaa gtatgccagt tcattatgat agtcaattag ataccactct atttggcaaa 3000

aagtcatctc cccttactga gtctggtgga cctctgagct tgagtgaaga aaataatgat 3060

tcaaagttgt tagaatcagg tttaatgaat agccaagaaa gttcatgggg aaaaaatgta 3120

tcgtcaacag agagtggtag gttatttaaa gggaaaagag ctcatggacc tgctttgttg 3180

actaaagata atgccttatt caaagttagc atctctttgt taaagacaaa caaaacttcc 3240

aataattcag caactaatag aaagactcac attgatggcc catcattatt aattgagaat 3300

agtccatcag tctggcaaaa tatattagaa agtgacactg agtttaaaaa agtgacacct 3360

ttgattcatg acagaatgct tatggacaaa aatgctacag ctttgaggct aaatcatatg 3420

tcaaataaaa ctacttcatc aaaaaacatg gaaatggtcc aacagaaaaa agagggcccc 3480

attccaccag atgcacaaaa tccagatatg tcgttcttta agatgctatt cttgccagaa 3540

tcagcaaggt ggatacaaag gactcatgga aagaactctc tgaactctgg gcaaggcccc 3600

agtccaaagc aattagtatc cttaggacca gaaaaatctg tggaaggtca gaatttcttg 3660

tctgagaaaa acaaagtggt agtaggaaag ggtgaattta caaaggacgt aggactcaaa 3720

gagatggttt ttccaagcag cagaaaccta tttcttacta acttggataa tttacatgaa 3780

aataatacac acaatcaaga aaaaaaaatt caggaagaaa tagaaaagaa ggaaacatta 3840

atccaagaga atgtagtttt gcctcagata catacagtga ctggcactaa gaatttcatg 3900

aagaaccttt tcttactgag cactaggcaa aatgtagaag gttcatatga cggggcatat 3960

gctccagtac ttcaagattt taggtcatta aatgattcaa caaatagaac aaagaaacac 4020

acagctcatt tctcaaaaaa aggggaggaa gaaaacttgg aaggcttggg aaatcaaacc 4080

aagcaaattg tagagaaata tgcatgcacc acaaggatat ctcctaatac aagccagcag 4140

aattttgtca cgcaacgtag taagagagct ttgaaacaat tcagactccc actagaagaa 4200

acagaacttg aaaaaaggat aattgtggat gacacctcaa cccagtggtc caaaaacatg 4260

aaacatttga ccccgagcac cctcacacag atagactaca atgagaagga gaaaggggcc 4320

attactcagt ctcccttatc agattgcctt acgaggagtc atagcatccc tcaagcaaat 4380

agatctccat tacccattgc aaaggtatca tcatttccat ctattagacc tatatatctg 4440

accagggtcc tattccaaga caactcttct catcttccag cagcatctta tagaaagaaa 4500

gattctgggg tccaagaaag cagtcatttc ttacaaggag ccaaaaaaaa taacctttct 4560

ttagccattc taaccttgga gatgactggt gatcaaagag aggttggctc cctggggaca 4620

agtgccacaa attcagtcac atacaagaaa gttgagaaca ctgttctccc gaaaccagac 4680

ttgcccaaaa catctggcaa agttgaattg cttccaaaag ttcacattta tcagaaggac 4740

ctattcccta cggaaactag caatgggtct cctggccatc tggatctcgt ggaagggagc 4800

cttcttcagg gaacagaggg agcgattaag tggaatgaag caaacagacc tggaaaagtt 4860

ccctttctga gagtagcaac agaaagctct gcaaagactc cctccaagct attggatcct 4920

cttgcttggg ataaccacta tggtactcag ataccaaaag aagagtggaa atcccaagag 4980

aagtcaccag aaaaaacagc ttttaagaaa aaggatacca ttttgtccct gaacgcttgt 5040

gaaagcaatc atgcaatagc agcaataaat gagggacaaa ataagcccga aatagaagtc 5100

acctgggcaa agcaaggtag gactgaaagg ctgtgctctc aaaacccacc agtcttgaaa 5160

cgccatcaac gggaaataac tcgtactact cttcagtcag atcaagagga aattgactat 5220

gatgatacca tatcagttga aatgaagaag gaagattttg acatttatga tgaggatgaa 5280

aatcagagcc cccgcagctt tcaaaagaaa acacgacact attttattgc tgcagtggag 5340

aggctctggg attatgggat gagtagctcc ccacatgttc taagaaacag ggctcagagt 5400

ggcagtgtcc ctcagttcaa gaaagttgtt ttccaggaat ttactgatgg ctcctttact 5460

cagcccttat accgtggaga actaaatgaa catttgggac tcctggggcc atatataaga 5520

gcagaagttg aagataatat catggtaact ttcagaaatc aggcctctcg tccctattcc 5580

ttctattcta gccttatttc ttatgaggaa gatcagaggc aaggagcaga acctagaaaa 5640

aactttgtca agcctaatga aaccaaaact tacttttgga aagtgcaaca tcatatggca 5700

cccactaaag atgagtttga ctgcaaagcc tgggcttatt tctctgatgt tgacctggaa 5760

aaagatgtgc actcaggcct gattggaccc cttctggtct gccacactaa cacactgaac 5820

cctgctcatg ggagacaagt gacagtacag gaatttgctc tgtttttcac catctttgat 5880

gagaccaaaa gctggtactt cactgaaaat atggaaagaa actgcagggc tccctgcaat 5940

atccagatgg aagatcccac ttttaaagag aattatcgct tccatgcaat caatggctac 6000

ataatggata cactacctgg cttagtaatg gctcaggatc aaaggattcg atggtatctg 6060

ctcagcatgg gcagcaatga aaacatccat tctattcatt tcagtggaca tgtgttcact 6120

gtacgaaaaa aagaggagta taaaatggca ctgtacaatc tctatccagg tgtttttgag 6180

acagtggaaa tgttaccatc caaagctgga atttggcggg tggaatgcct tattggcgag 6240

catctacatg ctgggatgag cacacttttt ctggtgtaca gcaataagtg tcagactccc 6300

ctgggaatgg cttctggaca cattagagat tttcagatta cagcttcagg acaatatgga 6360

cagtgggccc caaagctggc cagacttcat tattccggat caatcaatgc ctggagcacc 6420

aaggagccct tttcttggat caaggtggat ctgttggcac caatgattat tcacggcatc 6480

aagacccagg gtgcccgtca gaagttctcc agcctctaca tctctcagtt tatcatcatg 6540

tatagtcttg atgggaagaa gtggcagact tatcgaggaa attccactgg aaccttaatg 6600

gtcttctttg gcaatgtgga ttcatctggg ataaaacaca atatttttaa ccctccaatt 6660

attgctcgat acatccgttt gcacccaact cattatagca ttcgcagcac tcttcgcatg 6720

gagttgatgg gctgtgattt aaatagttgc agcatgccat tgggaatgga gagtaaagca 6780

atatcagatg cacagattac tgcttcatcc tactttacca atatgtttgc cacctggtct 6840

ccttcaaaag ctcgacttca cctccaaggg aggagtaatg cctggagacc tcaggtgaat 6900

aatccaaaag agtggctgca agtggacttc cagaagacaa tgaaagtcac aggagtaact 6960

actcagggag taaaatctct gcttaccagc atgtatgtga aggagttcct catctccagc 7020

agtcaagatg gccatcagtg gactctcttt tttcagaatg gcaaagtaaa ggtttttcag 7080

ggaaatcaag actccttcac acctgtggtg aactctctag acccaccgtt actgactcgc 7140

taccttcgaa ttcaccccca gagttgggtg caccagattg ccctgaggat ggaggttctg 7200

ggctgcgagg cacaggacct ctactgaggg tggccactgc agcacctgcc actgccgtca 7260

cctctccctc ctcagctcca gggcagtgtc cctccctggc ttgccttcta cctttgtgct 7320

aaatcctagc agacactgcc ttgaagcctc ctgaattaac tatcatcagt cctgcatttc 7380

tttggtgggg ggccaggagg gtgcatccaa tttaacttaa ctcttaccta ttttctgcag 7440

ctgctcccag attactcctt ccttccaata taactaggca aaaagaagtg aggagaaacc 7500

tgcatgaaag cattcttccc tgaaaagtta ggcctctcag agtcaccact tcctctgttg 7560

tagaaaaact atgtgatgaa actttgaaaa agatatttat gatgttaaca tttcaggtta 7620

agcctcatac gtttaaaata aaactctcag ttgtttatta tcctgatcaa gcatggaaca 7680

aagcatgttt caggatcaga tcaatacaat cttggagtca aaaggcaaat catttggaca 7740

atctgcaaaa tggagagaat acaataacta ctacagtaaa gtctgtttct gcttccttac 7800

acatagatat aattatgtta tttagtcatt atgaggggca cattcttatc tccaaaacta 7860

gcattcttaa actgagaatt atagatgggg ttcaagaatc cctaagtccc ctgaaattat 7920

ataaggcatt ctgtataaat gcaaatgtgc atttttctga cgagtgtcca tagatataaa 7980

gccatttggt cttaattctg accaataaaa aaataagtca ggaggatgca attgttgaaa 8040

gctttgaaat aaaataacaa tgtcttcttg aaatttgtga tggccaagaa agaaaatgat 8100

gatgacatta ggcttctaaa ggacatacat ttaatatttc tgtggaaata tgaggaaaat 8160

ccatggttat ctgagatagg agatacaaac tttgtaattc taataatgca ctcagtttac 8220

tctctccctc tactaatttc ctgctgaaaa taacacaaca aaaatgtaac aggggaaatt 8280

atataccgtg actgaaaact agagtcctac ttacatagtt gaaatatcaa ggaggtcaga 8340

agaaaattgg actggtgaaa acagaaaaaa cactccagtc tgccatatca ccacacaata 8400

ggatcccccc tcttgccctc cacccccata agattgtgaa gggtttactg ctccttccat 8460

ctgcctgacc ccttcactat gactacacag aatctcctga tagtaaaggg ggctggaggc 8520

aaggataagt tatagagcag ttggaggaag catccaaaga ttgcaaccca gggcaaatgg 8580

aaaacaggag atcctaatat gaaagaaaaa tggatcccaa tctgagaaaa ggcaaaagaa 8640

tggctacttt tttctatgct ggagtatttt ctaataatcc tgcttgaccc ttatctgacc 8700

tctttggaaa ctataacata gctgtcacag tatagtcaca atccacaaat gatgcaggtg 8760

caaatggttt atagccctgt gaagttctta aagtttagag gctaacttac agaaatgaat 8820

aagttgtttt gttttatagc ccggtagagg agttaacccc aaaggtgata tggttttatt 8880

tcctgttatg tttaacttga taatcttatt ttggcattct tttcccattg actatataca 8940

tctctatttc tcaaatgttc atggaactag ctcttttatt ttcctgctgg tttcttcagt 9000

aatgagttaa ataaaacatt gacacatac 9029

<210> 2
<211> 2804
<212> DNA
<213> Homo sapiens

<400> 2
accactttca caatctgcta gcaaaggtta tgcagcgcgt gaacatgatc atggcagaat 60

caccaggcct catcaccatc tgccttttag gatatctact cagtgctgaa tgtacagttt 120

ttcttgatca tgaaaacgcc aacaaaattc tgaatcggcc aaagaggtat aattcaggta 180

aattggaaga gtttgttcaa gggaaccttg agagagaatg tatggaagaa aagtgtagtt 240

ttgaagaagc acgagaagtt tttgaaaaca ctgaaagaac aactgaattt tggaagcagt 300

atgttgatgg agatcagtgt gagtccaatc catgtttaaa tggcggcagt tgcaaggatg 360

acattaattc ctatgaatgt tggtgtccct ttggatttga aggaaagaac tgtgaattag 420

atgtaacatg taacattaag aatggcagat gcgagcagtt ttgtaaaaat agtgctgata 480

acaaggtggt ttgctcctgt actgagggat atcgacttgc agaaaaccag aagtcctgtg 540

aaccagcagt gccatttcca tgtggaagag tttctgtttc acaaacttct aagctcaccc 600

gtgctgagac tgtttttcct gatgtggact atgtaaattc tactgaagct gaaaccattt 660

tggataacat cactcaaagc acccaatcat ttaatgactt cactcgggtt gttggtggag 720

aagatgccaa accaggtcaa ttcccttggc aggttgtttt gaatggtaaa gttgatgcat 780

tctgtggagg ctctatcgtt aatgaaaaat ggattgtaac tgctgcccac tgtgttgaaa 840

ctggtgttaa aattacagtt gtcgcaggtg aacataatat tgaggagaca gaacatacag 900

agcaaaagcg aaatgtgatt cgaattattc ctcaccacaa ctacaatgca gctattaata 960

agtacaacca tgacattgcc cttctggaac tggacgaacc cttagtgcta aacagctacg 1020

ttacacctat ttgcattgct gacaaggaat acacgaacat cttcctcaaa tttggatctg 1080

gctatgtaag tggctgggga agagtcttcc acaaagggag atcagcttta gttcttcagt 1140

accttagagt tccacttgtt gaccgagcca catgtcttcg atctacaaag ttcaccatct 1200

ataacaacat gttctgtgct ggcttccatg aaggaggtag agattcatgt caaggagata 1260

gtgggggacc ccatgttact gaagtggaag ggaccagttt cttaactgga attattagct 1320

ggggtgaaga gtgtgcaatg aaaggcaaat atggaatata taccaaggta tcccggtatg 1380

tcaactggat taaggaaaaa acaaagctca cttaatgaaa gatggatttc caaggttaat 1440

tcattggaat tgaaaattaa cagggcctct cactaactaa tcactttccc atcttttgtt 1500

agatttgaat atatacattc tatgatcatt gctttttctc tttacagggg agaatttcat 1560

attttacctg agcaaattga ttagaaaatg gaaccactag aggaatataa tgtgttagga 1620

aattacagtc atttctaagg gcccagccct tgacaaaatt gtgaagttaa attctccact 1680

ctgtccatca gatactatgg ttctccacta tggcaactaa ctcactcaat tttccctcct 1740

tagcagcatt ccatcttccc gatcttcttt gcttctccaa ccaaaacatc aatgtttatt 1800

agttctgtat acagtacagg atctttggtc tactctatca caaggccagt accacactca 1860

tgaagaaaga acacaggagt agctgagagg ctaaaactca tcaaaaacac tactcctttt 1920

cctctaccct attcctcaat cttttacctt ttccaaatcc caatccccaa atcagttttt 1980

ctctttctta ctccctctct ccctttttacc ctccatggtc gttaaaggag agatggggag 2040

catcattctg ttatacttct gtacacagtt atacatgtct atcaaaccca gacttgcttc 2100

catagtggag acttgctttt cagaacatag ggatgaagta aggtgcctga aaagtttggg 2160

ggaaaagttt ctttcagaga gttaagttat tttatatata taatatatat ataaaatata 2220

taatatacaa tataaatata tagtgtgtgt gtgtatgcgt gtgtgtagac acacacgcat 2280

acacacatat aatggaagca ataagccatt ctaagagctt gtatggttat ggaggtctga 2340

ctaggcatga tttcacgaag gcaagattgg catatcattg taactaaaaa agctgacatt 2400

gacccagaca tattgtactc tttctaaaaa taataataat aatgctaaca gaaagaagag 2460

aaccgttcgt ttgcaatcta cagctagtag agactttgag gaagaattca acagtgtgtc 2520

ttcagcagtg ttcagagcca agcaagaagt tgaagttgcc tagaccagag gacataagta 2580

tcatgtctcc tttaactagc ataccccgaa gtggagaagg gtgcagcagg ctcaaaggca 2640

taagtcattc caatcagcca actaagttgt ccttttctgg tttcgtgttc accatggaac 2700

attttgatta tagttaatcc ttctatcttg aatcttctag agagttgctg accaactgac 2760

gtatgtttcc ctttgtgaat taataaactg gtgttctggt tcat 2804

<210> 3
<211> 1286
<212> DNA
<213> Hepatitis B virus

<400> 3
gtcgagtata aaaacaatga gatctttgtt gatcttggtt ttgtgtttct tgccattggc 60

tgctttgggt aaggttcgac aaggcatggg aggttggtct tccaaacctc gacaaggcat 120

ggggacgaat ctttctgttc ccaatcctct gggattcttt cccgatcacc agttggaccc 180

tgcgttcgga gccaactcaa acaatccaga ttgggacttc aaccccaaca aggatcaatg 240

gccagaggca aatcaggtag gagcgggagc attcgggcca gggttcaccc caccacacgg 300

cggtcttttg gggtggagcc ctcaggctca gggcatattg acaacagtgc cagcagcacc 360

tcctcctgcc tccaccaatc ggcagtcagg aagacagcct actcccatct ctccacctct 420

aagagacagt catcctcagg ccatgcagtg gaattccaca acattccacc aagctctgct 480

agatcccaga gtgaggggcc tatattttcc tgctggtggc tccagttccg gaacagtaaa 540

ccctgttccg actactgcct cacccatatc tggggaccct gcaccgaaca tggagaacac 600

aacatcagga ttcctaggac ccctgctcgt gttacaggcg gggtttttct tgttgacaag 660

aatcctcaca ataccacaga gtctagactc gtggtggact tctctcaatt ttctaggggg 720

agcacccacg tgtcctggcc aaaattcgca gtccccaacc tccaatcact caccaacctc 780

ttgtcctcca atttgtcctg gctatcgctg gatgtgtctg cggcgtttta tcatattcct 840

cttcatcctg ctgctatgcc tcatcttctt gttggttctt ctggactacc aaggtatgtt 900

gcccgtttgt cctctacttc caggaacatc aaccaccagc acggggccat gcaagacctg 960

cacgattcct gctcaaggaa cctctatgtt tccctcttgt tgctgtacaa aaccttcgga 1020

cggaaactgc acttgtattc ccatcccatc atcctgggct ttcgcaagat tcctatggga 1080

gtgggcctca gtccgtttct cctggctcag tttactagtg ccatttgttc agtggttcgt 1140

agggctttcc cccactgttt ggctttcagt tatatggatg atgtggtatt gggggccaag 1200

tctgtacaac atcttgagtc ccttttacc tctattacca attttctttt gtctttgggt 1260

atacatttaa attgaattga attgaa 1286

<210> 4
<211> 2351
<212> PRT
<213> Homo sapiens

<400> 4

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1 5 10 15

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
20 25 30

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
35 40 45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
50 55 60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65 70 75 80

Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
85 90 95

Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
100 105 110

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
115 120 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
130 135 140

```
Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145                 150                 155                 160

Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
                165                 170                 175

Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
            180                 185                 190

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
        195                 200                 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
    210                 215                 220

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225                 230                 235                 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
                245                 250                 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
            260                 265                 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
        275                 280                 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
    290                 295                 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305                 310                 315                 320

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
                325                 330                 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
            340                 345                 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
        355                 360                 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
```

```
    370                 375                 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385                 390                 395                 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
                405                 410                 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
            420                 425                 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
        435                 440                 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
    450                 455                 460

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465                 470                 475                 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
                485                 490                 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
            500                 505                 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
        515                 520                 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
    530                 535                 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545                 550                 555                 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
                565                 570                 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
            580                 585                 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
        595                 600                 605
```

```
Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
    610                 615                 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625                 630                 635                 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
                645                 650                 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
            660                 665                 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
        675                 680                 685

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
    690                 695                 700

Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705                 710                 715                 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
                725                 730                 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
            740                 745                 750

Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Ser Arg His Pro
        755                 760                 765

Ser Thr Arg Gln Lys Gln Phe Asn Ala Thr Thr Ile Pro Glu Asn Asp
    770                 775                 780

Ile Glu Lys Thr Asp Pro Trp Phe Ala His Arg Thr Pro Met Pro Lys
785                 790                 795                 800

Ile Gln Asn Val Ser Ser Ser Asp Leu Leu Met Leu Leu Arg Gln Ser
                805                 810                 815

Pro Thr Pro His Gly Leu Ser Leu Ser Asp Leu Gln Glu Ala Lys Tyr
            820                 825                 830

Glu Thr Phe Ser Asp Asp Pro Ser Pro Gly Ala Ile Asp Ser Asn Asn
        835                 840                 845
```

```
Ser Leu Ser Glu Met Thr His Phe Arg Pro Gln Leu His His Ser Gly
    850                 855                 860

Asp Met Val Phe Thr Pro Glu Ser Gly Leu Gln Leu Arg Leu Asn Glu
865                 870                 875                 880

Lys Leu Gly Thr Thr Ala Ala Thr Glu Leu Lys Lys Leu Asp Phe Lys
                885                 890                 895

Val Ser Ser Thr Ser Asn Asn Leu Ile Ser Thr Ile Pro Ser Asp Asn
            900                 905                 910

Leu Ala Ala Gly Thr Asp Asn Thr Ser Ser Leu Gly Pro Pro Ser Met
        915                 920                 925

Pro Val His Tyr Asp Ser Gln Leu Asp Thr Thr Leu Phe Gly Lys Lys
    930                 935                 940

Ser Ser Pro Leu Thr Glu Ser Gly Gly Pro Leu Ser Leu Ser Glu Glu
945                 950                 955                 960

Asn Asn Asp Ser Lys Leu Leu Glu Ser Gly Leu Met Asn Ser Gln Glu
                965                 970                 975

Ser Ser Trp Gly Lys Asn Val Ser Ser Thr Glu Ser Gly Arg Leu Phe
            980                 985                 990

Lys Gly Lys Arg Ala His Gly Pro  Ala Leu Leu Thr Lys  Asp Asn Ala
        995                 1000                1005

Leu Phe  Lys Val Ser Ile Ser  Leu Leu Lys Thr Asn  Lys Thr Ser
    1010                1015                1020

Asn Asn  Ser Ala Thr Asn Arg  Lys Thr His Ile Asp  Gly Pro Ser
    1025                1030                1035

Leu Leu  Ile Glu Asn Ser Pro  Ser Val Trp Gln Asn  Ile Leu Glu
    1040                1045                1050

Ser Asp  Thr Glu Phe Lys Lys  Val Thr Pro Leu Ile  His Asp Arg
    1055                1060                1065

Met Leu  Met Asp Lys Asn Ala  Thr Ala Leu Arg Leu  Asn His Met
```

```
    1070                1075                1080

Ser Asn  Lys Thr Thr Ser Ser  Lys Asn Met Glu Met  Val Gln Gln
    1085                1090                1095

Lys Lys  Glu Gly Pro Ile Pro  Pro Asp Ala Gln Asn  Pro Asp Met
    1100                1105                1110

Ser Phe  Phe Lys Met Leu Phe  Leu Pro Glu Ser Ala  Arg Trp Ile
    1115                1120                1125

Gln Arg  Thr His Gly Lys Asn  Ser Leu Asn Ser Gly  Gln Gly Pro
    1130                1135                1140

Ser Pro  Lys Gln Leu Val Ser  Leu Gly Pro Glu Lys  Ser Val Glu
    1145                1150                1155

Gly Gln  Asn Phe Leu Ser Glu  Lys Asn Lys Val Val  Val Gly Lys
    1160                1165                1170

Gly Glu  Phe Thr Lys Asp Val  Gly Leu Lys Glu Met  Val Phe Pro
    1175                1180                1185

Ser Ser  Arg Asn Leu Phe Leu  Thr Asn Leu Asp Asn  Leu His Glu
    1190                1195                1200

Asn Asn  Thr His Asn Gln Glu  Lys Lys Ile Gln Glu  Glu Ile Glu
    1205                1210                1215

Lys Lys  Glu Thr Leu Ile Gln  Glu Asn Val Val Leu  Pro Gln Ile
    1220                1225                1230

His Thr  Val Thr Gly Thr Lys  Asn Phe Met Lys Asn  Leu Phe Leu
    1235                1240                1245

Leu Ser  Thr Arg Gln Asn Val  Glu Gly Ser Tyr Asp  Gly Ala Tyr
    1250                1255                1260

Ala Pro  Val Leu Gln Asp Phe  Arg Ser Leu Asn Asp  Ser Thr Asn
    1265                1270                1275

Arg Thr  Lys Lys His Thr Ala  His Phe Ser Lys Lys  Gly Glu Glu
    1280                1285                1290
```

```
Glu Asn  Leu Glu Gly Leu Gly  Asn Gln Thr Lys Gln  Ile Val Glu
    1295                 1300                 1305

Lys Tyr  Ala Cys Thr Thr Arg  Ile Ser Pro Asn Thr  Ser Gln Gln
    1310                 1315                 1320

Asn Phe  Val Thr Gln Arg Ser  Lys Arg Ala Leu Lys  Gln Phe Arg
    1325                 1330                 1335

Leu Pro  Leu Glu Glu Thr Glu  Leu Glu Lys Arg Ile  Ile Val Asp
    1340                 1345                 1350

Asp Thr  Ser Thr Gln Trp Ser  Lys Asn Met Lys His  Leu Thr Pro
    1355                 1360                 1365

Ser Thr  Leu Thr Gln Ile Asp  Tyr Asn Glu Lys Glu  Lys Gly Ala
    1370                 1375                 1380

Ile Thr  Gln Ser Pro Leu Ser  Asp Cys Leu Thr Arg  Ser His Ser
    1385                 1390                 1395

Ile Pro  Gln Ala Asn Arg Ser  Pro Leu Pro Ile Ala  Lys Val Ser
    1400                 1405                 1410

Ser Phe  Pro Ser Ile Arg Pro  Ile Tyr Leu Thr Arg  Val Leu Phe
    1415                 1420                 1425

Gln Asp  Asn Ser Ser His Leu  Pro Ala Ala Ser Tyr  Arg Lys Lys
    1430                 1435                 1440

Asp Ser  Gly Val Gln Glu Ser  Ser His Phe Leu Gln  Gly Ala Lys
    1445                 1450                 1455

Lys Asn  Asn Leu Ser Leu Ala  Ile Leu Thr Leu Glu  Met Thr Gly
    1460                 1465                 1470

Asp Gln  Arg Glu Val Gly Ser  Leu Gly Thr Ser Ala  Thr Asn Ser
    1475                 1480                 1485

Val Thr  Tyr Lys Lys Val Glu  Asn Thr Val Leu Pro  Lys Pro Asp
    1490                 1495                 1500

Leu Pro  Lys Thr Ser Gly Lys  Val Glu Leu Leu Pro  Lys Val His
    1505                 1510                 1515
```

```
Ile Tyr  Gln Lys Asp Leu Phe  Pro Thr Glu Thr Ser  Asn Gly Ser
    1520                 1525                 1530

Pro Gly  His Leu Asp Leu Val  Glu Gly Ser Leu Leu  Gln Gly Thr
    1535                 1540                 1545

Glu Gly  Ala Ile Lys Trp Asn  Glu Ala Asn Arg Pro  Gly Lys Val
    1550                 1555                 1560

Pro Phe  Leu Arg Val Ala Thr  Glu Ser Ser Ala Lys  Thr Pro Ser
    1565                 1570                 1575

Lys Leu  Leu Asp Pro Leu Ala  Trp Asp Asn His Tyr  Gly Thr Gln
    1580                 1585                 1590

Ile Pro  Lys Glu Glu Trp Lys  Ser Gln Glu Lys Ser  Pro Glu Lys
    1595                 1600                 1605

Thr Ala  Phe Lys Lys Lys Asp  Thr Ile Leu Ser Leu  Asn Ala Cys
    1610                 1615                 1620

Glu Ser  Asn His Ala Ile Ala  Ala Ile Asn Glu Gly  Gln Asn Lys
    1625                 1630                 1635

Pro Glu  Ile Glu Val Thr Trp  Ala Lys Gln Gly Arg  Thr Glu Arg
    1640                 1645                 1650

Leu Cys  Ser Gln Asn Pro Pro  Val Leu Lys Arg His  Gln Arg Glu
    1655                 1660                 1665

Ile Thr  Arg Thr Thr Leu Gln  Ser Asp Gln Glu Glu  Ile Asp Tyr
    1670                 1675                 1680

Asp Asp  Thr Ile Ser Val Glu  Met Lys Lys Glu Asp  Phe Asp Ile
    1685                 1690                 1695

Tyr Asp  Glu Asp Glu Asn Gln  Ser Pro Arg Ser Phe  Gln Lys Lys
    1700                 1705                 1710

Thr Arg  His Tyr Phe Ile Ala  Ala Val Glu Arg Leu  Trp Asp Tyr
    1715                 1720                 1725

Gly Met  Ser Ser Ser Pro His  Val Leu Arg Asn Arg  Ala Gln Ser
```

```
    1730                1735                1740

Gly Ser  Val Pro Gln Phe Lys  Lys Val Val Phe Gln  Glu Phe Thr
    1745                1750                1755

Asp Gly  Ser Phe Thr Gln Pro  Leu Tyr Arg Gly Glu  Leu Asn Glu
    1760                1765                1770

His Leu  Gly Leu Leu Gly Pro  Tyr Ile Arg Ala Glu  Val Glu Asp
    1775                1780                1785

Asn Ile  Met Val Thr Phe Arg  Asn Gln Ala Ser Arg  Pro Tyr Ser
    1790                1795                1800

Phe Tyr  Ser Ser Leu Ile Ser  Tyr Glu Glu Asp Gln  Arg Gln Gly
    1805                1810                1815

Ala Glu  Pro Arg Lys Asn Phe  Val Lys Pro Asn Glu  Thr Lys Thr
    1820                1825                1830

Tyr Phe  Trp Lys Val Gln His  His Met Ala Pro Thr  Lys Asp Glu
    1835                1840                1845

Phe Asp  Cys Lys Ala Trp Ala  Tyr Phe Ser Asp Val  Asp Leu Glu
    1850                1855                1860

Lys Asp  Val His Ser Gly Leu  Ile Gly Pro Leu Leu  Val Cys His
    1865                1870                1875

Thr Asn  Thr Leu Asn Pro Ala  His Gly Arg Gln Val  Thr Val Gln
    1880                1885                1890

Glu Phe  Ala Leu Phe Phe Thr  Ile Phe Asp Glu Thr  Lys Ser Trp
    1895                1900                1905

Tyr Phe  Thr Glu Asn Met Glu  Arg Asn Cys Arg Ala  Pro Cys Asn
    1910                1915                1920

Ile Gln  Met Glu Asp Pro Thr  Phe Lys Glu Asn Tyr  Arg Phe His
    1925                1930                1935

Ala Ile  Asn Gly Tyr Ile Met  Asp Thr Leu Pro Gly  Leu Val Met
    1940                1945                1950
```

```
Ala Gln  Asp Gln Arg Ile Arg  Trp Tyr Leu Leu Ser  Met Gly Ser
    1955                 1960                 1965

Asn Glu  Asn Ile His Ser Ile  His Phe Ser Gly His  Val Phe Thr
    1970                 1975                 1980

Val Arg  Lys Lys Glu Glu Tyr  Lys Met Ala Leu Tyr  Asn Leu Tyr
    1985                 1990                 1995

Pro Gly  Val Phe Glu Thr Val  Glu Met Leu Pro Ser  Lys Ala Gly
    2000                 2005                 2010

Ile Trp  Arg Val Glu Cys Leu  Ile Gly Glu His Leu  His Ala Gly
    2015                 2020                 2025

Met Ser  Thr Leu Phe Leu Val  Tyr Ser Asn Lys Cys  Gln Thr Pro
    2030                 2035                 2040

Leu Gly  Met Ala Ser Gly His  Ile Arg Asp Phe Gln  Ile Thr Ala
    2045                 2050                 2055

Ser Gly  Gln Tyr Gly Gln Trp  Ala Pro Lys Leu Ala  Arg Leu His
    2060                 2065                 2070

Tyr Ser  Gly Ser Ile Asn Ala  Trp Ser Thr Lys Glu  Pro Phe Ser
    2075                 2080                 2085

Trp Ile  Lys Val Asp Leu Leu  Ala Pro Met Ile Ile  His Gly Ile
    2090                 2095                 2100

Lys Thr  Gln Gly Ala Arg Gln  Lys Phe Ser Ser Leu  Tyr Ile Ser
    2105                 2110                 2115

Gln Phe  Ile Ile Met Tyr Ser  Leu Asp Gly Lys Lys  Trp Gln Thr
    2120                 2125                 2130

Tyr Arg  Gly Asn Ser Thr Gly  Thr Leu Met Val Phe  Phe Gly Asn
    2135                 2140                 2145

Val Asp  Ser Ser Gly Ile Lys  His Asn Ile Phe Asn  Pro Pro Ile
    2150                 2155                 2160

Ile Ala  Arg Tyr Ile Arg Leu  His Pro Thr His Tyr  Ser Ile Arg
    2165                 2170                 2175
```

```
Ser Thr  Leu Arg Met Glu Leu  Met Gly Cys Asp Leu  Asn Ser Cys
    2180                 2185                 2190

Ser Met  Pro Leu Gly Met Glu  Ser Lys Ala Ile Ser  Asp Ala Gln
    2195                 2200                 2205

Ile Thr  Ala Ser Ser Tyr Phe  Thr Asn Met Phe Ala  Thr Trp Ser
    2210                 2215                 2220

Pro Ser  Lys Ala Arg Leu His  Leu Gln Gly Arg Ser  Asn Ala Trp
    2225                 2230                 2235

Arg Pro  Gln Val Asn Asn Pro  Lys Glu Trp Leu Gln  Val Asp Phe
    2240                 2245                 2250

Gln Lys  Thr Met Lys Val Thr  Gly Val Thr Thr Gln  Gly Val Lys
    2255                 2260                 2265

Ser Leu  Leu Thr Ser Met Tyr  Val Lys Glu Phe Leu  Ile Ser Ser
    2270                 2275                 2280

Ser Gln  Asp Gly His Gln Trp  Thr Leu Phe Phe Gln  Asn Gly Lys
    2285                 2290                 2295

Val Lys  Val Phe Gln Gly Asn  Gln Asp Ser Phe Thr  Pro Val Val
    2300                 2305                 2310

Asn Ser  Leu Asp Pro Pro Leu  Leu Thr Arg Tyr Leu  Arg Ile His
    2315                 2320                 2325

Pro Gln  Ser Trp Val His Gln  Ile Ala Leu Arg Met  Glu Val Leu
    2330                 2335                 2340

Gly Cys  Glu Ala Gln Asp Leu  Tyr
    2345                 2350
```

<210> 5
<211> 461
<212> PRT
<213> Homo sapiens

<400> 5

```
Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
1               5                  10                  15
```

```
Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr Val Phe Leu
            20                  25                  30

Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg Tyr Asn
        35                  40                  45

Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg Glu Cys
    50                  55                  60

Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe Glu Asn
65                  70                  75                  80

Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly Asp Gln
                85                  90                  95

Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp Asp Ile
            100                 105                 110

Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys Asn Cys
        115                 120                 125

Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe
    130                 135                 140

Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr Glu Gly
145                 150                 155                 160

Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val Pro Phe
                165                 170                 175

Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr Arg Ala
            180                 185                 190

Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu Ala Glu
        195                 200                 205

Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn Asp Phe
    210                 215                 220

Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp
225                 230                 235                 240

Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile
```

```
                245                 250                 255

Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly
            260                 265                 270

Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu Thr Glu
        275                 280                 285

His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn
    290                 295                 300

Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu
305                 310                 315                 320

Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile
                325                 330                 335

Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr
            340                 345                 350

Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val
        355                 360                 365

Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Arg
    370                 375                 380

Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His
385                 390                 395                 400

Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val
                405                 410                 415

Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly
            420                 425                 430

Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser
        435                 440                 445

Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr
    450                 455                 460
```