

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
6 April 2006 (06.04.2006)

PCT

(10) International Publication Number
WO 2006/036879 A2

(51) International Patent Classification:
C12Q 1/70 (2006.01) *C12Q 1/68* (2006.01)

Bicetre (FR). **LEMON, Stanley, M.** [US/US]; 1517 Bayou Shore Dr., Galveston, TX 77551 (US).

(21) International Application Number:
PCT/US2005/034390

(74) Agent: **SHISHIMA, Gina, N.**; Fulbright & Jaworski L.L.P., 600 Congress Avenue, Suite 2400, Austin, TX 78701 (US).

(22) International Filing Date:
27 September 2005 (27.09.2005)

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US (patent), UZ, VC, VN, YU, ZA, ZM, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/613,266 27 September 2004 (27.09.2004) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/613,266 (CIP)
Filed on 27 September 2004 (27.09.2004)

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicants (*for all designated States except US*): **BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM** [US/US]; 201 W. 7th St., Austin, TX 78701 (US). **INSTITUT PASTEUR** [FR/FR]; 25-28, rue du Docteur Roux, F-75724 Paris Cedex 15 (FR).

(72) Inventors; and

Published:

(75) Inventors/Applicants (*for US only*): **MARTIN, Annette** [FR/FR]; 201 rue d'Alésia, F-75014 Paris (FR). **GHIBAUDO, David, V.** [FR/FR]; 2 Impasse de Muguet, Bâtiment A, F-21850 Saint Apollinaire (FR). **COHEN, Lisette** [FR/FR]; 39 rue Danton, F-94270 Le Kremlin

— 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.

(54) Title: CHIMERIC GB VIRUS B (GBV-B)

(57) Abstract: The present invention relates generally to the fields of biochemistry, molecular biology, and virology. More particularly, it relates to the production and use of GB virus B (GBV-B)/HCV chimeras. The invention involves nucleic acid constructs and compositions encoding GBV-B/HCV chimera. The chimeric viruses may be employed to study GBV-B and related hepatitis family members, such as hepatitis C virus. The invention thus includes methods of preparing GBV-B/HCV chimeric sequences, constructs, and viruses, as well as methods of employing these compositions.

WO 2006/036879 A2

DESCRIPTION

CHIMERIC GB VIRUS B (GBV-B)

BACKGROUND OF THE INVENTION

[0001] This application claims priority to U.S. Provisional Patent Application serial number 60/613,266, filed September 27, 2004, which is incorporated herein by reference in its entirety.

A. Field of the Invention

[0002] The present invention relates generally to the fields of biochemistry, molecular biology, and virology. More particularly, it relates to compositions and methods related to GBVirus-B (GBV-B) / Hepatitis C virus (HCV) chimeric viruses, polynucleotides, and proteins.

B. Description of Related Art

[0003] Chronic hepatitis C is a major threat to the public health. Serologic surveys suggest that as many as 3.9 million Americans are chronically infected with the responsible virus, hepatitis C virus (HCV) (Alter, 1997). These individuals are at increased risk of developing progressive hepatic fibrosis leading to cirrhosis and loss of hepatocellular function, as well as hepatocellular carcinoma. The course of chronic hepatitis C is typically lengthy, often extending over decades, with insidious clinical progression usually occurring in the absence of symptoms. Nonetheless, liver disease due to HCV results in the death of 8,000-10,000 Americans annually, and chronic hepatitis C is the most common cause of liver transplantation within the U.S.

[0004] Therefore, HCV is a major public health problem. However, therapy for chronic hepatitis C is problematic. Recombinant interferon- α is approved for treatment of chronic hepatitis C (Consensus Development Panel, 1997). The benefit of interferon- α results primarily from its antiviral properties and its ability to inhibit production of virus by infected hepatocytes (Neumann *et al.*, 1998). Nonetheless, even under optimal therapeutic regimens, the majority of patients with chronic hepatitis C fail to eliminate the virus or resolve their liver

disease. Treatment failures are especially common in persons infected with genotype 1 HCV, unfortunately the most prevalent genotype in the U.S. Thus, there is an urgent need to better understand the virus and develop better treatment. Unfortunately, technical difficulties in working with HCV have made it necessary to use infectious surrogate viruses in efforts to develop treatments and vaccines for HCV.

[0005] Scientists' efforts to better understand HCV and to develop new drugs for treatment of hepatitis C have been stymied by two overwhelming technical deficiencies: first, the nonexistence of a high permissive cell line that supports replication of the virus and second, the absence of a permissive animal species other than chimpanzees, which are endangered and therefore available on a limited basis.

[0006] Presently, those who are working on HCV treatment and prevention are employing an infectious chimeric virus of sindbis and HCV and/or an infectious clone of pestiviruses as surrogate virus models in HCV drug discovery efforts, due to the above technical difficulties of working with HCV. Alternatively, they are using isolated proteins or RNA segments of HCV for biochemical and structural studies. This approach precludes functional studies of virus replication and its inhibition.

[0007] GBV-B is a hepatotropic flavivirus that has a unique phylogenetic relationship to human HCV and strong potential to serve as a surrogate virus in drug discovery efforts related to hepatitis C antiviral drug development. GBV-B causes acute hepatitis in experimentally infected tamarins (Simons *et al.*, 1995; Schlauder *et al.*, 1995; Karayiannis *et al.*, 1989) and can serve as a surrogate virus for HCV in drug discovery efforts. GBV-B virus is much closer in sequence and biological properties than the above-described models. It will be easier to make biologically relevant chimeras between HCV and GBV-B than by using more distantly related viruses. GBV-B is hepatotropic (as is HCV), whereas the viruses used in these competing technologies are not. In view of the above, an infectious clone of GBV-B would be useful to those working on HCV treatment and prevention.

[0008] Unfortunately, the use of GBV-B as a surrogate or model for HCV has not been possible in the past, because no infectious molecular clone of GBV-B virus genome could be prepared. It is now known that this obstacle was encountered because the GBV-B genome was believed to be 259 nucleotides shorter than its actual length (Muerhoff *et al.*, 1995; Simons *et al.*, 1995). Others, previous to the inventors, had failed to realize that the 3' sequence of GBV-

B was missing from the prior sequences. Without this 3' sequence, it is not possible to prepare an infectious GBV-B molecular clone.

BRIEF SUMMARY OF THE INVENTION

[0009] As discussed above, an infectious molecular clone of GBV-B or GBV-B/HCV chimera would be very useful for the development of HCV preventative and therapeutic treatments. In particular, chimeric viruses, polynucleotides, and/or proteins are used as compositions and in the methods of the exemplified invention. The construction of an infectious molecular clone may require the newly determined 3' sequence to be included in order for the clone to be viable. The inventors have elucidated the previously unrecognized 3' terminal sequence of GBV-B (SEQ ID NO:1). This sequence has been reproducibly recovered from tamarin serum containing GBV-B RNA, in RT-PCR protocols using several different primer sets, and as a fusion with previously reported 5' GBV-B sequences. The newly identified 3' sequence is not included in published reports of the GBV-B sequence, nor described in patents relating to the original identification of the viral sequence (see U.S. Patent No. 5,807,670 and references therein). In various embodiments of the invention an infectious clone involving the 3' terminal sequence from GVB-B may or may not be required.

[0010] The invention has utility in that the inclusion of the sequence may be necessary, if desired, to construct an infectious molecular GBV-B clone. Such clones clearly have the potential to be constructed as chimeras including relevant hepatitis C virus sequences in lieu of the homologous GBV-B sequence, providing unique tools for drug discovery efforts. A full-length molecular clone of GBV-B was constructed, as described in later sections of this specification.

[0011] GBV-B can be used as a model for HCV, and the GBV-B genome, polynucleotides, and/or polypeptides can be used in the construction of chimeric viral RNAs, DNAs and/or polypeptides containing sequences of both HCV and GBV-B. Such chimeric viruses or molecules enable the investigation of the mechanisms for the different biological properties of these viruses and encoded proteins, and to discover and investigate potential inhibitors of specific HCV activities (*e.g.*, proteinase) required for HCV replication. However, certain aspects of this work is dependent upon construction of an infectious clone of GBV-B, which is itself dependent on the incorporation of the correct 3' terminal nucleotide sequence within this clone. GBV-B has unique advantages over HCV in terms of its ability to replicate

and cause liver disease in tamarins, which present fewer restrictions to research than chimpanzees, the only nonhuman primate species known to be permissive for HCV.

[0012] Embodiments of the invention include a chimeric GBV-B/HCV virus, a chimeric polynucleotide and/or an encoded chimeric polypeptide thereof. In certain aspects, the invention is an isolated polynucleotide encoding a chimeric GBV-B/HCV virus (*i.e.*, a virus containing GBV-B nucleic acid sequences with one or more corresponding HCV nucleic acid sequences replacing GBV-B sequences), chimeric GBV-B/HCV polynucleotide (*e.g.*, a nucleic acid containing GVB-B sequences and all or part of a nucleic acid segment encoding an HCV core region), and/or chimeric GBV-B/HCV polypeptide (*e.g.*, all or part of a chimeric core protein). It will be understood that a cognizable sequence (nucleotide or amino acid) from GBV-B or HCV refers to a sequence that can be recognized as from one or the other. "A corresponding sequence" refers to a sequence that corresponds in number and/or homology to the replaced sequence.

[0013] Typically, segment(s) of GBV-B genome, polynucleotide and/or encoded polypeptide are replaced by a corresponding segment(s) of HCV. In some embodiments of a chimeric virus or polynucleotide, for example, the virus or polynucleotide contains in addition to GBV-B sequences, an HCV nucleic acid segment encoding all or part of one or more of the following HCV proteins: a core protein; an E1 protein; an E2 protein; a p7 protein; an E1 and E2 protein; a core, E1, and E2 proteins; a core, E1, E2, and p7 proteins; NS2; NS3; NS4; NS4A; NS4b; NS5A; NS5B and various combinations and/or permutations thereof and may further contain a protease site such as, but not limited to Ubi. In further aspects, the polynucleotide of the invention may encode a polyprotein with one or more heterologous protease sites. Embodiments cover 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more protease sites that may be recognized by the same or difference proteases. A protease site, such as a Ubi site, may be located between protein regions or within a leader sequence of a protein, such as any of those discussed above or herein. For example, a NS3 protein may have a heterologous protease cleavage site and in particular an ubiquitin protease cleavage site.

[0014] A polynucleotide and/or polypeptide of the invention may comprise, comprise at least, or comprise at most 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99%, or any value derivable therefrom of an HCV genome, nucleic acid sequence or amino acid sequence. In certain non-limiting embodiments, the

polynucleotide may have a sequence set forth in SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, or SEQ ID NO:26. The polynucleotide may be a DNA or an RNA. The polynucleotide of the invention may be comprised in a plasmid.

[0015] In further embodiments the invention includes a hepatotropic virus. The hepatotropic virus may be a chimeric GBV-B/HCV virus as exemplified herein. Typically the hepatotropic virus will propagate *in vivo*. In particular aspects the chimeric virus will propagate in a primate or primate cells. In a preferred embodiment, the chimeric virus will propagate in a tamarin or cells derived there from.

[0016] In still a further embodiment of the invention includes a method of producing a chimeric virus. The method may include a) introducing into a host cell a viral expression construct comprising a polynucleotide encoding a chimeric GBV-B/HCV virus; and b) culturing said host cell under conditions permitting production of a chimeric virus from the construct. A host cell may be a prokaryotic cell, a eukaryotic cell, an animal, and more preferably a tamarin. The polynucleotide may comprise synthetic or recombinant RNA or DNA. The method may further comprising the step of isolating virus from a host or host cell. Preferably the virus is purified to homogeneity.

[0017] Other embodiments of the invention include methods for identifying or assessing the effectiveness a compound active against a viral infection comprising: a) providing a virus, polynucleotide, and/or polypeptide expressed from an expression construct comprising a chimeric GBV-B/HCV virus or polynucleotide; b) contacting the virus, polynucleotide, and/or polypeptide with a candidate substance; and c) comparing or assessing the infectious ability of the virus or effect on a polynucleotide or polypeptide in the presence of the candidate substance with the infectious ability of the virus or the characteristics of a polynucleotide or polypeptide in a similar system in the absence of the candidate substance.

[0018] In addition, an infectious molecular clone of GBV-B or a GVB-B/HCV chimera is expected to have utility in liver-specific gene expression or in gene therapy. This application might be enhanced by the inclusion of HCV genomic sequence. Further, an infectious GBV-B/HCV chimera expressing HCV polypeptides (*e.g.*, envelope proteins) can have utility as a vaccine immunogen for hepatitis C.

[0019] A full-length cDNA copy of the GBV-B genome or a GBV-B/HCV chimera may be constructed to contain the newly identified 3' terminal sequences. RNA transcribed from this cDNA copy of the genome would be infectious when inoculated into the liver of a GBV-B permissive tamarin, giving rise to rescued GBV-B or GBV-B/HCV virus particles. A chimeric molecule may be constructed from this infectious GBV-B clone in which all or part of the HCV 5' NTR, C region sequence or encoded protein, E1 region sequence or encoded protein, E2 region sequence or encoded protein, p7 region sequence or encoded protein, NS2 region sequence or encoded protein, NS3 region sequence or encoded protein, NS4A region sequence or encoded protein, NS4B region sequence or encoded protein, NS5A region sequence or encoded protein, NS5B region sequence or encoded protein, 3' NTR or various combinations and permutations would be placed in frame or in an operative position in lieu of the homologous GBV-B sequence, and this chimeric cDNA would be used to generate infectious GBV-B/HCV chimeric viruses by intrahepatic inoculation of synthetic RNA in tamarins. Published studies indicate that the GBV-B and HCV proteinases have closely related substrate recognition and cleavage properties, making such chimeras highly likely to be viable. These newly generated chimeric GBV-B/HCV viruses could be used in preclinical assessment of candidate HCV NS3 proteinase inhibitors as well as molecules that alter various characteristics of other polynucleotides or polypeptides of HCV.

[0020] Therefore, aspects of the present invention encompass an isolated polynucleotide encoding a 3' sequence of the GBV-B genome. The polynucleotide may include the sequence identified as SEQ ID NO:1. It is contemplated that the polynucleotide may be a DNA molecule or it can be an RNA molecule. It is further contemplated that expression constructs may contain a polynucleotide that has a stretch of contiguous nucleotides from SEQ ID NO:1 and/or SEQ ID NO:2, for example, lengths of 50, 100, 150, 250, 500, 1000, 5000, as well as the entire length of SEQ ID NO:1 or 2, are considered appropriate. Such polynucleotides may also be contained in other constructs of the invention or be used in the methods of the invention. Polynucleotides, such as chimeric GBV-B/HCV polynucleotides, employing sequences from SEQ ID NO:1 may alternatively contain sequences from SEQ ID NO:2 in the constructs and methods of the present invention.

[0021] The invention is also understood as covering a viral expression construct that includes a polynucleotide encoding a 3' sequence of the GBV-B genome. This expression construct is further understood to contain the sequence identified as SEQ ID NO:1. The present

invention contemplates the expression construct as a plasmid or as a virus. Furthermore, the expression construct can express GBV-B sequences; alternatively it may express sequences from a chimeric GBV-B/HCV virus.

[0022] The identification and isolation of a 3' sequence of GBV-B additionally provides a method of producing a virus, particularly a full-length virus, by introducing into a host cell an expression construct containing a polynucleotide encoding at least a 3' sequence of GBV-B and by culturing the host cell under conditions permitting production of a virus from the construct. This method can be practiced using a prokaryotic cell as a host cell, or by using a eukaryotic cell as a host cell. Furthermore, the eukaryotic cell can be located within an animal.

[0023] A method of producing virus according to the claimed invention can also be employed using a polynucleotide that contains synthetic RNA and/or synthetic DNA. Moreover, a step can be added to the method by also isolating any virus produced from the host cell. The virus can then be purified to homogeneity.

[0024] Additional examples of the claimed invention include a method for identifying a compound active against a viral infection by providing a virus, polynucleotide, and/or polypeptide expressed from an expression construct, which may or may not contain a 3' sequence of a GBV-B virus, by contacting the virus, polynucleotide, and/or polypeptide with a candidate substance; and by comparing the infectious ability of the virus or function of the polynucleotide or polypeptide in the presence of the candidate substance with the infectious ability or function in a similar system in the absence of the candidate substance. It is contemplated that the invention can be practiced using GBV-B virus or a GBV-B/HCV chimera or various segments thereof, including nucleic acid or peptide segment(s).

[0025] The present invention can also be understood to provide a compound active against a viral infection, *e.g.*, HCV infection, identified by providing a virus, polynucleotide, and/or polypeptide expressed from a viral construct containing a 3' sequence of a GBV-B virus; contacting the virus, polynucleotide, and/or polypeptide with a candidate substance; and comparing the infectious ability of the virus or the function of the polynucleotide or polypeptide in the presence of the candidate substance with a similar system in the absence of the candidate substance. In some embodiments an active compound is identified using a GBV-B virus, while in other embodiments an active compound is identified using a GBV-B/HCV chimera or polynucleotide or polypeptide segments thereof.

[0026] In various embodiments of the invention, a GBV-B polynucleotide may encode a GBV-B/HCV chimera that includes at least part of a 5' NTR sequence derived from a HCV 5' NTR. The 5' NTR may comprise at least one domain derived from the 5' NTR of HCV. In certain embodiments, the GVB-B/HCV chimera may include at least domain III of the 5' NTR derived from the 5' NTR of HCV. In yet other embodiments the infectious GBV-B clone may comprise domain III of the 5' NTR of HCV, which may or may not include one or more structural or non-structural genes of HCV also incorporated into the chimeric virus. The portions of the 5' NTR of the GVB-B/HCV chimeras will generally be replaced by analogous sequences from the 5' NTR of HCV. It will be understood that the portions or parts of the 5' NTR of GBV-B that may be replaced include all or part of domain I (including sub-region Ia and Ib of GBV-B), domain II, domain III, domain IV, or any combination thereof. Any combination of 5' NTR domains of GBV-B may be replaced with an analogous region of HCV. In certain embodiments, the replacement of a GBV-B region may be accompanied by the deletion of the 5' NTR GBV-B domain Ib region. In addition, any one, two, or three of the 5' NTR domains of GBV-B may be replaced in any combination with analogous sequences from HCV.

[0027] In further embodiments of the invention, a polynucleotide encoding a GBV-B/HCV chimera including a 5' NTR domain III sequence derived from a HCV 5' NTR may be propagated *in vivo*, in particular, in the liver of an appropriate host.

[0028] Various other embodiments may include isolated polynucleotides comprising a chimeric GBV-B genome, wherein at least part, but not all of a 5' NTR sequence is derived from a HCV 5' NTR. The polynucleotides may be synthetic RNA, RNA, DNA or the like.

[0029] Some embodiments include one or more virus, one or more hepatotropic virus, and/or one or more viral expression constructs comprising a chimeric GBV-B polynucleotides including at least a part of the 5' NTR sequence is derived from a HCV 5' NTR.

[0030] Methods of producing a chimeric GBV-B virus encoding at least part of a 5' NTR sequence derived from a HCV 5' NTR sequence comprising introducing into a host cell a viral expression construct comprising a chimeric GBV-B polynucleotide encoding at least part of a 5' NTR sequence derived from a HCV 5' NTR sequence and culturing said host cell under conditions permitting production of a virus from said construct are contemplated. The method may use a host cell that is a eukaryotic cell and the host cell may be in an animal. The method

may further include the step of isolating virus from said host cell and in particular purify the virus to homogeneity.

[0031] In addition, methods for identifying a compound active against a viral infection comprising are contemplated. The methods may include providing a virus expressed from a viral construct comprising at least part of a 5' NTR derived from a HCV 5' NTR, as described herein; contacting said virus with a candidate substance; and comparing the infectious ability of the virus in the presence of said candidate substance with the infectious ability of the virus in a similar system in the absence of said candidate substance. Each of the embodiments may use or include any of the 5' NTR chimeras described herein.

[0032] It is contemplated that any method or composition described herein can be implemented with respect to any other method or composition described herein.

[0033] Other embodiments of the invention may include a compound active against a viral infection identified according to the method described above.

[0034] The use of the term "or" in the claims is used to mean "and/or" unless explicitly indicated to refer to alternatives only or the alternatives are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and "and/or."

[0035] Throughout this application, the term "about" is used to indicate that a value includes the standard deviation of error for the device or method being employed to determine the value.

[0036] It is specifically contemplated that any embodiments described in the Examples section are included as an embodiment of the invention.

[0037] As used herein the specification, "a" or "an" may mean one or more. As used herein in the claim(s), when used in conjunction with the word "comprising," the words "a" or "an" may mean one or more than one. As used herein "another" may mean at least a second or more.

[0038] Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating preferred embodiments of the

invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0040] FIG. 1. Schematic representation of full-length, chimeric GBV-B/HCV cDNAs containing substitution of (C)-E1-E2-p7/p13 coding sequences. cDNA sequences from GBV-B and HCV are represented by blue and grey boxes, respectively, and the corresponding nontranslated regions (NTR) and encoded proteins are indicated on top of or below each genome. The boundaries of the substituted sequences encoding E1-E2-p7/p13 or C-E1-E2-p7/p13 are indicated by their respective amino acid positions within parental polyproteins for each chimeric GBV-B / HCV cDNA constructed (as named on the left). Restriction sites that have been used to linearize cDNAs prior to in vitro transcription are positioned within the cDNAs and indicated by arrows.

[0041] FIG. 2. Schematic representation of full-length, chimeric GBV-B/HCV cDNAs containing substitution of (C)-E1-E2-p7/p13-NS2 coding sequences. cDNA sequences from GBV-B and HCV are represented by blue and grey boxes, respectively, and the corresponding nontranslated regions (NTR) and encoded proteins are indicated on top of each genome. In each chimeric GBV-B/ HCV cDNA (as named on the left), the substituted sequences encoding E1-E2-p7/p13-NS2 or C-E1-E2-p7/p13-NS2 are C-terminally fused to sequences of the protease domain of NS3 (NS3pro) derived from the same virus, followed by the ubiquitin gene (Ubi, red box). The location of the cleavage carried out by the deubiquitination enzyme is indicated by a dark red arrow. Restriction sites that have been used to linearize cDNAs prior to in vitro transcription are positioned within the cDNAs and indicated by arrows.

[0042] FIGs. 3A and 3B. Comparative analysis of translational efficiencies of chimeric RNAs containing heterologous 5'nontranslated region (NTR) and core-coding sequences with chimeric RNAs containing homologous such sequences. (3A) Schematic

representation of RNAs transcribed in vitro from chimeric cDNAs linearized at the BamHI or AvrII restriction sites in the HCV or GBV-B backbones, respectively. (3B) Chimeric, truncated RNAs (as indicated on the top of the gel) were translated in vitro in rabbit reticulocyte lysates in the presence of [35S]Met and the resulting products were separated by SDS-PAGE. Polypeptides are marked by a star (as in panel A) at the right of each lane in the gel, and identified on either side of the gel. The number of methionine residues (# Met) present in each polypeptide is indicated at the bottom of the gel for quantitation purposes (see text).

[0043] FIGs. 4A and 4B. Analysis of the signal peptidase-mediated proteolytic cleavages at heterologous junctions within chimeric polyproteins containing substitutions of (C)-E1-E2-p7/p13 sequences. Panel (4A) shows data relative to the heterologous C / E1 junction, whereas panel (4B) shows data relative to the heterologous p13 / NS2 junction. In panel (4A), RNAs were transcribed from parental or chimeric cDNAs linearized within the N-terminal NS3 coding sequence at the BstZ17I restriction site in HCV or the AflIII site in GBV-B (see FIG. 1). In panel (4B), RNAs were transcribed from cDNAs linearized either within the NS4B coding sequence, or within the NS3 coding sequence (site BstZ17I), as indicated below the gel. Corresponding, encoded polypeptide precursors are schematically represented at the bottom of each panel. RNAs were translated in rabbit reticulocyte lysates in the presence of canine microsomal membranes and either [35S]Cys (4A) or [35S]Met (4B). Resulting polypeptides were separated by 16% (4A) or 14% (4B) SDS-PAGE. GBV-B (marked "GB") and HCV (marked "HC") proteins are identified on both sides of the gels.

[0044] FIGs. 5A and 5B. Analysis of the signal peptidase-mediated proteolytic cleavages at heterologous junctions within chimeric polyproteins containing substitutions of E1-E2-p7/p13-NS2 sequences. (5A) Schematic representation of chimeric polypeptide precursors translated from RNAs transcribed in vitro from cDNAs linearized at the indicated restriction sites (see also FIG. 2). Cleavages carried out by either cellular signal peptidases (◇), viral NS2-3 proteinase (arrow), or viral NS3/4A proteinase (arrow) within these polypeptides are indicated. NS3pro represents the N-terminal third of NS3 that corresponds to the protease domain of NS3. (5B) RNAs transcribed from cDNA templates indicated above each lane of the gels were translated in rabbit reticulocyte lysates in the presence of [35S]Met, and in the presence (left panel) or in the absence (right panel) of canine microsomal membranes. Resulting polypeptides were separated by 14% SDS-PAGE. GBV-B (marked as "GB") and HCV (marked as "HC") polypeptides, as well as fusion polypeptides with ubiquitin (Ubi) are identified on both sides of

the gels. Proteins with identical electrophoresis mobilities (NS3HC, E2HC, NS3GB) are indicated by special signs at the right side of each lane.

[0045] FIG. 6A and 6B. Analysis of the replication capacities in cell culture of chimeric genomes generated in the backbone of HCV A. cDNAs from either infectious genotype 1a HCV (Yanagi *et al.*, 1997), cell culture-adapted genotype 1a HCV (HCV A; Yi and Lemon, 2004), chimeric HCA /C-p13GB or HCA /E1-p13GB constructs in the backbone of HCV A (A), or chimeric HCA /C-NS3pro GB-Ubi or HCA /E1-NS3pro GB-Ubi constructs in the backbone of HCV A (B), were linearized at the XbaI restriction site prior to in vitro transcription to generate corresponding full-length RNAs. 2×10^6 Huh-7 cells were transfected with 5 μ g of each RNA by electroporation (240 volts, 900 μ F). Non-adapted, infectious genotype 1a HCV RNA, which is not capable of replication in cell culture, was used as a negative control. Total RNA was extracted from transfected cells at 4 days post-transfection, quantified by optical density, and 7.5 μ g of these RNAs were loaded on a denaturing agarose gel, transferred to a Nylon membrane, and subjected to detection by Northern blot with an [α -32P]-UTP-labeled riboprobe of negative polarity specific for HCV 3' sequences. Known quantities of in vitro transcribed viral HCV A RNAs (10^7 à 10^9 genome equivalents (ge)) were mixed with RNAs isolated from mock-transfected cells and run in parallel on the gel to serve as a size marker and for quantitation purposes. The amount of viral RNAs present in each extract was quantified by densitometry after analysis of the Northern blots with a PhosphorImager (Molecular Dynamics) and normalized with respect to amounts of housekeeping beta-actin mRNAs present in the same samples and quantified with an [α -32P]-UTP-labeled specific riboprobe (bottom images). Cellular and viral RNAs of interest are identified by arrows.

[0046] FIGs. 7A-7D. Identification of chimeric virus-like particles by electron microscopy. Cellular extracts from insect cells infected with recombinant baculoviruses expressing parental C-p7HC (7A, 7B) or chimeric CGB/E1-p7HC (7C, 7D) structural protein precursors were prepared and fractionated on sucrose gradients to isolate virus-like particles (VLPs). After immunoblotting with specific antibodies directed to all three structural proteins (Core of HCV or GBV-B, E1 and E2 of HCV), 2 fractions (#12+13) were found to contain CHC (7A, 7B) or CGB (14C, 14D), as well as E1HC and E2HC. These fractions were pooled, diluted in PBS, and centrifugated to eliminate sucrose and concentrate the immuno-reactive material. After immunogold labeling with antibodies specific for E1HC (A4; panels 7B, 7D), or antibodies specific for E2HC that recognize a functionally-folded form of E2 (H53; panels 7A, 7C),

followed by staining with uranyl acetate, observation of the material by electron microscopy revealed the presence of VLPs with both parental (7A, 7B) and chimeric (7C, 7D) constructs. Scale bars at the bottom of the pictures correspond to 100 nm.

[0047] FIG. 8 Shows cDNA of chimeric constructs of the invention.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

I. GVB-B VIRUS

[0048] The GBV-B genome structure is very similar to hepatitis C and these viruses share approximately 25% nucleotide identity (Simons *et al.*, 1995; Muerhoff *et al.*, 1995). As indicated above, this makes GBV-B more closely related to HCV than any other known virus. GBV-B genomic RNA is about 9.5 kb in length (Muerhoff *et al.*, 1995) with a structured 5' noncoding region that contains an IRES that shares many structural features with the HCV IRES (Honda *et al.*, 1996; Rijnbrand *et al.*, 1999). As in HCV, this IRES drives the cap-independent translation of a long open reading frame. The polyprotein expressed from this reading frame appears to be organized identically to that of HCV, and processed to generate proteins with functions similar to those of HCV (Muerhoff *et al.*, 1995). In fact, the major serine proteinases of these viruses (NS3) have been shown to have similar cleavage specificities (Scarselli *et al.*, 1997). Finally, like HCV and distinct from the pestiviruses, the genomic RNA of GBV-B has a poly(U) tract located near its 3' terminus (Simons *et al.*, 1995; Muerhoff *et al.*, 1995). In addition, unreported sequences located at the extreme 3' end of the genome have been identified. This work indicates that the GBV-B RNA, like that of HCV (HCV (Tanaka *et al.*, 1995; Kolykhalov *et al.*, 1996), terminates in a lengthy run of heterogeneous bases (310 nts in GBV-B) possessing a readily apparent secondary structure.

[0049] The HCV structural region (C E1 E2 p7) polypeptide contains four internally located endoplasmic reticulum (ER) signal peptide sequences, which lead the nascent HCV polyprotein to translocate across the ER membrane. Host signal peptidases cleave the signal peptides at their C terminus in the lumen of the ER and release the mature viral proteins, including the capsid protein C, the two envelope glycoproteins E1 and E2, and the p7 polypeptide, immediately upstream of sequences of NS2 nonstructural protein. Signal peptides remain N terminally linked to the C terminus of E1, E2, and p7, and are likely to serve as an ER membrane anchor for those proteins, due to the hydrophobicity of their amino acid sequence (Op

De Beeck *et al.*, 2001). In contrast, the signal peptide present at the C terminus of the core protein is cleaved off in the mature form of the core protein (McLauchlan *et al.*, 2002).

[0050] The hydropathy profile of the N terminal part of the GBV B polyprotein is very similar to that of the HCV structural precursor. The inventors have determined by sequencing the boundaries of the GBV-B core, E1, E2, and NS2 proteins and demonstrated the existence of a 13 kDa- protein between E2 and NS2 that is partially homologous to HCV p7 (Ghibaud *et al.*, 2004).

[0051] Studies using transient viral and non-viral expression systems have shown that HCV glycoproteins can follow two different folding pathways, leading to the formation of either noncovalent heterodimers, the probable native complexes, or disulfide bond aggregates, likely representing dead-end products. The folding of E1 is slow and co-expression of E2 is necessary for the proper folding of E1. In addition, transmembrane domains of both E1 and E2 are involved in heterodimerization and contain E2 retention signals. The native E1-E2 complexes identified in the ER most probably represent a prebudding form of HCV glycoprotein heterodimers (Dubuisson, 2000).

[0052] HCV p7 is a small transmembrane protein that functions as an ion channel (Carrere-Kremer *et al.*, 2002; Griffin *et al.*, 2003) and that has been shown to be necessary in the virus life cycle (Sakai *et al.*, 2003). GBV-B p13 is likely to share ion channel activity, but this remains presently unexplored.

[0053] The HCV and GBV-B NS2 sequences are located downstream of p7 and p13 sequences, respectively, and are cleaved at their N terminus by a cellular signal peptidase, as mentioned. These proteins are very hydrophobic and contain several putative transmembrane domains (Yamaga and Ou, 2002). HCV NS2 is not required for RNA replication, as it has been shown that subgenomic replicons containing only the NS3 to NS5B sequences of HCV, hence devoid of the NS2 sequence, are capable of replication in Huh7 cells (Lohmann *et al.*, 1999). The only known function of HCV NS2 is to carry a proteolytic activity that is responsible for cis-cleavage at the NS2/NS3 junction (Grakoui *et al.*, 1993; Pieroni *et al.*, 1997). The exact nature of this proteolytic activity is still controversial, and it is not known whether it is a metalloproteinase or a cysteine protease. Because only the C-terminal half of NS2 is required for the NS2/NS3 cleavage, it is very likely that NS2 plays another, yet undefined but critical role in the viral life cycle. There is a strong possibility that it may play a role in infectious viral particle

assembly and/or particle final maturation steps and export, based on analogy with other flaviviruses (Kummerer and Rice, 2002; Agapov *et al.*, 2004). The mechanisms underlying HCV particle assembly, budding and release remain, however, very poorly understood in the absence of a cell culture system that supports HCV replication.

[0054] One goal was to create chimeric GBV-B/HCV genomes that encode GBV-B sequences necessary for tamarin host-range specificity and HCV sequences substituted to GBV-B sequences that are not specifically required for virus replication in tamarins. Since GBV-B host-range determinants are unknown, chimeric cDNAs were constructed in which the replication unit, comprising sequences encoding all nonstructural proteins as well as 5' and 3' nontranslated regions (NTRs), was derived from GBV-B, and sequences encoding envelope glycoproteins were derived from HCV. Similarly, complementary genomes designed to encode GBV B envelope proteins in an HCV backbone were also constructed.

[0055] In designing such genomes, it was considered likely that proper glycoprotein folding and particle assembly would require the two homologous E1 and E2 proteins of GBV-B, like with HCV. Therefore, sequences encoding both E1 and E2 were considered as a unit in the substitutions described herein. In the absence of information on the virus-specific role of p7/p13 proteins, sequences of these proteins were also exchanged together with that of E1-E2 proteins.

[0056] Second, whether there are virus-specific interactions between core (C) protein and the viral RNA or between core and envelope (E1, E2) proteins is unknown for these flaviviruses. Only one report described a specific interaction between HCV C and E1 proteins (Lo *et al.*, 1996). One piece of evidence suggesting that core protein/RNA interaction may be more stringent than core/envelope protein interaction comes from the fact that it is possible to substitute envelope proteins (prM-E) only, but not capsid and envelope proteins together (C-prM-E), between viruses of the flavivirus genus without disrupting genome infectivity (Pletnev and Men, 1998; Chambers *et al.*, 1999). Third, whether cellular signalases would be able to process heterologous junctions that were created in the chimeric polyproteins are assessed.

[0057] Chimeric cDNAs were engineered in which the sequences encoding E1-E2-p13 only or C-E1-E2-p13 were exchanged. For that purpose, a mutagenesis strategy based on overlapping extension PCR was used. Two chimeric cDNAs were thus obtained in the backbone of an infectious GBV-B molecular clone (Martin *et al.*, 2003), in which E1-E2-p13 or C-E1-E2-

p13 sequences were replaced by E1-E2-p7 or C-E1-E2-p7 sequences of the H77 strain of HCV genotype 1a (Yanagi *et al.*, 1997) (GB/E1-p7HC, GB/C-p7HC; FIG. 1). GBV-B protein boundaries were derived from protein sequencing (Ghibaud *et al.*, 2004). Complementary constructs containing GBV-B E1-E2-p13 or C-E1-E2-p13 sequences in the backbone of the infectious HCV 1a molecular clone (gift of R. Purcell, N.I.A.I.D., Bethesda) were also engineered (HC/E1-p13GB, HC/C-p13GB; FIG. 1).

[0058] Chimeric GBV-B or HCV cDNAs were also engineered in which sequences encoding E1-E2-p7/p13-NS2 were replaced by analogous sequences from the other virus. To avoid problems at the fusion site between heterologous NS2 and NS3 sequences, the sequence of the ubiquitin gene may be inserted such as to cleave the chimeric polyprotein at the first amino acid of the downstream NS3 sequence. To avoid fusion of the ubiquitin sequence to the upstream heterologous NS2 protein and to release mature NS2, sequences of the proteinase domain of NS3 from the same virus, which are known to be sufficient for cis-cleavage at the NS2/NS3 site in HCV (Thibeault *et al.*, 2001), were fused to the NS2 sequences. Hence, chimeric GBV-B cDNAs pGB/E-NS3proHC-Ubi or pGB/C-NS3proHC-Ubi encoding (C)-E1-E2-p7-NS2-NS3pro proteins of HCV followed by Ubiquitin, as well as chimeric HCV cDNAs pHC/E1-NS3proGB-Ubi or pHC/C-NS3proHC-Ubi encoding (C)-E1-E2-p13-NS2-NS3pro proteins of GBV-B followed by Ubiquitin were generated (FIG. 2).

II. NUCLEIC ACIDS

[0059] The present invention provides a nucleic acid sequence encoding a 3' sequence of the GBV-B genome (SEQ ID NO:1). It should be clear that the present invention is not limited to the specific nucleic acids disclosed herein. As discussed below, a "3' sequence of the GBV-B genome" may contain a variety of different bases and yet still be functionally indistinguishable from the sequences disclosed herein.

A. Polynucleotides Encoding the 3' Sequence of the GBV-B Genome

[0060] A 3' sequence of the GBV-B genome disclosed in SEQ ID NO:1 is one aspect of the present invention. Nucleic acids according to the present invention may encode the 3' sequence of the GBV-B genome set forth in SEQ ID NO:1, the entire GBV-B genome, or any other fragment of a 3' sequence of the GBV-B genome set forth herein. The nucleic acid may be

derived from genomic RNA as cDNA, *i.e.*, cloned directly from the genome of GBV-B. cDNA may also be assembled from synthetic oligonucleotide segments.

[0061] It also is contemplated that a 3' sequence of the GBV-B genome may be represented by natural variants that have slightly different nucleic acid sequences but, nonetheless, maintain the same general structure and perform the same function in RNA replication.

[0062] As used in this application, the term "a nucleic acid encoding a 3' sequence of the GBV-B genome" refers to a nucleic acid molecule that may be isolated free of total viral nucleic acid. In preferred embodiments, the invention concerns nucleic acid sequences essentially as set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, and SEQ ID NO:12. The term "as set forth in SEQ ID NO:1" means that the nucleic acid sequence substantially corresponds to a portion of SEQ ID NO:1. It is contemplated that the techniques and methods described in this disclosure may apply to any of the sequences contained herein, including SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, and SEQ ID NO:12.

[0063] Allowing for the degeneracy of the genetic code, sequences that have at least about 50%, usually at least about 60%, more usually about 70%, most usually about 80%, preferably at least about 90% and most preferably about 95% of nucleotides that are identical to the nucleotides of SEQ ID NO:1 will be sequences that are "as set forth in SEQ ID NO:1." Sequences that are essentially the same as those set forth in SEQ ID NO:1 may also be functionally defined as sequences that are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO:1 under standard conditions.

[0064] The nucleic acid segments and polynucleotides of the present invention include those encoding biologically functional equivalent 3' sequences of the GBV-B genome. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques or may be introduced randomly and screened later for the desired function, as described below.

[0065] 3' sequence of the GBV-B genome sequences also are provided. Each of the foregoing is included within all aspects of the following description. The present invention concerns cDNA segments reverse transcribed from GBV-B genomic RNA (referred to as "DNA"). As used herein, the term "polynucleotide" refers to an RNA or DNA molecule that may be isolated free of other RNA or DNA of a particular species.

[0066] "Isolated substantially away from other coding sequences" means that the 3' sequence of the GBV-B genome forms the significant part of the RNA or DNA segment and that the segment does not contain large portions of naturally-occurring coding RNA or DNA, such as large fragments or other functional genes or cDNA noncoding regions. Of course, this refers to the polynucleotide as originally isolated, and does not exclude genes or coding regions later added to the it by the hand of man.

[0067] In certain other embodiments, the invention concerns isolated DNA segments (cDNA segments reverse transcribed from GVB-B genomic RNA) and recombinant vectors that include within their sequence a nucleic acid sequence essentially as set forth in SEQ ID NO:1. The term "essentially as set forth in SEQ ID NO:1" is used in the same sense as described above.

[0068] It also will be understood that nucleic acid sequences may include additional residues, such as additional 5' or 3' sequences, and still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological activity. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include additional various non-coding sequences flanking either of the 5' or 3' portions of the coding region, which are known to occur within viral genomes.

[0069] Sequences that are essentially the same as those set forth in SEQ ID NO:1 also may be functionally defined as sequences that are capable of hybridizing to a nucleic acid segment containing the complement of SEQ ID NO:1 under relatively stringent conditions. Suitable relatively stringent hybridization conditions will be well known to those of skill in the art.

[0070] The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other RNA or DNA sequences, such as

promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

[0071] For example, nucleic acid fragments may be prepared that include a short contiguous stretch identical to or complementary to SEQ ID NO:1, such as about 15-24 or about 25-34 nucleotides and that are up to about 259 nucleotides being preferred in certain cases. Other stretches of contiguous sequence that may be identical or complementary to any of the sequences disclosed herein, including the SEQ ID NOS. include the following ranges of nucleotides: 50-9,399, 100-9,000, 150-8,000, 200-7,000, 250-6,000, 300-5,000, 350-4,000, 400-3,000, 450-2,000, 500-1000. RNA and DNA segments with total lengths of about 1,000, about 500, about 200, about 100 and about 50 base pairs in length (including all intermediate lengths) are also contemplated to be useful.

[0072] In a non-limiting example, one or more nucleic acid constructs may be prepared that include a contiguous stretch of nucleotides identical to or complementary to SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, or SEQ ID NO:12. Such a stretch of nucleotides, or a nucleic acid construct, may be about, or at least about, 3, about 4, about 5, about 6, about 7, about 8,, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 31, about 32, about 33, about 34, about 35, about 36, about 37, about 38, about 39 about 40, about 45, about 50, about 55, about 60, about 65, about 70, about 75, about 80, about 85, about 90, about 95, about 100, about 105, about 110, about 115, about 120, about 125, about 130, about 135, about 140, about 145, about 150, about 155, about 160, about 165, about 170, about 175, about 180, about 185, about 190, about 195, about 200, about 210, about 220, about 230, about 240, about 250, about 260, about 270, about 280, about 290, about 300, about 310, about 320, about 330, about 340, about 350, about 360, about 370, about 380, about 390, about 400, about 410, about 420, about 430, about 440, about 450, about 460, about 470, about 480, about 490, about 500, about 510, about 520, about 530, about 540, about 550, about 560, about 570, about 580, about 590, about 600, about 610, about 618, about 650, about 700, about 750, about 1,000, about 2,000, about

3,000, about 4,000, about 5,000, about 6,000, about 7,000, about 8,000, about 9,000, about 9,100, about 9,200, about 9,300, about 9,399, about 9,400, about 9,500, about 9,600, about 9,700, about 9,800, about 9,900, about 10,000, about 15,000, about 20,000, about 30,000, about 50,000, about 100,000, about 250,000, about 500,000, about 750,000, to about 1,000,000 nucleotides in length, as well as constructs of greater size, up to and including chromosomal sizes (including all intermediate lengths and intermediate ranges), given the advent of nucleic acids constructs such as a yeast artificial chromosome are known to those of ordinary skill in the art.

[0073] It will be readily understood that “intermediate lengths,” in these contexts means any length between the quoted ranges, such as 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, etc.; 100, 101, 102, 103, etc.; 150, 151, 152, 153, etc.; including all integers through the 200-500; 500-1,000; 1,000-2,000; ranges, up to and including sequences of about 1,001, 1,250, 1,500, and the like.

[0074] The various probes and primers designed around the disclosed nucleotide sequences of the present invention may be of any length. By assigning numeric values to a sequence, for example, the first residue is 1, the second residue is 2, *etc.*, an algorithm defining all primers can be proposed:

$$n \text{ to } n + y$$

[0075] where n is an integer from 1 to the last number of the sequence and y is the length of the primer minus one, where $n + y$ does not exceed the last number of the sequence. Thus, for a 20-mer, the probes correspond to bases 1 to 20, 2 to 21, 3 to 22 ... and so on. For a 30-mer, the probes correspond to bases 1 to 30, 2 to 31, 3 to 32 ... and so on. For a 35-mer, the probes correspond to bases 1 to 35, 2 to 36, 3 to 37 ... and so on.

B. Oligonucleotide Probes and Primers

[0076] Naturally, the present invention also encompasses RNA and DNA segments that are complementary, or essentially complementary, to the sequence set forth in SEQ ID NO:1. Nucleic acid sequences that are “complementary” are those that are capable of base-pairing according to the standard Watson-Crick complementary rules. As used herein, the term

“complementary sequences” means nucleic acid sequences that are substantially complementary, as may be assessed by the same nucleotide comparison set forth above, or as defined as being capable of hybridizing to the nucleic acid segment of SEQ ID NO. 1 under relatively stringent conditions such as those described herein. Such sequences may encode the entire 3' sequence of the GBV-B genome or functional or non-functional fragments thereof.

[0077] Alternatively, the hybridizing segments may be shorter oligonucleotides. Sequences of 17 bases long should occur only once in the human genome and, therefore, suffice to specify a unique target sequence. Although shorter oligomers are easier to make and increase *in vivo* accessibility, numerous other factors are involved in determining the specificity of hybridization. Both binding affinity and sequence specificity of an oligonucleotide to its complementary target increases with increasing length. It is contemplated that exemplary oligonucleotides of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225 or more base pairs will be used, although others are contemplated. Longer polynucleotides encoding 250, 500, 1000, 1212, 1500, 2000, 2500, 3000 or 3431 bases and longer are contemplated as well. Such oligonucleotides will find use, for example, as probes in Southern and Northern blots and as primers in amplification reactions.

[0078] Suitable hybridization conditions will be well known to those of skill in the art. In certain applications, for example, substitution of amino acids by site-directed mutagenesis, it is appreciated that lower stringency conditions are required. Under these conditions, hybridization may occur even though the sequences of probe and target strand are not perfectly complementary but are mismatched at one or more positions. Conditions may be rendered less stringent by increasing salt concentration and decreasing temperature. For example, a medium stringency condition could be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37°C to about 55°C, while a low stringency condition could be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Thus, hybridization conditions can be readily manipulated and thus will generally be a method of choice depending on the desired results.

[0079] In other embodiments, hybridization may be achieved under conditions of, for example, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 10 mM dithiothreitol, at temperatures between approximately 20°C to about 37°C. Other hybridization conditions

utilized could include approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, at temperatures ranging from approximately 40°C to about 72°C. Formamide and SDS also may be used to alter the hybridization conditions.

[0080] One method of using probes and primers of the present invention is in the search for other viral sequences related to GBV-B or, more particularly, homologs of the GBV-B sequence. By varying the stringency of hybridization, and the region of the probe, different degrees of homology may be discovered.

[0081] Another way of exploiting probes and primers of the present invention is in site-directed, or site-specific, mutagenesis. The technique provides a ready ability to prepare and test sequence variants, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into complementary DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences that encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

[0082] The technique typically employs a bacteriophage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids are also routinely employed in site directed mutagenesis, which eliminates the step of transferring the gene of interest from a phage to a plasmid.

[0083] In general, site-directed mutagenesis is performed by first obtaining a single-stranded vector, or melting of two strands of a double stranded vector which includes within its sequence a DNA sequence encoding the desired protein. An oligonucleotide primer bearing the desired mutated sequence is synthetically prepared. This primer is then annealed with the single-stranded DNA preparation, taking into account the degree of mismatch when selecting hybridization conditions, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated

sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement. There are newer and simpler site-directed mutagenesis techniques that can also be employed for this purpose. These include procedures marketed in kit form that are readily available to one of ordinary skill in the art.

[0084] The preparation of sequence variants of the selected gene using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting, as there are other ways in which sequence variants of genes may be obtained. For example, recombinant vectors encoding the desired gene may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

C. Antisense Constructs

[0085] In certain embodiments of the invention, the use of antisense constructs of the 3' sequence of the GBV-B genome is contemplated.

[0086] Antisense methodology takes advantage of the fact that nucleic acids tend to pair with "complementary" sequences. By complementary, it is meant that polynucleotides are those which are capable of base-pairing according to the standard Watson-Crick complementarity rules. That is, the larger purines will base pair with the smaller pyrimidines to form combinations of guanine paired with cytosine (G:C) and adenine paired with either thymine (A:T) in the case of DNA, or adenine paired with uracil (A:U) in the case of RNA. Inclusion of less common bases such as inosine, 5-methylcytosine, 6-methyladenine, hypoxanthine and others in hybridizing sequences does not interfere with pairing.

[0087] Targeting double-stranded (ds) DNA with polynucleotides leads to triple-helix formation; targeting RNA will lead to double-helix formation. Antisense polynucleotides, when introduced into a target cell, specifically bind to their target polynucleotide and interfere with transcription, RNA processing, transport, translation and/or stability. Antisense RNA constructs, or DNA encoding such antisense RNAs, may be employed to inhibit gene transcription or translation or both within a host cell, either *in vitro* or *in vivo*, such as within a host animal, including a human subject.

[0088] Antisense constructs could be used to block early steps in the replication of GBV-B and related viruses, by annealing to 3' terminal sequences and blocking their role in negative-strand initiation.

[0089] As stated above, "complementary" or "antisense" means polynucleotide sequences that are substantially complementary over their entire length and have very few base mismatches. For example, sequences of 15 bases in length may be termed complementary when they have complementary nucleotides at 13 or 14 positions. Naturally, sequences which are completely complementary will be sequences which are entirely complementary throughout their entire length and have no base mismatches. Other sequences with lower degrees of homology also are contemplated. For example, an antisense construct which has limited regions of high homology but also contains a non-homologous region (*e.g.*, ribozyme; see below) could be designed. These molecules, though having less than 50% homology, would bind to target sequences under appropriate conditions.

D. Amplification and PCR™

[0090] The present invention utilizes amplification techniques in a number of its embodiments. Nucleic acids used as a template for amplification are isolated from cells contained in the biological sample, according to standard methodologies (Sambrook *et al.*, 1989). The nucleic acid may be genomic DNA or RNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to convert the RNA to a complementary DNA using reverse transcriptase (RT). In one embodiment, the RNA is genomic RNA and is used directly as the template for amplification. In others, genomic RNA is first converted to a complementary DNA sequence (cDNA) and this product is amplified according to protocols described below.

[0091] Pairs of primers that selectively hybridize to nucleic acids corresponding to GBV-B sequences are contacted with the isolated nucleic acid under conditions that permit selective hybridization. The term "primer," as defined herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded or single-stranded form, although the single-stranded form is preferred.

[0092] Once hybridized, the nucleic acid:primer complex is contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplification, also referred to as "cycles," are conducted until a sufficient amount of amplification product is produced.

[0093] Next, the amplification product is detected. In certain applications, the detection may be performed by visual means. Alternatively, the detection may involve indirect identification of the product via chemiluminescence, radioactive scintigraphy of incorporated radiolabel or fluorescent label or even via a system using electrical or thermal impulse signals.

[0094] A number of template dependent processes are available to amplify the marker sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCR™) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, and each incorporated herein by reference in entirety.

[0095] Briefly, in PCR™, two or more primer sequences are prepared that are complementary to regions on opposite complementary strands of the marker sequence. An excess of deoxynucleoside triphosphates are added to a reaction mixture along with a DNA polymerase, *e.g.*, *Taq* polymerase. If the marker sequence is present in a sample, the primers will bind to the marker and the polymerase will cause the primers to be extended along the marker sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the marker to form reaction products, excess primers will bind to the marker and to the reaction products and the process is repeated.

[0096] A reverse transcriptase PCR™ amplification procedure may be performed in order to quantify the amount of mRNA amplified. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook *et al.*, 1989. Alternative methods for reverse transcription utilize thermostable, RNA-dependent DNA polymerases. These methods are described in WO 90/07641, filed December 21, 1990, incorporated herein by reference. Polymerase chain reaction methodologies are well known in the art.

[0097] Another method for amplification is the ligase chain reaction ("LCR"), disclosed in EPA No. 320 308, incorporated herein by reference in its entirety. U.S. Patent 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence.

[0098] Qbeta Replicase, described in PCT Application No. PCT/US87/00880, incorporated herein by reference, also may be used as still another amplification method in the present invention.

[0099] An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-triphosphates in one strand of a restriction site also may be useful in the amplification of nucleic acids in the present invention.

[0100] Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.*, nick translation. A similar method, called Repair Chain Reaction (RCR), involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases can be added as biotinylated derivatives for easy detection. A similar approach is used in SDA. Target specific sequences also can be detected using a cyclic probe reaction (CPR). In CPR, a probe having 3' and 5' sequences of non-specific DNA and a middle sequence of specific RNA is hybridized to DNA that is present in a sample. Upon hybridization, the reaction is treated with RNase H, and the products of the probe identified as distinctive products that are released after digestion. The original template is annealed to another cycling probe and the reaction is repeated.

[0101] Still another amplification methods described in GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR-like, template- and enzyme-dependent synthesis. The primers may be modified by labeling with a capture moiety (*e.g.*, biotin) and/or a detector moiety (*e.g.*, enzyme). In the latter application, an excess of labeled probes are added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

[0102] Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR Gingeras *et al.*, PCT Application WO 88/10315, incorporated herein by reference.

[0103] Davey *et al.*, EPA No. 329 822 (incorporated herein by reference in its entirety) disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention.

[0104] Miller *et al.*, PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target ssDNA followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, *i.e.*, new templates are not produced from the resultant RNA transcripts. Other amplification methods include "RACE" and "one-sided PCR" (Frohman, 1990 incorporated by reference).

[0105] Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, also may be used in the amplification step of the present invention.

[0106] Following any amplification, it may be desirable to separate the amplification product from the template and the excess primer for the purpose of determining whether specific amplification has occurred. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods. See Sambrook *et al.*, 1989.

[0107] Alternatively, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography that may be used in the present invention: adsorption, partition, ion-exchange and molecular sieve, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography.

[0108] Amplification products must be visualized in order to confirm amplification of the marker sequences. One typical visualization method involves staining of a gel with ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the amplification products can then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

[0109] In one embodiment, visualization is achieved indirectly. Following separation of amplification products, a labeled nucleic acid probe is brought into contact with the

amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, and the other member of the binding pair carries a detectable moiety.

[0110] In one embodiment, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art and can be found in many standard books on molecular protocols. See Sambrook *et al.*, 1989. Briefly, amplification products are separated by gel electrophoresis. The gel is then contacted with a membrane, such as nitrocellulose or nylon, permitting transfer of the nucleic acid and non-covalent binding. Subsequently, the membrane is incubated with a chromophore-conjugated probe that is capable of hybridizing with a target amplification product. Detection is by exposure of the membrane to x-ray film or ion-emitting detection devices.

[0111] One example of the foregoing is described in U.S. Patent 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.

E. Expression Constructs

[0112] In some embodiments of the present invention, an expression construct that encodes a 3' sequence of GBV-B is utilized. The term "expression construct" is meant to include any type of genetic construct containing a nucleic acid coding for a gene product in which part or all of the nucleic acid encoding sequence is capable of being transcribed. The transcript may be translated into a protein, but it need not be. Expression includes both transcription of a gene and translation of mRNA into a gene product. Expression may also include only transcription of the nucleic acid encoding a gene of interest.

[0113] In some constructs, the nucleic acid encoding a gene product is under transcriptional control of promoter and/or enhancer. The term promoter will be used here to refer to a group of transcriptional control modules that are clustered around the initiation site for RNA polymerase II. Much of the thinking about how promoters are organized derives from analyses of several viral promoters, including those for the HSV thymidine kinase (tk) and SV40

early transcription units. These studies have shown that promoters are composed of discrete functional modules, each consisting of approximately 7-20 bp of nucleic acids, and containing one or more recognition sites for transcriptional activator or repressor proteins.

[0114] At least one module in each promoter functions to position the start site for RNA synthesis. The best known example of this is the TATA box, but in some promoters lacking a TATA box, such as the promoter for the mammalian terminal deoxynucleotidyl transferase gene and the promoter for the SV40 late genes, a discrete element overlying the start site itself helps to fix the place of initiation.

[0115] Additional promoter elements regulate the frequency of transcriptional initiation. Typically, these are located in the region 30-110 bp upstream of the start site, although a number of promoters have recently been shown to contain functional elements downstream of the start site as well. The spacing between promoter elements frequently is flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. In the tk promoter, the spacing between promoter elements can be increased to 50 bp apart before activity begins to decline. Depending on the promoter, it appears that individual elements can function either co-operatively or independently to activate transcription.

[0116] Enhancers were originally detected as genetic elements that increased transcription from a promoter located at a distant position on the same molecule of DNA. This ability to act over a large distance had little precedent in classic studies of prokaryotic transcriptional regulation. Subsequent work showed that regions of nucleic acids with enhancer activity are organized much like promoters. That is, they are composed of many individual elements, each of which binds to one or more transcriptional proteins.

[0117] The basic distinction between enhancers and promoters is operational. An enhancer region as a whole must be able to stimulate transcription at a distance; this need not be true of a promoter region or its component elements. On the other hand, a promoter must have one or more elements that direct initiation of RNA synthesis at a particular site and in a particular orientation, whereas enhancers lack these specificities. Promoters and enhancers are often overlapping and contiguous, often seeming to have a very similar modular organization.

F. Host Cells and Permissive Cells

[0118] As used herein, the terms “cell,” “cell line,” and “cell culture” may be used interchangeably. All of these terms also include their progeny, which is any and all subsequent generations. It is understood that all progeny may not be identical due to deliberate or inadvertent mutations. In the context of the present invention, “host cell” refers to a prokaryotic or eukaryotic cell, and it includes any transformable organisms that is capable of replicating a vector or virus and/or expressing viral proteins. A host cell can, and has been, used as a recipient for vectors, including viral vectors. A host cell may be “transfected” or “transformed,” which refers to a process by which exogenous nucleic acid is transferred or introduced into the host cell. A transformed cell includes the primary subject cell and its progeny. A “permissive cell” refers to a cell that supports the replication of a given virus and consequently undergoes cell lysis. In the context of the present invention, such a virus would include HCV, GBV-B, or other hepatitis viruses. In a “nonpermissive cell,” productive infection does not result, but the cell may become stably transformed. In some embodiments, methods employ permissive cells that are a cell line derived from liver cells (liver cell line).

[0119] Host cells may be derived from prokaryotes or eukaryotes, depending upon whether the desired result is replication of the vector or expression of part or all of the vector-encoded nucleic acid sequences. Numerous cell lines and cultures are available for use as a host cell, and they can be obtained through the American Type Culture Collection (ATCC), which is an organization that serves as an archive for living cultures and genetic materials (www.atcc.org). An appropriate host can be determined by one of skill in the art based on the vector backbone and the desired result. A plasmid or cosmid, for example, can be introduced into a prokaryote host cell for replication of many vectors. Bacterial cells used as host cells for vector replication and/or expression include DH5 α , JM109, and KC8, as well as a number of commercially available bacterial hosts such as SURE[®] Competent Cells and SOLOPACK[™] Gold Cells (STRATAGENE[®], La Jolla). Alternatively, bacterial cells such as *E. coli* LE392 could be used as host cells for phage viruses.

[0120] Examples of eukaryotic host cells for replication and/or expression of a vector include HeLa, NIH3T3, Jurkat, 293, Cos, CHO, Saos, and PC12. Many host cells from various cell types and organisms are available and would be known to one of skill in the art. Similarly, a viral vector or virus or virus particle may be used in conjunction with either a

eukaryotic or prokaryotic host cell, particularly one that is permissive for replication or expression of the vector. It is contemplated that the present invention includes vectors composed of viral sequences, viruses, and viral particles in the methods of the present invention, and that they may be used interchangeably in these methods, depending on their utility.

[0121] Some vectors may employ control sequences that allow it to be replicated and/or expressed in both prokaryotic and eukaryotic cells. One of skill in the art would further understand the conditions under which to incubate all of the above described host cells to maintain them and to permit replication of a vector. Also understood and known are techniques and conditions that would allow large-scale production of vectors, as well as production of the nucleic acids encoded by vectors and their cognate polypeptides, proteins, or peptides.

III. PHARMACEUTICAL COMPOSITIONS

[0122] The present invention encompasses the use of a 3' sequence of GBV-B in the production of or use as a vaccine to combat HCV infection. Compositions of the present invention comprise an effective amount of GBV-B clone as a therapeutic dissolved or dispersed in a pharmaceutically acceptable carrier or aqueous medium. The phrases "pharmaceutically or pharmacologically acceptable" refer to molecular entities and compositions that do not produce an adverse, allergic or other untoward reaction when administered to an animal, or a human, as appropriate.

[0123] As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions. For human administration, preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biologics standards.

[0124] The biological material should be extensively dialyzed to remove undesired small molecular weight molecules and/or lyophilized for more ready formulation into a desired vehicle, where appropriate. The active compounds will then generally be formulated for parenteral administration, *e.g.*, formulated for injection via the intravenous, intramuscular, sub-

cutaneous, intralesional, or even intraperitoneal routes. The preparation of an aqueous composition that contains GVB-B nucleic acid sequences as an active component or ingredient will be known to those of skill in the art in light of the present disclosure. Typically, such compositions can be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for using to prepare solutions or suspensions upon the addition of a liquid prior to injection can also be prepared; and the preparations can also be emulsified.

[0125] The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions; formulations including sesame oil, peanut oil or aqueous propylene glycol; and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi.

[0126] Solutions of the active compounds as free base or pharmacologically acceptable salts can be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

[0127] A GBV-B clone of the present invention can be formulated into a composition in a neutral or salt form. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the protein) and are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. In terms of using peptide therapeutics as active ingredients, the technology of U.S. Patents 4,608,251; 4,601,903; 4,599,231; 4,599,230; 4,596,792; and 4,578,770, each incorporated herein by reference, may be used.

[0128] The carrier also can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the

required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

[0129] Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof. The preparation of more, or highly, concentrated solutions for direct injection is also contemplated, where the use of DMSO as solvent is envisioned to result in extremely rapid penetration, delivering high concentrations of the active agents to a small area.

[0130] Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms, such as the type of injectable solutions described above, but drug release capsules and the like also can be employed.

[0131] For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, sterile aqueous media which can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage could be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-

1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject.

[0132] In addition to the compounds formulated for parenteral administration, such as intravenous or intramuscular injection, other pharmaceutically acceptable forms include, *e.g.*, tablets or other solids for oral administration; liposomal formulations; time release capsules; and any other form currently used, including cremes.

[0133] Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders. In certain defined embodiments, oral pharmaceutical compositions will comprise an inert diluent or assimilable edible carrier, or they may be enclosed in hard or soft shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compounds may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 0.1% of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 2 to about 75% of the weight of the unit, or preferably between 25-60%. The amount of active compounds in such therapeutically useful compositions is such that a suitable dosage will be obtained.

[0134] The tablets, troches, pills, capsules and the like may also contain the following: a binder, as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active

compounds sucrose as a sweetening agent methyl and propylparabens as preservatives, a dye and flavoring, such as cherry or orange flavor.

EXAMPLES

[0135] The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLE 1: Infectious GBV-B Genome

[0136] The inventors have elucidated a previously unrecognized 3' terminal sequence of GBV-B (SEQ ID NO 1). This sequence was reproducibly recovered from tamarin serum containing GBV-B RNA, in RT-PCR nucleic acid amplification procedures using several different primer sets, and as a fusion with previously reported 5' GBV-B sequences.

[0137] There is information in the published literature reporting the putative sequences of the 5' and 3' termini of the GBV-B genome. The nucleic acid sequences of these termini were reportedly determined by ligating the ends of the viral RNA together, amplifying the sequence in the region of the resulting junction by reverse-transcription polymerase chain reaction (RT-PCR), and sequencing of the cDNA amplification product across the junction. However, the inventors believed these results required confirmation. Of particular concern was the fact that the 3' terminus appeared to be shorter than the equivalent region of other viruses in the family *Flaviviridae* (especially within the genus *Hepacivirus*) and that the reported 3' sequence lacked a defined RNA hairpin structure such as those present in these related viruses. Additional novel sequences at the 3' end of the GBV-B genome were investigated using a serum sample collected from a tamarin that was experimentally infected with virus. Amplification was used to determine the sequence of the 3' end.

[0138] First, serum (50 μ L) known to contain GBV-B RNA by RT-PCR assay was extracted with Trizol, and the RNA was washed and dried. A synthetic oligonucleotide was then ligated to the 3' end of the viral RNA. The oligonucleotide, AATTCGGCCCTGCAGGCCACAACAGTC (SEQ ID NO:27), which was phosphorylated at the 5' end and chemically blocked at the 3' end, was ligated to the RNA essentially using the method described by Kolykhalov *et al.* (1996). The RNA was initially dissolved in DMSO and the following additions were made: Tris-Cl, pH 7.5 (10 mM), $MgCl_2$ (10 mM), DTT (5 mM), hexamine cobalt chloride (1 mM), 10 pmol oligo and 8 U T4 ligase. The final concentration of DMSO was 30% in a final volume of 10 μ L. The ligation reaction was incubated for 4 or 20 hours at 19° C. 1 μ L of the ligation reaction was used directly to make cDNA, using a primer complementary to the ligated oligonucleotide and the Superscript 2 system, in a final volume of 15 μ L. 1 μ L of cDNA was amplified using the Advantage cDNA system (Clontech) and two additional oligonucleotide primers. These primers included one that was complementary to the ligated oligonucleotide (*i.e.*, "negative sense") and a positive-sense primer located near the 3' end of the reported GBV-B sequence. A product approximately 290 bases in length was obtained, and this was gel purified and directly sequenced. Sequencing was done in both directions using the oligonucleotide primers employed for the amplification; 259 bases that had not been previously reported were identified as fused to the sequence that had been previously described as the 3' terminus of the viral genome.

[0139] To ensure that this novel 3' sequence from viral RNA could be reproducibly amplified, an additional 10 μ L of infected tamarin serum was extracted using Trizol. cDNA was prepared by reverse transcription using an oligonucleotide primer complementary to the penultimate 3' 25 bases of the novel sequence. Amplification was then done by PCR using the primer previously utilized for cDNA synthesis and a positive-sense primer mapping within the previously published GBV-B sequence. In the initial studies, although a product was readily detected, DNA sequencing showed that this product was missing all of the sequence distal to the poly-U tract. Carrying out the cDNA synthesis in the presence of DMSO circumvented this problem. A cDNA product of approximately 290 bases was obtained. This was sequenced and shown to consist of the 5' primer, 20 bases of the published GBV-B sequence, and 259 bases of the novel sequence obtained in the preceding studies and containing the sequence of the 3' primer. The sequence of the 3' end of GBV-B is shown in SEQ ID NO: 1. The presence of a

predicted hairpin structure at the extreme 3' end of this novel sequence is consistent with its location at the 3' terminus of the viral RNA.

[0140] The GBV-B cDNA (synthesis described above) was used as a template for PCR amplification of the 3' 1553 nucleotides (nts) of the GBV-B genome. This PCR amplification product was gel purified and cloned into plasmid DNA using the "Perfectly Blunt Cloning Kit" (Novagene).

[0141] **Construction of an Infectious GBV-B Clone** - The elucidation of a 3' sequence of the GBV-B genome will allow those of skill in the art to construct and validate an infectious molecular clone of GBV-B. This will be done using the following procedures.

[0142] A full-length cDNA copy of the GBV-B genome containing the newly identified 3' terminal sequences was constructed. RNA transcribed from this cDNA copy of the genome will be infectious when inoculated into the liver of a GBV-B permissive tamarin, giving rise to rescued GBV-B virus particles.

[0143] A 1:1000 dilution of GBV-B infectious tamarin serum was obtained. This material was used as a source of viral RNA for the amplification of GBV-B nucleic acid sequences by reverse-transcription polymerase chain reaction. For amplification of previously reported segments of the GBV-B genome, 250 μ L of the diluted serum was extracted with Trizol using the manufacturer's instructions. The final RNA pellet was dissolved in 10 μ L of a 100 mM DTT buffer containing 5% RNasin. This material was converted into cDNA using Superscript 2 reverse transcriptase and oligonucleotide primers designed to be complementary to the reported GBV-B RNA sequence and to contain unique restriction sites. This cDNA was amplified using the Advantage cDNA kit (Clontech) employing the cDNA primer (negative sense) as the downstream primer and a similar positive-sense upstream primer, again containing a unique restriction site. The published sequence of GBV-B allowed for the selection of primers in convenient areas of the genome containing unique restriction sites. Using this general strategy, the inventors amplified segments of the reported GBV-B genome representing: (1) nucleotides (nts) 1-1988, using an upstream primer containing a T7 RNA polymerase promoter and a BamHI site upstream of nt 1, and a downstream primer containing a unique EcoRI site (nt 1978); (2) nts 1968-5337, using a downstream primer containing a unique ClaI site at position 5027; (3) nts 5317- 7837, using a downstream primer containing a SalI site at nt 7847; and, (4)

nts 7837-9143, using a downstream primer containing an added Xho1 site. It was found necessary to use different PCR conditions for each primer set.

[0144] The RT-PCR products generated in these reactions were cloned into plasmid DNA after gel purification, using the "Perfectly Blunt Cloning Kit" (Novagene). Ten bacterial colonies from each of the four RT-PCR products were analyzed for insert size by restriction endonuclease digestion using EcoR1, the sites for this enzyme being located on either side of the insert in the resulting plasmids. For three of the RT-PCR amplicons, 9 of 10 colonies contained plasmids with the correct size insert. The EcoR1-Cla1 amplicon generated only 1/10 colonies with a correct size insert. Thus, 30 additional colonies were examined, yielding two more clones with insert of the correct size. For each of these plasmids, simple restriction patterns were obtained using two restriction enzymes. As these appeared to be correct, the plasmid DNAs were subjected to sequencing using an ABI automatic sequencer.

[0145] **Nucleotide sequence of the cloned GBV-B cDNA** - The 5' region of the cloned sequence revealed a relatively long nontranslated region corresponding to the published sequence of the GBV-B 5'NTR, which includes an IRES. This region was followed by a long open reading frame. Near the 3' end of the genome a poly-U tract was identified; however, this was shorter than the published 3' homopolymeric poly-U region. The sequence from these clones was compared with those in the GenBank database (Accession U22304, "Hepatitis GB virus B polypeptide complete genome"). Twenty-two nucleotide differences were identified, of which 14 gave rise to amino acid changes (Table 1). In order to determine whether these changes were genuine or RT-PCR artifacts, which could have been introduced due to the very small amount of material from which these sequences were amplified, segments of the genome containing these changes were reamplified using a serum sample from an independently infected tamarin. Of the 14 changes noted in the original cDNA clones, 12 were not present in these newly amplified sequences and thus were probably RT-PCR artifacts (Table 2.). A particularly interesting difference from the published GenBank sequence, however, which was present in both the original clones as well as a repeat amplification, was a two-nucleotide substitution that obliterated the Sal1 site present in the published sequence.

Table 1. Differences in the amino acid sequences of GBV-B cDNA clones and the GBV-B sequence reported by Simons *et al.* (1995).

GBV-B Protein	AMINO ACID Δ FROM ABBOTT SEQUENCE	RT-PCR Products from Tamarin 12024 (PCR Reaction #)
Core	G ₉₉ →S	G (38.1a)
E1	V ₃₉₅ →I	V (40.2a)
E2	D ₇₀₃ →N	D (42.3a)
E2	P ₇₀₆ →Q	H (42.3a)
E2	A ₇₂₈ →V	A (42.3a)
NS2	L ₇₉₁ →F	F(42.3a)
NS2	T ₈₀₄ →A	A(42.3a)
NS5A	L ₁₉₉₀ →M	L (46.5a)
NS5A	I ₂₀₈₂ →T	I (46.5a)
NS5A	S ₂₁₇₄ →P	(not done)
NS5A	G ₂₂₂₈ →E	E (48.6a)
NS5A	T ₂₂₃₃ →S	T (48.6a)
NS5A	A ₂₂₃₆ →V	A (48.6a)
NS5B	V ₂₈₃₃ →I	V (50.7a)

[0146] **Construction of a full-length GBV-B cDNA clone.** - The four GBV-B cDNA inserts described above were cloned into Bluescript ks⁺ using unique restriction sites. Since the unique SalI site that was reported to be present in the published GBV-B sequence (nt position 7847) was absent in these cDNA clones, this restriction site was created by engineering two silent nucleotide changes using the “Quick Change” mutagenesis system (Stratagene). Although the most 5' clones (nucleotides 1-7847) could be readily constructed, attempts to add the remaining 3' clones were unsuccessful due to rearrangements and deletions. This problem was overcome by use of pACNR1180, a plasmid that had been used to construct an infectious clone of yellow fever virus. Finally, the most 3' 771 nucleotides of GBV-B were excised from the plasmid containing the novel, previously unreported 3' sequence, and inserted into the truncated assembled GBV-B cDNA construct to complete the 3' end. The 3' terminus of this full-length cDNA was then subjected to DNA sequencing to confirm its integrity. Extensive restriction digests indicated that this construct had the characteristics of a full-length cDNA copy of GBV-B virus. Because there is not yet an understanding of which cultured cells (if any)

might be permissive for GBV-B replication, the infectivity of the synthetic GBV-B RNA will be assessed by injecting the RNA directly into the liver of a susceptible tamarin.

[0147] Alternatively, an infectious full-length clone can be produced by the following protocol. A plasmid will be made containing a cassette including the 5' and 3' ends of the virus flanked by appropriate restriction sites. These constructs have been shown to efficiently translate reporter genes, with transcription taking place via a T7 promoter placed immediately upstream of the 5'NTR (*e.g.*, *see* Rijnbrand *et al.*, 1999). The major portion of the GBV-B genome would then be amplified by long range RT-PCR. This method is now well established for hepatitis C virus and other flaviviruses (Teller *et al.*, 1996), and it has been used successfully also to amplify rhinovirus RNA. Briefly this technique uses "Superscript" reverse transcriptase to synthesize cDNA and a mixture of "KlenTaq 1", and "DeepVent" polymerases to amplify this cDNA. Primers that can be used will contain restriction sites to allow cloning of the RT-PCR products into the cassette vector. After being transformed into suitable competent bacteria, extensive restriction analysis will enable us to determine which clones contain inserts that are of full length and which have a high probability of being correct. Apparent full-length clones will be analyzed further by coupled transcription-translation using the Promega "TnT" system, with the addition of microsomal membranes to allow the cleavage of the structural proteins by cellular signalase enzymes. Clones which appear correct by restriction analysis and which produce GBV-B proteins, in particular the protein coded for by the extreme 3' end of the genome, NS5B, will be selected, and RNA will be transcribed from these clones using the Ambion MegaScript system. "Correct" looking clones (>10) can be injected directly into a tamarin liver at several sites. A successful infection will be determined as described below. If a positive signal is detected the entire genome will be amplified and sequenced to determine which plasmid the virus originated from.

[0148] **Rescue of Infectious GVB-B.** Infectious GBV-B will be rescued from synthetic genome-length RNA following its injection into the liver of tamarins (*Saquinus sp.*). In past studies, HAV from synthetic RNA in owl monkeys has been recovered (*Aotus trivirgatus*) (Shaffer *et al.*, 1995), and more recently, the recovery of virus from a chimpanzee injected intrahepatically with RNA transcribed from a full-length genotype 1b HCV cDNA clone was reported (Beard *et al.*, 1999).

[0149] RNA will be prepared for these studies using the T7 MegaScript kit (Ambion) and a total of 10 µg of plasmid DNA as template. An aliquot of the reaction products will be utilized to ensure the integrity of the RNA by electrophoresis in agarose-formaldehyde gels. The remainder of the transcription reaction mix will be frozen at -80°C until its injection, without further purification, into the liver of a tamarin. Because of the small size of the tamarin, the RNA will be injected under direct visualization following a limited incision and exposure of the liver. Under similar conditions, in other primate species, RT-PCR-detectable viral RNA or cDNA has not been detected in serum samples collected within days of this procedure in the absence of viral replication (Kolykhalov *et al.*, 1997; Yanagi *et al.*, 1998; Beard *et al.*, 1999). Thus, the appearance of RNA in serum collected subsequently from these tamarins will be strong evidence for the replication competence of the synthetic RNA. Serum will be collected weekly for six weeks, then every other week for an additional 6 weeks from inoculated animals. In addition to RT-PCR for detection of viral RNA, alanine aminotransferase (ALT) levels will be measured as an indicator of liver injury and to assess liver histology in punch biopsies taken at the time of ALT elevation. Maximum viremia and an acute phase ALT response is expected to occur around 14-28 days post-inoculation of infectious RNA (Simons *et al.*, 1995; Schlauder *et al.*, 1995; Karayiannis *et al.*, 1989). Transfections will be considered to have failed to give rise to infectious virus if RNA is not detected in the serum within 12 weeks of inoculation. Successfully infected animals will be followed with twice weekly bleeds until resolution of the viremia, or for 6 months, whichever is longest.

EXAMPLE 2: GBV-B/HCV Chimeras

[0150] The GBV-B genome can be used as the acceptor molecule in the construction of chimeric viral RNAs containing sequences of both HCV and GBV-B. Such constructs will allow one to investigate the mechanisms for the different biological properties of these viruses and to discover and investigate potential inhibitors of specific HCV activities (*e.g.*, proteinase) required for HCV replication. Different classes of chimeric viruses are contemplated. These include: (a) replacement of the GBV-B IRES with that of HCV; and (b) replacement of the NS3 major serine proteinase and helicase, and (c) the replacement of the NS5B RNA-dependent RNA polymerase with the homologous proteins of HCV.

[0151] The chimeric constructs described in the following sections will be made by PCR mutagenesis, using high fidelity polymerases and oligonucleotide primers designed to

include the specific fusions of GBV-B and HCV sequences (Landt *et al.*, 1990). First round PCR reactions will create the desired fusion, and generate a new “primer” to be used in a second PCR reaction spanning the region to a convenient unique restriction site. PCR cycles will be kept to the minimum number necessary for successful amplification, and all segments of viral sequence that are amplified by PCR will be subjected to DNA sequencing to exclude the presence of unwanted PCR-introduced errors. Sequencing will be accomplished at UTMB’s core Recombinant DNA Laboratory. Amplified segments will be kept to the minimum by the exchange of cloned cDNA segments spanning convenient restriction sites in subgenomic clones, and where necessary PCR artifacts can be corrected by site-directed mutagenesis (QuickChange mutagenesis kit, Stratagene).

[0152] A number of viable positive-strand RNA virus chimeras have been constructed previously in which IRES elements have been swapped between different viruses. Most of these chimeras have involved the exchange of IRES elements between picornaviruses. Others have been successful in constructing viable poliovirus chimeras containing the HCV IRES in place of the native poliovirus IRES (Zhao *et al.*, 1999; Lu and Wimmer, 1996). A similar rhinovirus 14 chimera containing the HCV IRES has been constructed, although its replication phenotype is not as robust as the poliovirus chimera described by Lu and Wimmer (Lu and Wimmer, 1996). More importantly, Frolov *et al.* (Frolov *et al.*, 1998) recently reported chimeric flaviviruses in which the HCV IRES was inserted into the genetic background of a pestivirus, bovine viral diarrhea virus (BVDV) in lieu of the homologous BVDV sequence. Although these viable chimeric polioviruses and pestiviruses replicate in cell cultures, they are poor surrogates for HCV in animal models as neither virus is hepatotropic or causes liver disease. Importantly, Frolov *et al.* (Frolov *et al.*, 1998) demonstrated quite convincingly that the requirement for cis-acting replication signals at the 5’ terminus of the pestivirus genome was limited to a short tetranucleotide sequence. This requirement presumably reflects the need for the complement of this sequence at the 3’ end of the negative strand during initiation of positive-strand RNA synthesis. The work of Frolov *et al.* shows that the IRES of BVDV does not contain necessary replication signals, or that if these are present within the BVDV IRES they can be complemented with similar signals in either the HCV or encephalomyocarditis virus (EMCV, a picornavirus) IRES sequence. Since GBV-B and HCV are more closely related to each other than BVDV and HCV, these studies provide strong support for the viability of chimeras containing the HCV IRES in the background of GBV-B.

[0153] Construction of a viable IRES chimera will be enhanced by studies that have documented the sequence requirements and secondary structures of the IRES elements of both HCV and GBV-B (*see* Lemon and Honda, 1997; Honda *et al.*, 1996; Rijnbrand *et al.*, 1999). To a considerable extent, the work of Frolov *et al.* (Frolov *et al.*, 1998) was guided by studies of the HCV IRES structure. More recently, these studies have been extended to include a detailed mutational analysis of the GBV-B IRES. The results of these studies indicate that the functional IRES of GBV-B extends from the 5' end of structural domain II (nt 62) to the initiator AUG codon (nt 446). This segment of the full-length GBV-B clone will be replaced with HCV sequence extending from 5' end of the analogous domain II within the HCV IRES (nt 42) to the initiator codon at the 5' end of the HCV open reading frame (nt 341) to construct the candidate chimera, "GB/C:IRES". The source of HCV cDNA for these studies will be the infectious HCV clone, pCV-H77C, which contains the sequence of the genotype 1a Hutchinson strain virus (Yanagi *et al.*, 1998), whose infectivity in a chimpanzee following intrahepatic inoculation with synthetic RNA transcribed from pCV-H77C has been confirmed.

[0154] This GB/C:IRES construct will retain two upstream hairpins within the GBV-B sequence (stem-loops Ia and Ib), and it is thus analogous to the viable "BVDV+HCVdelB2B3H1" chimera of Frolov *et al.* (Frolov *et al.*, 1998). A second chimera can be constructed in which the entire HCV 5' nontranslated RNA will be inserted in lieu of nts 62-446 of the GBV-B genome ("GB/C:5'NTR"). This construct will add to the inserted HCV sequence the most 5' stem-loop from HCV (stem-loop I). A similar insertion was shown to substantially increase the replication capacity of BVDV+HCVdelB2B3H1 by Frolov *et al.* (Frolov *et al.*, 1998), providing a replication phenotype similar to wild-type BVDV in cell culture.

[0155] It is important to point out that there is strong evidence from multiple lines of investigation indicating that it will not be necessary to include coding sequence in these IRES chimeras. This is the case even though Reynolds *et al.* (Reynolds *et al.*, 1995) have argued that the HCV IRES extends past the initiator codon, and into the core-coding region of that virus. Although Lu and Wimmer (Lu and Wimmer, 1996) found it necessary to include HCV core sequence to obtain a viable chimeric poliovirus, the BVDV chimeras reported by Frolov *et al.* (Frolov *et al.*, 1998) did not contain any HCV coding sequence. This discrepancy may be explained by the observation that the only downstream requirement for full activity of both the GBV-B and HCV IRES elements is the presence of an unstructured RNA segment (Honda *et al.*,

1996; Rijnbrand *et al.*, 1999). Presumably, this facilitates interaction of the viral RNA with the 40S ribosome subunit in the early steps of cap-independent translation (Honda *et al.*, 1996). The 5' GBV-B coding sequence fulfills this criterion (Rijnbrand *et al.*, 1999).

[0156] ***In vitro* Characterization of the Translational Activity of IRES Chimeras.** - The fidelity of the genome-length chimeric constructs will be confirmed by sequencing any DNA segments that have been subjected to PCR during the construction process, as well as confirming sequence at the junction sites. In addition, the translational activity of synthetic RNA derived from these constructs will be assessed and compared to the translational activity of the wild-type GBV-B and HCV RNAs. These studies will be carried out in a cell-free translation assay utilizing rabbit reticulocyte lysates (Rijnbrand *et al.*, 1999). Synthetic RNA will be produced by runoff T7 RNA polymerase transcription using as template ClaI-digested plasmid DNA (BamHI digestion in the case of the genome HCV construct) (T7 Megascript kit, Ambion). ³H-UTP will be added to the reaction mix to allow for quantification of the RNA product. Reticulocyte lysates (Promega) will be programmed for translation by the addition of RNA (at least 50% full-length as determined by agarose gel electrophoresis) at 20, 40 and 80 µg/ml, and translation reactions will be supplemented with microsomal membranes (Promega). ³⁵S-Methionine-labelled translation products will be separated by SDS-PAGE, and the quantity of E1 protein produced from each RNA determined by PhosphorImager analysis (Molecular Dynamics). Comparisons of the activity of the HCV IRES in the background of GBV-B and HCV will take into account differences in the methionine content of the E1 proteins of these viruses. Based on previous studies of both the GBV-B and HCV IRES elements (Honda *et al.*, 1996; Rijnbrand *et al.*, 1999), it is expected that these studies will confirm that the HCV IRES will retain nearly full activity when placed within the GBV-B background.

[0157] ***In vivo* Characterization of IRES Chimeras.** Synthetic RNAs produced from each of the two chimeric GBV-B/HCV constructs (GB/C:IRES and GB/C:5'NTR) will be tested for their ability to induce infection and cause liver disease in susceptible tamarins. These studies will be carried out as described in Example 2. GB/C:5'NTR may generate viremia and liver injury more closely resembling that observed with wild-type GBV-B infection (Frolov *et al.*, 1998).

[0158] **Chimeric GBV-B/HCV with HCV NS3 in a GBV-B Background.** Chimeric flaviviruses containing the HCV NS3 serine proteinase/helicase within the GBV-B

background are also contemplated within the present invention. The construction of chimeric flaviviruses containing specific heterologous functional polyprotein domains, however, poses a number of special problems. Unlike the situation with the IRES, where the relevant RNA segments appear to have a unique function restricted to cap-independent translation initiation and interact with host cell macromolecules, viral proteins often have multiple functions and may form specific macromolecular complexes with other viral proteins that are essential for virus replication (Lindenbach and Rice, 1999). Furthermore, such chimeric polyproteins must be amenable to efficient processing by the viral proteinases (NS2/NS3 or NS3). This requires knowledge of the proteinase cleavage specificities as well as specific sites of proteolytic cleavage. Although to date there have been no published studies of the processing of the GBV-B polyprotein, the relatively close relationship between GBV-B and HCV, about 30% overall amino acid identity within the polyprotein (Muerhoff *et al.*, 1995), allows good computer predictions of the alignments of these proteins. The crystallographic structures of both the proteinase and helicase domains of the HCV NS3 protein have been solved (Yao *et al.*, 1997). Thus, both linear alignments and models of the 3D structure of the NS3 proteins of these viruses can provide guidelines for designing specific chimeric fusions that are likely to preserve function.

[0159] **NS3 Proteinase-Domain Chimeras.** In HCV, NS3 contains the major serine proteinase that is responsible for most cleavage events in the processing of the nonstructural proteins, *i.e.*, those that occur at the NS3/4A, 4A/4B, 4B/5A and 5A/5B junctions. The active proteinase domain of HCV is located within the amino terminal third of the NS3 protein (residues 1-181), which shares 31% amino acid identity with the analogous segment of the GBV-B polyprotein (GBV-B vs HCV-BK) (Muerhoff *et al.*, 1995). Importantly, the active site of this proteinase appears to be particularly well conserved in GBV-B. The GBV-B proteinase maintains the residues that are responsible for catalysis and zinc binding in the HCV enzyme (Muerhoff *et al.*, 1995), and unlike the NS3 proteinases of some other flaviviruses preserves the Phe-154 residue that determines in part the S₁ specificity pocket of the enzyme and the preference of the HCV proteinase for substrates with a cysteine residue at the P1 position (Scarselli *et al.*, 1997). Thus, it is not surprising that the relevant proteolytic cleavage sites within the GBV-B polyprotein that are predicted from alignments with the HCV polyprotein all possess a Cys residue at this position. Of greatest significance for the proposed studies, however, is the work of Scarselli *et al.* (Scarselli *et al.*, 1997) who demonstrated that the GBV-B

NS3 proteinase is able to effectively process the polyprotein of HCV in studies carried out *in vitro*. Using synthetic peptide substrates, these investigators demonstrated that the enzymatic activities of the GBV-B proteinase (residues 1-181) had kinetic parameters similar to the HCV proteinase on NS4A/4B and NS4B/4A substrates HCV. They did not possess reagents allowing a determination of whether the HCV proteinase is able to cleave a GBV-B substrate, but their results indicate that these viral proteinases share important functional properties. Therefore, these similarities suggest that the HCV proteinase could function in lieu of the GBV-B proteinase if used to replace this segment of an infectious GBV-B clone. In addition, studies with sindbis/HCV chimeras have shown that the HCV proteinase can cleave within the framework of a sindbis polyprotein (Filocamo *et al.*, 1997).

[0160] In considering the design of these NS3 proteinase chimeras, there are two additional important considerations. First, in HCV, the cleavage between NS2 and NS3 occurs *in cis*, as the result of a zinc-dependent metalloproteinase that spans the NS2/NS3 junction (Hijikata *et al.*, 1993). As only the NS3 sequences will initially be exchanged, the viability of the resulting chimeras will be dependent upon preservation of the *cis*-active cleavage across a chimeric NS2/NS3 proteinase domain. The alignment of GBV-B and HCV sequences shows that residues in HCV that have been shown by Grakoui *et al.*, 1993, to be essential for the NS2/NS3 cleavage are conserved in GBV-B (Muerhoff *et al.*, 1995). Additional chimeras that will include the relevant carboxyl-terminal portion of NS2 can also be created.

[0161] A second important consideration is that the mature HCV NS3 proteinase functions as a noncovalent assembly of the NS3 proteinase domain and the amino terminal portion of NS4A, a proteinase accessory factor. The details of this association are well known, and have been studied at the crystallographic level (Kim *et al.*, 1996). The N-terminal domain of the folded proteinase contains eight β strands, including one contributed directly by the NS4A peptide backbone. X-ray studies have shown that this array of β strands gives rise to a much more ordered N-terminus. Thus, the presence of the NS4A strand seems likely to contribute to the structure of the substrate-binding pocket. It is not known whether the NS3 proteinase of GBV-B also requires a similar interaction with NS4A of that virus for complete activity, or, if so, whether the NS4A of GBV-B could substitute for NS4A of HCV in forming the fully active NS3 proteinase of HCV. The predicted GBV-B NS4A molecule is 54 amino acid residues in length (Simons, *et al.*, 1995; Muerhoff *et al.*, 1995), just as in HCV. However, the level of amino acid homology between the NS4A molecules is not especially high, and the potential interaction with

either NS3 molecule cannot be predicted from this sequence on the basis of available knowledge. To overcome this potential problem, chimeras will be created in which not only the NS3 proteinase domain of GBV-B is replaced, but also the relevant NS4A segment as well, with homologous segments of the HCV polyprotein. The interaction of the HCV NS3 and NS4A domains represents a unique target for antiviral drug design, and it would be beneficial to have this specific interaction present in any virus to be used as a surrogate for HCV in the evaluation of candidate antiviral inhibitors of HCV proteinase *in vivo*.

[0162] The NS3 proteinase chimeras that can be made include “GB/C:NS3^P”, which will contain the sequence encoding the first 181 amino acid residues of the HCV NS3 molecule in lieu of that encoding the first 181 residues of GBV-B NS3, and “GB/C:NS3^P4A”, which will include the same NS3 substitution as well as the HCV sequence encoding the amino-terminal segment of NS4A that forms the interaction with NS3. The precise NS4A sequence to be included in the latter chimera will be based on the modeling studies, which may also suggest more effective fusions of the NS3 proteinase domain of HCV with the downstream NS3 helicase domain of GBV-B. The source of HCV cDNA for these studies will be the infectious HCV clone, pCV-H77C, which contains the sequence of the genotype 1a Hutchinson strain virus (Yanagi *et al.*, 1998).

[0163] **NS3 Helicase Domain Chimeras.** In addition to serine proteinase activity located within the amino-third of NS3, the downstream carboxy-terminal two-thirds of the molecule contains an RNA helicase activity. These two functional domains appear to be separated by a flexible spacer, within which the fusion of HCV proteinase or helicase domain sequences with GBV-B sequence will be placed. The exact role of the helicase in the HCV life-cycle is not known, but it is almost certainly required for dsRNA strand-separation during some phase of viral RNA synthesis. The helicase domains of GBV-B and HCV are remarkably well conserved, with some regions within the helicase showing as much as 55% amino acid identity (Muerhoff *et al.*, 1995). The GBV-B helicase is more closely related to the HCV helicase than all other flaviviral NS3 helicases, and it preserves many residues found within the conserved helicase motifs of HCV. Thus the HCV NS3 helicase may be capable of functioning when placed within the polyprotein of GBV-B, and such a chimeric virus may be capable of replication. Residues 182-620 of the GBV-B NS3 molecule will be substituted with the analogous segment of HCV. A chimera will also be made in which the entire NS3 and amino terminal NS4 protein sequences of GBV-B is replaced with the homologous HCV sequences

("GB/C:NS3-4A"). The latter construct will thus represent a dual proteinase-helicase chimera. As with the proteinase chimeras, the HCV cDNA will be derived from pCV-H77C (Yanagi *et al.*, 1998).

[0164] ***In vitro* Characterization of NS3 and NS3-NS4A Chimera.** Prior to being evaluated for infectivity in susceptible tamarins, RNAs produced *in vitro* from these clones will be characterized *in vitro*. This evaluation will be restricted to a documentation of the proper processing of the expressed polyprotein (*i.e.*, NS2/NS3 and NS3 proteinase functions), since there are no relevant assays that can determine whether the helicase or RNA-dependent RNA polymerase activities in these polyproteins are sufficient for virus replication. The proteolytic processing of the polyprotein is important, however, as it may be altered either by inclusion of the heterologous HCV NS3 proteinase in lieu of the natural GBV-B protease, or by a change in the folding of the polyprotein induced by inclusion of HCV sequence anywhere within the polyprotein. These studies will be carried out in cell-free coupled transcription/translation assays ("TnT" system, Promega) supplemented with microsomal membranes. Template DNAs will be digested with Sall, which restricts the cDNA within the NS5B coding region. ³⁵S-methionine-labelled translation products will be separated by SDS-PAGE, and the mature NS3 protein identified by its apparent molecular mass. The NS3 and NS5B proteins will be identified by immunoblot analysis using rabbit antisera to the GBV-B NS3 and NS5B proteins. Generation of a mature ~68 kDa NS3 protein will provide proof of both the *cis*-active NS2/NS3 cleavage and the NS3-mediated cleavage of NS3/NS4A. Similarly, identification of a mature, processed NS5B molecule will provide further support for the activity of the NS3 proteinase. Controls for these studies will be the wild-type GBV-B polyprotein expressed in similar fashion from the full-length GBV-B clone. If necessary to more clearly demonstrate the processing of the nonstructural proteins in these constructs, subclones representing the nonstructural region of the chimeric sequences could be produced.

[0165] ***In vivo* Characterization of NS3 and NS3-4A Chimeras.** Synthetic RNAs produced from each of the chimeric GBV-B/HCV constructs described in the preceding section will be tested for their ability to induce infection and cause liver disease in susceptible tamarins. These studies will be carried out using the approach described above.

EXAMPLE 3: Chimeric GBV-B/HCV Containing HCV NS5B in a GBV-B Background

[0166] The HCV NS5B molecule contains an RNA-dependent RNA polymerase that plays a central role in replication of the virus. Although this molecule represents a prime target for drug discovery efforts, it has proven difficult to express NS5B in a form that retains enzymatic activity specific for HCV RNA as a substrate. Thus, relatively little is known of the functional activity of the HCV replicase, including structure-function relationships of NS5B. Despite this, the NS5B proteins of GBV-B and HCV appear to be functionally closely related, as they share as much as 43% amino acid identity (Muerhoff *et al.*, 1995). A more important question may be whether an RNA dependent RNA polymerase can act on foreign substrates. However, published work has shown that *in vitro* purified HCV polymerase has very little specificity for its template, using hepatitis C or globin message with equal fidelity (Behrens *et al.*, 1996; Al *et al.*, 1998). This finding is very similar to that obtained with picornaviral polymerases, where it has been known for many years that *in vitro* the enzyme exhibits very little specificity. It has always been considered highly likely that this situation would not pertain *in vivo* where it was thought that the interaction of viral or cellular factors with the 3' end of the genome would generate template specificity. However, recent reports have shown that the removal of the entire 3' untranslated sequence (leaving, however, the poly(A) region present) from both the poliovirus and rhinovirus genome does not completely abrogate the infectivity of the virus (Todd *et al.*, 1997). Furthermore, virus, which was recovered after the initial transfections, was shown to have recovered much of the infectivity of the original virus (Todd *et al.*, 1997). The mechanism for this recovery of infectivity is at present unknown, but these results suggest that the HCV polymerase may be able to function to replicate infectious GBV-B/HCV NS5B chimeras.

[0167] Thus a chimeric genome-length virus can be created in which the NS5B coding sequence of HCV (amino acids 2422-3014, 593 residues) is inserted within the background of GBV-B in lieu of its native RNA-dependent RNA polymerase (amino acids 2274-2864, 591 residues). This chimeric virus would be valuable for animal studies of candidate antiviral inhibitors of HCV RNA synthesis.

[0168] This NS5B chimera would be evaluated to determine that there was proper proteolytic processing of the polyprotein. This would be accomplished by expression of the chimeric polyprotein in a coupled translation-transcription reaction, followed by immunoblot

analysis for the mature NS5B protein, as described for the NS3 and NS3-4A chimeras in the preceding section. If these results confirm that the GB/C:5B chimeric polyprotein is processed with release of NS5B, studies in tamarins would progress to determine whether synthetic RNA transcribed from the clone is infectious and capable of causing liver disease in intrahepatically inoculated animals. These studies would be carried out as described above.

[0169] A chimeric molecule can be constructed from an infectious GBV-B clone in which the HCV NS3 proteinase or proteinase/helicase sequence would be placed in frame in lieu of the homologous GBV-B sequence, and this chimeric cDNA would be used to generate infectious GBV-B/HCV chimeric viruses by intrahepatic inoculation of synthetic RNA in tamarins. Published studies indicate that the GBV-B and HCV proteinases have closely related substrate recognition and cleavage properties, likely making such chimeras viable and capable of initiating viral replication in appropriate cell types.

EXAMPLE 4: Chimeric Viruses Containing HCV Structural Proteins within a GBV-B Genetic Background, and GBV-B Structural Proteins within an HCV Background

[0170] It is well documented that the structural proteins of one flavivirus may in some cases be substituted for those from another member of the family. Such chimeric viruses have been recovered from viruses as distantly related to each other as dengue virus and tick-borne encephalitis virus (Pletnev *et al.*, 1992). More recently, the prM and E proteins of Japanese encephalitis virus have been used to replace the equivalent proteins in a vaccine strain of yellow fever virus to produce a JE/YF chimera (Chambers *et al.*, 1999). These observations suggest that chimeras in which the structural proteins of HCV have been used to replace the homologous proteins of GBV-B may well be viable and capable of replication. The isolation of a chimeric virus containing HCV structural proteins, but having the growth characteristics of GBV-B virus, could answer many fundamental questions concerning the structure and interaction of these proteins in HCV. They would also be useful in addressing the nature of the immune response to HCV structural proteins in infected primates (Farci *et al.*, 1992). More to the point of this application, the availability of such chimeric viruses would allow studies of candidate HCV vaccines to be carried out in the tamarin model. This would be a major advance, because at present such studies are limited to chimpanzees (Choo *et al.*, 1994).

[0171] The basis for the difference in the host ranges of HCV and GBV-B is completely unknown. Among many other possibilities, it is conceivable that the host range is

dependent upon the availability of a specific receptor(s). If this were the case, host range might be dependent upon the envelope proteins that must interact with the putative cellular receptor. Thus, a chimeric virus containing the envelope proteins of HCV within the genetic background of GBV-B might be noninfectious in tamarins (but potentially infectious in chimpanzees). Thus, a finding that both structural protein chimeras are noninfectious in the tamarin, may require the construction of complementary chimeras in which the relevant GBV-B structural proteins will be inserted into the background of an infectious HCV clone. If inclusion of the GBV-B envelope proteins within the backbone of HCV confers on the resulting chimera the ability to replicate in tamarins, it will confirm an important role for the structural proteins in defining the different host ranges of these viruses. More importantly, the resulting virus would be an exceptionally valuable resource for future studies as it would contain all of the nonstructural replication elements, as well as the 5' and 3' nontranslated regions, of HCV. Such a virus would allow the tamarin model to be used to address many unresolved issues in HCV biology and pathogenesis.

[0172] **Construction and Evaluation of Structural Protein Chimeras.** In designing structural protein chimeras, it is important to note that the two envelope proteins of HCV, E1 and E2, form noncovalent heterodimeric complexes that are likely to be important in the assembly of infectious virus particles. This is not known to be the case with the envelope proteins of GBV-B, but it is likely given similarities in the sizes and hydropathy profiles of these proteins (Simons *et al.*, 1995; Muerhoff *et al.*, 1995). Accordingly, the E1 and E2 proteins will be replaced as a unit, and chimeras containing only one of these proteins from the heterologous virus will generally not be produced. First, a chimera will be created where the E1 and E2 regions of GBV-B virus are replaced with those of HCV, "GB/C:E1-2". The source of HCV cDNA for these constructions will be pCV-H77C (Yanagi *et al.*, 1998). A chimera will also be made in which the core protein, in addition to the envelope proteins, is replaced with the homologous proteins of HCV ("GB/C:Co-E2"). Additional chimeras will be made to determine whether tamarins can be infected with chimeras containing the GBV-B structural proteins within the genetic background of HCV. These will include "C/GB:E1-2" and "C/GB:Co-E2". The backbone for these chimeras will be pCV-H77C, the infectious genotype 1a cDNA clone developed in the Purcell laboratory at NIAID (Yanagi *et al.*, 1998).

[0173] The specific amino acid sequences of GBV-B to be replaced with the homologous segments of HCV have been determined by alignments of the GBV-B and HCV sequences, coupled with the location of signalase cleavage sites predicted to be present within

the amino terminal third of the GBV-B polyprotein using the computer algorithm of Von Heijne. These predicted signalase cleavages lie between residues 156/157 (core/E1), aa 348-349 (E1/E2) and 732/733 (E2/NS2) in the GBV-B sequence. Thus, the chimera GB/C:E1-2 will contain sequence encoding HCV aa 192-809 in lieu of that encoding aa 157-732 in GBV-B, while the insertion in the GB/C:Co-E2 chimera will extend from the initiator AUG codon (aa 1) to residue 809 in HCV, and will be spliced into GBV-B in lieu of the segment encoding aa 1-732 in the GBV-B clone. The complementary chimeras to be constructed within the background of HCV will involve exchanges of the same segments of the genomes.

METHODS

[0174] Infectious, genomic-length cDNAs of GBV-B (Martin *et al.*, 2003) and the H77 isolate of HCV genotype 1a (obtained from Dr. R.H. Purcell, N.I.H., USA, under MTA, Yanagi *et al.*, 1997), both cloned downstream of the T7 RNA polymerase promoter, were used as backbones to construct chimeric cDNAs.

[0175] To carry out exact substitutions of GBV-B sequences by analogous HCV sequences or vice-versa, a PCR-based fusion strategy was used, involving the synthesis of 3 PCR fragments each with 27-29 nucleotide (nt)-overlaps. For example, a central PCR fragment corresponding to the GBV-B sequence to be inserted, as well as 2 PCR fragments corresponding to HCV sequences framing the sequence to substitute, each containing a unique restriction site for cloning purposes, were synthesized from appropriate cDNA templates. A final chimeric PCR fragment was generated using a mixture of the 3 fragments with overlaps. After digestion at the unique sites at the 5' and 3' ends, the chimeric fragment was cloned between the same sites in the GBV-B parental cDNA.

[0176] **pHC/C-p13^{GB} et pHC/E1-p13^{GB}** - Restriction enzymes *Age*I (nt 155) and *Bst*Z17I (nt 3643) of pCV-H77C (Yanagi *et al.*, 1997) within the HCV cDNA were used to insert chimeric HCV/GBV-B/HCV fragments that included sequences coding for E1-E2-p13 (nts 914-2614) or C-E1-E2-p13 (nts 446-2614) of GBV-B, generating plasmids pHC/E1-p13^{GB} and pHC/C-p13^{GB}, respectively (FIG. 1).

[0177] **pGB/C-p7^{HC} et pGB/E1-p7^{HC}** - Restriction enzymes *Ssp*I (nt 11613, in the vector sequences upstream of the T7 promoter) and *Afl*III (nt 3414, in the GBV-B cDNA) of pGBV-B/2 (Martin *et al.*, 2003) were used to insert chimeric GBV-B/HCV/GBV-B fragments

that included sequences coding for E1-E2-p7 (nts 915-2768) or C-E1-E2-p7 (nts 342-2768) of HCV, generating plasmids pGB/E1-p7^{HC} and pGB/C-p7^{HC}, respectively (FIG. 1).

[0178] **pHC/C-NS3_{pro}^{GB}-Ubi et pHC/E1-NS3_{pro}^{GB}-Ubi** - Chimeric PCR fragments spanning nts 1714-3835 of the GBV-B cDNA, followed by the sequence of the ubiquitin gene and nts 3420-3784 of the HCV cDNA were generated by a PCR-based fusion strategy. The cDNA fragments between restriction sites *PmeI* and *SgrAI*, in the GBV-B E2 sequence and in the HCV NS3 sequence, respectively, of plasmids pHC/C-p13^{GB} and pHC/E1-p13^{GB} were replaced by the newly synthesized PCR fragment encoding (E2Δ-NS2 NS3_{pro})^{GB}-Ubi-(NS3Δ)^{HC} (Δ denoting a truncated form of the protein), generating plasmids pHC/C-NS3_{pro}^{GB}-Ubi and pHC/E1-NS3_{pro}^{GB}-Ubi, respectively (FIG. 2).

[0179] **pGB/C-NS3_{pro}^{HC}-Ubi et pGB/E1-NS3_{pro}^{HC}-Ubi** - Chimeric PCR fragments spanning nts 2225-3986 of the HCV cDNA, followed by the sequence of the ubiquitin gene and nts 3266-4678 of the GBV-B cDNA were generated by a PCR-based fusion strategy. The cDNA fragments between restriction sites *SacI* and *BsaBI*, in the HCV E2 sequence and in the GBV-B NS3 sequence, respectively, of plasmids pHC/C-p13^{GB} and pHC/E1-p13^{GB} were replaced by the newly synthesized PCR fragment encoding (E2Δ-NS2 NS3_{pro})^{HC}-Ubi-(NS3Δ)^{GB} (Δ denoting a truncated form of the protein), generating plasmids pGB/C-NS3_{pro}^{HC}-Ubi and pGB/E1-NS3_{pro}^{HC}-Ubi, respectively (FIG. 2).

[0180] In all chimeric cDNAs generated in the backbone of the GBV-B cDNA, the 3 *XhoI* sites present in HCV C and E1 sequences were eliminated, so that the *XhoI* site located downstream of the 3' end of the cDNA remains unique in the corresponding chimeric plasmids and could be used to linearize cDNAs prior to *in vitro* transcription. Similarly, in the cDNAs generated in the HCV backbone, the *XbaI* site was destroyed in the GBV-B E2 sequence. This was carried out by a PCR-based fusion strategy that introduced silent changes within the existing restriction sites.

RESULTS

[0181] Analysis of the translational competence of chimeric genomes. - For both HCV and GBV-B, IRES-driven polyprotein translation initiation is modulated by the nature and probably the structure of the sequence present downstream of the initiating AUG (Rijnbrand *et al.*, 2001).

[0182] The inventors therefore determined whether the fusion of heterologous 5'NTR and core sequences in the case of chimeric genomes GB/E1-p7^{HC}, HC/E1-p13^{GB}, GB/E1-NS3_{pro}^{HC}-Ubi, and HC/E1-NS3_{pro}^{GB}-Ubi could alter the translational competence of these genomes.

[0183] To study this, the inventors used an *in vitro* translation system in rabbit reticulocyte lysates programmed with RNAs transcribed from truncated cDNA templates that had been linearized within the E1 coding sequence, either at the *AvrII* restriction site (in GBV-B) or the *BamHI* site (in HCV) (FIG. 3A)]. Hence, a unique, short polypeptide was synthesized in the absence of canine microsomal membranes in each case. The efficiencies of translation directed from IRESes present in GB/C-p7^{HC} and GB/E1-p7^{HC} were compared, the latter containing an IRES similar to parental GBV-B IRES, and from IRESes present in HC/C-p13^{GB} and HC/E1-p13^{GB}, that contains an IRES similar to that of parental HCV.

[0184] Serial dilutions of *in vitro* transcribed RNA were quantified precisely on agarose gels and 10 ng/μl of RNA were used to program *in vitro* translation reactions in the presence of [³⁵S]Met and in the absence of microsomal membranes. The translation of chimeric 5'NTR-C^{GB}/E1Δ^{HC} RNA (FIG. 3) generated a 34 kDa product, that corresponds to C^{GB} fused to 148 amino acids of E1^{HC}, whereas the translation of chimeric 5'NTR^{GB}/C-E1Δ^{HC} RNA produced a 37 kDa polypeptide corresponding to C^{HC} fused to the same 148 residues of E1^{HC} (FIG. 3B). Comparative quantitations of the resulting products were carried out by densitometry after analysis of the gel with a PhosphorImager (Molecular Dynamics) and corrected with respect to the number of methionine residues present in each polypeptide.

[0185] The inventors found that all constructs were translation-competent. However, the translational efficiency of 5'NTR^{GB}/C-E1Δ^{HC} RNA, that contains heterologous 5'NTR and core sequences, was decreased by 55% with respect to that of 5'NTR-C^{GB}/E1Δ^{HC} RNA. Similarly, the translational efficiency of 5'NTR^{HC}/C-E1Δ^{GB} RNA was decreased by 45% with respect to that of the 5'NTR-C^{HC}/E1Δ^{GB} (FIG. 3B). Therefore, both parental GBV-B and HCV IRESes followed by homologous core sequences are more efficient in directing translation than a chimeric IRES composed of GBV-B 5'NTR sequences followed by HCV core sequences, or an HCV IRES followed by GBV-B core sequences. Although it is not impossible that a two-fold decrease in the translational capacities of chimeric genomes that contain heterologous

5'NTR and core sequences may affect the overall viral production, the data essentially show that all chimeric genomes are translationally competent.

[0186] Analysis of the processing at heterologous junctions within the chimeric polyproteins. - [0191] To monitor proteolytic processing of chimeric GBV-B/HCV structural precursors by cellular signalases, *in vitro* translation reactions in rabbit reticulocyte lysates were programmed with subgenomic RNAs in the presence of either ³⁵S-methionine (FIG. 4B) or ³⁵S-cysteine (FIG. 4A), as well as canine pancreatic microsomal membranes. Since the predicted GBV-B core protein lacks methionine residues with the exception of the initiating methionine residue, ³⁵S-cysteine was utilized to visualize this polypeptide.

[0187] Parental or chimeric cDNA templates in the GBV-B or HCV backbones were linearized with restriction enzymes *Afl*III or *Bst*Z17I, respectively. Such linearized cDNA templates were transcribed *in vitro* using T7 RNA polymerase, generating subgenomic RNAs with a coding capacity corresponding to C-E1-E2-NS2 followed by a short N-terminal segment of NS3 (FIG. 4A). Processing of chimeric precursors HC/E1-p13^{GB} and HC/C-p13^{GB} generated two polypeptides with respective electrophoretic mobilities indistinguishable from those of parental GBV-B E1 and E2 proteins. In addition to envelope glycoproteins, cleavage of HC/E1-p13^{GB} precursor also yielded a polypeptide that co-migrated with the core protein originating from parental HCV precursor (C^{HC}) by SDS-PAGE, exhibiting a ~21 kDa apparent molecular weight (FIG. 4A). This is consistent with what has been described in the literature for the mature HCV core protein (Yasui *et al.*, 1998; McLauchlan *et al.*, 2002). In addition, cleavage of HC/C-p13^{GB} generated a pair of polypeptides that co-migrated with polypeptides presumably corresponding to an immature form of GBV-B core protein containing the E1 signal peptide and the mature form (FIG. 4A).

[0188] Conversely, processing of chimeric precursors containing HCV structural proteins in the GBV-B backbone (GB/E1-p7^{HC}, GB/C-p7^{HC}) proved to generate polypeptides with expected electrophoretic mobilities, identical to those of parental proteins with respect to HCV E1 and E2 glycoproteins and GBV-B core protein (GB/E1-p7^{HC}, (FIG. 4A)). It should be stressed that, whether derived from parental HCV or chimeric GBV-B/HCV precursors, HCV glycoprotein E1 is present in two distinct forms that are likely to reflect various degrees of glycosylation. The reason why HCV core protein was observed only in its mature form, while GBV-B analogous protein seemed to be present in what the inventors anticipate to represent both

immature and mature forms is unclear at present, but might reflect more efficient protein maturation in this system in the case of HCV. Overall, this study strongly suggests that there is no defect in cleavages at heterologous C/E1 or homologous E1/E2 junctions in the chimeric substrates.

[0189] To analyze processing at the heterologous junction engineered between p13 and NS2 in chimeric polyproteins within the HCV backbone, RNA templates were used that can encode a precursor comprising C-NS3 sequences followed by an N-terminal part of NS4B. For that matter, cDNAs were linearized with *BsmI* prior to RNA transcription (FIG. 4B). *In vitro* translation reactions carried out in the presence of microsomal membranes showed that chimeric precursors HC/E1-p13^{GB} and HC/C-p13^{GB} released GBV-B E1 and E2 proteins, as well as HCV core protein in the case of chimera HC/E1-p13^{GB}, as expected from results above. In addition, a polypeptide of approximately 23 kDa, similar to that derived from the processing of HCV native precursor was observed (FIG. 4B). This 23 kDa-polypeptide was identified as HCV NS2 protein on the basis of its absence in the pattern of translated products generated from an HCV precursor that does not include an intact NS3 proteinase domain, as required for the NS2/NS3 cleavage to occur (FIG. 4B). The release of NS2^{HC} is further substantiated by the concomittent production of a polypeptide with an apparent molecular weight of ~67 kDa, that was identified as HCV NS3 protein (FIG. 4B) by immunoprecipitation with relevant antibodies.

[0190] These studies with HCV chimeras containing GBV-B structural proteins strongly suggest that there is accurate processing at the heterologous p13GB/NS2HC junction. Similarly, although GBV-B NS2 was not observed in such *in vitro* translation patterns, the converse p7HC/NS2GB junction appeared to be accurately processed in chimeras constructed in the GBV-B backbone since polypeptides analogous to HCV E2 and GBV-B NS3 proteins were released from these chimeric precursors.

[0191] Similarly, a detailed study of the proteolytic processing of the chimeric polyproteins expressed from RNAs GB/C-NS3_{pro}^{HC}-Ubi, GB/E1-NS3_{pro}^{HC}-Ubi, HC/C-NS3_{pro}^{GB}-Ubi, and HC/E1-NS3_{pro}^{GB}-Ubi has been carried out (for an exmaple, see FIG. 5). In particular, these studies demonstrated that there is efficient cleavage at the engineered NS2/NS3pro and Ubi/NS3 junctions.

[0192] Analysis of the replication capacity of the chimeric genomes. - In the absence of reverse genetic systems in cell culture for either HCV or GBV-B, the inventors were

studied the replication competence of chimeric RNAs generated in the HCV backbone (HC/E1-p13^{GB} et HC/C-p13^{GB}). Sequences coding for structural proteins (C, E1, E2), as well as those coding for p7 and NS2 do not appear to be involved in RNA replication, since a subgenomic HCV RNA devoid of all these sequences does replicate in Huh-7 cells. The inventors thus hypothesized that the substitution of HCV sequences coding for (C)-E1-E2-p7-(NS2) by analogous sequences from GBV-B would not dramatically impair the replication of chimeric HCV/GBV-B genomes. However, the inventors analyzed the impact of decreased translational efficiency of some genomes, or potential impairment of polyprotein cleavage kinetics on genome replication.

[0193] Replication-competent, bicistronic RNAs in Huh-7 cells have essentially been described in the context of HCV genotype 1b (Lohmann *et al.*, 1999). However, Yi *et al.* recently demonstrated that a monocistronic RNA derived from the H77 strain of HCV genotype 1a (H77c/QR/VI/KR/KR^{5A}/SI) was capable of replication in Huh-7 cells (Yi and Lemon, 2004). This RNA (which is referred to as "HCV^A" for adapted HCV) contains 5 coding mutations within NS3, NS4A and NS5A coding sequences with respect to the corresponding infectious RNA (Yanagi *et al.*, 1997), that confer to HCV^A a robust replication phenotype in cell culture. Four days after transfection of Huh-7 cells with RNA transcripts derived from HCV^A, approximately 3×10^7 genome equivalents per μg of total cellular RNA were detected, whereas RNA from the non-adapted HCV molecular clone remained undetectable (FIG. 6).

[0194] In order to study the replication capacity of chimeric genomes generated in the HCV backbone, chimeric GBV-B/HCV sequences from HC/E1-p13^{GB} and HC/C-p13^{GB} constructs were transferred into the backbone of the adapted HCV genome (HCV^A), thus generating corresponding chimeric HC^A/GBV-B cDNAs. Synthetic RNAs were transcribed in vitro from HC^A/E1-p13^{GB}, HC^A/C-p13^{GB}, HCV^A, and non-adapted HCV cDNAs that have been linearized at the *Xba*I site and 5 μg of these RNAs used to transfect 2×10^6 Huh-7 cells. At 4 days post-transfection, total cellular RNAs were prepared and 7.5 μg of RNA, as measured by optical density, were loaded on a denaturing agarose gel and analyzed by Northern blot with an [α -³²P]-UTP riboprobe of negative polarity specific for the 3' end of the HCV genome. Housekeeping β -actin mRNA was also monitored in each sample in order to normalize quantitations of viral RNAs with respect to fixed amounts of cellular RNA.

[0195] As shown in FIG. 6A, chimeric HC^A/E1-p13^{GB} RNA replicated twice as efficiently than chimeric HC^A/C-p13^{GB} which encodes an heterologous core protein, but both RNAs replicate as robustly, if not better, than parental HCV^A RNA. The RNA detected does not reflect residual input RNAs but are clearly derived from *de novo* synthesis since no RNA was detected after transfection of non-adapted HCV RNAs.

[0196] These results show that the substitution of HCV sequences coding for C-E1-E2-p7 or E1-E2-p7 by GBV-B analogous sequences does not alter the RNA replication capacity. However, it is interesting to note that heterologous C and 5'NTR sequences result in a 50% decrease in replication efficiency, when compared to homologous C and 5'NTR sequences (HCA/C-p13GB versus HCA/E1-p13GB).

[0197] The replication capacity of HC/E1-NS3_{pro}^{GB}-Ubi and HC/C-NS3_{pro}^{GB}-Ubi chimeric RNAs were analyzed in cell culture after transferring chimeric sequences into the backbone of HCV^A. Northern blots revealed that the HC^A/E1-NS3_{pro}^{GB}-Ubi chimeric RNA replicated as efficiently as the parental HCV^A genome, whereas the HC^A/C-NS3_{pro}^{GB}-Ubi chimeric genome replicated at approximately 50-60 % of the parental genome (FIG. 6B). These results demonstrate that these two other chimeric genomes are also replication-competent, although the fusion of an heterologous core sequence to 5'NTR resulted in a decrease in replication efficiency.

[0198] Altogether, these data suggest that the co-substitution of the core coding sequence with those of envelope proteins is somewhat detrimental to RNA replication, whether the substitution involves NS2 sequence or not. This could be the result of a decreased translational capacity of these chimeric RNAs. In addition, the co-substitution of NS2 sequence and/or the insertion of the ubiquitin sequence also resulted in a two-fold decrease in genome replication capacity.

[0199] In the absence of any available model system in the laboratory to study GBV-B RNA replication, the replication capacity of the chimeras constructed in the GBV-B backbone could not be assessed.

[0200] Assembly of proteins derived from chimeric structural precursors expressing heterologous capsid and envelope proteins into virus-like particles. - In order to investigate whether heterologous capsid and envelope proteins of GBV-B and HCV could

assemble to form viral particles (question relevant to chimeras GB/E1-p7^{HC}, HC/E1-p13^{GB}, GB/E1-NS3_{pro}^{HC}-Ubi, HC/E1-NS3_{pro}^{GB}-Ubi), an expression system based on recombinant baculoviruses was used that allowed the production of virus-like particles, as initially described for HCV by Baumert *et al.* (1998). For that purpose, cDNA sequences encoding chimeric GBV-B/HCV structural precursors corresponding to either GBV-B core protein followed by HCV E1-E2-p7 or, conversely, HCV core protein followed by GBV-B E1-E2-p13 were cloned into transfer plasmid pVL1392 (Pharmingen). Correspondingly, parental GBV-B and HCV cDNA sequences encoding C-p13 or C-p7 precursors, respectively, were cloned in pVL1392 downstream of the baculovirus DNA polyhedrin promoter. After homologous recombination between the recombinant pVL1392 DNAs and baculovirus DNA in Sf9 cells, two parental GBV-B and HCV recombinant baculoviruses, Bac-C-p13^{GB} and Bac-C-p7^{HC}, as well as two chimeric GBV-B/HCV recombinant baculoviruses, Bac-C^{GB}/E1-p7^{HC} and Bac-C^{HC}/E1-p13^{GB} were thus obtained.

[0201] The inventors sought virus-like particle (VLP) formation in Sf9 cells infected with recombinant baculoviruses expressing either chimeric GBV-B/HCV structural precursors or parental, HCV or GBV-B precursors. Since nothing was known about the capability of a C-p13 precursor of GBV-B, such as the one expressed here via a recombinant baculovirus, to drive assembly of VLPs either in insect or in mammalian cells, the inventors first focused on a comparative analysis of C^{GB}/E1-p7^{HC} and HCV precursors.

[0202] Cytoplasmic extracts prepared at 3 days after infection of Sf9 cells with parental Bac-C-p7^{HC} were concentrated through a 30% sucrose cushion, then fractionated on a 20-60 % sucrose gradient. The polypeptide content of each fraction of the sucrose gradients was examined by immunoblotting with a mixture of monoclonal antibodies specific for HCV C, E1 or E2. Fractions 12-13 contained all three structural polypeptides, core and both glycoproteins, indicative of the likelihood to find assembled VLPs in these fractions. To further demonstrate the presence of VLPs in fractions 12-13, a pool of these fractions was concentrated and dialyzed for sucrose elimination prior to analysis by electron microscopy. After negative staining of such preparations, spherical structures approximately 55 nm in diameter were observed. Further immunogold labeling with either monoclonal antibody A4 directed to HCV E1, or monoclonal antibody H53 that recognizes HCV E2 in a conformation-dependent fashion, was performed. Most spherical structures were heavily and specifically marked by gold grains with both A4 and H53 as primary antibodies, but not with monoclonal antibodies specific for HCV core, strongly suggesting that they represent VLPs with both glycoproteins at their surface (FIGs. 7A-7B).

[0203] Similar analyses were performed with chimeric virus Bac-C^{GB}/E1-p7^{HC}-infected cell extracts. Immunoblot probing of each sucrose fraction with a mixture of anti-HCV E1 and E2 antibodies showed that the vast majority of E1 and E2 envelope proteins were present in the same fractions (#12-13) as those containing HCV VLPs from Bac-C-p7^{HC}-infected cells. Furthermore, immunoblotting with anti-GBV-B core polyclonal antibodies revealed that GBV-B core co-localized with HCV envelope proteins in the same fractions. Immune electron microscopy work demonstrated particles of similar shape and sizes (61 nm in diameter) by negative staining as VLPs derived from the HCV structural precursor. Most chimeric particles, with the exception of those with a smooth surface that looked like empty structures were covered with gold grains upon immunogold labeling with A4 and H53 antibodies (FIGs. 7C-7D).

[0204] The facts that the shape and sizes of chimeric GBV-B/HCV VLPs are close to those of HCV VLPs produced in an analogous system, that these chimeric VLPs exhibit both HCV envelope proteins at their surface, and that GBV-B core protein co-localized in the relevant fractions, are all supportive of the capability of GBV-B core protein to assemble with HCV envelope proteins and form VLPs.

[0205] Analysis of the infectivity of genome-length, chimeric RNAs in tamarins. - The promising results obtained in translation, processing, replication and assembly studies of the structural chimeras prompted the assay of the infectivity of full-length chimeric genomes in tamarins.

[0206] RNAs were transcribed *in vitro* from chimeric cDNAs that had previously been linearized at the unique *Xho*I or *Xba*I sites downstream of the 3' end of the viral cDNAs in GBV-B or HCV backbones, respectively. Chimeric GB/E1-p7^{HC} and GB/C-p7^{HC} RNAs were then inoculated separately into the liver of a single GBV-B naïve tamarin (*S. oedipus*), while a mixture of chimeric HC/E1-p13^{GB} and HC/C-p13^{GB} RNAs was inoculated to another single animal. A mixture of the four chimeric RNAs, GB/E1-NS3_{pro}^{HC}-Ubi, GB/C-NS3_{pro}^{HC}-Ubi, HC/E1-NS3_{pro}^{GB}-Ubi, HC/C-NS3_{pro}^{GB}-Ubi was inoculated to two animals (*S. mystax*).

[0207] Viral replication was monitored by periodic testing of tamarin sera for genomic RNA using appropriate GBV-B (primers and probe in the NS5A coding region) or HCV (primers and probe in the 5'NTR) real-time, quantitative RT-PCR assays. Onset of hepatitis was followed by testing for serum transaminase (ALT) levels.

[0208] In contrast to viral replication that developed over weeks 1 to 10-20 after inoculation of GBV-B wild-type RNA (Martin *et al.*, 2003), no sign of viral replication was detected in animals inoculated with either GB/E1-p7^{HC}, GB/C-p7^{HC}, GB/E1-NS3_{pro}^{HC}-Ubi + GB/C-NS3_{pro}^{HC}-Ubi, nor with HC/E1-p13^{GB} or HC/C-p13^{GB}.

[0209] In one animal out of the two animals inoculated with the 4 genomes containing substitutions extending into the NS2 sequence, a low-level replication (100-1000 genome equivalents/ml) of either chimera HC/E1-NS3proGB-Ubi or HC/C-NS3proGB-Ubi was reproducibly detected at weeks two and four post-inoculation, but no robust replication was further detected. This is reminiscent of what has been observed with the GBV-B IRES chimera, which inoculation resulted in such a low-level replication at early time-points, followed by undetectable replication and a sudden raise in virus replication starting at week 12 post-inoculation. This reflected the requirement for adaptive mutations that arose in one animal out of two inoculated with this chimeric RNA.

[0210] In order to check upon such a requirement for adaptive changes in the genome of the structural chimeras, other animals will be inoculated with chimeras HC/E1-NS3proGB-Ubi and HC/C-NS3proGB-Ubi, individually or mixed, but in the absence of other chimeras in the backbone of GBV-B.

EXAMPLE 5: Further Characterization of Rescued Chimeric Viruses.

[0211] Where infection with chimeric viruses is induced in animals that are injected within the liver with synthetic RNA, this virus will be passaged in GBV-B naïve tamarins to further characterize the nature of the infection induced by the chimera. This will be accomplished by taking a pool of the 3 highest titer GBV-B RNA-containing serum specimens from the animal that was successfully transfected with RNA, and inoculating 1 mL of a 1:100 dilution of this pool intravenously into two susceptible animals. These animals will be monitored for infection and liver disease. These animals will be followed until resolution of the viremia and appearance of antibodies detectable in immunoblots with GST-NS3 protein expressed in *E. coli*, or for at least 6 months should an animal sustain a chronic infection. RT-PCR amplification of chimeric segments of the genome may be employed to determine whether the altered phenotype results from mutations within the heterologous portion of the genome.

EXAMPLE 6: Use of GBV-B as Model for HCV

[0212] GBV-B and/or GBV-B/HCV chimeras can be used as a model for HCV. Such studies will allow one to investigate the mechanisms for the different biological properties of these viruses and to discover and investigate potential inhibitors of specific HCV activities (e.g., proteinase) required for HCV replication. GBV-B/HCV viruses may be used in preclinical testing of candidate HCV NS3 proteinase inhibitors or other inhibitors of HCV.

[0213] **Candidate Substances** - As used herein the term "candidate substance" refers to any molecule that is capable of modulating HCV NS3 proteinase activity or any other activity related to HCV infection. The candidate substance may be a protein or fragment thereof, a small molecule inhibitor, or even a nucleic acid molecule. It may prove to be the case that the most useful pharmacological compounds for identification through application of the screening assay will be compounds that are structurally related to other known modulators of HCV NS3 proteinase activity. The active compounds may include fragments or parts of naturally-occurring compounds or may be only found as active combinations of known compounds that are otherwise inactive. However, prior to testing of such compounds in humans or animal models, it will be necessary to test a variety of candidates to determine which ones have potential.

[0214] Accordingly, the active compounds may include fragments or parts of naturally-occurring compounds or may be found as active combinations of known compounds that are otherwise inactive. As such, the present invention provides screening assays to identify agents that are capable of inhibiting proteinase activity in a cell infected with chimeric GBV-B/HCV viruses containing the HCV proteinase. It is proposed that compounds isolated from natural sources, such as animals, bacteria, fungi, plant sources, including leaves and bark, and marine samples may be assayed as candidates for the presence of potentially useful pharmaceutical agents. It will be understood that the pharmaceutical agents to be screened could also be derived or synthesized from chemical compositions or man-made compounds. Thus, it is understood that the candidate substance identified by the present invention may be polypeptide, polynucleotide, small molecule inhibitors or any other compounds that may be designed through rational drug design starting from known inhibitors of proteinases or from structural studies of the HCV proteinase.

[0215] The candidate screening assays are simple to set up and perform. Thus, in assaying for a candidate substance, after obtaining a chimeric GBV-B/HCV virus with infectious

properties, a candidate substance can be incubated with cells infected with the virus, under conditions that would allow measurable changes in infection by the virus to occur. In this fashion, one can measure the ability of the candidate substance to prevent or inhibit viral replication, in relationship to the replication ability of the virus in the absence of the candidate substance. In this fashion, the ability of the candidate inhibitory substance to reduce, abolish, or otherwise diminish viral infection may be determined.

[0216] "Effective amounts" in certain circumstances are those amounts effective to reproducibly reduce infection by the virus in comparison to the normal infection level. Compounds that achieve significant appropriate changes in activity will be used. Candidate compounds can be administered by any of a wide variety of routes, such as intravenously, intraperitoneally, intramuscularly, orally, or any other route typically employed.

[0217] It will, of course, be understood that all the screening methods of the present invention are useful in themselves notwithstanding the fact that effective candidates may not be found. The invention provides methods of screening for such candidates, not solely methods of finding them.

[0218] ***In vitro* Assays** - In one particular embodiment, the invention encompasses *in vitro* screening of candidate substances. Using a cell line that can propagate GBV-B in culture, *in vitro* screening can be used such that GBV-B or HCV virus production or some indicator of viremia is monitored in the presence of candidate compounds. A comparison between the absence and presence of the candidate can identify compounds with possible preventative and therapeutic value.

[0219] ***In Vivo* Assays** - The present invention also encompasses the use of various animal models to test for the ability of candidate substances to inhibit infection by HCV. This form of testing may be done in tamarins.

[0220] The assays previously described could be extended to whole animal studies in which the chimeric virus could be used to infect a GBV-B permissive primate, such as a tamarin. One would then look for suppression of viral replication in the animal, and a possible impact on liver disease related to replication of the infectious chimeric virus. The advantage of this *in vivo* assay over present available assays utilizing HCV infection in chimpanzees is the reduced cost and greater availability of GBV-B permissive nonhuman primate species.

[0221] Treatment of animals with test compounds will involve the administration of the compound, in an appropriate form, to the animal. Administration will be by any route that could be utilized for clinical or non-clinical purposes, including but not limited to oral, nasal, buccal, rectal, vaginal or topical. Alternatively, administration may be by intratracheal instillation, bronchial instillation, intradermal, subcutaneous, intramuscular, intraperitoneal or intravenous injection. Specifically contemplated are systemic intravenous injection, intraperitoneal injection, and oral administration.

[0222] Determining the effectiveness of a compound *in vivo* may involve a variety of different criteria. Such criteria include, but are not limited to, survival, reduction of rate of infection, arrest or slowing of infection, elimination of infection, increased activity level, improvement in liver function, and improved food intake.

EXAMPLE 7: Use of Infectious GBV-B/HCV Chimeras as Vaccines

[0223] Infectious GBV-B/HCV chimeras expressing HCV envelope proteins will have utility as a vaccine immunogen for hepatitis C. Such clones clearly have the potential to be constructed as chimeras including relevant hepatitis C virus sequences in lieu of the homologous GBV-B sequence, providing unique tools for drug discovery efforts.

[0224] Chimeric viruses containing the envelope proteins of hepatitis C virus (as described in the attached) would confer the antigenic characteristics of hepatitis C virus on the chimera. These chimeras may have the ability to replicate in chimpanzees (and thus humans) by virtue of the fact that the chimeric envelope is now able to interact with the human hepatocyte cell surface, a necessary first step in virus replication. Therefore, the chimeric virus, while able to infect and replicate in humans, may not cause much or any disease-the reasoning here is that the genetic backbone of the chimera that encodes the nonstructural proteins of GBV-B has not evolved for replication in human cells and thus may not replicate well. Thus, the chimera may have limited replication ability, cause no disease, but still elicit immunity to the surface envelope proteins of HCV and thus have potential as a hepatitis C vaccine. These chimeras can be tested for their ability to promote immunity to HCV through an immune response.

EXAMPLE 8:

[0225] With regard to chimeric genomes derived from the HCV and comprising the sequences coding E1-E2-p13-NS2-NS3_{pro} or C-E1-E2-p13-NS2-NS3_{pro} of GBV-B instead of E1-

E2-p7-NS2 sequences or C-E1-E2-p7-NS2 of the HCV, which had given encouraging results, the nucleotide sequence of both chimeric cDNAs HC/E1-NS3_{pro}GB-Ubi and HC/C-NS3_{pro}GB-Ubi was confirmed. Indeed, the plasmids containing these cDNA are very low copy number and it is difficult to propagate them in great quantity in the bacteria. In order to avoid any risk of appearance of specific changes during the amplification of the plasmids in the bacterium, the inventors sequenced the plasmid preparation used for *in vitro* transcription studies. Instead of working with two independent clones for each construction (the initial strategy to minimize the risks of undesirable change in a molecular clone), one validated clone was used and thus limit the quantities of RNA inoculated to the animal. Once the sequences of the two cDNA were confirmed relative to the approximately 10250 bases, RNA was prepared by *in vitro* transcription using the RNA polymerase of the phage T7, whose promoter is cloned upstream of the target cDNA.

[0226] The resulting RNA, whose quality and quantity were controlled, were sent to the laboratory of Dr. Lanford (Southwest Foundation for Biomedical Research, San Antonio, Texas) for inoculation in the liver of two tamarins (T16472, T16467). These two animals had been inoculated (4 and 2 years before, respectively) with other chimeric genomes derived from GBV-B, with no sign of viremia being detected during the 22 week observation period. The inventors chose to use two animals in order to optimize the chances to select changes "adaptation" conferring an effective replication at this experimental host. Such changes are probably necessary to obtaining a chimeric virus with robust replicative capacities, with respect to the results of the first inoculations of these chimeric genomes. There was no trace of GBV-B sequences in the serum of the animals taken one week before the inoculation. A mixture of approximately 100 µg of each RNA HC/E1-NS3_{pro}GB-Ubi and HC/C-NS3_{pro}GB-Ubi was then inoculated intrahepatically. The quality of the inoculated RNA was checked the day of the inoculation by agarose gel. Samples of serum were prepared to 0, 1, 2 days, then every two weeks after the inoculation and were tested by specific quantitative RT-PCR for the genome of the HCV (primers and probes typically hybridized in the 5' non-coding region of the genome) to seek the presence of viral particles. Assays to detect the rise in serum transaminases allows characterization of possible hepatitis on the biochemical level. The serums from 0 to 16 weeks post-inoculation were tested. The results obtained to date seem to indicate that a viremy (2×10^4 genome equivalent/ml to the maximum) could be detected at 8 to 12 weeks post-inoculation in the two animals, but no signal was then detected in 14 and 16 weeks post-inoculation.

[0227] The inventors will pass these inoculums to naive tamarins and examine whether one can thus force the adaptation of such a chimeric virus. With regard to the chimeric genomes derived from GBV-B and comprising the sequences coding E1-E2-p7-NS2-NS3_{pro} or C-E1-E2-p7-NS2-NS3_{pro} of the HCV instead of E1-E2-p13-NS2 sequences or C-E1-E2-p13-NS2 of GBV-B, cDNA GB/E1-NS3_{pro}^{HC}-Ubi and GB/C-NS3_{pro}^{HC}-Ubi nucleotide substitution allowing the transition from the residue Val(2236) of NS5A (coded by the clone) to the Ala residue, whose codon is systematically found in the genome of all the viruses resulting from our molecular clone, and who thus confers a replicative advantage on the virus. The nucleotide sequence of both new cDNA construct also was sequenced in its entirety, in order to exclude any potential undesirable change which could have appeared during amplifications of plasmids in the bacterium. Repair of the plasmid GB/C-NS3_{pro}^{HC}-Ubi was carried out by exchange of a restriction fragment containing the parental residue, then the sequence of cDNA of the new clone was confirmed. A mixture of the two chimeric RNA derived from these modified GBV-B clones were inoculated intrahepatically in two animals to test their infectivity. The inventors envision the use of two other animals for the evaluation of other chimeric genomes targeting non-structural proteins and the effects on replication of the viral RNA. They are genomes derived from GBV-B and having the sequences from the protein p7 HCV instead of whole or part from the protein p13 from GBV-B. The protein p7 of the HCV has an ion channel function (Pavlovic et al, 2003, Proc Natl Acad Sci USA 100:6104; Premkumar et al., 2004, FEBS Lett 557:99), potentially implied in the stages of synthesis and/or export of the virions. The existence of a similar protein for GBV-B (p13) has been shown, but of double size, composed of two parts equipped with potentially distinct functions (Ghibaud et al., 2004, J Biol Chem 279:24965). A cDNA for the sequences coding p13 or coding the final half C of p13 (homologous with p7) were substituted by the sequence of p7 of the HCV. The chimeric viruses having such genomes would constitute invaluable tools, as well for the study of the role of these candidates in the infectious cycle of the hepacivirus, as for research the new antiviral ones targeting p7 HCV and being able to interfere with the formation of the infectious particles. Such research is difficult with HCV, whose only experimental host is the chimpanzee.

[0228] All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred

embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents, which are both chemically and physiologically related, may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

[0229] The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

U.S. Patent 4,578,770

U.S. Patent 4,596,792

U.S. Patent 4,599,230

U.S. Patent 4,599,231

U.S. Patent 4,601,903

U.S. Patent 4,608,251

U.S. Patent 4,683,195

U.S. Patent 4,683,202

U.S. Patent 4,800,159

U.S. Patent 4,883,750

U.S. Patent 5,279,721

U.S. Patent 5,807,670

EPA No. 320,308

EPA No. 329,822

GB Application No. 2,202,328

PCT/US87/00880

PCT/US89/01025

WO 88/10315

WO 89/06700

WO 90/07641

Agapov *et al.*, *J. Virol.*, 78:2414-2425, 2004.

Al *et al.*, *Virus Res.*, 53:141-149, 1998.

Alter *et al.*, *Hepatology*, 26:625-655, 1997.

Baumert *et al.* *J. Virol.*, 72:3827-3836, 1998.

Beames *et al.*, *J. Virol.*, 74:11764-11772, 2000.

Beard *et al.*, *Hepatology*, 1999.

Behrens *et al.*, *EMBO. J.*, 15:12-22, 1996.

Carrere-Kremer *et al.*, *J. Virol.*, 76:3720-3730, 2002.

- Chambers *et al.*, *J. Virol.*, 73:3095-3101, 1999.
- Chambers *et al.*, *J. Virol.*, 73:3095-3101, 1999.
- Choo *et al.*, *Proc. Natl. Acad. Sci. USA*, 91:1294-1298, 1994.
- Consensus Development Panel. National Institutes of Health Consensus Development Conference Panel statement: Management of hepatitis C. *Hepatology*, 26:Supplement 1[3], 2S-10S, 1997.
- Dubuisson, *Curr. Top. Microbiol. Immunol.*, 242:135-148, 2000.
- Farci *et al.*, *J. Infectious Diseases*, 165:1006-1011, 1992.
- Farci *et al.*, *Science*, 258:135-140, 1992.
- Filocamo *et al.*, *J. Virol.*, 71:1417-27, 1997.
- Frohman, *PCR Protocols: A Guide To Methods And Applications*, Academic Press, New York, 1990.
- Frolov *et al.*, *RNA*, 4:1418-1435, 1998.
- Ghibaud *et al.*, *J. Biol. Chem.*, 279:24965-24975, 2004.
- Ghibaud *et al.*, *J. Biol. Chem.*, 279:24965-24975, 2004.
- Grakoui *et al.*, *J. Virol.*, 67(3):1385-95, 1993.
- Grakoui *et al.*, *J. Virol.*, 67(5):2832-2843, 1993.
- Grakoui *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:10583-10587, 1993.
- Griffin *et al.*, *FEBS Lett.*, 535:34-38, 2003.
- Hijikata *et al.*, *J. Virol.*, 67:4665-4675, 1993.
- Honda *et al.*, *RNA*, 2:955-956, 1996.
- Inchauspe *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:10292-10296, 1991.
- Karayiannis *et al.*, *Hepatology*, 9:186-192, 1989.
- Kim *et al.*, *Cell*, 87:343-355, 1996.
- Kolykhalov *et al.*, *J. Virol.*, 70:3363-3371, 1996.
- Kolykhalov *et al.*, *Science*, 277:570-574, 1997.
- Kummerer and Rice, *J. Virol.*, 76:4773-4784, 2002.
- Landt *et al.*, *Gene*, 96:125-128, 1990.
- Lemon and Honda, *Seminars in Virology*, 8:274-288, 1997.
- Lindenbach and Rice, *J. Virol.*, 73:4611-4621, 1999.
- Lo *et al.*, *J. Virol.*, 70:5177-5182, 1996.
- Lohmann *et al.*, *Science*, 285:110-113, 1999.
- Lohmann *et al.*, *Science*, 285:110-3, 1999.

- Lu and Wimmer, *Proc. Natl. Acad. Sci. USA*, 93:1412-1417, 1996.
- Martin *et al.*, *Proc. Natl. Acad. Sci. USA*, 100:9962-9967, 2003.
- McLauchlan *et al.*, *EMBO J.*, 21:3980-3988, 2002.
- McLauchlan *et al.*, *EMBO J.*, 21:3980-3988, 2002.
- Muerhoff *et al.*, *J. Virol.*, 69:5621-5630, 1995.
- Neumann *et al.*, *Science*, 282:103-107, 1998.
- Op De Beeck *et al.*, *J. Gen. Virol.*, 82:2589-2595, 2001.
- Pieroni *et al.*, *J. Virol.*, 71:6373-6380, 1997.
- Pletnev and Men, *Proc. Natl. Acad. Sci. USA*, 95:1746-1751, 1998.
- Pletnev *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:10532-10536, 1992.
- Remington's Pharmaceutical Sciences, 15th Edition, pages 1035-1038 and 1570-1580.
- Reynolds *et al.*, *EMBO J.*, 14:6010-6020, 1995.
- Rijnbrand *et al.*, Mutational analysis of the GB virus B internal ribosome entry site. Submitted for publication, 1999.
- Rijnbrand *et al.*, *RNA*, 7:585-597, 2001.
- Sakai *et al.*, *Proc. Natl. Acad. Sci. USA*, 100:11646-11651, 2003.
- Sambrook, *et al.*, , *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring, Harbor, N.Y., 1989.
- Scarselli *et al.*, *J. Virol.*, 71 4985-4989, 1997.
- Schlauder *et al.*, *J. Med. Virol.*, 46:81-90, 1995.
- Shaffer *et al.*, *J. Virol.*, 69:600-6604, 1995.
- Simons, *et al.*, *Proc. Natl. Acad. Sci.*, 92:3401-3405, 1995.
- Tanaka *et al.*, *Biochem. Biophys. Res. Comm.*, 215:744-749, 1995.
- Teller *et al.*, *J. Clin. Micro.*, 34:3085-3091, 1996.
- Thibeault *et al.*, *J. Biol. Chem.*, 276:46678-46684, 2001.
- Todd *et al.*, *J. Virol.*, 71:8868-74, 1997.
- Yamaga and Ou, *Biol. Chem.*, 277:33228-33234, 2002.
- Yanagi *et al.*, *Proc. Natl. Acad. Sci. USA*, 16:8738-8743, 1998.
- Yanagi *et al.*, *Proc. Natl. Acad. Sci. USA*, 94:8738-8743, 1997.
- Yao *et al.*, *Nature Structural Biology*, 4: 463-467, 1997.
- Yasui *et al.*, *J. Virol.*, 72:6048-6055, 1998.
- Yi and Lemon, *J. Virol.*, 78:7904-7915, 2004.
- Zhao *et al.*, *J. Virol.*, 73:1546-1554, 1999.

CLAIMS

What is claimed is:

1. A chimeric hepatotropic recombinant virus comprising part of a GBV-B polynucleotide and part of the polynucleotide sequence derived from HCV.
2. The hepatotropic virus of claim 1, wherein the nucleic acid segment of HCV encodes a core protein.
3. The hepatotropic virus of claim 1, wherein the nucleic acid segment of HCV encodes an E1 protein.
4. The hepatotropic virus of claim 3, wherein the nucleic acid segment of HCV encodes an E2 protein.
5. The hepatotropic virus of claim 1, wherein the nucleic acid segment of HCV encodes a p7 protein.
6. The hepatotropic virus of claim 1, wherein the nucleic acid segment of HCV encodes an E1 and E2 protein.
7. The hepatotropic virus of claim 1, wherein the nucleic acid segment of HCV encodes a core, E1, and E2 proteins.
8. The hepatotropic virus of claim 1, wherein the nucleic acid segment of HCV encodes a core, E1, E2, and p7 proteins.
9. The hepatotropic virus of claim 1 further comprising a NS2 protein.
10. The hepatotropic virus of claim 1, further comprising a NS3 protein having a heterologous protease cleavage site.
11. The hepatropic virus according to claim 1 further comprising a pair of NS3 having a heterologous protease cleavage site.
12. The hepatotropic virus of claim 1, wherein the heterologous cleavage site is an ubiquitin protease cleavage site.

13. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 1% of an HCV genome.
14. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 5% of an HCV genome.
15. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 10% of an HCV genome.
16. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 20% of an HCV genome.
17. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 30% of an HCV genome.
18. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 40% of an HCV genome.
19. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 50% of an HCV genome.
20. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 60% of an HCV genome.
21. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 70% of an HCV genome.
22. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 80% of an HCV genome.
23. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 90% of an HCV genome.
24. The hepatotropic virus of claim 1, wherein the polynucleotide comprises at least 95% of an HCV genome.

25. The hepatotropic virus of claim 1, wherein the polynucleotide has a sequence set forth in SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, or SEQ ID NO:26.
26. The hepatotropic virus of claim 1, wherein the virus propagates *in vivo*.
27. A method of producing a chimeric virus comprising:
- a) introducing into a host cell an expression construct comprising a polynucleotide encoding a chimeric GBV-B/HCV virus; and
 - b) culturing said host cell under conditions permitting production of a chimeric virus from the construct.
28. The method of claim 27, wherein the host cell is a prokaryotic cell.
29. The method of claim 27, wherein the host cell is a eukaryotic cell.
30. The method of claim 29, wherein the host cell is in an animal.
31. The method of claim 30, wherein the host cell is in a tamarin.
32. The method of claim 27, wherein the polynucleotide comprises synthetic RNA.
33. The method of claim 27, wherein the polynucleotide comprises synthetic DNA.
34. The method of claim 27, further comprising the step of isolating virus from the host cell.
35. The method of claim 34, wherein said virus is purified to homogeneity.
36. A method for identifying a compound active against a viral infection comprising:
- a) providing an expression construct comprising a polynucleotide that when expressed produces a chimeric GBV-B/HCV virus;

- b) contacting the virus with a candidate substance; and
 - c) comparing the infectious ability of the virus in the presence of the candidate substance with the infectious ability of the virus in a similar system in the absence of the candidate substance.
37. A method of producing a virus comprising:
- a) introducing into a host cell an expression construct comprising a chimeric GBV-B polynucleotide encoding at least part of an HCV sequence; and
 - b) culturing said host cell under conditions permitting production of a virus from the construct.
38. The method of claim 37, wherein said host cell is a eukaryotic cell.
39. The method of claim 38, wherein said host cell is in an animal.
40. The method of claim 37, wherein said polynucleotide comprises synthetic RNA.
41. The method of claim 37, further comprising the step of isolating virus from said host cell.
42. The method of claim 41, wherein said virus is purified to homogeneity.
43. A compound active against a viral infection identified according to a method comprising:
- a) providing a virus expressed from an construct comprising GBV-B/HCV chimera;
 - b) contacting the virus with a candidate substance; and
 - c) comparing the infectious ability of the virus in the presence of the candidate substance with the infectious ability of the virus in a similar system in the absence of the candidate substance.

44. The polynucleotide of claim 1 wherein the polynucleotide as a set forth in SEQ ID 23, SEQ ID 24, SEQ ID 25, SEQ ID 19, SEQ ID 20, SEQ ID 21, SEQ ID 22, SEQ ID 24, or SEQ ID 26.

45. The polynucleotide of claim 1 wherein the HCV or GVB-B nucleotide sequences comprise at least one of the sequences of SEQ ID 23, SEQ ID 24, SEQ ID 25, SEQ ID 19, SEQ ID 20, SEQ ID 21, SEQ ID 22, SEQ ID 24, or SEQ ID 26 or a fragment thereof which said fragment leads to a recombinant chimeric virus of claim 1.

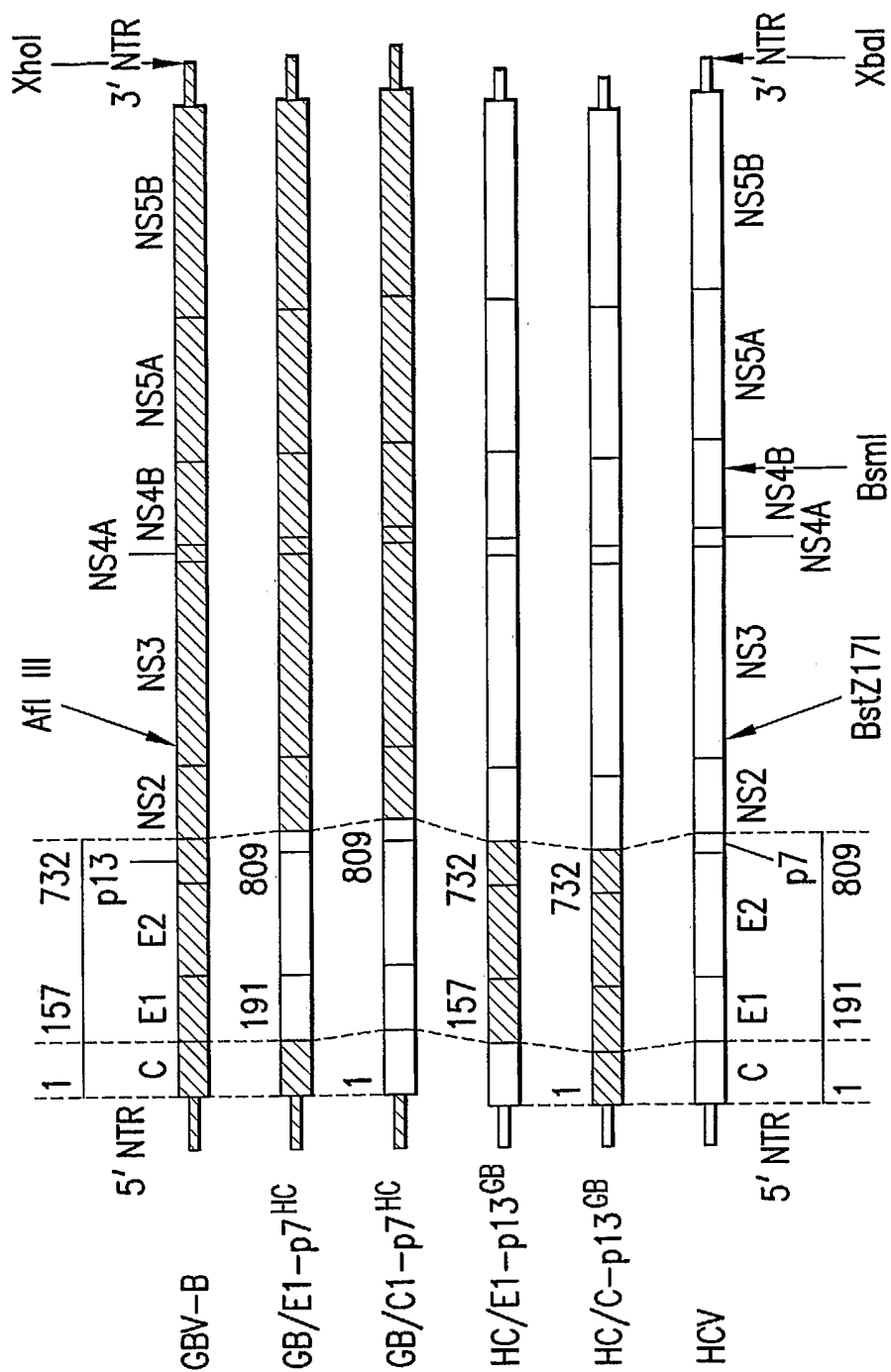


FIG. 1

2/9

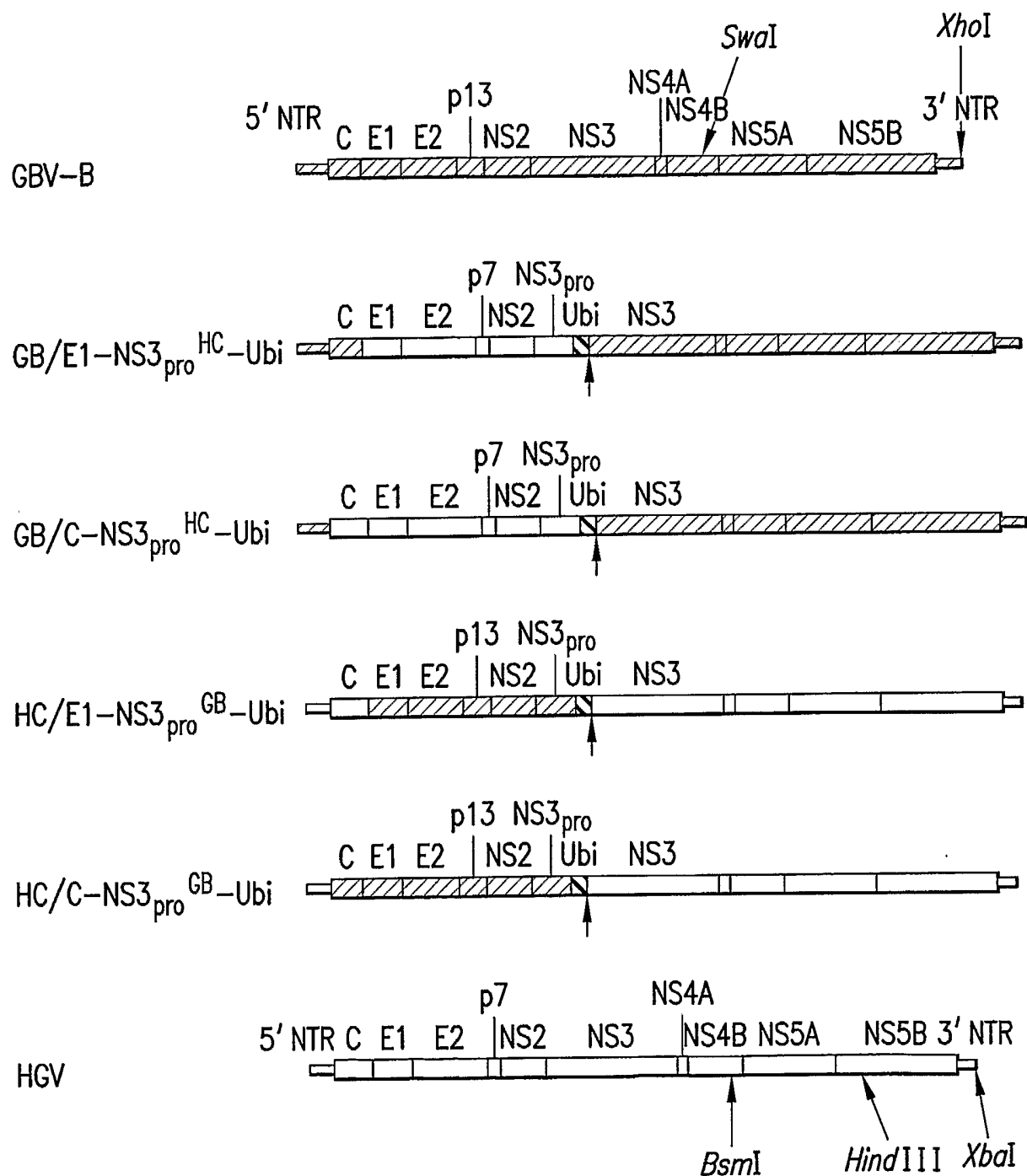


FIG. 2

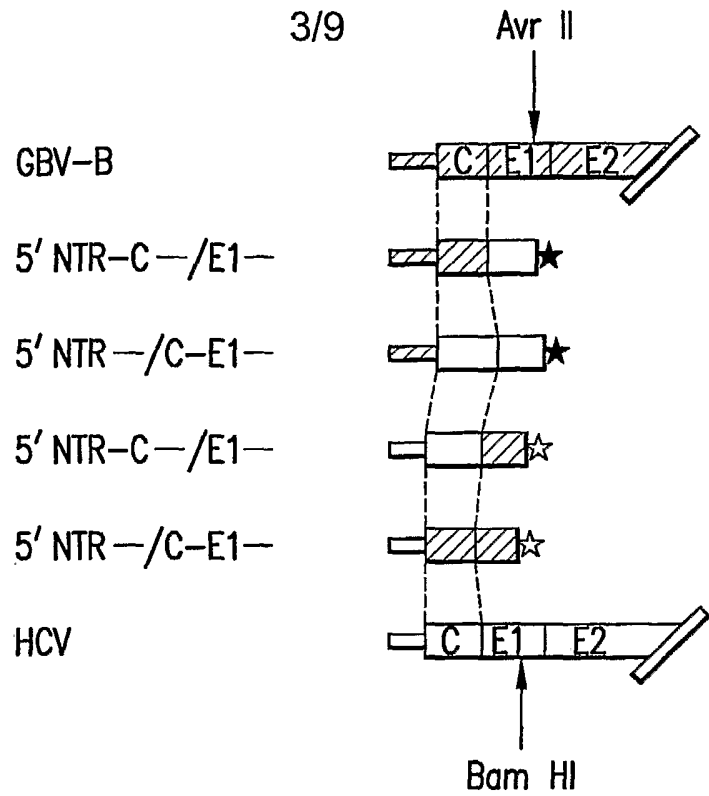


FIG.3A

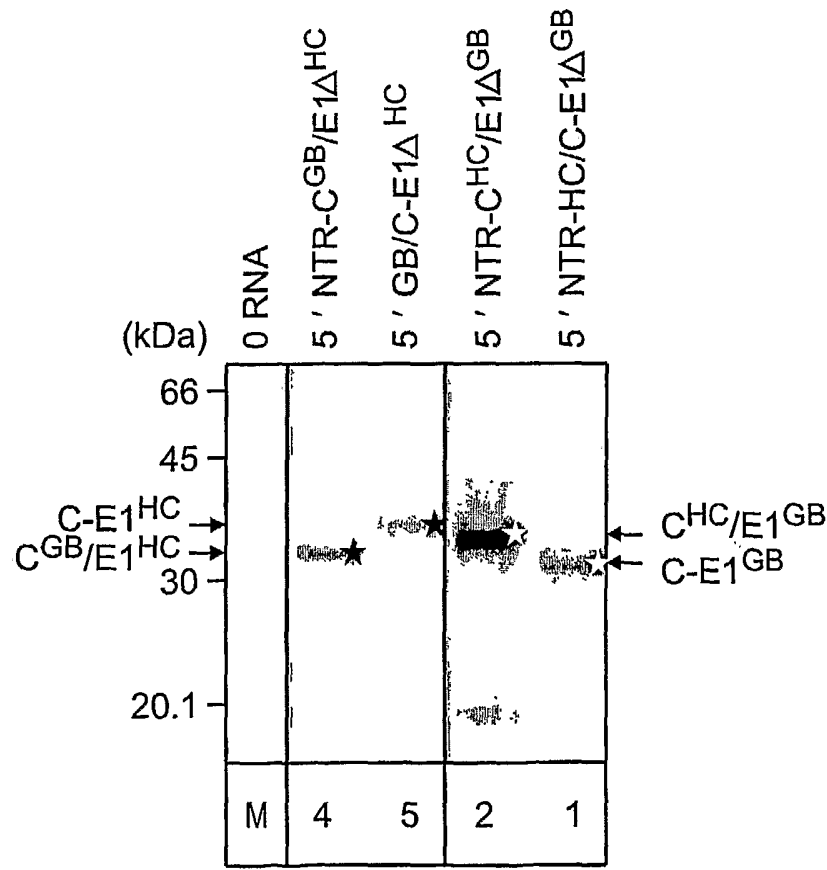


FIG.3B

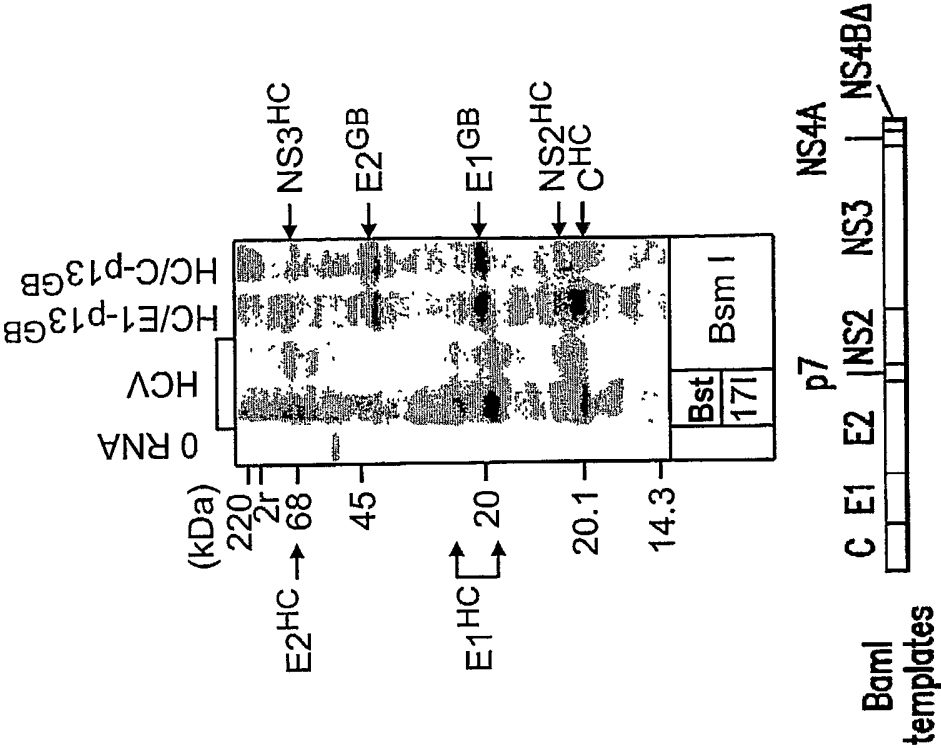


FIG. 4B

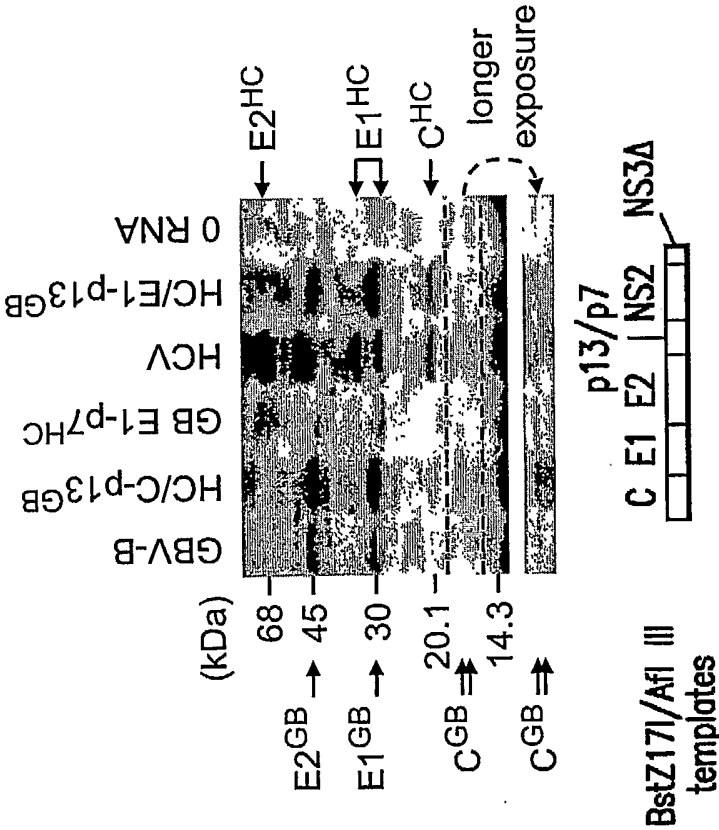


FIG. 4A

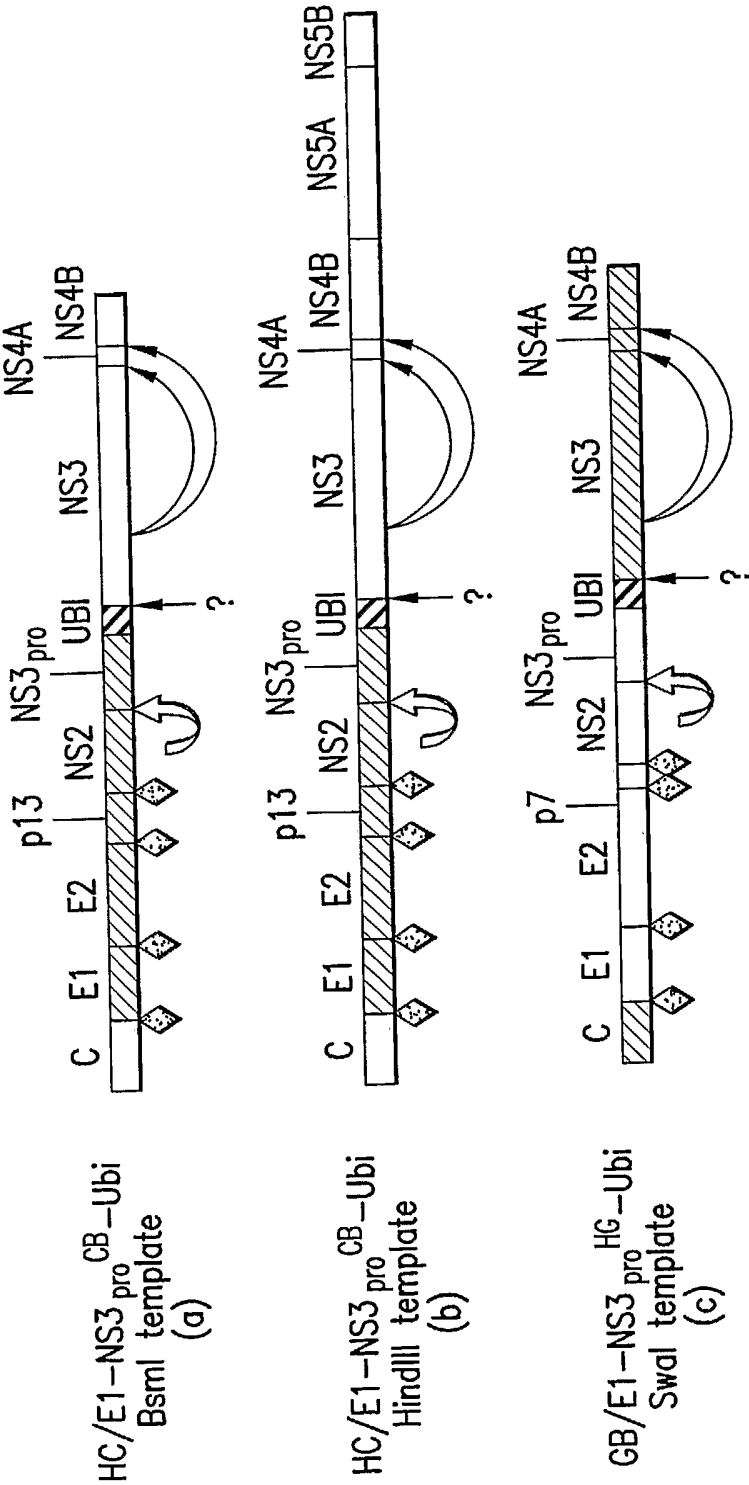


FIG.5A

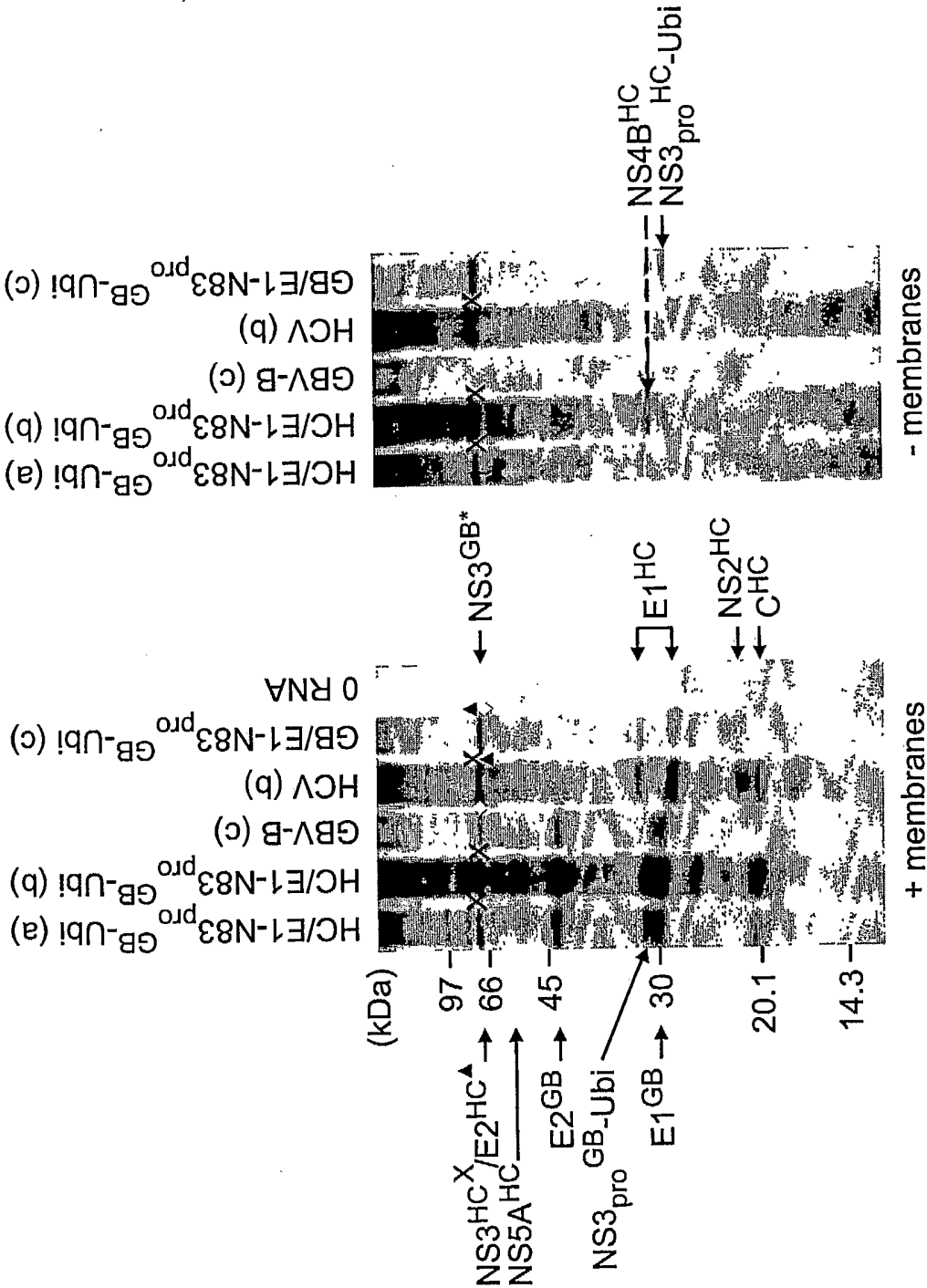


FIG.5B

7/9

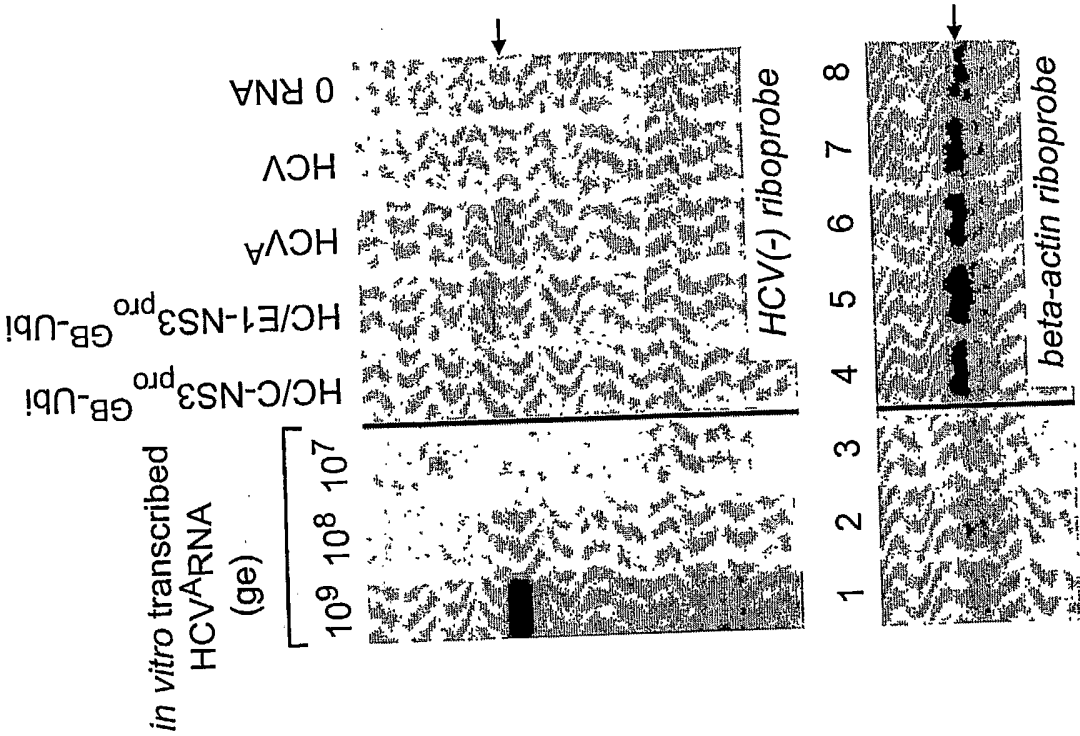


FIG. 6B

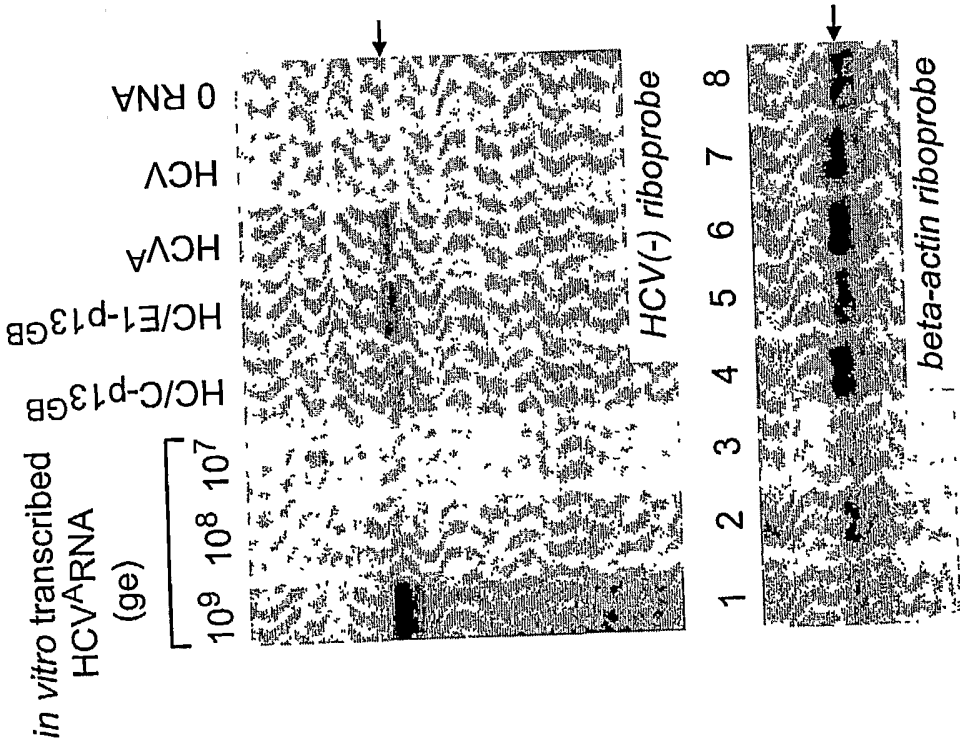


FIG. 6A

8/9

Parental Bac-C-p7^{HC}

anti-E2 (H53)

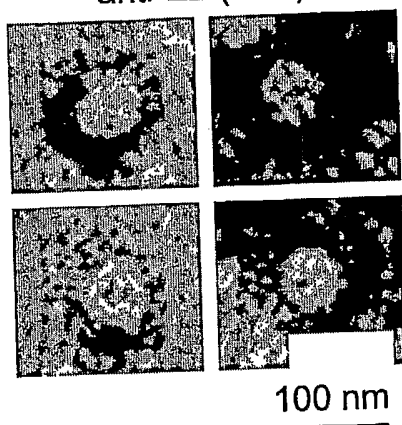


FIG.7A

Parental Bac-C-p7^{HC}

anti-E1 (A4)

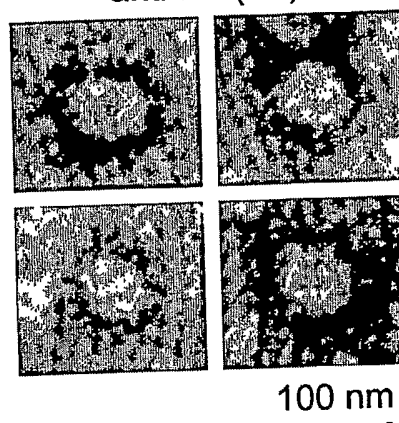


FIG.7B

Chimeric Bac-C^{GB}/E1-p7^{HC}

anti-E2 (H53)

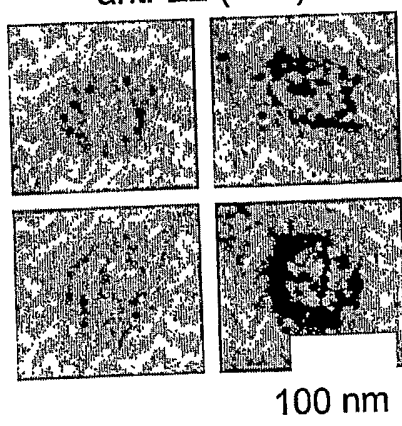


FIG.7C

Chimeric Bac-C^{GB}/E1-p7^{HC}

anti-E1 (A4)

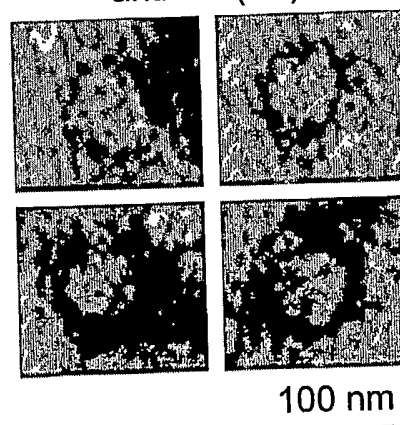


FIG.7D

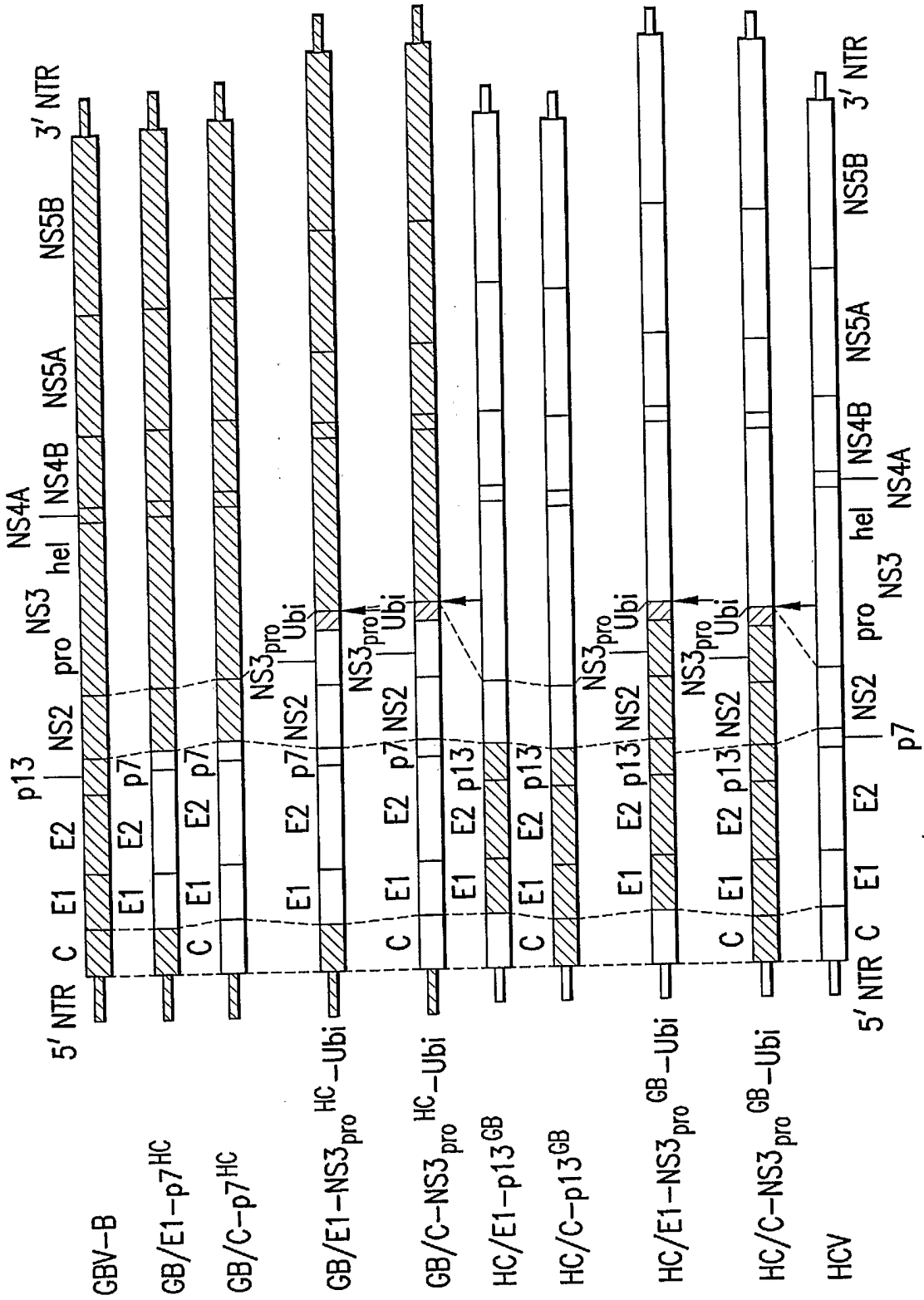


FIG.8

SEQUENCE LISTING

<110> MARTIN, Annette
GHIBAUDO, David
COHEN, Lisette
LEMON, STANLEY M.

<120> Chimeric GB Virus B (GBV-B)

<130> UTFG:270WO

<140> UNKNOWN

<141> 2005-09-27

<150> 60/613,266

<151> 2004-09-27

<160> 27

<170> PatentIn Ver. 2.1

<210> 1

<211> 260

<212> RNA

<213> Unknown

<220>

<223> Description of Unknown Organism: GB VIRUS B

<400> 1

```
gaguuugcga ccauggugga ucagaaccgu uucgggugaa gccauggucu gaaggggaug 60
acgucccuuc uggcucaucc acaaaaaccg ucucgggugg gugaggaguc cuggcugugu 120
gggaagcagu caguauaaau cccgucgugu guggugacgc cucacgacgu auuuguccgc 180
ugugcagagc guaguaccaa gggcugcacc ccgguuuuug uuccaagcgg agggcaaccc 240
ccgcuuggaa uaaaaaacug                                     260
```

<210> 2

<211> 9399

<212> DNA

<213> GBV-A-like virus

<400> 2

```
accacaaaca ctccagtttg ttacactccg ctaggaatgc tcctggagca cccccctag 60
cagggcgtag gggatttccc ctgcccgtct gcagaagggt ggagccaacc accttagtat 120
gtaggcgtag ggactcatga cgctcgcttg atgacaagcg ccaagcttga ctggatggc 180
cctgatgggc gttcatgggt tcggtggttg tggcgcttta ggcagcctcc acgcccacca 240
cctcccagat agagcggcgg cactgtaggg aagaccgggg accggtcact accaaggacg 300
cagacctctt tttgagtatc acgcctccgg aagtagtttg gcaagcccac ctatatgtgt 360
tgggatgggt ggggttagcc atccataccg tactgcctga tagggtcctt gcgaggggat 420
ctgggagtct cgtagaccgt agcacatgcc tgttatttct actcaaaca gtcctgtacc 480
tgcgcccaga acgcgcaaga acaagcagac gcaggcttca tatcctgtgt ccattaaaac 540
atctgttgaa aggggacaac gagcaaagcg caaagtccag cgcgatgctc ggcctcgtaa 600
ttacaaaatt gctggtatcc atgatggctt gcagacattg gctcaggctg ctttgccagc 660
tcattggttg ggacgccaag accctcgcca taagtctcgc aatccttgaa tccttctgga 720
ttaccctttg ggggtgattg gtgatgttac aactcacaca cctctagtag gcccgctggg 780
ggcaggagcg gtcgttcgac cagtctgcca gatagtacgc ttgctggagg atggagtcaa 840
ctgggctact ggttggttcg gtgtccacct ttttgtggta tgtctgctat ctttggcctg 900
tcctgtagtg ggggcgcggg tcaactgacc agacacaaat accacaatcc tgaccaattg 960
ctgccagcgt aatcaggtta tctattgttc tccttccact tgcctacacg agcctggttg 1020
tgtgatctgt gcggacgagt gctgggttcc cgccaatccg tacatctcac acccttccaa 1080
ttggactggc acggactcct tcttggctga ccacattgat tttgttatgg gcgctcttgt 1140
gacctgtgac gcccttgaca ttggtgagtt gtgtggtgctg tgtgtattag tcggtgactg 1200
```

gcttgtcagg	caactggctta	ttcacataga	cctcaatgaa	actggtactt	gttacctgga	1260
agtgccact	ggaatagatc	ctgggttcc	aggggtttatc	gggtggatgg	ccggcaaggt	1320
cgaggctgtc	atcttcttga	ccaaactggc	ttcacaagta	ccatacgcta	ttgcgactat	1380
gttttagcagt	gtacactacc	tggcggttgg	cgctctgac	tactatgcct	ctcggggcaa	1440
gtggatcag	ttgctcctag	cgcttatgct	ttacatagaa	gcgacctctg	gaaaccctat	1500
cagggtgccc	actggatgct	caatagctga	gttttgctcg	cctttgatga	taccatgtcc	1560
ttgccactct	tatttgagtg	agaatgtgtc	agaagtcat	tgttacagtc	caaagtggac	1620
caggcctatc	actctagagt	ataacaactc	catatcttgg	taccctata	caatccctgg	1680
tgcgagggga	tgtatggtta	aattcaaaaa	taacacatgg	ggttgctgcc	gtattcgcaa	1740
tgtgccatcg	tactgacta	tgggcactga	tgcagtgtgg	aacgacactc	gcaacactta	1800
cgaagtatgc	ggtgtaacac	catggctaac	aaccgcatgg	cacaacggct	cagccctgaa	1860
attggctata	ttacaatacc	ctgggtctaa	agaaatgttt	aaacctcata	attggatgtc	1920
aggccatttg	tattttgagg	gatcagatac	ccctatagtt	tacttttatg	accctgtgaa	1980
ttccactctc	ctaccaccgg	agaggtgggc	taggttgccc	ggtacccac	ctgtggtacg	2040
tggttcttgg	ttacaggttc	cgcaaggggt	ttacagtgat	gtgaaagacc	tagccacagg	2100
attgatcacc	aaagacaaag	cctggaaaaa	ttatcaggtc	ttatattccg	ccacgggtgc	2160
tttgtctctt	acgggagtta	ccaccaaggc	cgtggtgcta	attctgttgg	ggttgtgtgg	2220
cagcaagtat	cttatttttag	cctacctctg	ttacttgtcc	ctttgttttg	ggcgcgcttc	2280
tggttaccct	ttgcgtcctg	tgtcccatc	ccagtcgtat	ctccaagctg	gctgggatgt	2340
tttgtctaaa	gctcaagtag	ctccttttgc	tttgattttc	ttcatctggt	gctatctccg	2400
ctgcaggcta	cgttatgctg	ccctttttagg	gtttgtgccc	atggctgcgg	gcttgcccct	2460
aactttcttt	gttgcagcag	ctgctgccc	accagattat	gactggtggg	tgcgactgct	2520
agtggcagg	ttagttttgt	gggccggccg	taaccgtggt	caccgcatag	ctctgcttgt	2580
aggtccttgg	cctctggtag	cgcttttaac	cctcttgcat	ttggttacgc	ctgcttcagc	2640
ttttgatacc	gagataattg	gagggctgac	aataccacct	gtagtagcat	tagttgtcat	2700
gtctcgtttt	ggcttctttg	ctcacttggt	acctcgctgt	gctttagtta	actcctatct	2760
ttggcaacgt	tgggagaatt	ggttttggaa	cgttacacta	agaccggaga	ggtttttcc	2820
tgtgctgggt	tgtttccccg	gtgcgacata	tgacgcgctg	gtgactttct	gtgtgtgtca	2880
cgtagctctt	ctatgtttta	catccagtg	agcatcgctc	tttgggactg	actctagggt	2940
tagggcccat	agaatgttgg	tgcgtctcgg	aaagtgtcat	gcttggtatt	ctcattatgt	3000
tcttaagt	ttcctcttag	tgtttggtga	gaatggtgtg	tttttctata	agcacttgca	3060
tggatgatgc	ttgcctaattg	attttgcctc	gaaactacca	ttgcaagagc	catttttccc	3120
ttttgaaggc	aaggcaagg	tctataggaa	tgaaggga	cgcttggcgt	gtggggacac	3180
ggttgatgg	ttgcocgttg	ttgcgcgtct	oggcgacctt	gttttcgcag	ggttggtctat	3240
gccgccagat	gggtgggcca	ttaccgcacc	ttttacgctg	cagtgtctct	ctgaacgtgg	3300
cacgctgtca	gcgatggcag	tggatcatgac	tggatatagac	ccccgaactt	ggactggaac	3360
tatcttcaga	ttaggatctc	tggccactag	ctacatggga	tttgtttgtg	acaacgtgtt	3420
gtatactgct	caccatggca	gcaagggg	cgggttggtc	catcccacag	gctctataca	3480
cccaataa	ggtgacgcgg	ctaataacca	ggacatctat	caaccacct	gtggagctgg	3540
gtcccttact	cgggtgctct	gcggggagac	caaggggtat	ctggtaacac	gactggggtc	3600
attggttgag	gtcaacaaat	ccgatgaccc	ttattggtgt	gtgtgcgggg	cccttcccat	3660
ggctgttgcc	aagggttctt	cagggtgccc	gattctgtgc	tcctccgggc	atgttatttg	3720
gatgttcacc	gctgctagaa	attctggcgg	ttcagtcagt	cagattagg	ttaggccgtt	3780
ggtgtgtgct	ggataccatc	cccagtacac	agcacatgcc	actcttgata	caaaacctac	3840
tgtgccta	gagtattcag	tgcataattt	aattgcccc	actggcagcg	gcaagtcaac	3900
caaattacca	ctttcttata	tgcaggagaa	gtatgaggtc	ttggtcctaa	atcccaattg	3960
ggctacaaca	gcataaatgc	caaagtacat	gcacgcgacg	tacggcgtga	atccaaattg	4020
ctattttta	ggcaaatgta	ccaacacagg	ggcttcactt	acgtacagca	catatggcat	4080
gtacctgacc	ggagcatgtt	cccggaaacta	tgatgtaatc	atttgtgacg	aatgccatgc	4140
taccgatgca	accaccgtgt	tgggcatttg	aaaggctcta	accgaagctc	catccaaaaa	4200
tgttaggcta	gtggttcttg	ccacggctac	ccccctgga	gtaatcccta	caccacatgc	4260
caacataact	gagattcaat	taaccgatga	aggcactatc	ccctttcatg	gaaaaaagat	4320
taaggaggaa	aatctgaaga	aaggagagaca	ccttatcttt	gaggctacca	aaaaaactg	4380
tgatgagctt	gctaacgagt	tagctcgaaa	gggaataaca	gctgtctctt	actatagggg	4440
atgtgacatc	tcaaaaatcc	ctgagggcga	ctgtgtagta	gttgccactg	atgccttgtg	4500
tacaggttac	actggtgact	ttgattccgt	gtatgactgc	agcctcatgg	tagaaggcac	4560
atgccatgtt	gaccttgacc	ctactttcac	catgggtgtt	cgtgtgtg	gggtttcagc	4620
aatagttaaa	ggccagcgta	ggggccgcac	aggccgtggg	agagctggca	tataactacta	4680
tgtagacggg	agttgtaccc	cttcgggtat	ggttcctgaa	tgcaacattg	ttgaagcctt	4740
cgacgcagcc	aaggcatggt	atggtttgtc	atcaacagaa	gctcaacta	ttctggacac	4800
ctatcgcacc	caacctgggt	tacctgcgat	aggagcaaat	ttggacgagt	gggctgatct	4860
cttttctatg	gtcaaccccg	aaccttcatt	tgtcaatact	gcaaaaagaa	ctgctgacaa	4920
ttatgttttg	ttgactgcag	cccaactaca	actgtgtcat	cagtatggct	atgctgctcc	4980

caatgacgca	ccacgggtggc	agggagcccc	gcttgggaaa	aaaccttgtg	gggttctgtg	5040
gcgcttgac	ggcgctgacg	cctgtcctgg	cccagagccc	agcgaggtga	ccagatacca	5100
aatgtgcttc	actgaagtca	atacttctgg	gacagccgca	ctcgctgttg	gcgttggagt	5160
ggctatggct	tatctagcca	ttgacacttt	tggcgccact	tgtgtgcggc	gttgcctggc	5220
tattgcatca	gtccctaccg	gtgctactgt	cgccccagtg	gttgacgaag	aagaaatcgt	5280
ggaggagtgt	gcacattca	ttcccttggg	ggccatgggt	gctgcaatcg	ataagctgaa	5340
gagtacaatc	accacaacta	gtcccttcac	attggaaaac	gcccctgaaa	aacttaacac	5400
ctttcttggg	cctcatgcag	ctacaatcct	tgctatcata	gagtattgct	gtggtttagt	5460
cactttacct	gacaatccct	ttgcatcatg	cgtgtttgct	ttcattgcgg	gtattactac	5520
cccactacct	cacaagatca	aaatgttcct	gtcattatct	ggaggcgcaa	ttgcgtccaa	5580
gcttacagac	gctagaggcg	cactggcggt	catgatggcc	ggggctgcgg	gaacagctct	5640
tggtagatgg	acatcggtgg	gttttgtctt	tgacatgcta	ggcggctatg	ctgccgcctc	5700
atccactgct	tgcttgacat	ttaaagtctt	gatgggtgag	tggccacta	tggatcagct	5760
tgctgggtta	gtctactccg	cggtcaatcc	ggcgcgagga	gttgtgggcg	tcttgtcagc	5820
ttgtgcaatg	tttgctttga	caacagcagg	gccagatcac	tggcccaaca	gacttcttac	5880
tatgcttgct	aggagcaaca	ctgtatgtaa	tgagtacttt	attgccactc	gtgacatccg	5940
caggaagata	ctgggcattc	tggaggcatc	taccccctgg	agtgtcatat	cagcttgcac	6000
ccgttggtct	cacacccoga	cggaggatga	ttgcggcctc	attgcttggg	gtctagagat	6060
ttggcagtat	gtgtgcaatt	tctttgtgat	ttgctttaag	gtccttaaag	ctggagttca	6120
gagcatgggt	aacattcctg	gttgcctttt	ctacagctgc	cagaaggggt	acaagggccc	6180
ctggatttga	tcagggtatg	tccaagcacg	ctgtccatgc	ggtgctgaac	tcactctttc	6240
tggtgagaat	ggttttgcaa	aactttacaa	aggaccocaga	acttgttcaa	attactggag	6300
aggggctgtt	ccagtcaacg	ctaggctgtg	tgggtcggct	agaccggacc	caactgattg	6360
gactagtctt	gtcgtcaatt	atggcggttag	ggactactgt	aaatatgaga	aaatgggaga	6420
tcacattttt	gttacagcag	tatcctctcc	aaatgtctgt	ttcaccacag	tgcccccaac	6480
cttgagagct	gcagtggccg	tggacggcgt	acaggttcag	tggtatctag	gtgagcccaa	6540
aactccttgg	acgacatctg	cttgcctgta	cggctcctgac	ggtaagggta	aaactgttaa	6600
gcttcccttc	cgcggtgacg	gtcacacacc	tgggtgtgog	atgcaactta	atttgcgtga	6660
tgcaacttgag	acaaatgact	gtaattccac	aaacaacact	cctagtgatg	aagccgcagt	6720
gtccgctctt	gttttcaaac	aggagtgtcg	gcgtacaaac	caattgcttg	aggcaatttc	6780
agctggcggt	gacaccacca	aactgccagc	cccctccatc	gaagaggtag	tggtaagaaa	6840
gcgccagttc	cgggcaagaa	ctgggttcgt	taccttgcct	ccccctccga	gatccgtccc	6900
aggagtgtca	tgtcctgaaa	gcctgcaacg	aagtgaaccc	ttagaaggtc	cttcaaacct	6960
ccctccttca	ccactgttc	tacagtggc	catgcgatg	cccctgttgg	gagcgggtga	7020
gtgtaaccct	ttcaactgcaa	ttggatgtgc	aatgaccgaa	acaggcgag	gccctgatga	7080
tttaccacgt	taccctccca	aaaaggagggt	ctctgaatgg	tcagacgaaa	gttggctcac	7140
ggctacaaac	gtttccagct	acgttactgg	ccccccgtac	cctaagatac	ggggaaagga	7200
ttccactcag	tcagcccccg	ccaaacggcc	tacaaaaaag	aagttgggaa	agagttagtt	7260
ttcgtgcagc	atgagctaca	cctggaccga	cgtgattagc	ttcaaaactg	cttctaaagt	7320
tctgtctgca	actcgggcca	tactagtgg	tttccctcaa	caaagatcat	tgggtgatgt	7380
gactgagccg	cgggatgcgg	agcttagaaa	acaaaaagtc	actattaata	gacaacctct	7440
gttcccccca	tcataaccaca	agcaagttag	attggctaag	gaaaaagctt	caaaagtgtg	7500
cgggtgcatg	tgggactatg	atgaagtagc	agctcacacg	ccctctaagt	ctgctaagtc	7560
ccacatcaact	ggccttcggg	gcactgatgt	tcgttctgga	gcagcccgcg	aggctgttct	7620
ggacttgacg	aagtgtgtcg	aggcagggtga	gataccgagt	cattatcggc	aaactgtgat	7680
agttccaaaag	gaggaggtct	tcgtgaagac	ccccagaaa	ccaacaaaga	aacccccagg	7740
gcttatctcg	tacccccacc	ttgaaatgag	atgtgttgag	aagatgtact	acggctcagg	7800
tgctcctgac	gtagttaaag	ctgtcatggg	agatgcgtac	gggtttgtag	atccacgtac	7860
ccgtgtcaag	cgtctgttgt	cgatgtggtc	accgatgca	gtcggagcca	catgcgatac	7920
agtgtgtttt	gacagtacca	tcacacccga	ggatatcatg	gtggagacag	acatctactc	7980
agcagctaaa	ctcagtgacc	aacaccgagc	tggcattcac	accattgcga	ggcagttata	8040
cgtgtggagga	ccgatgatcg	cttatgatgg	ccgagagatc	ggatatcgta	ggtgtaggtc	8100
ttccggcgct	tatactacct	caagttccaa	cagtttgacc	tgctggctga	aggtaaattgc	8160
tgacgccgaa	caggctggca	tgaagaacct	tcgcttccct	atttgcggcg	atgattgcac	8220
cgtaatttgg	aagagcgccg	gagcagatgc	agacaaacaa	gcaatgcgtg	tctttgctag	8280
ctggatgaag	gtgatgggtg	caccacaaga	ttgtgtgcoct	caacccaaat	acagtttggg	8340
agaattaaca	tcatgctcat	caaagtgtac	ctctggaatt	acaaaaagtg	gcaagcctta	8400
ctactttctt	acaagagatc	ctcgtatccc	ccttggcagg	tgctctgccg	agggctctgg	8460
atacaacccc	agtgtgtcgt	ggattgggta	tctaatacat	cactacccat	gtttgtgggt	8520
tagccgtgtg	ttggctgtcc	atttcatgga	gcagatgctc	tttgaggaca	aacttcccga	8580
gacggtgacc	tttgactggg	atgggaaaaa	ttatacgggtg	cctgtagaag	atctgccag	8640
catcattgct	ggtgtgcacg	gtattgaggc	tttctcgggtg	gtgcgctaca	ccaacgctga	8700
gatcctcaga	gtttcccaat	cactaacaga	catgaccatg	ccccccctgc	gagcctggcg	8760

```

aaagaaagcc agggcggtcc tcgccagcgc caagaggcgt ggccggagcac acgcaaaatt 8820
ggctcgcttc cttctctggc atgctacatc tagacctcta ccagatttgg ataagacgag 8880
cgtggctcgg tacaccactt tcaattattg tgatgtttac tccccggagg gggatgtgtt 8940
tattacacca cagagaagat tgcagaagtt tcttgtgaag tatttggctg tcattgtttt 9000
tgccctaggg ctcatgtctg ttggattagc catcagctga acccccaaat tcaaaattaa 9060
ctaacagttt tttttttttt tttttttttt agggcagcgg caacagggga gaccccgggc 9120
ttaacgaccc cgccgatgtg agtttgccga ccatggtgga tcagaaccgt ttcgggtgaa 9180
gccatggtct gaaggggatg acgtcccttc tggtcatcc acaaaaaccg tctcgggttg 9240
gtgaggagtc ctggctgtgt gggaagcagt cagtataatt cccgtcgtgt gtggtgacgc 9300
ctcacgacgt atttgtccgc tgtgcagagc gtagtaccac gggctgcacc ccggtttttg 9360
ttccaagcgg agggcaaccc ccgcttgga ttaaaaact 9399

```

<210> 3
 <211> 48
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer

<400> 3
 cgggatcccc taatacgact cactatagac cacaaacact ccagtttg 48

<210> 4
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer

<400> 4
 gtggaattca cagcgtcata 20

<210> 5
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer

<400> 5
 tgtgaattcc actctcctac c 21

<210> 6
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Primer

<400> 6
 ttatcgattg cagcaaccat g 21

<210> 7
 <211> 39

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 7
catggttgct gcaatcgata agctgaagag tacaataac 39

<210> 8
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 8
gacaacagac gcttgacacg 20

<210> 9
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 9
ctgtcatggg agatgcgtac 20

<210> 10
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 10
cgagctcgag cacatgcggg ggctggttaag cccgggggtct cc 42

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 11
gacaacagac gcttgacacg 20

<210> 12
<211> 108
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 12

```
ccgactcgag aattcggccc tgcaggccac aacagtctcg cgagttttta attccaagcg 60
ggggttgccc tccgcttgga acaaaaaccg ggggtgcagcc cttggtac 108
```

<210> 13

<211> 9599

<212> DNA

<213> Hepatitis C virus

<220>

<221> CDS

<222> (342)..(9377)

<400> 13

```
gccagcccc tgatgggggc gacactccac catgaatcac tcccctgtga ggaactactg 60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgtcgtgcag cctccaggac 120
ccccctccc gggagagcca tagtgggtctg cggaaccggg gagtacaccg gaattgccag 180
gacgaccggg tcctttcttg gataaaccgg ctcaatgcct ggagatttgg gcgtgcccc 240
gcaagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtgggtact gcctgatagg 300
gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac c atg agc acg aat cct 356
                                     Met Ser Thr Asn Pro
                                     1             5
```

```
aaa cct caa aga aaa acc aaa cgt aac acc aac cgt cgc cca cag gac 404
Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp
                10                15                20
```

```
gtc aag ttc ccg ggt ggc ggt cag atc gtt ggt gga gtt tac ttg ttg 452
Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu
                25                30                35
```

```
ccg cgc agg ggc cct aga ttg ggt gtg cgc gcg acg agg aag act tcc 500
Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser
                40                45                50
```

```
gag cgg tcg caa cct cga ggt aga cgt cag cct atc ccc aag gca cgt 548
Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg
                55                60                65
```

```
cgg ccc gag ggc agg acc tgg gct cag ccc ggg tac cct tgg ccc ctc 596
Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu
                70                75                80                85
```

```
tat ggc aat gag ggt tgc ggg tgg gcg gga tgg ctc ctg tct ccc cgt 644
Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg
                90                95                100
```

```
ggc tct cgg cct agc tgg ggc ccc aca gac ccc cgg cgt agg tcg cgc 692
Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg
                105                110                115
```

```
aat ttg ggt aag gtc atc gat acc ctt acg tgc ggc ttc gcc gac ctc 740
Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
                120                125                130
```

atg ggg tac ata ccg ctc gtc ggc gcc cct ctt gga ggc gct gcc agg	788
Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg	
135 140 145	
gcc ctg gcg cat ggc gtc cgg gtt ctg gaa gac ggc gtg aac tat gca	836
Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala	
150 155 160 165	
aca ggg aac ctt cct ggt tgc tct ttc tct atc ttc ctt ctg gcc ctg	884
Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu	
170 175 180	
ctc tct tgc ctg act gtg ccc gct tca gcc tac caa gtg cgc aat tcc	932
Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr Gln Val Arg Asn Ser	
185 190 195	
tcg ggg ctt tac cat gtc acc aat gat tgc cct aac tcg agt att gtg	980
Ser Gly Leu Tyr His Val Thr Asn Asp Cys Pro Asn Ser Ser Ile Val	
200 205 210	
tac gag gcg gcc gat gcc atc ctg cac act ccg ggg tgt gtc cct tgc	1028
Tyr Glu Ala Ala Asp Ala Ile Leu His Thr Pro Gly Cys Val Pro Cys	
215 220 225	
gtt cgc gag ggt aac gcc tcg agg tgt tgg gtg gcg gtg acc ccc acg	1076
Val Arg Glu Gly Asn Ala Ser Arg Cys Trp Val Ala Val Thr Pro Thr	
230 235 240 245	
gtg gcc acc agg gac ggc aaa ctc ccc aca acg cag ctt cga cgt cat	1124
Val Ala Thr Arg Asp Gly Lys Leu Pro Thr Thr Gln Leu Arg Arg His	
250 255 260	
atc gat ctg ctt gtc ggg agc gcc acc ctc tgc tcg gcc ctc tac gtg	1172
Ile Asp Leu Leu Val Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val	
265 270 275	
ggg gac ctg tgc ggg tct gtc ttt ctt gtt ggt caa ctg ttt acc ttc	1220
Gly Asp Leu Cys Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe	
280 285 290	
tct ccc agg cgc cac tgg acg acg caa gac tgc aat tgt tct atc tat	1268
Ser Pro Arg Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr	
295 300 305	
ccc ggc cat ata acg ggt cat cgc atg gca tgg gat atg atg atg aac	1316
Pro Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn	
310 315 320 325	
tgg tcc cct acg gca gcg ttg gtg gta gct cag ctg ctc cgg atc cca	1364
Trp Ser Pro Thr Ala Ala Leu Val Val Ala Gln Leu Leu Arg Ile Pro	
330 335 340	
caa gcc atc atg gac atg atc gct ggt gct cac tgg gga gtc ctg gcg	1412
Gln Ala Ile Met Asp Met Ile Ala Gly Ala His Trp Gly Val Leu Ala	
345 350 355	
ggc ata gcg tat ttc tcc atg gtg ggg aac tgg gcg aag gtc ctg gta	1460
Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu Val	
360 365 370	
gtg ctg ctg cta ttt gcc ggc gtc gac gcg gaa acc cac gtc acc ggg	1508
Val Leu Leu Leu Phe Ala Gly Val Asp Ala Glu Thr His Val Thr Gly	

375	380	385	
gga aat gcc ggc cgc acc acg gct ggg ctt gtt ggt ctc ctt aca cca Gly Asn Ala Gly Arg Thr Thr Ala Gly Leu Val Gly Leu Leu Thr Pro 390 395 400 405			1556
ggc gcc aag cag aac atc caa ctg atc aac acc aac ggc agt tgg cac Gly Ala Lys Gln Asn Ile Gln Leu Ile Asn Thr Asn Gly Ser Trp His 410 415 420			1604
atc aat agc acg gcc ttg aat tgc aat gaa agc ctt aac acc ggc tgg Ile Asn Ser Thr Ala Leu Asn Cys Asn Glu Ser Leu Asn Thr Gly Trp 425 430 435			1652
tta gca ggg ctc ttc tat caa cac aaa ttc aac tct tca ggc tgt cct Leu Ala Gly Leu Phe Tyr Gln His Lys Phe Asn Ser Ser Gly Cys Pro 440 445 450			1700
gag agg ttg gcc agc tgc cga cgc ctt acc gat ttt gcc cag ggc tgg Glu Arg Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Ala Gln Gly Trp 455 460 465			1748
ggt cct atc agt tat gcc aac gga agc ggc ctc gac gaa cgc ccc tac Gly Pro Ile Ser Tyr Ala Asn Gly Ser Gly Leu Asp Glu Arg Pro Tyr 470 475 480 485			1796
tgc tgg cac tac cct cca aga cct tgt ggc att gtg ccc gca aag agc Cys Trp His Tyr Pro Pro Arg Pro Cys Gly Ile Val Pro Ala Lys Ser 490 495 500			1844
gtg tgt ggc ccg gta tat tgc ttc act ccc agc ccc gtg gtg gtg gga Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val Val Gly 505 510 515			1892
acg acc gac agg tcg ggc gcg cct acc tac agc tgg ggt gca aat gat Thr Thr Asp Arg Ser Gly Ala Pro Thr Tyr Ser Trp Gly Ala Asn Asp 520 525 530			1940
acg gat gtc ttc gtc ctt aac aac acc agg cca ccg ctg ggc aat tgg Thr Asp Val Phe Val Leu Asn Asn Thr Arg Pro Pro Leu Gly Asn Trp 535 540 545			1988
ttc ggt tgt acc tgg atg aac tca act gga ttc acc aaa gtg tgc gga Phe Gly Cys Thr Trp Met Asn Ser Thr Gly Phe Thr Lys Val Cys Gly 550 555 560 565			2036
gcg ccc cct tgt gtc atc gga ggg gtg ggc aac aac acc ttg ctc tgc Ala Pro Pro Cys Val Ile Gly Gly Val Gly Asn Asn Thr Leu Leu Cys 570 575 580			2084
ccc act gat tgc ttc cgc aaa cat ccg gaa gcc aca tac tct cgg tgc Pro Thr Asp Cys Phe Arg Lys His Pro Glu Ala Thr Tyr Ser Arg Cys 585 590 595			2132
ggc tcc ggt ccc tgg att aca ccc agg tgc atg gtc gac tac ccg tat Gly Ser Gly Pro Trp Ile Thr Pro Arg Cys Met Val Asp Tyr Pro Tyr 600 605 610			2180
agg ctt tgg cac tat cct tgt acc atc aat tac acc ata ttc aaa gtc Arg Leu Trp His Tyr Pro Cys Thr Ile Asn Tyr Thr Ile Phe Lys Val 615 620 625			2228
agg atg tac gtg gga ggg gtc gag cac agg ctg gaa gcg gcc tgc aac			2276

Arg 630	Met	Tyr	Val	Gly	Gly 635	Val	Glu	His	Arg	Leu 640	Glu	Ala	Ala	Cys	Asn 645	
tgg	acg	cgg	ggc	gaa	cgc	tgt	gat	ctg	gaa	gac	agg	gac	agg	tcc	gag	2324
Trp	Thr	Arg	Gly	Glu 650	Arg	Cys	Asp	Leu	Glu 655	Asp	Arg	Asp	Arg	Ser	Glu 660	
ctc	agc	ccg	ttg	ctg	ctg	tcc	acc	aca	cag	tgg	cag	gtc	ctt	ccg	tgt	2372
Leu	Ser	Pro	Leu 665	Leu	Leu	Ser	Thr	Thr 670	Gln	Trp	Gln	Val	Leu 675	Pro	Cys	
tct	ttc	acg	acc	ctg	cca	gcc	ttg	tcc	acc	ggc	ctc	atc	cac	ctc	cac	2420
Ser	Phe	Thr	Thr 680	Leu	Pro	Ala	Leu 685	Ser	Thr	Gly	Leu	Ile 690	His	Leu	His	
cag	aac	att	gtg	gac	gtg	cag	tac	ttg	tac	ggg	gta	ggg	tca	agc	atc	2468
Gln	Asn	Ile	Val	Asp	Val	Gln 700	Tyr	Leu	Tyr	Gly	Val 705	Gly	Ser	Ser	Ile	
gcg	tcc	tgg	gcc	att	aag	tgg	gag	tac	gtc	gtt	ctc	ctg	ttc	ctt	ctg	2516
Ala	Ser	Trp	Ala	Ile 710	Lys 715	Trp	Glu	Tyr	Val	Val 720	Leu	Leu	Phe	Leu	Leu 725	
ctt	gca	gac	gcg	cgc	gtc	tgc	tcc	tgc	ttg	tgg	atg	atg	tta	ctc	ata	2564
Leu	Ala	Asp	Ala	Arg 730	Val	Cys	Ser	Cys	Leu 735	Trp	Met	Met	Leu	Leu 740	Ile	
tcc	caa	gcg	gag	gcg	gct	ttg	gag	aac	ctc	gta	ata	ctc	aat	gca	gca	2612
Ser	Gln	Ala	Glu 745	Ala	Ala	Leu	Glu	Asn 750	Leu	Val	Ile	Leu 755	Asn	Ala	Ala	
tcc	ctg	gcc	ggg	acg	cac	ggc	ctt	gtg	tcc	ttc	ctc	gtg	ttc	ttc	tgc	2660
Ser	Leu	Ala	Gly 760	Thr	His	Gly	Leu 765	Val	Ser	Phe	Leu	Val 770	Phe	Phe	Cys	
ttt	gcg	tgg	tat	ctg	aag	ggc	agg	tgg	gtg	ccc	gga	gcg	gtc	tac	gcc	2708
Phe	Ala	Trp	Tyr	Leu	Lys	Gly 780	Arg	Trp	Val	Pro	Gly 785	Ala	Val	Tyr	Ala	
ctc	tac	ggg	atg	tgg	cct	ctc	ctc	ctg	ctc	ctg	ctg	gcg	ttg	cct	cag	2756
Leu	Tyr	Gly	Met	Trp	Pro	Leu 795	Leu	Leu	Leu	Leu 800	Leu	Ala	Leu	Pro	Gln 805	
cgg	gca	tac	gca	ctg	gac	acg	gag	gtg	gcc	gcg	tcg	tgt	ggc	ggc	gtt	2804
Arg	Ala	Tyr	Ala	Leu 810	Asp	Thr	Glu	Val	Ala 815	Ala	Ser	Cys	Gly	Gly	Val 820	
gtt	ctt	gtc	ggg	tta	atg	gcg	ctg	act	ctg	tcg	cca	tat	tac	aag	cgc	2852
Val	Leu	Val	Gly 825	Leu	Met	Ala	Leu 830	Thr	Leu	Ser	Pro	Tyr	Tyr 835	Lys	Arg	
tat	atc	agc	tgg	tgc	atg	tgg	tgg	ctt	cag	tat	ttt	ctg	acc	aga	gta	2900
Tyr	Ile	Ser	Trp 840	Cys	Met	Trp	Trp 845	Leu	Gln	Tyr	Phe	Leu 850	Thr	Arg	Val	
gaa	gcg	caa	ctg	cac	gtg	tgg	gtt	ccc	ccc	ctc	aac	gtc	cgg	ggg	ggg	2948
Glu	Ala	Gln	Leu 855	His	Val	Trp	Val 860	Pro	Pro	Leu	Asn 865	Val	Arg	Gly	Gly	
cgc	gat	gcc	gtc	atc	tta	ctc	atg	tgt	gta	gta	cac	ccg	acc	ctg	gta	2996
Arg	Asp	Ala	Val	Ile 870	Leu	Met	Cys 875	Val	Val 880	His	Pro	Thr	Leu	Val 885		

ttt gac atc acc aaa cta ctc ctg gcc atc ttc gga ccc ctt tgg att	3044
Phe Asp Ile Thr Lys Leu Leu Leu Ala Ile Phe Gly Pro Leu Trp Ile	
890 895 900	
ctt caa gcc agt ttg ctt aaa gtc ccc tac ttc gtg cgc gtt caa ggc	3092
Leu Gln Ala Ser Leu Leu Lys Val Pro Tyr Phe Val Arg Val Gln Gly	
905 910 915	
ctt ctc cgg atc tgc gcg cta gcg cgg aag ata gcc gga ggt cat tac	3140
Leu Leu Arg Ile Cys Ala Leu Ala Arg Lys Ile Ala Gly Gly His Tyr	
920 925 930	
gtg caa atg gcc atc atc aag tta ggg gcg ctt act ggc acc tat gtg	3188
Val Gln Met Ala Ile Ile Lys Leu Gly Ala Leu Thr Gly Thr Tyr Val	
935 940 945	
tat aac cat ctc acc cct ctt cga gac tgg gcg cac aac ggc ctg cga	3236
Tyr Asn His Leu Thr Pro Leu Arg Asp Trp Ala His Asn Gly Leu Arg	
950 955 960 965	
gat ctg gcc gtg gct gtg gaa cca gtc gtc ttc tcc cga atg gag acc	3284
Asp Leu Ala Val Ala Val Glu Pro Val Val Phe Ser Arg Met Glu Thr	
970 975 980	
aag ctc atc acg tgg ggg gca gat acc gcc gcg tgc ggt gac atc atc	3332
Lys Leu Ile Thr Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile	
985 990 995	
aac ggc ttg ccc gtc tct gcc cgt agg ggc cag gag ata ctg ctt ggg	3380
Asn Gly Leu Pro Val Ser Ala Arg Arg Gly Gln Glu Ile Leu Leu Gly	
1000 1005 1010	
cca gcc gac gga atg gtc tcc aag ggg tgg agg ttg ctg gcg ccc atc	3428
Pro Ala Asp Gly Met Val Ser Lys Gly Trp Arg Leu Leu Ala Pro Ile	
1015 1020 1025	
acg gcg tac gcc cag cag acg aga ggc ctc cta ggg tgt ata atc acc	3476
Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr	
1030 1035 1040 1045	
agc ctg act ggc cgg gac aaa aac caa gtg gag ggt gag gtc cag atc	3524
Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val Gln Ile	
1050 1055 1060	
gtg tca act gct acc caa acc ttc ctg gca acg tgc atc aat ggg gta	3572
Val Ser Thr Ala Thr Gln Thr Phe Leu Ala Thr Cys Ile Asn Gly Val	
1065 1070 1075	
tgc tgg act gtc tac cac ggg gcc gga acg agg acc atc gca tca ccc	3620
Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr Ile Ala Ser Pro	
1080 1085 1090	
aag ggt cct gtc atc cag atg tat acc aat gtg gac caa gac ctt gtg	3668
Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val	
1095 1100 1105	
ggc tgg ccc gct cct caa ggt tcc cgc tca ttg aca ccc tgt acc tgc	3716
Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser Leu Thr Pro Cys Thr Cys	
1110 1115 1120 1125	
ggc tcc tcg gac ctt tac ctg gtc acg agg cac gcc gat gtc att ccc	3764
Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro	
1130 1135 1140	

gtg cgc cgg cga ggt gat agc agg ggt agc ctg ctt tcg ccc cgg ccc	3812
Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro	
1145 1150 1155	
att tcc tac ttg aaa ggc tcc tcg ggg ggt ccg ctg ttg tgc ccc gcg	3860
Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ala	
1160 1165 1170	
gga cac gcc gtg ggc cta ttc agg gcc gcg gtg tgc acc cgt gga gtg	3908
Gly His Ala Val Gly Leu Phe Arg Ala Ala Val Cys Thr Arg Gly Val	
1175 1180 1185	
gct aaa gcg gtg gac ttt atc cct gtg gag aac cta ggg aca acc atg	3956
Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn Leu Gly Thr Thr Met	
1190 1195 1200 1205	
aga tcc ccg gtg ttc acg gac aac tcc tct cca cca gca gtg ccc cag	4004
Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln	
1210 1215 1220	
agc ttc cag gtg gcc cac ctg cat gct ccc acc ggc agc ggt aag agc	4052
Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser	
1225 1230 1235	
acc aag gtc ccg gct gcg tac gca gcc cag ggc tac aag gtg ttg gtg	4100
Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val	
1240 1245 1250	
ctc aac ccc tct gtt gct gca acg ctg ggc ttt ggt gct tac atg tcc	4148
Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser	
1255 1260 1265	
aag gcc cat ggg gtt gat cct aat atc agg acc ggg gtg aga aca att	4196
Lys Ala His Gly Val Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile	
1270 1275 1280 1285	
acc act ggc agc ccc atc acg tac tcc acc tac ggc aag ttc ctt gcc	4244
Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala	
1290 1295 1300	
gac ggc ggg tgc tca gga ggt gct tat gac ata ata att tgt gac gag	4292
Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu	
1305 1310 1315	
tgc cac tcc acg gat gcc aca tcc atc ttg ggc atc ggc act gtc ctt	4340
Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu	
1320 1325 1330	
gac caa gca gag act gcg ggg gcg aga ctg gtt gtg ctc gcc act gct	4388
Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala	
1335 1340 1345	
acc cct ccg ggc tcc gtc act gtg tcc cat cct aac atc gag gag gtt	4436
Thr Pro Pro Gly Ser Val Thr Val Ser His Pro Asn Ile Glu Glu Val	
1350 1355 1360 1365	
gct ctg tcc acc acc gga gag atc ccc ttt tac ggc aag gct atc ccc	4484
Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro	
1370 1375 1380	
ctc gag gtg atc aag ggg gga aga cat ctc atc ttc tgc cac tca aag	4532
Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys	

1385	1390	1395	
aag aag tgc gac gag ctc gcc gcg aag ctg gtc gca ttg ggc atc aat			4580
Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn			
1400	1405	1410	
gcc gtg gcc tac tac cgc ggt ctt gac gtg tct gtc atc ccg acc agc			4628
Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser			
1415	1420	1425	
ggc gat gtt gtc gtc gtg tcg acc gat gct ctc atg act ggc ttt acc			4676
Gly Asp Val Val Val Val Ser Thr Asp Ala Leu Met Thr Gly Phe Thr			
1430	1435	1440	1445
ggc gac ttc gac tct gtg ata gac tgc aac acg tgt gtc act cag aca			4724
Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr			
1450	1455	1460	
gtc gat ttc agc ctt gac cct acc ttt acc att gag aca acc acg ctc			4772
Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Leu			
1465	1470	1475	
ccc cag gat gct gtc tcc agg act caa cgc cgg ggc agg act ggc agg			4820
Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg			
1480	1485	1490	
ggg aag cca ggc atc tat aga ttt gtg gca ccg ggg gag cgc ccc tcc			4868
Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser			
1495	1500	1505	
ggc atg ttc gac tcg tcc gtc ctc tgt gag tgc tat gac gcg ggc tgt			4916
Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys			
1510	1515	1520	1525
gct tgg tat gag ctc acg ccc gcc gag act aca gtt agg cta cga gcg			4964
Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala			
1530	1535	1540	
tac atg aac acc ccg ggg ctt ccc gtg tgc cag gac cat ctt gaa ttt			5012
Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe			
1545	1550	1555	
tgg gag ggc gtc ttt acg ggc ctc act cat ata gat gcc cac ttt tta			5060
Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu			
1560	1565	1570	
tcc cag aca aag cag agt ggg gag aac ttt cct tac ctg gta gcg tac			5108
Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Pro Tyr Leu Val Ala Tyr			
1575	1580	1585	
caa gcc acc gtg tgc gct agg gct caa gcc cct ccc cca tcg tgg gac			5156
Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp			
1590	1595	1600	1605
cag atg tgg aag tgt ttg atc cgc ctt aaa ccc acc ctc cat ggg cca			5204
Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro			
1610	1615	1620	
aca ccc ctg cta tac aga ctg ggc gct gtt cag aat gaa gtc acc ctg			5252
Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Leu			
1625	1630	1635	
acg cac cca atc acc aaa tac atc atg aca tgc atg tcg gcc gac ctg			5300

Thr His Pro Ile Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu	
1640	1645 1650
gag gtc gtc acg agc acc tgg gtg ctc gtt ggc ggc gtc ctg gct gct	5348
Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala	
1655	1660 1665
ctg gcc gcg tat tgc ctg tca aca ggc tgc gtg gtc ata gtg ggc agg	5396
Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg	
1670	1675 1680 1685
atc gtc ttg tcc ggg aag ccg gca att ata cct gac agg gag gtt ctc	5444
Ile Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu	
1690	1695 1700
tac cag gag ttc gat gag atg gaa gag tgc tct cag cac tta ccg tac	5492
Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr	
1705	1710 1715
atc gag caa ggg atg atg ctc gct gag cag ttc aag cag aag gcc ctc	5540
Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu	
1720	1725 1730
ggc ctc ctg cag acc gcg tcc cgc cat gca gag gtt atc acc cct gct	5588
Gly Leu Leu Gln Thr Ala Ser Arg His Ala Glu Val Ile Thr Pro Ala	
1735	1740 1745
gtc cag acc aac tgg cag aaa ctc gag gtc ttt tgg gcg aag cac atg	5636
Val Gln Thr Asn Trp Gln Lys Leu Glu Val Phe Trp Ala Lys His Met	
1750	1755 1760 1765
tgg aat ttc atc agt ggg ata caa tac ttg gcg ggc ctg tca acg ctg	5684
Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu	
1770	1775 1780
cct ggt aac ccc gcc att gct tca ttg atg gct ttt aca gct gcc gtc	5732
Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val	
1785	1790 1795
acc agc cca cta acc act ggc caa acc ctc ctc ttc aac ata ttg ggg	5780
Thr Ser Pro Leu Thr Thr Gly Gln Thr Leu Leu Phe Asn Ile Leu Gly	
1800	1805 1810
ggg tgg gtg gct gcc cag ctc gcc gcc ccc ggt gcc gct act gcc ttt	5828
Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe	
1815	1820 1825
gtg ggt gct ggc cta gct ggc gcc gcc atc ggc agc gtt gga ctg ggg	5876
Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly	
1830	1835 1840 1845
aag gtc ctc gtg gac att ctt gca ggg tat ggc gcg ggc gtg gcg gga	5924
Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly	
1850	1855 1860
gct ctt gta gca ttc aag atc atg agc ggt gag gtc ccc tcc acg gag	5972
Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu	
1865	1870 1875
gac ctg gtc aat ctg ctg ccc gcc atc ctc tcg cct gga gcc ctt gta	6020
Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val	
1880	1885 1890

gtc ggt gtg gtc tgc gca gca ata ctg cgc cgg cac gtt ggc ccg ggc	6068
Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly	
1895 1900 1905	
gag ggg gca gtg caa tgg atg aac cgg cta ata gcc ttc gcc tcc cgg	6116
Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg	
1910 1915 1920 1925	
ggg aac cat gtt tcc ccc acg cac tac gtg ccg gag agc gat gca gcc	6164
Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala	
1930 1935 1940	
gcc cgc gtc act gcc ata ctc agc agc ctc act gta acc cag ctc ctg	6212
Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu	
1945 1950 1955	
agg cga ctg cat cag tgg ata agc tcg gag tgt acc act cca tgc tcc	6260
Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser	
1960 1965 1970	
ggt tcc tgg cta agg gac atc tgg gac tgg ata tgc gag gtg ctg agc	6308
Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser	
1975 1980 1985	
gac ttt aag acc tgg ctg aaa gcc aag ctc atg cca caa ctg cct ggg	6356
Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly	
1990 1995 2000 2005	
att ccc ttt gtg tcc tgc cag cgc ggg tat agg ggg gtc tgg cga gga	6404
Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Arg Gly Val Trp Arg Gly	
2010 2015 2020	
gac ggc att atg cac act cgc tgc cac tgt gga gct gag atc act gga	6452
Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly	
2025 2030 2035	
cat gtc aaa aac ggg acg atg agg atc gtc ggt cct agg acc tgc agg	6500
His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg	
2040 2045 2050	
aac atg tgg agt ggg acg ttc ccc att aac gcc tac acc acg ggc ccc	6548
Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro	
2055 2060 2065	
tgt act ccc ctt cct gcg ccg aac tat aag ttc gcg ctg tgg agg gtg	6596
Cys Thr Pro Leu Pro Ala Pro Asn Tyr Lys Phe Ala Leu Trp Arg Val	
2070 2075 2080 2085	
tct gca gag gaa tac gtg gag ata agg cgg gtg ggg gac ttc cac tac	6644
Ser Ala Glu Glu Tyr Val Glu Ile Arg Arg Val Gly Asp Phe His Tyr	
2090 2095 2100	
gta tcg ggt atg act act gac aat ctt aaa tgc ccg tgc cag atc cca	6692
Val Ser Gly Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Ile Pro	
2105 2110 2115	
tcg ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc cta cac agg ttt	6740
Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe	
2120 2125 2130	
gcg ccc cct tgc aag ccc ttg ctg cgg gag gag gta tca ttc aga gta	6788
Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val	
2135 2140 2145	

gga ctc cac gag tac ccg gtg ggg tcg caa tta cct tgc gag ccc gaa	6836
Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu	
2150 2155 2160 2165	
ccg gac gta gcc gtg ttg acg tcc atg ctc act gat ccc tcc cat ata	6884
Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile	
2170 2175 2180	
aca gca gag gcg gcc ggg aga agg ttg gcg aga ggg tca ccc cct tct	6932
Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser	
2185 2190 2195	
atg gcc agc tcc tcg gct agc cag ctg tcc gct cca tct ctc aag gca	6980
Met Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala	
2200 2205 2210	
act tgc acc gcc aac cat gac tcc cct gac gcc gag ctc ata gag gct	7028
Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala	
2215 2220 2225	
aac ctc ctg tgg agg cag gag atg ggc ggc aac atc acc agg gtt gag	7076
Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu	
2230 2235 2240 2245	
tca gag aac aaa gtg gtg att ctg gac tcc ttc gat ccg ctt gtg gca	7124
Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala	
2250 2255 2260	
gag gag gat gag cgg gag gtc tcc gta cct gca gaa att ctg cgg aag	7172
Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg Lys	
2265 2270 2275	
tct cgg aga ttc gcc cgg gcc ctg ccc gtc tgg gcg cgg ccg gac tac	7220
Ser Arg Arg Phe Ala Arg Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr	
2280 2285 2290	
aac ccc ccg cta gta gag acg tgg aaa aag cct gac tac gaa cca cct	7268
Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro	
2295 2300 2305	
gtg gtc cat ggc tgc ccg cta cca cct cca cgg tcc cct cct gtg cct	7316
Val Val His Gly Cys Pro Leu Pro Pro Pro Arg Ser Pro Pro Val Pro	
2310 2315 2320 2325	
ccg cct cgg aaa aag cgt acg gtg gtc ctc acc gaa tca acc cta tct	7364
Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser	
2330 2335 2340	
act gcc ttg gcc gag ctt gcc acc aaa agt ttt ggc agc tcc tca act	7412
Thr Ala Leu Ala Glu Leu Ala Thr Lys Ser Phe Gly Ser Ser Ser Thr	
2345 2350 2355	
tcc ggc att acg ggc gac aat acg aca aca tcc tct gag ccc gcc cct	7460
Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro Ala Pro	
2360 2365 2370	
tct ggc tgc ccc ccc gac tcc gac gtt gag tcc tat tct tcc atg ccc	7508
Ser Gly Cys Pro Pro Asp Ser Asp Val Glu Ser Tyr Ser Ser Met Pro	
2375 2380 2385	
ccc ctg gag ggg gag cct ggg gat ccg gat ctc agc gac ggg tca tgg	7556
Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp	

2390	2395	2400	2405	
tcg acg gtc agt agt ggg gcc gac acg gaa gat gtc gtg tgc tgc tca				7604
Ser Thr Val Ser Ser Gly Ala Asp Thr Glu Asp Val Val Cys Cys Ser	2410	2415	2420	
atg tct tat tcc tgg aca ggc gca ctc gtc acc ccg tgc gct gcg gaa				7652
Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu	2425	2430	2435	
gaa caa aaa ctg ccc atc aac gca ctg agc aac tcg ttg cta cgc cat				7700
Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His	2440	2445	2450	
cac aat ctg gtg tat tcc acc act tca cgc agt gct tgc caa agg cag				7748
His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln	2455	2460	2465	
aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc cat tac cag				7796
Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln	2470	2475	2480	2485
gac gtg ctc aag gag gtc aaa gca gcg gcg tca aaa gtg aag gct aac				7844
Asp Val Leu Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn	2490	2495	2500	
ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca cat tca gcc				7892
Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala	2505	2510	2515	
aaa tcc aag ttt ggc tat ggg gca aaa gac gtc cgt tgc cat gcc aga				7940
Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg	2520	2525	2530	
aag gcc gta gcc cac atc aac tcc gtg tgg aaa gac ctt ctg gaa gac				7988
Lys Ala Val Ala His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp	2535	2540	2545	
agt gta aca cca ata gac act acc atc atg gcc aag aac gag gtt ttc				8036
Ser Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe	2550	2555	2560	2565
tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt ctc atc gtg				8084
Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val	2570	2575	2580	
ttc ccc gac ctg ggc gtg cgc gtg tgc gag aag atg gcc ctg tac gac				8132
Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp	2585	2590	2595	
gtg gtt agc aag ctc ccc ctg gcc gtg atg gga agc tcc tac gga ttc				8180
Val Val Ser Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe	2600	2605	2610	
caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa gcg tgg aag				8228
Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys	2615	2620	2625	
tcc aag aag acc ccg atg ggg ttc tcg tat gat acc cgc tgt ttt gac				8276
Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp	2630	2635	2640	2645
tcc aca gtc act gag agc gac atc cgt acg gag gag gca att tac caa				8324

Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln	
2650	2655 2660
tgt tgt gac ctg gac ccc caa gcc cgc gtg gcc atc aag tcc ctc act	8372
Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr	
2665	2670 2675
gag agg ctt tat gtt ggg ggc cct ctt acc aat tca agg ggg gaa aac	8420
Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn	
2680	2685 2690
tgc ggc tac cgc agg tgc cgc gcg agc ggc gta ctg aca act agc tgt	8468
Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys	
2695	2700 2705
ggt aac acc ctc act tgc tac atc aag gcc cgg gca gcc tgt cga gcc	8516
Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala	
2710	2715 2720 2725
gca ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac gac tta gtc	8564
Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val	
2730	2735 2740
gtt atc tgt gaa agt gcg ggg gtc cag gag gac gcg gcg agc ctg aga	8612
Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg	
2745	2750 2755
gcc ttc acg gag gct atg acc agg tac tcc gcc ccc ccc ggg gac ccc	8660
Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro	
2760	2765 2770
cca caa cca gaa tac gac ttg gag ctt ata aca tca tgc tcc tcc aac	8708
Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn	
2775	2780 2785
gtg tca gtc gcc cac gac ggc gct gga aag agg gtc tac tac ctt acc	8756
Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr	
2790	2795 2800 2805
cgt gac cct aca acc ccc ctc gcg aga gcc gcg tgg gag aca gca aga	8804
Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg	
2810	2815 2820
cac act cca gtc aat tcc tgg cta ggc aac ata atc atg ttt gcc ccc	8852
His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro	
2825	2830 2835
aca ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt agc gtc ctc	8900
Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu	
2840	2845 2850
ata gcc agg gat cag ctt gaa cag gct ctt aac tgt gag atc tac gga	8948
Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asn Cys Glu Ile Tyr Gly	
2855	2860 2865
gcc tgc tac tcc ata gaa cca ctg gat cta cct cca atc att caa aga	8996
Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg	
2870	2875 2880 2885
ctc cat ggc ctc agc gca ttt tca ctc cac agt tac tct cca ggt gaa	9044
Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu	
2890	2895 2900

```

atc aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gtc ccg ccc ttg 9092
Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu
      2905                      2910                      2915

cga gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg ctt ctg tcc 9140
Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser
      2920                      2925                      2930

aga gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc aac tgg gca 9188
Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala
      2935                      2940                      2945

gta aga aca aag ctc aaa ctc act cca ata gcg gcc gct ggc cgg ctg 9236
Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Arg Leu
      2950                      2955                      2960                      2965

gac ttg tcc ggt tgg ttc acg gct ggc tac agc ggg gga gac att tat 9284
Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr
      2970                      2975                      2980

cac agc gtg tct cat gcc cgg ccc cgc tgg ttc tgg ttt tgc cta ctc 9332
His Ser Val Ser His Ala Arg Pro Arg Trp Phe Trp Phe Cys Leu Leu
      2985                      2990                      2995

ctg ctc gct gca ggg gta ggc atc tac ctc ctc ccc aac cga tga 9377
Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
      3000                      3005                      3010

aggttgggggt aaacactccg gcctcttaag ccatttcctg tttttttttt tttttttttt 9437

ttttttttct tttttttttt ctttcctttc cttctttttt tcctttcttt ttcccttctt 9497

taatggtggc tccatcttag ccctagtcac ggctagctgt gaaaggtccg tgagccgcat 9557

gactgcagag agtgctgata ctggcctctc tgcagatcat gt 9599

```

<210> 14

<211> 3011

<212> PRT

<213> Hepatitis C virus

<400> 14

```

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

```

Gly	Val	Asn	Tyr	Ala	Thr	Gly	Asn	Leu	Pro	Gly	Cys	Ser	Phe	Ser	Ile	
				165					170					175		
Phe	Leu	Leu	Ala	Leu	Leu	Ser	Cys	Leu	Thr	Val	Pro	Ala	Ser	Ala	Tyr	
			180					185					190			
Gln	Val	Arg	Asn	Ser	Ser	Gly	Leu	Tyr	His	Val	Thr	Asn	Asp	Cys	Pro	
		195					200					205				
Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Ala	Asp	Ala	Ile	Leu	His	Thr	Pro	
		210				215					220					
Gly	Cys	Val	Pro	Cys	Val	Arg	Glu	Gly	Asn	Ala	Ser	Arg	Cys	Trp	Val	
225					230					235					240	
Ala	Val	Thr	Pro	Thr	Val	Ala	Thr	Arg	Asp	Gly	Lys	Leu	Pro	Thr	Thr	
				245					250					255		
Gln	Leu	Arg	Arg	His	Ile	Asp	Leu	Leu	Val	Gly	Ser	Ala	Thr	Leu	Cys	
			260				265						270			
Ser	Ala	Leu	Tyr	Val	Gly	Asp	Leu	Cys	Gly	Ser	Val	Phe	Leu	Val	Gly	
		275					280					285				
Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg	Arg	His	Trp	Thr	Thr	Gln	Asp	Cys	
		290				295					300					
Asn	Cys	Ser	Ile	Tyr	Pro	Gly	His	Ile	Thr	Gly	His	Arg	Met	Ala	Trp	
305					310					315					320	
Asp	Met	Met	Met	Asn	Trp	Ser	Pro	Thr	Ala	Ala	Leu	Val	Val	Ala	Gln	
				325					330					335		
Leu	Leu	Arg	Ile	Pro	Gln	Ala	Ile	Met	Asp	Met	Ile	Ala	Gly	Ala	His	
			340				345						350			
Trp	Gly	Val	Leu	Ala	Gly	Ile	Ala	Tyr	Phe	Ser	Met	Val	Gly	Asn	Trp	
		355					360					365				
Ala	Lys	Val	Leu	Val	Val	Leu	Leu	Leu	Phe	Ala	Gly	Val	Asp	Ala	Glu	
		370				375					380					
Thr	His	Val	Thr	Gly	Gly	Asn	Ala	Gly	Arg	Thr	Thr	Ala	Gly	Leu	Val	
385					390					395					400	
Gly	Leu	Leu	Thr	Pro	Gly	Ala	Lys	Gln	Asn	Ile	Gln	Leu	Ile	Asn	Thr	
				405					410					415		
Asn	Gly	Ser	Trp	His	Ile	Asn	Ser	Thr	Ala	Leu	Asn	Cys	Asn	Glu	Ser	
			420					425					430			
Leu	Asn	Thr	Gly	Trp	Leu	Ala	Gly	Leu	Phe	Tyr	Gln	His	Lys	Phe	Asn	
		435					440					445				
Ser	Ser	Gly	Cys	Pro	Glu	Arg	Leu	Ala	Ser	Cys	Arg	Arg	Leu	Thr	Asp	
		450				455				460						
Phe	Ala	Gln	Gly	Trp	Gly	Pro	Ile	Ser	Tyr	Ala	Asn	Gly	Ser	Gly	Leu	
465					470					475					480	
Asp	Glu	Arg	Pro	Tyr	Cys	Trp	His	Tyr	Pro	Pro	Arg	Pro	Cys	Gly	Ile	
				485					490					495		
Val	Pro	Ala	Lys	Ser	Val	Cys	Gly	Pro	Val	Tyr	Cys	Phe	Thr	Pro	Ser	
			500					505					510			
Pro	Val	Val	Val	Gly	Thr	Thr	Asp	Arg	Ser	Gly	Ala	Pro	Thr	Tyr	Ser	
		515					520						525			
Trp	Gly	Ala	Asn	Asp	Thr	Asp	Val	Phe	Val	Leu	Asn	Thr	Arg	Pro		
		530				535					540					
Pro	Leu	Gly	Asn	Trp	Phe	Gly	Cys	Thr	Trp	Met	Asn	Ser	Thr	Gly	Phe	
545					550					555					560	
Thr	Lys	Val	Cys	Gly	Ala	Pro	Pro	Cys	Val	Ile	Gly	Gly	Val	Gly	Asn	
				565					570					575		
Asn	Thr	Leu	Leu	Cys	Pro	Thr	Asp	Cys	Phe	Arg	Lys	His	Pro	Glu	Ala	
			580					585					590			
Thr	Tyr	Ser	Arg	Cys	Gly	Ser	Gly	Pro	Trp	Ile	Thr	Pro	Arg	Cys	Met	
		595					600					605				
Val	Asp	Tyr	Pro	Tyr	Arg	Leu	Trp	His	Tyr	Pro	Cys	Thr	Ile	Asn	Tyr	
		610				615					620					
Thr	Ile	Phe	Lys	Val	Arg	Met	Tyr	Val	Gly	Gly	Val	Glu	His	Arg	Leu	
625					630					635					640	
Glu	Ala	Ala	Cys	Asn	Trp	Thr	Arg	Gly	Glu	Arg	Cys	Asp	Leu	Glu	Asp	
				645					650					655		
Arg	Asp	Arg	Ser	Glu	Leu	Ser	Pro	Leu	Leu	Leu	Ser	Thr	Thr	Gln	Trp	

Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu Phe Arg Ala Ala Val
 1170 1175 1180
 Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn
 1185 1190 1195 1200
 Leu Gly Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro
 1205 1210 1215
 Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr
 1220 1225 1230
 Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly
 1235 1240 1245
 Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe
 1250 1255 1260
 Gly Ala Tyr Met Ser Lys Ala His Gly Val Asp Pro Asn Ile Arg Thr
 1265 1270 1275 1280
 Gly Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr
 1285 1290 1295
 Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile
 1300 1305 1310
 Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly
 1315 1320 1325
 Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val
 1330 1335 1340
 Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Ser His Pro
 1345 1350 1355 1360
 Asn Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr
 1365 1370 1375
 Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile
 1380 1385 1390
 Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val
 1395 1400 1405
 Ala Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser
 1410 1415 1420
 Val Ile Pro Thr Ser Gly Asp Val Val Val Val Ser Thr Asp Ala Leu
 1425 1430 1435 1440
 Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr
 1445 1450 1455
 Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile
 1460 1465 1470
 Glu Thr Thr Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg
 1475 1480 1485
 Gly Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro
 1490 1495 1500
 Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys
 1505 1510 1515 1520
 Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr
 1525 1530 1535
 Val Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln
 1540 1545 1550
 Asp His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile
 1555 1560 1565
 Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Pro
 1570 1575 1580
 Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro
 1585 1590 1595 1600
 Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro
 1605 1610 1615
 Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln
 1620 1625 1630
 Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Thr Cys
 1635 1640 1645
 Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly
 1650 1655 1660
 Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val

1665 1670 1675 1680
 Val Ile Val Gly Arg Ile Val Leu Ser Gly Lys Pro Ala Ile Ile Pro
 1685 1690 1695
 Asp Arg Glu Val Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ser
 1700 1705 1710
 Gln His Leu Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe
 1715 1720 1725
 Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg His Ala Glu
 1730 1735 1740
 Val Ile Thr Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Val Phe
 1745 1750 1755 1760
 Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala
 1765 1770 1775
 Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala
 1780 1785 1790
 Phe Thr Ala Ala Val Thr Ser Pro Leu Thr Thr Gly Gln Thr Leu Leu
 1795 1800 1805
 Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly
 1810 1815 1820
 Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly
 1825 1830 1835 1840
 Ser Val Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly
 1845 1850 1855
 Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu
 1860 1865 1870
 Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser
 1875 1880 1885
 Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg
 1890 1895 1900
 His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile
 1905 1910 1915 1920
 Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro
 1925 1930 1935
 Glu Ser Asp Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr
 1940 1945 1950
 Val Thr Gln Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys
 1955 1960 1965
 Thr Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile
 1970 1975 1980
 Cys Glu Val Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met
 1985 1990 1995 2000
 Pro Gln Leu Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Arg
 2005 2010 2015
 Gly Val Trp Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly
 2020 2025 2030
 Ala Glu Ile Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly
 2035 2040 2045
 Pro Arg Thr Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala
 2050 2055 2060
 Tyr Thr Thr Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Lys Phe
 2065 2070 2075 2080
 Ala Leu Trp Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Arg Val
 2085 2090 2095
 Gly Asp Phe His Tyr Val Ser Gly Met Thr Thr Asp Asn Leu Lys Cys
 2100 2105 2110
 Pro Cys Gln Ile Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val
 2115 2120 2125
 Arg Leu His Arg Phe Ala Pro Cys Lys Pro Leu Leu Arg Glu Glu
 2130 2135 2140
 Val Ser Phe Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu
 2145 2150 2155 2160
 Pro Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr
 2165 2170 2175

Asp Pro Ser His Ile Thr Ala Glu Ala Gly Arg Arg Leu Ala Arg
 2180 2185 2190
 Gly Ser Pro Pro Ser Met Ala Ser Ser Ala Ser Gln Leu Ser Ala
 2195 2200 2205
 Pro Ser Leu Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala
 2210 2215 2220
 Glu Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn
 2225 2230 2235 2240
 Ile Thr Arg Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe
 2245 2250 2255
 Asp Pro Leu Val Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala
 2260 2265 2270
 Glu Ile Leu Arg Lys Ser Arg Arg Phe Ala Arg Ala Leu Pro Val Trp
 2275 2280 2285
 Ala Arg Pro Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro
 2290 2295 2300
 Asp Tyr Glu Pro Pro Val His Gly Cys Pro Leu Pro Pro Pro Arg
 2305 2310 2315 2320
 Ser Pro Pro Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr
 2325 2330 2335
 Glu Ser Thr Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Lys Ser Phe
 2340 2345 2350
 Gly Ser Ser Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser
 2355 2360 2365
 Ser Glu Pro Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Val Glu Ser
 2370 2375 2380
 Tyr Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu
 2385 2390 2395 2400
 Ser Asp Gly Ser Trp Ser Thr Val Ser Ser Gly Ala Asp Thr Glu Asp
 2405 2410 2415
 Val Val Cys Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr
 2420 2425 2430
 Pro Cys Ala Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn
 2435 2440 2445
 Ser Leu Leu Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser
 2450 2455 2460
 Ala Cys Gln Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu
 2465 2470 2475 2480
 Asp Ser His Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ala Ser
 2485 2490 2495
 Lys Val Lys Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr
 2500 2505 2510
 Pro Pro His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val
 2515 2520 2525
 Arg Cys His Ala Arg Lys Ala Val Ala His Ile Asn Ser Val Trp Lys
 2530 2535 2540
 Asp Leu Leu Glu Asp Ser Val Thr Pro Ile Asp Thr Thr Ile Met Ala
 2545 2550 2555 2560
 Lys Asn Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro
 2565 2570 2575
 Ala Arg Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys
 2580 2585 2590
 Met Ala Leu Tyr Asp Val Val Ser Lys Leu Pro Leu Ala Val Met Gly
 2595 2600 2605
 Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu
 2610 2615 2620
 Val Gln Ala Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp
 2625 2630 2635 2640
 Thr Arg Cys Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu
 2645 2650 2655
 Glu Ala Ile Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala
 2660 2665 2670
 Ile Lys Ser Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn

2675 2680 2685
 Ser Arg Gly Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val
 2690 2695 2700
 Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg
 2705 2710 2715 2720
 Ala Ala Cys Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys
 2725 2730 2735
 Gly Asp Asp Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp
 2740 2745 2750
 Ala Ala Ser Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala
 2755 2760 2765
 Pro Pro Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr
 2770 2775 2780
 Ser Cys Ser Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg
 2785 2790 2795 2800
 Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala
 2805 2810 2815
 Trp Glu Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile
 2820 2825 2830
 Ile Met Phe Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His
 2835 2840 2845
 Phe Phe Ser Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asn
 2850 2855 2860
 Cys Glu Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro
 2865 2870 2875 2880
 Pro Ile Ile Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser
 2885 2890 2895
 Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu
 2900 2905 2910
 Gly Val Pro Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg
 2915 2920 2925
 Ala Arg Leu Leu Ser Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr
 2930 2935 2940
 Leu Phe Asn Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala
 2945 2950 2955 2960
 Ala Ala Gly Arg Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser
 2965 2970 2975
 Gly Gly Asp Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Phe
 2980 2985 2990
 Trp Phe Cys Leu Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu
 2995 3000 3005
 Pro Asn Arg
 3010

<210> 15

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<400> 15

gaaggaggga ggtttgaagg a

21

<210> 16

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<400> 16

ccagttccgg gcaagaact

19

<210> 17

<211> 13042

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<400> 17

gccagcccc	tgatgggggc	gacactccac	catgaatcac	tcccctgtga	ggaactactg	60
tcttcacgca	gaaagcgtct	agccatggcg	ttagtatgag	tgtcgtgcag	cctccaggac	120
ccccctccc	gggagagcca	tagtggctcg	cggaaaccgt	gagtacaccg	gaattgccag	180
gacgaccggg	tcctttcttg	gataaaccgg	ctcaatgcct	ggagatttgg	gcgtgcccc	240
gcaagactgc	tagccgagta	gtgttgggtc	gcgaaaggcc	ttgtggtact	gcctgatagg	300
gtgcttgcga	gtgccccggg	aggtctcgta	gaccgtgcac	catgcctgtt	atttctactc	360
aaacaagtcc	tgtacctgcg	cccagaacgc	gcaagaacaa	gcagacgcag	gcttcataatc	420
ctgtgtccat	taaaacatct	gttgaaaggg	gacaacgagc	aaagcgcaaa	gtccagcgcg	480
atgctcggcc	tcgtaattac	aaaattgctg	gtatccatga	tggcttgca	acattggctc	540
aggctgcttt	gccagctcat	ggttggggac	gccaagaccc	tcgccataag	tctcgcaatc	600
ttggaatcct	tctggattac	cctttggggt	ggattggtga	tgttacaact	cacacacctc	660
tagtagggcc	gctggtggca	ggagcggtcg	tgcaccagt	ctgccagata	gtacgcttgc	720
tggaggatgg	agtcaactgg	gctactgggt	ggttcgggtg	ccaccttttt	gtggtatgtc	780
tgctatcttt	ggcctgtccc	tgtagtgggg	cgcgggtcac	tgaccagac	acaaatacca	840
caatcctgac	caattgctgc	cagcgtaatc	aggttatcta	ttgttctcct	tccacttgcc	900
tacacgagcc	tggttgtgtg	atctgtgcgg	acgagtgcctg	ggttcccgcc	aatccgtaca	960
tctcacaccc	ttccaattgg	actggcacgg	actccttctt	ggctgaccac	attgattttg	1020
ttatggggcg	tcttgtgacc	tgtgacgccc	ttgacattgg	tgagttgtgt	ggtgcgtgtg	1080
tattagtccg	tgactggctt	gtcaggcact	ggcttattca	catagacctc	aatgaaactg	1140
gtacttgtaa	cctggaagtg	cccactggaa	tagatcctgg	gttcctaggg	tttatcgggt	1200
ggatggccgg	caaggtcgag	gctgtcatct	tcttgaccaa	actggcttca	caagtaccat	1260
acgctattgc	gactatgttt	agcagtgtac	actacctggc	ggttggcgct	ctgatctact	1320
atgcctctcg	gggcaagtgg	tatcagttgc	tcctagcgct	tatgctttac	atagaagcga	1380
cctctggaaa	ccctatcagg	gtgcccactg	gatgctcaat	agctgagttt	tgtcgcctt	1440
tgatgatacc	atgtccttgc	cactcttatt	tgagtggagaa	tgtgtcagaa	gtcatttgtt	1500
acagtccaaa	gtggaccagg	cctatcactc	tggagtataa	caactccata	tcttgggtacc	1560
cctatacaat	ccctggtgcg	aggggatgta	tggctaaatt	caaaaataac	acatgggggt	1620
gctgccgtat	tcgcaatgtg	ccatcgtact	gcactatggg	caactgatga	gtgtggaactg	1680
acactcgcaa	cacttacgaa	gtatgcggtg	taacaccatg	gctaacaacc	gcatggcaca	1740
acggctcagc	cctgaaattg	gctatattac	aataccctgg	gtctaaagaa	atgtttaaac	1800
ctcataaattg	gatgtcaggc	catttgtatt	ttgagggatc	agataccctc	atagtttact	1860
tttatgaccc	tgtgaattcc	actctcctac	caccggagag	gtgggctagg	ttgcccggta	1920
ccccacctgt	ggtacgtggg	tcttggttac	aggttccgca	agggttttac	agtgatgtga	1980
aagacctagc	cacaggattg	atcaccaaag	acaaagcctg	gaaaaattat	caggtcttat	2040
attcogccac	gggtgctttg	tctcttacgg	gagttaccac	caaggccgtg	gtgctaattc	2100
tgttgggggt	gtgtggcagc	aagtatctta	ttttagccta	cctctgttac	ttgtcccttt	2160
gttttggggc	cgcttctggg	taccctttgc	gtcctgtgct	cccatcccag	tcgtatctcc	2220
aagctggctg	ggatgttttg	tctaaagctc	aagtagctcc	ttttgctttg	attttcttca	2280
tctgttgeta	tctccgctgc	aggctacgtt	atgctgccct	tttaggggtt	gtgcccatgg	2340
ctgcgggctt	gcccctaact	ttctttgttg	cagcagctgc	tgcccaacca	gattatgact	2400
ggtgggtgcg	actgctagtg	gcagggttag	ttttgtgggc	cggccgtaac	cgtggtcacc	2460
gcatagtctt	gcttgtagggt	ccttggcctc	tggtagcgct	tttaaccctc	ttgcatttgg	2520
ttacgcctgc	ttcagctttt	gataccgaga	taattggagg	gctgacaata	ccacctgtag	2580
tagcattagt	tgtcatgtct	cgttttggct	tctttgctca	cctgtttacct	cgctgtgctt	2640

tagttaactc	ctatccttgg	caacgttggg	agaattgggt	ttggaacggt	acactaagac	2700
cggagaggtt	tttcccttgg	ctgggttgggt	tccccgggtgc	gacatatgac	gcgctgggtga	2760
ctttctgtgt	gtgtcacgta	gctcttctat	gtttaacatc	cagtgcagca	tcgttctttg	2820
ggactgactc	tagggttagg	gcccatagaa	tgttgggtgcg	tctcggaag	tgctcatgctt	2880
ggtattctca	ttatgttctt	aagtttttcc	tcttagtggt	tggtgagaat	gggtgtgtttt	2940
tctataagca	cttgcattgg	gatgtcttgc	ctaattgattt	tgcctcgaaa	ctaccattgc	3000
aagagccatt	tttccctttt	gaaggcaagg	caagggtcta	taggaatgaa	ggaagacgct	3060
tggcgtgtgg	ggacacggtt	gatggtttgc	ccgttgttgc	gcgtctcggc	gaccttgttt	3120
tcgcagggtt	ggctatgcgg	ccagatgggt	gggccattac	cgcacctttt	acgctgcagt	3180
gtctctctga	acgtggcacg	ctgtcagcga	tggcagtgggt	catgactgggt	atagaccccc	3240
gaacttggac	tggaaactatc	ttcagattag	gatctctggc	cactagctac	atgggattcg	3300
tttgtgacaa	cgtgttgtat	actgctcacc	atggcagcaa	ggggcgccgg	ttgggtcatc	3360
ccacaggctc	catacaccca	ataaccgttg	acggcgctaa	tgaccaggac	atctatcaac	3420
caccatgtgg	agctgggtcc	cttactcgggt	gctcttgcgg	ggagaccaag	gggtatcttg	3480
taacacgact	ggggtcattg	gttgagggtca	acaaatccga	tgaccttat	tggtgtgtgt	3540
gcggggccct	tcccatgggt	gttgccaagg	gttcttcagg	tgccccgatt	ctgtgtctct	3600
ccgggcatgt	tattgggatg	ttcaccgctg	ctagaaattc	tggcggttca	gtcagtcaga	3660
ttagggttag	gocgttgggtg	tgtgtctggat	accatcccca	gtacacagca	catgccactc	3720
ttgatacaaa	aacgcgtatg	cagatcttctg	tgaagacct	gacgggcaag	accatcactc	3780
ttgaggtcga	gcccagtgac	accatcgaga	atgtcaaggc	caagatccaa	gacaaggaag	3840
gcatccacc	tgaccagcag	aggctgatat	tcgcgggcaa	acagctggag	gatggccgca	3900
ccctgtccga	ctacaacatc	cagaaagagt	ccaccttgca	cctgggtgctg	cgtctccgcg	3960
gtggcgcgcc	catcacggcg	tacgcccagc	agacgagagg	cctcctagggt	tgtataatca	4020
ccagcctgac	tggccgggac	aaaaaccaag	tggaggggtga	ggtccagatc	gtgtcaactg	4080
ctacccaaac	cttccctggca	acgtgcatca	atgggggtatg	ctggactgtc	taccacgggg	4140
ccggaacgag	gaccatcgca	tcacccaagg	gtcctgtcat	ccagatgtat	accaatgtgg	4200
accaagacct	tgtgggctgg	cccgtctctc	aagggtcccg	ctcattgaca	ccctgtacct	4260
gcggctcctc	ggacctttac	ctgggtcacga	ggcacgcgca	tgtcattccc	gtgcgcgggc	4320
gaggtgatag	caggggtagc	ctgctttcgc	cccggcccat	ttcctacttg	aaaggctcct	4380
cggggggtcc	gctgttgtgc	cccgcgggac	acgcgctggg	cctattcagg	gccgcgggtgt	4440
gcaccctggg	agtggctaaa	gcgggtggact	ttatccctgt	ggagaacctc	gggacaacca	4500
tgagatcccc	ggtgttcacg	gacaactcct	ctccaccagc	agtgcgccag	agcttccagg	4560
tggcccacct	gcatgctccc	accggcagcg	ctaaagacac	caagggtccc	gctgcgtacg	4620
cagcccaggg	ctacaagggtg	ttgggtgctca	acccctctgt	tgtgcgaacg	ctgggctttg	4680
gtgcttacat	gtccaaggcc	catgggggttg	atcctaatat	caggacgggg	gtgagaacaa	4740
ttaccactgg	cagccccatc	acgtactcca	cctacggcaa	gttcccttgc	gacggcgggt	4800
gctcaggagg	tgcttatgac	ataataattt	gtgacgagtg	ccactccacg	gatgccacat	4860
ccatcttggg	catcggcact	gtccttgacc	aagcagagac	tgcgggggcg	agactgggtg	4920
tgctcgccac	tgctacccct	ccgggtcccg	tcactgtgtc	ccatccctaac	atcgaggagg	4980
ttgctctgtc	caccaccgga	gagatccct	tttacggcaa	ggctatcccc	ctcgagggtga	5040
tcaagggggg	aagacatctc	atcttctgcc	actcaaagaa	gaagtgcgac	gagctcgccg	5100
cgaagctggt	cgcattgggc	atcaatgccg	tggcctaacta	ccgcggtctt	gacgtgtctg	5160
tcaccccgac	cagcggcgat	gttgtcgtcg	tgtcgaccga	tgctctcatg	actggcttta	5220
ccggcgactt	cgactctgtg	atagactgca	acacgtgtgt	cactcagaca	gtcgatttca	5280
gccttgacct	tacctttacc	attgagacaa	ccacgctccc	ccaggatgct	gtctccagga	5340
ctcaacgcgg	gggcaggact	ggcaggggat	agccaggcat	ctatagattt	gtgcgcggcg	5400
gggagcgccc	ctccggcactg	ttcgactcgt	cgctcctctg	tgagtgtat	gacgacgggt	5460
gtgcttggta	tgagctcacg	cccgcgcgaga	ctacagttag	gctacgagcg	tacatgaaca	5520
ccccggggct	tcccggtgtg	caggaccatc	ttgaattttg	ggagggcgtc	tttacggggc	5580
tactcatat	agatgccac	tttttatccc	agacaaagca	gagtggggag	aactttcctt	5640
acctggtagc	gtaccaagcc	accgtgtgcg	ctagggtctca	agccccctcc	ccatcggtgg	5700
accagatgtg	gaagtgtttg	atccgcctta	aaccaccct	ccatggggca	acacccttgc	5760
tatacagact	gggcgctgtt	cagaatgaag	tcacctgac	gcacccaatc	accaaataca	5820
tcatgacatg	catgtcggcc	gacctggagg	tgtcacagag	cacctgggtg	ctcgttggcg	5880
gcgtcctggc	tgtctctggc	gcgtattgcc	tgtcaacagg	ctgcgtggtc	atagtgggca	5940
ggatcgtctt	gtccgggaag	ccggcaatta	tacctgacag	ggaggttctc	taccaggagt	6000
tcgatgagat	ggaagagtgc	tctcagcact	taccgtacat	cgagcaaggg	atgatgctcg	6060
ctgagcagtt	caagcagaag	gccctcggcc	tcctgcagac	cgcgtcccgc	catgcagagg	6120
ttatcacccc	tgtgttccag	accaactggc	agaaactcga	ggtctttttg	gcgaagcaca	6180
tgtggaattt	catcagtggg	atacaatact	tggcgggcct	gtcaacgctg	cctggtaacc	6240
ccgccattgc	ttcattgatg	gctttttacag	ctgcgtcac	cagccacta	accactggcc	6300
aaacctcct	cttcaacata	ttgggggggt	gggtggctgc	ccagctcgcc	gccccgggtg	6360
ccgctactgc	ctttgtgggt	gctggccctag	ctggcgccgc	catcggcagc	gttggaactg	6420

ggaaggtcct	cgtggacatt	ottgcagggg	atggcgcggg	cgtggcgggg	gctctttag	6480
cattcaagat	catgagcggg	gaggtccctt	ccacggagga	cctgggtcaat	ctgctgccc	6540
ccatcctctc	gcctggagcc	cttgtagtgc	gtgtggtctg	cgcagcaata	ctgcccggc	6600
acgttggccc	gggagggggg	gcagtgaact	ggatgaaccg	gctaataagg	ttcgccctcc	6660
gggggaacca	tgtttccccc	acgcactacg	tgcggagag	cgatgcagcc	gcccgcgtca	6720
ctgccatact	cagcagccct	actgtaacct	agctcctgag	gcgactgcat	cagtggataa	6780
gctcggagtg	taccactcca	tgctccgggt	cctgggctaag	ggacatctgg	gactggatat	6840
gcgaggtgct	gagcgacttt	aagacctggc	tgaagcccaa	gctcatgcc	caactgcctg	6900
ggattccctt	tgtgtcctgc	cagcgcggtt	ataggggggt	ctggcgagga	gacggcatta	6960
tgcacactcg	ctgccactgt	ggagctgaga	tactgggaca	tgtcaaaaac	gggacgatga	7020
ggatcgtcgg	tcctaggacc	tgcaggaaca	tgtggagtgg	gacgttcccc	attaacgcct	7080
acaccacggg	cccctgtact	ccccttcctg	cgcgaacta	taagttcgcg	ctgtggagg	7140
tgtctgcaga	ggaatacgtg	gagataaggc	gggtggggga	cttccactac	gtatcgggt	7200
tgactactga	caatcttaaa	tgcccgtgcc	agatcccatc	gcccgaattt	ttcacagaat	7260
tggaacgggt	gcgcctacac	aggtttgcgc	ccccttgcaa	gcccttgctg	cgggaggagg	7320
tatcattcag	agtaggactc	cacgagtaac	cgggtgggtc	gcaattacct	tgcgagccc	7380
aaccggacgt	agccgtgttg	acgtccatgc	tactgatcc	ctcccatata	acagcagagg	7440
cggcggggag	aaggttggcg	agaggggtac	ccccttctat	ggccagctcc	tcggctagcc	7500
agctgtccgc	tccatctctc	aaggcaactt	gcaccgccaa	ccatgactcc	cctgacggcg	7560
agctcataga	gcctaacctc	ctgtggaggc	aggagatggg	cggcaacatc	accagggttg	7620
agtcagagaa	caaagtgggtg	attctggact	ccttcgatcc	gcttgtggca	gaggaggatg	7680
agcgggagg	ctccgtacct	gcagaaatcc	tcgggaagtc	tcggagattc	gcccggggcc	7740
tgcccgtctg	ggcgcgcccg	gactacaacc	ccccgctagt	agagacgtgg	aaaaagcctg	7800
actacgaacc	acctgtggtc	catggctgcc	cgtaccacc	tcacgggtcc	cctcctgtgc	7860
ctccgcctcg	gaaaaagcgt	acgggtggtc	tcaccgaatc	aaccctatct	actgccttgg	7920
ccgagcttgc	cacaaaaggt	tttggcagct	cctcaacttc	cggcattacg	ggcgacaata	7980
cgacaacatc	ctctgagccc	gccccttctg	gctgcccccc	cgactccgac	gttgagtcct	8040
attcttccat	gccccccctg	gagggggagc	ctggggatcc	ggatctcagc	gacgggtcat	8100
ggtcgacgg	cagtagtggg	gcccacacgg	aagatgtcgt	gtgctgctca	atgtcttatt	8160
cctggacagg	cgcactcgtc	accccgctgc	ctggggaaga	acaaaaactg	cccatcaacg	8220
cactgagcaa	ctcgttgcta	cgccatcaca	atctggtgta	ttccaccact	tcacgcagt	8280
cttgccaaag	gcagaagaaa	gtcacatttg	acagactgca	agttctggac	agccattacc	8340
aggacgtgct	caaggaggtc	aaagcagcgg	cgtcaaaagt	gaaggctaac	ttgctattcc	8400
tagaggaagc	ttgcagcctg	acgccccac	attcagccaa	atccaagttt	ggctatgggg	8460
caaaagacgt	ccgttgccat	gccagaaagg	ccgtagccca	catcaactcc	gtgtggaaag	8520
accttctgga	agacagtgt	acaccaatag	acactaccat	catggccaag	aacgaggttt	8580
tctgogttca	gcctgagaag	gggggtcgta	agccagctcg	tctcatcgct	ttccccgacc	8640
tgggcgtgcg	cgtgtgcgag	aagatggccc	tgtacgacgt	ggttagcaag	ctccccctgg	8700
ccgtgatggg	aagctcctac	ggattccaat	actcaccagg	acagcgggtt	gaattcctcg	8760
tgcaagcgtg	gaagtccaag	aagaccccg	tggggttctc	gtatgatacc	cgctgttttg	8820
actccacagt	cactgagagc	gacatccgta	cggaggaggc	aattttacaa	tggtgtgacc	8880
tggaaccccc	agcccgctg	gccatcaagt	ccctcactga	gaggctttat	gttggggggc	8940
ctcttaccaa	ttcaaggggg	gaaaactgcg	gctaccgcag	gtgccgcgcg	agcggcgtag	9000
tgacaactag	ctgtggtaac	accctcactt	gctacatcaa	ggcccgggca	gcctgtcgag	9060
ccgcagggct	ccaggactgc	accatgctcg	tgtgtggcga	cgacttagtc	gttatctgtg	9120
aaagtgcggg	ggtcacggag	gacgcggcga	gacctgagag	cctcacggag	gctatgacca	9180
ggtactccgc	ccccccggg	gacccccac	acacagaata	cgacttgagg	cttataacat	9240
catgctctct	caacgtgtca	gtcgccacg	acggcgctgg	aaagagggtc	tactacctta	9300
cccgtgacct	tacaaccccc	ctcgcgagag	ccgcgtggga	gacagcaaga	cacactccag	9360
tcaattcctg	gctaggcaac	ataatcatgt	ttgccccac	actgtgggcg	aggatgatac	9420
tgatgaccca	tttcttttag	gtcctcatag	ccagggatca	gcttgaacag	gctcttaact	9480
gtgagatcta	cggagcctgc	tactccatag	aaccactgga	tctacctcca	atcattcaaa	9540
gactccatgg	cctcagcgca	ttttcactcc	acagttactc	tccagggtga	atcaatagg	9600
tggcgcgatg	cctcagaaaa	cttgggggtc	cgccttgcg	agcttggaga	caccggggcc	9660
ggagcgtccg	cgtatggctt	ctgtccagag	gaggcagggc	tgccatatgt	ggcaagtacc	9720
tcttcaactg	ggcagtaaga	acaaagctca	aactcactcc	aatagcggcc	gctggccggc	9780
tggaacttgc	cggttgggtc	acggctggct	acagcggggg	agacatttat	cacagcgtgt	9840
ctcatgccc	gccccgctgg	ttctggtttt	gcctactcct	gctcgtgca	ggggtaggca	9900
tctacctcct	ccccaaccca	tgaaggtttg	ggtaaactct	ccggcctctt	aagccatttc	9960
ctgttttttt	tttttttttt	tttttttttt	tctttttttt	tttctttcct	ttccttcttt	10020
ttttcctttt	tttttccctt	ctttaatggt	ggctccatct	tagccctagt	cacggctagc	10080
tgtgaaaggt	cogtgagccg	catgactgca	gagagtgtct	atactggcct	ctctgcagat	10140
catgttctag	actggaattc	gtcgacgagc	tccctatagt	gagtcgtatt	agaggccgac	10200

```

ttggccaaat  tcgtaatcat  ggtcatagct  gtttcctgtg  tgaaattgtt  atccgctcac  10260
aattccacac  aacatacgag  ccggaagcat  aaagtgtaaa  gcctggggtg  cctaatgagt  10320
gagctaaact  acattaattg  cgttgcgctc  actgcccgtc  ttccagtcgg  gaaacctgtc  10380
gtgccagctg  cattaatgaa  tcggccaacg  cgcggggaga  ggcggtttgc  gtattgggcg  10440
ctcttcogct  tcctcgctca  ctgactcgct  gcgctcggtc  gttcggtgc  ggcgagcggt  10500
atcagctcac  tcaaaggcgg  taatacgggt  atccacagaa  tcaggggata  acgcaggaaa  10560
gaacatgtga  gcaaaaggcc  agcaaaaggc  caggaaccgt  aaaaaggccg  cgttgctggc  10620
gtttttccat  aggctccgcc  ccctgacga  gcatcacaaa  aatcgacgct  caagtcagag  10680
gtggcgaaac  ccgacaggac  tataaagata  ccaggcggtt  cccctggaa  gctccctcgt  10740
gcgctctcct  gttccgacct  tgccgcttac  cggatacctg  tccgcctttc  tcccttcggg  10800
aagcgtggcg  ctttctcata  gctcacgctg  taggtatctc  agttcggtgt  aggtcgttcg  10860
ctccaagctg  ggctgtgtgc  acgaaccccc  cgttcagccc  gaccgctgcg  ccttatccgg  10920
taactatcgt  cttgagtcca  acccggttaag  acacgactta  tcgccactgg  cagcagccac  10980
tggtaacagg  attagcagag  cgaggtaggt  aggcggtgct  acagagttct  tgaagtgggtg  11040
gcctaactac  ggctacacta  gaaggacagt  atttggtatc  tgcgctctgc  tgaagccagt  11100
taccttcgga  aaaagagttg  gtagctcttg  atccggcaaa  caaacaccg  ctggtagcgg  11160
tggttttttt  gtttgcaagc  agcagattac  gcgcagaaaa  aaaggatctc  aagaagatcc  11220
tttgatcttt  tctacggggt  ctgacgctca  gtggaacgaa  aactcacgtt  aagggatttt  11280
ggtcatgaga  ttatcaaaaa  ggatcttcac  ctagatcctt  ttaaatttaa  aatgaagttt  11340
taaatcaatc  taaagtatat  atgagtaaac  ttggtctgac  agttaccaat  gcttaatacag  11400
tgaggcacct  atctcagcga  tctgtctatt  tcgttcattc  atagttgcct  gactccccgt  11460
cgtgtagata  actacgatac  gggagggtct  accatctggc  cccagtgtgc  caatgatacc  11520
gcgagaccca  cgctcaccgg  ctccagattt  atcagcaata  aaccagccag  ccggaagggc  11580
cgagcgcaga  agtggctctg  caactttatc  cgcctccatc  cagtctatta  attgttgccg  11640
ggaagctaga  gtaagtagtt  cgccagttaa  tagtttgctc  aacgttggtg  ccattgctac  11700
aggcatcgtg  gtgtcacgct  cgtcgtttgg  tatggcttca  ttcagctccg  gttcccaacg  11760
atcaaggcga  gttacatgat  ccccatgtt  gtgcaaaaaa  gcggttagct  ccttcgggtc  11820
tccgatcgtt  gtcagaagta  agttggccgc  agtggtatca  ctcatgggta  tggcagcact  11880
gcataattct  cttactgtca  tgccatccgt  aagatgcttt  tctgtgactg  gtgagtactc  11940
aaccaagtca  ttctgagaat  agtgtatgcg  gcgaccgagt  tgctcttgcc  cggcgtcaat  12000
acgggataat  accgcgccac  atagcagaac  tttaaaagtg  ctcatcattg  gaaaacgttc  12060
ttcggggcga  aaactctcaa  ggatcttacc  gctgttgaga  tccagttcga  tgtaaccac  12120
tcgtgcaccc  aactgatctt  cagcatcttt  tactttcacc  agcgtttctg  ggtgagcaaa  12180
aacaggaagg  caaaatgccg  caaaaaaggg  aataaggggc  acacggaaat  gttgaatact  12240
catactcttc  ctttttcaat  attattgaag  catttatcag  ggttattgtc  tcatgagcgg  12300
atacatattt  gaatgtattt  agaaaaataa  acaaataggg  gttccgcgca  catttccccg  12360
aaaagtgcc  cctgacgcgc  cctgtagcgg  cgcattaagc  gcggcgggtg  tgggtggttac  12420
gcgcagcgtg  accgctacac  ttgccagcgc  cctagcgcgc  gctcctttcg  ctttcttccc  12480
ttcctttctc  gccacgttcg  ccggctttcc  ccgtcaagct  ctaaatacgg  gcatcccttt  12540
agggttccga  tttagtgtct  tacggcacct  cgaccccaaa  aaacttgatt  agggtagtgg  12600
ttcacgtagt  gggccatcgc  cctgatagac  ggtttttcgc  cctttgacgt  tggagtccac  12660
gttctttaat  agtggactct  tgttccaaac  tggacaacaa  ctcaacccta  tctcggtcta  12720
ttcttttgat  ttataaggga  ttttgccgat  ttcggcctat  tggttaaaaa  atgagctgat  12780
ttaacaaaaa  tttaacgcga  attttaacaa  aatattaaca  aaatattaac  gtttacaatt  12840
tcccatcgc  cattcaggct  gcgcaactgt  tgggaagggc  gatcgggtgc  ggcctcttcg  12900
ctattacgcc  agctggcgaa  agggggatgt  gctgcaaggc  gattaagttg  ggtaacgcca  12960
gggttttccc  agtcacgacg  ttgtaaaacg  acggccagtg  ccaagctgac  ttggtcagcg  13020
gocgctaata  cgactacta  ta  13042

```

<210> 18

<211> 13147

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<400> 18

```

gccagccccc  tgatgggggc  gacactccac  catgaatcac  tcccctgtga  ggaactactg  60
tcttcacgca  gaaagcgtct  agccatggcg  ttagtatgag  tgtcgtgcag  cctccaggac  120
ccccctccc  gggagagcca  tagtggtctg  cggaaaccgg  gagtacaccg  gaattgccag  180

```

gacgaccggg	tcctttcttg	gataaaccgg	ctcaatgcct	ggagatttgg	gcgtgcccc	240
gcaagactgc	tagccgagta	gtgttgggtc	gcgaaaggcc	ttgtgggtact	gcctgatagg	300
gtgcttgoga	gtgccccggg	aggtctcgta	gaccgtgcac	catgagcacg	aatcctaaac	360
ctcaaagaaa	aaccaaactg	aacaccaacc	gtcgcccaca	ggacgtcaag	ttccccgggtg	420
gcggtcagat	cgttgggtgga	gtttacttgt	tgccgcgcag	gggccctaga	ttgggtgtgc	480
gcgcgacgag	gaagacttcc	gagcggtcgc	aacctcgagg	tagacgtcag	cctatcccca	540
aggcacgtcg	gcccgaaggc	aggacctggg	ctcagcccgg	gtacccttgg	cccctctatg	600
gcaatgaggg	ttgcgggtgg	gcgggatggc	tcctgtctcc	ccgtggctct	cggcctagct	660
ggggcccccac	agacccccgg	cgtaggtcgc	gcaatttggg	taaggctcatc	gataccctta	720
cgtgcggctt	cgcgcacctc	atggggtaca	taccgctcgt	cggcgcccct	cttgaggcg	780
ctgccagggc	cctgggcgat	ggcgtccggg	ttctggaaga	cggcgtgaac	tatgcaacag	840
ggaaaccttcc	tggttgctct	ttctctatct	tccttctggc	cctgctctct	tgcctgactg	900
tgccgccttc	agccgcgcgg	gtcactgacc	cagacacaaa	taccacaatc	ctgaccaatt	960
gctgccagcg	taatcagggt	atctattgtt	ctccttccac	ttgcctacac	gagcctgggt	1020
gtgtgatctg	tgccgacgag	tgcctgggtc	ccgccaatcc	gtacatctca	cacccttcca	1080
attggactgg	cacggactcc	ttcttggctg	accacattga	ttttgttatg	ggcgtctctg	1140
tgacctgtga	cgccttggac	attgggtgag	tgtgtgggtc	gtgtgtatta	gtcgggtgact	1200
ggcttgtcag	gcactggctt	attcacatag	acctcaatga	aactgggtact	tgttacctgg	1260
aagtggccac	tggaaatagat	cctgggttcc	tagggtttat	cgggtggatg	gccggcaagg	1320
tcgaggctgt	catcttcttg	accaaactgg	cttcacaagt	accatacgct	attgcgacta	1380
tgtttagcag	tgtacactac	ctggcggttg	gcgctctgat	ctactatgcc	tctcggggca	1440
agtggatatca	gttgctccta	gcgcttatgc	tttacctaga	agcgacctct	ggaaacccta	1500
tcagggtgcc	cactggatgc	tcaatagctg	agttttgtct	gcctttgatg	ataccatgtc	1560
cttgccactc	ttattttagt	gagaatgtgt	cagaagtcat	ttgttacagt	ccaaagtggg	1620
ccaggcctat	cactctagag	tataacaact	ccatatcttg	gtacccctat	acaatccctg	1680
gtgcgagggg	atgtatgggt	aaattcaaaa	ataacacatg	gggttgctgc	cgtattcgca	1740
atgtgccatc	gtactgcact	atgggcactg	atgcagtgtg	gaacgacact	cgcaacactt	1800
acgaagtatg	cgggtgtaaca	ccatggctaa	caaccgcatg	gcacaacggc	tcagccctga	1860
aattggctat	attacaatac	cctgggtcta	aagaaatgtt	taaacctcat	aattggatgt	1920
caggccattt	gtatttttag	ggatcagata	cccctatagt	ttacttttat	gaccctgtga	1980
attccactct	cctaccacgg	gagagggtgg	ctagggttgc	cggtagccca	cctgtgggtac	2040
gtggttcttg	gttacagggt	ccgcaagggt	tttacagtga	tgtgaaagac	ctagccacag	2100
gattgatcac	caaagacaaa	gcctggaaaa	attatcaggt	cttatattcc	gccacgggtg	2160
ctttgtctct	tacgggaggt	accaccaagg	ccgtggtgct	aattctgttg	gggttgtgtg	2220
gcagcaagta	gttattttta	gcctacctct	gttacttgtc	cctttgtttt	gggcgcgctg	2280
ctggttaccc	tttgcgtcct	gtgctcccat	cccagtcgta	tctccaagct	ggctgggatg	2340
ttttgtctaa	agctcaagta	gtcctttttg	ccttgatttt	cttcatctgt	tgctatctcc	2400
gctgcaggct	acgttatgtc	gccccttttag	ggtttgtgcc	catggctgcg	ggcttgcccc	2460
taactttctt	tgttgcagca	gctgctgccc	aaccagatta	tgactgggtg	gtgcgactgc	2520
tagtggcagg	gttagttttg	tgggcccggc	gtaaccgtgg	tcaccgcata	gctctgcttg	2580
taggtccttg	gcctctggta	gcgcttttaa	ccctcttgca	tttggttacg	cctgcttcag	2640
cttttgatac	cgagataaatt	ggagggtgta	caataccacc	tgtagtagca	ttagttgtca	2700
tgtctcgttt	tggcttcttt	gctcacttgt	tacctcgctg	tgcttttagt	aactcctatc	2760
tttggcaacg	ttggggagaat	tggtttttga	acgttacact	aagaccggag	aggtttttcc	2820
ttgtgctgg	ttgtttcccc	ggtgcgacat	atgacgcgct	ggtgactttc	tgtgtgtgtc	2880
acgtagctct	tctatgttta	acatccagtg	cagcatcggt	ctttgggact	gactctaggg	2940
ttagggccca	tagaatgttg	gtgcgtctcg	gaaagtgtca	tgcttgggtat	tctcattatg	3000
ttcttaagtt	tttctcttta	gtgtttgggt	agaattgggt	gtttttctat	aagcacttgc	3060
atgggtgatgt	cctgcctaatt	gattttgcct	cgaaactacc	attgcaagag	ccatttttcc	3120
cttttgaagg	caaggcaagg	gtctatagga	atgaagggaag	acgcttggcg	tgtggggaca	3180
cggttgatgg	tttgcocgtt	gttgcgcgtc	tgcgcgacct	tgttttcgca	gggttggcta	3240
tgccgccaga	tgggtggggc	attaccgcac	cttttacgct	gcagtgtctc	tctgaacgtg	3300
gcacgctgtc	agcgatggca	gtggtcatga	ctggatataga	ccccgaact	tggactggaa	3360
ctatcttcag	attaggatct	ctggccacta	gctacatggg	attcgtttgt	gacaacgtgt	3420
tgtatactgc	tcaccatggc	agcaaggggc	gccggttggc	tcatcccaca	ggctccatac	3480
acccaataac	cgttgacgog	gctaatagacc	aggacatcta	tcaaccacca	tgtggagctg	3540
ggtcccttac	tccgtgctct	tgcggggaga	ccaaggggta	tctggttaaca	cgactgggg	3600
cattggttga	ggtcaacaaa	tcogatgaac	cttattgggtg	tgtgtgcggg	gcccttccca	3660
tggctggttc	caagggttct	tcagggtgcc	cgattctgtg	ctcctccggg	catgttattg	3720
ggatgttcac	cgtgctaga	aattctggcg	gttcagtcag	tcagattagg	gttaggccgt	3780
tgggtgtgtc	tggataccat	ccccagtaca	cagcacatgc	cactcttgat	acaaaaacgc	3840
gtatgcagat	cttcgtgaag	acctgacggg	gcaagaccat	cactcttgag	gtcgaagcca	3900
gtgacaccat	cgagaatgtc	aaggccaaga	tccaagacaa	ggaaggcatc	ccacctgacc	3960

agcagaggct	gatattcgcg	ggcaaacagc	tggaggatgg	ccgcaccctg	tccgactaca	4020
acatccagaa	agagtcaccc	ttgcacctgg	tgctgcgtct	ccgcggtggc	gcgcccata	4080
cggcgtagcg	ccagcagacg	agaggcctcc	taggggtgat	aatcaccagc	ctgactggcc	4140
gggacaaaaa	ccaagtggag	ggtgaggtcc	agatcgtgtc	aactgctacc	caaacccttc	4200
tggcaacgtg	catcaatggg	gtatgctgga	ctgtctacca	cggggccgga	acgaggacca	4260
tcgcatcacc	caagggctcc	gtcatccaga	tgtataccaa	tgtggaccaa	gacctgtgtg	4320
gctggccccg	tcctcaaggt	ccccgctcat	tgacaccctg	tacctgcggc	tcctcggacc	4380
tttacctggg	cacgaggcac	gccgatgtca	ttcccgtgcg	ccggcgaggt	gatagcaggg	4440
gtagcctgct	ttcgccccgg	cccatttcct	acttgaaagg	ctcctcgggg	ggtccgctgt	4500
tgtgccccgc	gggacacgcc	gtgggcctat	tcagggccgc	ggtgtgcacc	cgtggagtgg	4560
ctaaagcggg	ggactttatc	cctgtggaga	acctagggac	aacctatgaga	tcctccggtg	4620
tcacggacaa	ctcctctcca	ccagcagtg	cccagagctt	ccaggtggcc	cacctgcatg	4680
ctcccacccg	cagcggtaag	agcaccaagg	tcctcggtgc	gtacgcagcc	cagggctaca	4740
aggtgtttgg	gtcacaaccc	tctgttgctg	caacgctggg	ctttgggtgct	tacatgtcca	4800
aggcccatgg	ggttgatcct	aatatcagga	ccgggggtgag	aacaattacc	actggcagcc	4860
ccatcacgta	ctccacctac	ggcaagttcc	ttgccgacgg	cgggtgctca	ggaggtgctt	4920
atgacataat	aattttgtgac	gagtgcact	ccacggatgc	cacatccatc	ttgggcatcg	4980
gcactgtcct	tgaccaagca	gagactgcgg	ggccgagact	ggttggtgctc	gccactgcta	5040
ccccctcggg	ctcgcgtact	gtgtcccatc	ctaaccatga	ggaggttgct	ctgtccacca	5100
ccggagagat	cccccttttac	ggcaaggcta	tcctccctga	ggtgatcaag	gggggaagac	5160
atctcatctt	ctgccactca	aagaagaagt	gcgacgagct	cgcgcgaag	ctggtcgcac	5220
tgggcatcaa	tgccgtggcc	tactaccgcg	gtcttgacgt	gtctgtcatc	ccgaccagcg	5280
gcgatgttgt	cgctcgtgtc	accgatgctc	tcatgactgg	ctttaccggc	gacttcgact	5340
ctgtgataga	ctgcaacacg	tgtgtcactc	agacagtoga	tttcagcctt	gacctacct	5400
ttaccattga	gacaaccacg	ctccccaggg	atgctgtctc	caggactcaa	cgccggggca	5460
ggactggcag	ggggaagcca	ggcatctata	gatttggtgg	accgggggag	cgccccctcg	5520
gcatgttcga	ctcgtccgtc	ctctgtgagt	gctatgaocg	gggctgtgct	tggtatgagc	5580
tcacgccccg	cgagactaca	gttaggctac	gagcgtacat	gaacaccccc	gggcttcccc	5640
tgtgccagga	ccatcttgaa	ttttgggagg	gcgtctttac	gggcctcact	catatagatg	5700
cccacttttt	atcccagaca	aagcagagtg	gggagaactt	tccttacctg	gtagcgtacc	5760
aagccaccgt	gtgcgctagg	gctcaagccc	ctcccccatc	gtgggaocag	atgtggaagt	5820
gtttgatccg	ccttaaacc	accctccatg	ggccaacacc	cctgctatac	agactgggcg	5880
ctgttcagaa	tgaagtcacc	ctgacgcacc	caatcaccaa	atacatcatg	acatgcatgt	5940
cggccgacct	ggaggtcgtc	acgagcact	gggtgctcgt	tggcgcgctc	ctggctgctc	6000
tggccgcgta	ttgcctgtca	acaggctcgc	tggtcatagt	gggcaggatc	gtcttgctcc	6060
ggaagccggc	aattatacct	gacagggagg	ttctctacca	ggagtctgat	gagatggaag	6120
agtgtctctc	gcacttaccc	tacatcgagc	aagggtatgat	gctcgtcgag	cagttcaagc	6180
agaaggccct	cggcctcctg	cagaccgcgt	cccgccatgc	agaggttatc	acccctgctg	6240
tccagaccaa	ctggcagaaa	ctcgaggtct	tttgggcgaa	gcacatgtgg	aatttcatca	6300
gtgggataca	atacttggcg	ggcctgtcaa	cgctgcctgg	taacccccgc	attgtctcat	6360
tgatggcttt	tacagctgcc	gtcaccagcc	cactaaccac	tggccaaacc	ctcctcttca	6420
acatattggg	ggggtgggtg	gctgccacgc	tcgccgcccc	cgggtgccgt	actgcctttg	6480
tgggtgctgg	cctagctggc	gccgccatcg	gcagcgttgg	actggggaag	gtcctcgtgg	6540
acattctttg	agggtatggc	gcgggcgtgg	cgggagctct	tgtagcattc	aagatcatga	6600
gcggtgaggt	ccccctccac	gaggacctgg	tcaatctgct	gccccccatc	ctctcgctcg	6660
gagcccttgt	agtcggtgtg	gtctgcgcag	caatactgcg	ccggcacgtt	ggccccggcg	6720
agggggcag	gcaatggatg	aaccggctaa	tagccttcgc	ctccccgggg	aacctgtttt	6780
ccccacgcga	ctacgtgccg	gagagcgatg	cagccgcccc	cgtcactgcc	atactcagca	6840
gcctcactgt	aaccagctc	ctgaggcgac	tgcatcagtg	gataagctcg	gagtgtacca	6900
ctccatgctc	cggttcctgg	ctaagggaca	tctgggactg	gatatgcgag	gtgctgagcg	6960
actttaagac	ctggctgaaa	gccaaagctc	tgccacaact	gcctgggatt	ccctttgtgt	7020
cctgccagcg	cgggtatagg	ggggtctggc	gaggagacgg	cattatgcac	actcgtgcc	7080
actgtggagc	tgagatcact	ggacatgtca	aaaacgggac	gatgaggatc	gtcgtgctca	7140
ggacctgcag	gaacatgtgg	agtgggacgt	tcctcattaa	cgcctacacc	acgggccccct	7200
gtactcccc	tcctgcgcgc	aactataagt	tcgcgctgtg	gaggggtgtc	gcagaggaat	7260
acgtggagat	aaggcgggtg	ggggacttcc	actacgtatc	gggtatgact	actgacaatc	7320
ttaaatgcc	gtgccagatc	ccatcgcccc	aattttttcac	agaattggac	ggggtgcgcc	7380
tacacagggt	tgcgccccct	tgcaagccct	tgctgcggga	ggaggtatca	ttcagagtag	7440
gactccacga	gtacccgggtg	gggtcgcaat	taccttgcca	gcccgaaccg	gacgtagccg	7500
tgttgacgtc	catgctcact	gatccctccc	atataacagc	agaggcgcc	gggagaaggt	7560
tggcgagagg	gtcaccocct	tctatggcca	gctcctcgga	tagccagctg	tcctctccat	7620
ctctcaaggc	aacttgacc	gccaaacctg	actccctga	cgcgcagctc	atagaggcta	7680
acctcctgtg	gaggcaggag	atgggcggca	acatcaccag	ggttgagtca	gagaacaaag	7740

tgggtgattct	ggactccttc	gatccgcttg	tggcagagga	ggatgagcgg	gaggtctccg	7800
tacctgcaga	aattctgcgg	aagtctcgga	gattcgcccg	ggccctgccc	gtctggggcg	7860
ggccggacta	caaccccccg	ctagtagaga	cgtggaaaaa	gcctgactac	gaaccacctg	7920
tgggtccatgg	ctgcccgcga	ccacctccac	ggtccctccc	tgtgcctccg	cctcggaaaa	7980
agcgtacggt	ggtcctcacc	gaatcaaccc	tatctactgc	cttgcccgag	cttgccacca	8040
aaagttttgg	cagctcctca	acttccggca	ttacggggcg	caatacgaca	acatcctctg	8100
agcccccccc	ttctggtctc	cccccgact	ccgacgttga	gtcctattct	tccatgcccc	8160
ccctggaggg	ggagcctggg	gatccggatc	tcagcgacgg	gtcatggctg	acggtcagta	8220
gtggggccga	cacggaagat	gtcgtgtgct	gctcaatgtc	ttattcctgg	acaggcgcac	8280
tcgtcaccac	gtgcgctgcg	gaagaacaaa	aactgcccac	caacgcactg	agcaactcgt	8340
tgctacgcca	tcacaatctg	gtgtattcca	ccacttcacg	cagtgtctgc	caaaggcaga	8400
agaaagtcac	atttgacaga	ctgcaagttc	tggacagcca	ttaccaggac	gtgctcaagg	8460
aggtcaaagc	agcggcgtca	aaagtgaagg	ctaacttgct	atccgtagag	gaagcttgca	8520
gcctgacgcc	cccacattca	gccaaatcca	agtttggtta	tggggcaaaa	gacgtccgtt	8580
gccatgccag	aaaggccgta	gccacatca	actccgtgtg	gaaagacctt	ctggaagaca	8640
gtgtaacacc	aatagacact	accatcatgg	ccaagaacga	ggttttctgc	gttcagcctg	8700
agaagggggg	tcgtaagcca	gctcgtctca	tcgtgttccc	cgacctgggc	gtgcgcgtgt	8760
gcgagaagat	ggccctgtac	gacgtgggta	gcaagctccc	cctggccgtg	atggggaagt	8820
cctacggatt	ccaatactca	ccaggacagc	gggttgaatt	cctcgtgcaa	gcgtggaagt	8880
ccaagaagac	cccgatgggg	ttctcgtatg	ataccogctg	ttttgactcc	acagtcaactg	8940
agagcgacat	ccgtacggag	gaggcaattt	accaatgttg	tgacctggac	ccccaaagccc	9000
gcgtggccat	caagtccctc	actgagaggc	tttatgttgg	gggccctctt	accaattcaa	9060
ggggggaaaa	ctgcggtac	cgaggtgcc	gcgcgagcgg	cgtactgaca	actagctgtg	9120
gtaacaccct	cacttgctac	atcaaggccc	gggcagcctg	tcgagccgca	gggctccagg	9180
actgcaccat	gctcgtgtgt	ggcgacgact	tagtcgttat	ctgtgaaaagt	gcgggggtcc	9240
aggaggacgc	ggcgagcctg	agagccttca	cggaggctat	gaccaggtag	tccgcccccc	9300
ccggggaccc	cccacaacca	gaatacgact	tggagcttat	aacatcatgc	tccccaacg	9360
tgctcagtcgc	ccacgacggc	gctggaaaga	gggtctacta	ccttaccctg	gaccctacaa	9420
ccccctcgc	gagagccgcg	tgggagacag	caagacacac	tccagtcaat	tccgtggctg	9480
gcaacataat	catgtttgcc	cccacactgt	ggcgagggat	gatactgatg	acctatttct	9540
tttagcgtcct	catagccagg	gatcagcttg	aacaggctct	taactgtgag	atctacggag	9600
cctgctactc	catagaacca	ctggatctac	ctccaatcat	tcaaagactc	catggcctca	9660
gcgcattttc	actccacagt	tactctccag	gtgaaatcaa	taggggtggc	gcatgcctca	9720
gaaaacttgg	ggtcccgcgc	ttgcgagctt	ggagacaccg	ggcccgagc	gtccgcgcta	9780
ggcttctgtc	cagaggaggc	agggtgcc	tatgtggcaa	gtacctcttc	aactgggcag	9840
taagaacaaa	gctcaaactc	actccaatag	cggccgctgg	ccggctggac	ttgtccggtt	9900
ggttcacggc	tggctacagc	gggggagaca	tttatcacag	cgtgtctcat	gcccggcccc	9960
gctggttctg	gttttgccca	ctcctgctcg	ctgcaggggt	aggcatctac	ctcctcccca	10020
accgatgaag	gttggggtaa	acactccggc	ctcttaagcc	atttcctgtt	tttttttttt	10080
tttttttttt	ttttttcttt	ttttttttct	ttcctttcct	tctttttttc	ctttcttttt	10140
cccttcttta	atggtggctc	catcttagcc	ctagtacagg	ctagctgtga	aaggctcogt	10200
agccgcatga	ctgcagagag	tgctgatact	ggcctctctg	cagatcatgt	tctagactgg	10260
aattcgtcga	cgagctccct	atagttagct	gtattagagg	ccgacttggc	caaattctga	10320
atcatggtca	tagctgtttc	ctgtgtgaaa	ttgttatccg	ctcacaattc	cacacaacat	10380
acgagccgga	agcataaagt	gtaaaagcct	gggtgcctaa	tgagttagct	aactcacatt	10440
aattgcgttg	cgctcactgc	ccgctttcca	ctcgggaaac	ctgtcgtgcc	agctgcatta	10500
atgaatcggc	caacgcgcgg	ggagaggcgg	tttgcgtatt	gggcgctctt	ccgcttcctc	10560
gctcactgac	tcgctgcgct	cggtcgttgc	gctgcggcga	gcggtatcag	ctcactcaaa	10620
ggcggttaata	cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	tgtgagcaaa	10680
aggccagcaa	aaggccagga	accgtaaaaa	ggccgcgttg	ctggcgtttt	tccataggct	10740
ccgccccctc	gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	gaaacccgac	10800
aggactataa	agataccagg	cgtttccccc	tggaaagctcc	ctcgtgcgct	ctcctgttcc	10860
gacctgccc	cttaccggat	acctgtccgc	ctttctccct	tcgggaaagc	tggcgctttc	10920
tcatagctca	cgctgtaggt	atctcagttc	ggtgtaggtc	gttcgctcca	agctgggctg	10980
tgtgcacgaa	cccccggttc	agccccagcg	ctgcgcctta	tccggttaact	atcgtcttga	11040
gtccaacccg	gtaagacacg	acttatcgcc	actggcagca	gccactggta	acaggattag	11100
cagagcgagg	tatgtaggcg	gtgctacaga	gttcttgaag	tgggtggccta	actacggcta	11160
cactagaagg	acagtatttg	gtatctgcgc	tctgtgaag	ccagttacct	toggaaaaag	11220
agttggtagc	ctttgatccg	gcaaaaaaac	caccgctggt	agcggtggtt	tttttggttt	11280
caagcgtagc	atttagcgca	gaaaaaaagg	atctcaagaa	gatcctttga	tcttttctac	11340
ggggtctgac	gctcagtggg	acgaaaactc	acgttaaggg	attttgggtca	tgagattatc	11400
aaaaaggatc	ttcacctaga	tcctttttaa	ttaaaaatga	agttttaaat	caatctaaag	11460
tatatatgag	taaacttggg	ctgacagtta	ccaatgctta	atcagtgagg	cacctatctc	11520

```

agcgatctgt ctatttcggt catccatagt tgccctgactc cccgtcgtgt agataaactac 11580
gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag acccacgctc 11640
accggctcca gattttatcag caataaacca gccagccgga agggccgagc gcagaagtgg 11700
tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag ctagagtaag 11760
tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt gctacaggca tcgtggtgtc 11820
acgctcgtcg tttggtatgg cttcattcag ctccggttcc caacgatcaa ggcgagttac 11880
atgatcccc atgttggtgca aaaaagcggg tagctccttc ggctcctcga tcgttgtcag 11940
aagtaagttg gccgcagtgt tatcactcat gggttatggca gcaactgcata attctcttac 12000
tgtcatgcca tccgtaagat gcttttctgt gactgggtgag tactcaacca agtcattctg 12060
agaatagtgt atgcggcgac cgagttgctc ttgcccgcg tcaatacggg ataataccgc 12120
gccacatagc agaactttaa aagtgtcat cattggaaaa cgttcttcgg ggcgaaaact 12180
ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg caccacactg 12240
atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa 12300
tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac tcttctcttt 12360
tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca tatttgaatg 12420
tatttagaaa aataaacaata taggggttcc gcgcacattt ccccgaaaag tgccacctga 12480
cgcgccctgt agcggcgcat taagcgcggc ggggtgtggtg gttacgcgca gcgtgacgcg 12540
tacacttgcc agcgccctag cgcccgctcc tttcgctttc ttcccttctt ttctcgccac 12600
gttcgccggc tttccccgtc aagctctaaa tcggggcatc cctttagggt tccgatttag 12660
tgctttacgg cactctgacc ccaaaaaact tgattagggt gatggttcac gtagtgggac 12720
atcgccctga tagacggttt ttcgcccttt gacgttggag tccacgttct ttaatagtgg 12780
actcttggtc caaactggaa caacactcaa ccctatctcg gtctattctt ttgatttata 12840
agggattttg ccgatttctg cctattgggt aaaaaatgag ctgatttaac aaaaatttaa 12900
cgcgaatttt aacaaaatat taacaaaata ttaacgttta caatttccca ttcgccattc 12960
aggctgcgca actggttgga agggcgatcg gtgcgggcct cttcgctatt acgccagctg 13020
gcgaaagggg gatgtgctgc aaggcgatta agttgggtaa cgccagggtt ttcccagtc 13080
cgacgttgta aaacgacggc cagtgcgaag ctgacttggt cagcggcgcg taatacgact 13140
cactata 13147

```

<210> 19

<211> 12851

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<220>

<221> CDS

<222> (446)..(10096)

<400> 19

```

accacaaaca ctccagtttg ttacactccg ctaggaatgc tcctggagca cccccctag 60
cagggcgtgg gggatttccc ctgcccgtct gcagaagggt ggagccaacc accttagtat 120
gtaggcggcg ggactcatga cgctcgctgt atgacaagcg ccaagcttga cttggatggc 180
cctgatgggc gttcatgggt tcggtggtgg tggcgcttta ggcagcctcc acgcccacca 240
cctcccagat agagcggcgg cactgtaggg aagaccgggg accggtcact accaaggacg 300
cagacctctt tttgagtatc acgcctccgg aagtagttgg gcaagccac ctatatgtgt 360
tgggatggtt ggggttagcc atccataccg tactgcctga tagggtcctt gcgaggggat 420
ctgggagtct cgtagaccgt agcac atg agc acg aat cct aaa cct caa aga 472
Met Ser Thr Asn Pro Lys Pro Gln Arg
1 5
aaa acc aaa cgt aac acc aac cgt cgc cca cag gac gtc aag ttc ccg 520

```

Lys 10	Thr	Lys	Arg	Asn	Thr 15	Asn	Arg	Arg	Pro	Gln 20	Asp	Val	Lys	Phe	Pro 25	
ggc	ggc	ggc	cag	atc	gtt	ggc	gga	gtt	tac	ttg	ttg	ccg	cgc	agg	ggc	568
Gly	Gly	Gly	Gln	Ile	Val	Gly	Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	
				30					35					40		
cct	aga	ttg	ggc	gtg	cgc	gcg	acg	agg	aag	act	tcc	gag	cgg	tcg	caa	616
Pro	Arg	Leu	Gly	Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser	Gln	
			45					50					55			
cct	cgt	ggc	aga	cgt	cag	cct	atc	ccc	aag	gca	cgt	cgg	ccc	gag	ggc	664
Pro	Arg	Gly	Arg	Arg	Gln	Pro	Ile	Pro	Lys	Ala	Arg	Arg	Pro	Glu	Gly	
		60					65					70				
agg	acc	tgg	gct	cag	ccc	ggg	tac	cct	tgg	ccc	ctc	tat	ggc	aat	gag	712
Arg	Thr	Trp	Ala	Gln	Pro	Gly	Tyr	Pro	Trp	Pro	Leu	Tyr	Gly	Asn	Glu	
	75					80					85					
ggc	tgc	ggg	tgg	gcg	gga	tgg	ctc	ctg	tct	ccc	cgt	ggc	tct	cgg	cct	760
Gly	Cys	Gly	Trp	Ala	Gly	Trp	Leu	Leu	Ser	Pro	Arg	Gly	Ser	Arg	Pro	
	90				95					100					105	
agc	tgg	ggc	ccc	aca	gac	ccc	cgg	cgt	agg	tcg	cgc	aat	ttg	ggc	aag	808
Ser	Trp	Gly	Pro	Thr	Asp	Pro	Arg	Arg	Arg	Ser	Arg	Asn	Leu	Gly	Lys	
				110					115					120		
gtc	atc	gat	acc	ctt	acg	tgc	ggc	ttc	gcc	gac	ctc	atg	ggg	tac	ata	856
Val	Ile	Asp	Thr	Leu	Thr	Cys	Gly	Phe	Ala	Asp	Leu	Met	Gly	Tyr	Ile	
			125					130					135			
ccg	ctc	gtc	ggc	gcc	cct	ctt	gga	ggc	gct	gcc	agg	gcc	ctg	gcg	cat	904
Pro	Leu	Val	Gly	Ala	Pro	Leu	Gly	Gly	Ala	Ala	Arg	Ala	Leu	Ala	His	
		140					145					150				
ggc	gtc	cgg	gtt	ctg	gaa	gac	ggc	gtg	aac	tat	gca	aca	ggg	aac	ctt	952
Gly	Val	Arg	Val	Leu	Glu	Asp	Gly	Val	Asn	Tyr	Ala	Thr	Gly	Asn	Leu	
	155					160					165					
cct	ggc	tgc	tct	ttc	tct	atc	ttc	ctt	ctg	gcc	ctg	ctc	tct	tgc	ctg	1000
Pro	Gly	Cys	Ser	Phe	Ser	Ile	Phe	Leu	Leu	Ala	Leu	Leu	Ser	Cys	Leu	
	170				175					180					185	
act	gtg	ccc	gct	tca	gcc	tac	caa	gtg	cgc	aat	tcc	tcg	ggg	ctt	tac	1048
Thr	Val	Pro	Ala	Ser	Ala	Tyr	Gln	Val	Arg	Asn	Ser	Ser	Gly	Leu	Tyr	
				190					195					200		
cat	gtc	acc	aat	gat	tgc	cct	aac	tca	agt	att	gtg	tac	gag	gcg	gcc	1096
His	Val	Thr	Asn	Asp	Cys	Pro	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Ala	
			205					210					215			
gat	gcc	atc	ctg	cac	act	ccg	ggg	tgt	gtc	cct	tgc	gtg	cgc	gag	ggc	1144
Asp	Ala	Ile	Leu	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Glu	Gly	
		220					225					230				
aac	gct	tcg	agg	tgt	tgg	gtg	gcg	gtg	acc	ccc	acg	gtg	gcc	acc	agg	1192
Asn	Ala	Ser	Arg	Cys	Trp	Val	Ala	Val	Thr	Pro	Thr	Val	Ala	Thr	Arg	
	235					240					245					
gac	ggc	aaa	ctc	ccc	aca	acg	cag	ctt	cga	cgt	cat	atc	gat	ctg	ctt	1240
Asp	Gly	Lys	Leu	Pro	Thr	Thr	Gln	Leu	Arg	Arg	His	Ile	Asp	Leu	Leu	
	250				255				260						265	

gtc ggg agc gcc acc ctc tgc tcg gcc ctc tac gtg ggg gac ctg tgc	1288
Val Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys	
270 275 280	
ggg tct gtc ttt ctt gtt ggt caa ctg ttt acc ttc tct ccc agg cgc	1336
Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg Arg	
285 290 295	
cac tgg acg acg caa gac tgc aat tgt tct atc tat ccc ggc cat ata	1384
His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His Ile	
300 305 310	
acg ggt cat cgc atg gca tgg gat atg atg atg aac tgg tcc cct acg	1432
Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr	
315 320 325	
gca gcg ttg gtg gta gct cag ctg ctc cgg atc cca caa gcc atc atg	1480
Ala Ala Leu Val Val Ala Gln Leu Leu Arg Ile Pro Gln Ala Ile Met	
330 335 340 345	
gac atg atc gct ggt gct cac tgg gga gtc ctg gcg ggc ata gcg tat	1528
Asp Met Ile Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala Tyr	
350 355 360	
ttc tcc atg gtg ggg aac tgg gcg aag gtc ctg gta gtg ctg ctg cta	1576
Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu Leu	
365 370 375	
ttt gcc ggc gtc gac gcg gaa acc cac gtc acc ggg gga aat gcc ggc	1624
Phe Ala Gly Val Asp Ala Glu Thr His Val Thr Gly Gly Asn Ala Gly	
380 385 390	
cgc acc acg gct ggg ctt gtt ggt ctc ctt aca cca ggc gcc aag cag	1672
Arg Thr Thr Ala Gly Leu Val Gly Leu Leu Thr Pro Gly Ala Lys Gln	
395 400 405	
aac atc caa ctg atc aac acc aac ggc agt tgg cac atc aat agc acg	1720
Asn Ile Gln Leu Ile Asn Thr Asn Gly Ser Trp His Ile Asn Ser Thr	
410 415 420 425	
gcc ttg aat tgc aat gaa agc ctt aac acc ggc tgg tta gca ggg ctc	1768
Ala Leu Asn Cys Asn Glu Ser Leu Asn Thr Gly Trp Leu Ala Gly Leu	
430 435 440	
ttc tat caa cac aaa ttc aac tct tca ggc tgt cct gag agg ttg gcc	1816
Phe Tyr Gln His Lys Phe Asn Ser Ser Gly Cys Pro Glu Arg Leu Ala	
445 450 455	
agc tgc cga cgc ctt acc gat ttt gcc cag ggc tgg ggt cct atc agt	1864
Ser Cys Arg Arg Leu Thr Asp Phe Ala Gln Gly Trp Gly Pro Ile Ser	
460 465 470	
tat gcc aac gga agc ggc ctc gac gaa cgc ccc tac tgc tgg cac tac	1912
Tyr Ala Asn Gly Ser Gly Leu Asp Glu Arg Pro Tyr Cys Trp His Tyr	
475 480 485	
cct cca aga cct tgt ggc att gtg ccc gca aag agc gtg tgt ggc ccg	1960
Pro Pro Arg Pro Cys Gly Ile Val Pro Ala Lys Ser Val Cys Gly Pro	
490 495 500 505	
gta tat tgc ttc act ccc agc ccc gtg gtg gtg gga acg acc gac agg	2008
Val Tyr Cys Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg	
510 515 520	

tcg ggc gcg cct acc tac agc tgg ggt gca aat gat acg gat gtc ttc	2056
Ser Gly Ala Pro Thr Tyr Ser Trp Gly Ala Asn Asp Thr Asp Val Phe	
525 530 535	
gtc ctt aac aac acc agg cca ccg ctg ggc aat tgg ttc ggt tgt acc	2104
Val Leu Asn Asn Thr Arg Pro Pro Leu Gly Asn Trp Phe Gly Cys Thr	
540 545 550	
tgg atg aac tca act gga ttc acc aaa gtg tgc gga gcg ccc cct tgt	2152
Trp Met Asn Ser Thr Gly Phe Thr Lys Val Cys Gly Ala Pro Pro Cys	
555 560 565	
gtc atc gga ggg gtg ggc aac aac acc ttg ctc tgc ccc act gat tgc	2200
Val Ile Gly Gly Val Gly Asn Asn Thr Leu Leu Cys Pro Thr Asp Cys	
570 575 580 585	
ttc cgc aaa cat ccg gaa gcc aca tac tct cgg tgc ggc tcc ggt ccc	2248
Phe Arg Lys His Pro Glu Ala Thr Tyr Ser Arg Cys Gly Ser Gly Pro	
590 595 600	
tgg att aca ccc agg tgc atg gtc gac tac ccg tat agg ctt tgg cac	2296
Trp Ile Thr Pro Arg Cys Met Val Asp Tyr Pro Tyr Arg Leu Trp His	
605 610 615	
tat cct tgt acc atc aat tac acc ata ttc aaa gtc agg atg tac gtg	2344
Tyr Pro Cys Thr Ile Asn Tyr Thr Ile Phe Lys Val Arg Met Tyr Val	
620 625 630	
gga ggg gtc gag cac agg ctg gaa gcg gcc tgc aac tgg acg cgg ggc	2392
Gly Gly Val Glu His Arg Leu Glu Ala Ala Cys Asn Trp Thr Arg Gly	
635 640 645	
gaa cgc tgt gat ctg gaa gac agg gac agg tcc gag ctc agc ccg ttg	2440
Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg Ser Glu Leu Ser Pro Leu	
650 655 660 665	
ctg ctg tcc acc aca cag tgg cag gtc ctt ccg tgt tct ttc acg acc	2488
Leu Leu Ser Thr Thr Gln Trp Gln Val Leu Pro Cys Ser Phe Thr Thr	
670 675 680	
ctg cca gcc ttg tcc acc ggc ctc atc cac ctc cac cag aac att gtg	2536
Leu Pro Ala Leu Ser Thr Gly Leu Ile His Leu His Gln Asn Ile Val	
685 690 695	
gac gtg cag tac ttg tac ggg gta ggg tca agc atc gcg tcc tgg gcc	2584
Asp Val Gln Tyr Leu Tyr Gly Val Gly Ser Ser Ile Ala Ser Trp Ala	
700 705 710	
att aag tgg gag tac gtc gtt ctc ctg ttc ctt ctg ctt gca gac gcg	2632
Ile Lys Trp Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala	
715 720 725	
cgc gtc tgc tcc tgc ttg tgg atg atg tta ctc ata tcc caa gcg gag	2680
Arg Val Cys Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu	
730 735 740 745	
gcg gct ttg gag aac ctc gta ata ctc aat gca gca tcc ctg gcc ggg	2728
Ala Ala Leu Glu Asn Leu Val Ile Leu Asn Ala Ala Ser Leu Ala Gly	
750 755 760	
acg cac ggt ctt gtg tcc ttc ctc gtg ttc ttc tgc ttt gcg tgg tat	2776
Thr His Gly Leu Val Ser Phe Leu Val Phe Phe Cys Phe Ala Trp Tyr	

765				770				775								
ctg	aag	ggt	agg	tgg	gtg	ccc	gga	gcg	gtc	tac	gcc	ctc	tac	ggg	atg	2824
Leu	Lys	Gly	Arg	Trp	Val	Pro	Gly	Ala	Val	Tyr	Ala	Leu	Tyr	Gly	Met	
		780					785					790				
tgg	cct	ctc	ctc	ctg	ctc	ctg	ctg	gcg	ttg	cct	cag	cgg	gca	tac	gca	2872
Trp	Pro	Leu	Leu	Leu	Leu	Leu	Leu	Ala	Leu	Pro	Gln	Arg	Ala	Tyr	Ala	
		795				800					805					
ctg	gac	acg	gag	gtg	gcc	gcg	tcg	tgt	ggc	ggc	gtt	gtt	ctt	gtc	ggg	2920
Leu	Asp	Thr	Glu	Val	Ala	Ala	Ser	Cys	Gly	Gly	Val	Val	Leu	Val	Gly	
810					815					820					825	
tta	atg	gcg	ctg	act	ctg	tcg	cca	tat	tac	aag	cgc	tat	atc	agc	tgg	2968
Leu	Met	Ala	Leu	Thr	Leu	Ser	Pro	Tyr	Tyr	Lys	Arg	Tyr	Ile	Ser	Trp	
				830					835					840		
tgc	atg	tgg	tgg	ctt	cag	tat	ttt	ctg	acc	aga	gta	gaa	gcg	caa	ctg	3016
Cys	Met	Trp	Trp	Leu	Gln	Tyr	Phe	Leu	Thr	Arg	Val	Glu	Ala	Gln	Leu	
			845					850					855			
cac	gtg	tgg	gtt	ccc	ccc	ctc	aac	gtc	cgg	ggg	ggg	cgc	gat	gcc	gtc	3064
His	Val	Trp	Val	Pro	Pro	Leu	Asn	Val	Arg	Gly	Gly	Arg	Asp	Ala	Val	
		860					865					870				
atc	tta	ctc	atg	tgt	gta	gta	cac	ccg	acc	ctg	gta	ttt	gac	atc	acc	3112
Ile	Leu	Leu	Met	Cys	Val	Val	His	Pro	Thr	Leu	Val	Phe	Asp	Ile	Thr	
	875					880					885					
aaa	cta	ctc	ctg	gcc	atc	ttc	gga	ccc	ctt	tgg	att	ctt	caa	gcc	agt	3160
Lys	Leu	Leu	Leu	Ala	Ile	Phe	Gly	Pro	Leu	Trp	Ile	Leu	Gln	Ala	Ser	
890					895					900					905	
ttg	ctt	aaa	gtc	ccc	tac	ttc	gtg	cgc	gtt	caa	ggc	ctt	ctc	cgg	atc	3208
Leu	Leu	Lys	Val	Pro	Tyr	Phe	Val	Arg	Val	Gln	Gly	Leu	Leu	Arg	Ile	
			910						915					920		
tgc	gcg	cta	gcg	cgg	aag	ata	gcc	gga	ggt	cat	tac	gtg	caa	atg	gcc	3256
Cys	Ala	Leu	Ala	Arg	Lys	Ile	Ala	Gly	Gly	His	Tyr	Val	Gln	Met	Ala	
			925					930					935			
atc	atc	aag	tta	ggg	gcg	ctt	act	ggc	acc	tat	gtg	tat	aac	cat	ctc	3304
Ile	Ile	Lys	Leu	Gly	Ala	Leu	Thr	Gly	Thr	Tyr	Val	Tyr	Asn	His	Leu	
		940					945					950				
acc	cct	ctt	cga	gac	tgg	gcg	cac	aac	ggc	ctg	cga	gat	ctg	gcc	gtg	3352
Thr	Pro	Leu	Arg	Asp	Trp	Ala	His	Asn	Gly	Leu	Arg	Asp	Leu	Ala	Val	
		955				960					965					
gct	gtg	gaa	cca	gtc	gtc	ttc	tcc	cga	atg	gag	acc	aag	ctc	atc	acg	3400
Ala	Val	Glu	Pro	Val	Val	Phe	Ser	Arg	Met	Glu	Thr	Lys	Leu	Ile	Thr	
970					975					980					985	
tgg	ggg	gca	gat	acc	gcc	gcg	tgc	ggt	gac	atc	atc	aac	ggc	ttg	ccc	3448
Trp	Gly	Ala	Asp	Thr	Ala	Ala	Cys	Gly	Asp	Ile	Ile	Asn	Gly	Leu	Pro	
				990					995					1000		
gtc	tct	gcc	cgt	agg	ggc	cag	gag	ata	ctg	ctt	ggg	cca	gcc	gac	gga	3496
Val	Ser	Ala	Arg	Arg	Gly	Gln	Glu	Ile	Leu	Leu	Gly	Pro	Ala	Asp	Gly	
			1005				1010						1015			
atg	gtc	tcc	aag	ggg	tgg	agg	ttg	ctg	gcg	ccc	atc	acg	gcg	tac	gcc	3544

Met Val Ser Lys Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ala	
1020 1025 1030	
cag cag acg aga ggc ctc cta ggg tgt ata atc acc agc ctg act ggc	3592
Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly	
1035 1040 1045	
cgg gac aaa aac caa gtg gag ggt gag gtc cag atc gtg tca act gct	3640
Arg Asp Lys Asn Gln Val Glu Gly Glu Val Gln Ile Val Ser Thr Ala	
1050 1055 1060 1065	
acc caa acc ttc ctg gca acg tgc atc aat ggg gta tgc tgg act gtc	3688
Thr Gln Thr Phe Leu Ala Thr Cys Ile Asn Gly Val Cys Trp Thr Val	
1070 1075 1080	
tac cac ggg gcc gga acg agg acc atc gca tca ccc aag ggt cct gtc	3736
Tyr His Gly Ala Gly Thr Arg Thr Ile Ala Ser Pro Lys Gly Pro Val	
1085 1090 1095	
atc cag atg tat acc aat gtg gac caa gac ctt gtg ggc tgg ccc gct	3784
Ile Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Pro Ala	
1100 1105 1110	
cct caa ggt tcc cgc tca ttg aca ccc tgt acc tgc ggc tcc tcg gac	3832
Pro Gln Gly Ser Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp	
1115 1120 1125	
ctt tac ctg gtc acg agg cac gcc gat gtc att ccc gtg cgc cgg cga	3880
Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg	
1130 1135 1140 1145	
ggt gat agc agg ggt agc ctg ctt tcg ccc cgg ccc att tcc tac ttg	3928
Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu	
1150 1155 1160	
aaa ggc tcc tcg ggg ggt ccg ctg ttg tgc ccc gcg gga cac gcc gtg	3976
Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala Val	
1165 1170 1175	
ggc cta ttc agg gcc gcg gtg tgc acc cgt gga gtg gct aaa gcg gtg	4024
Gly Leu Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val	
1180 1185 1190	
gac ttt atc cct gtg gag aac cta ggg aca acc atg aga tcc ccg gtg	4072
Asp Phe Ile Pro Val Glu Asn Leu Gly Thr Thr Met Arg Ser Pro Val	
1195 1200 1205	
ttc acg gac aac tcc tct acg cgt atg cag atc ttc gtg aag acc ctg	4120
Phe Thr Asp Asn Ser Ser Thr Arg Met Gln Ile Phe Val Lys Thr Leu	
1210 1215 1220 1225	
acg ggc aag acc atc act ctt gag gtc gag ccc agt gac acc atc gag	4168
Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Pro Ser Asp Thr Ile Glu	
1230 1235 1240	
aat gtc aag gcc aag atc caa gac aag gaa ggc atc cca cct gac cag	4216
Asn Val Lys Ala Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln	
1245 1250 1255	
cag agg ctg ata ttc gcg ggc aaa cag ctg gag gat ggc cgc acc ctg	4264
Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu	
1260 1265 1270	

tcc gac tac aac atc cag aaa gag tcc acc ttg cac ctg gtg ctg cgt	4312
Ser Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg	
1275 1280 1285	
ctc cgc ggt ggc gca cct ttt acg ctg cag tgt ctc tct gaa cgt ggc	4360
Leu Arg Gly Gly Ala Pro Phe Thr Leu Gln Cys Leu Ser Glu Arg Gly	
1290 1295 1300 1305	
acg ctg tca gcg atg gca gtg gtc atg act ggt ata gac ccc cga act	4408
Thr Leu Ser Ala Met Ala Val Val Met Thr Gly Ile Asp Pro Arg Thr	
1310 1315 1320	
tgg act gga act atc ttc aga tta gga tct ctg gcc act agc tac atg	4456
Trp Thr Gly Thr Ile Phe Arg Leu Gly Ser Leu Ala Thr Ser Tyr Met	
1325 1330 1335	
gga ttc gtt tgt gac aac gtg ttg tat act gct cac cat ggc agc aag	4504
Gly Phe Val Cys Asp Asn Val Leu Tyr Thr Ala His His Gly Ser Lys	
1340 1345 1350	
ggg cgc cgg ttg gct cat ccc aca ggc tcc ata cac cca ata acc gtt	4552
Gly Arg Arg Leu Ala His Pro Thr Gly Ser Ile His Pro Ile Thr Val	
1355 1360 1365	
gac gcg gct aat gac cag gac atc tat caa cca cca tgt gga gct ggg	4600
Asp Ala Ala Asn Asp Gln Asp Ile Tyr Gln Pro Pro Cys Gly Ala Gly	
1370 1375 1380 1385	
tcc ctt act cgg tgc tct tgc ggg gag acc aag ggg tat ctg gta aca	4648
Ser Leu Thr Arg Cys Ser Cys Gly Glu Thr Lys Gly Tyr Leu Val Thr	
1390 1395 1400	
cga ctg ggg tca ttg gtt gag gtc aac aaa tcc gat gac cct tat tgg	4696
Arg Leu Gly Ser Leu Val Glu Val Asn Lys Ser Asp Asp Pro Tyr Trp	
1405 1410 1415	
tgt gtg tgc ggg gcc ctt ccc atg gct gtt gcc aag ggt tct tca ggt	4744
Cys Val Cys Gly Ala Leu Pro Met Ala Val Ala Lys Gly Ser Ser Gly	
1420 1425 1430	
gcc ccg att ctg tgc tcc tcc ggg cat gtt att ggg atg ttc acc gct	4792
Ala Pro Ile Leu Cys Ser Ser Gly His Val Ile Gly Met Phe Thr Ala	
1435 1440 1445	
gct aga aat tct ggc ggt tca gtc agt cag att agg gtt agg ccg ttg	4840
Ala Arg Asn Ser Gly Gly Ser Val Ser Gln Ile Arg Val Arg Pro Leu	
1450 1455 1460 1465	
gtg tgt gct gga tac cat ccc cag tac aca gca cat gcc act ctt gat	4888
Val Cys Ala Gly Tyr His Pro Gln Tyr Thr Ala His Ala Thr Leu Asp	
1470 1475 1480	
aca aaa cct act gtg cct aac gag tat tca gtg caa att tta att gcc	4936
Thr Lys Pro Thr Val Pro Asn Glu Tyr Ser Val Gln Ile Leu Ile Ala	
1485 1490 1495	
ccc act ggc agc ggc aag tca acc aaa tta cca ctt tct tac atg cag	4984
Pro Thr Gly Ser Gly Lys Ser Thr Lys Leu Pro Leu Ser Tyr Met Gln	
1500 1505 1510	
gag aag tat gag gtc ttg gtc cta aat ccc agt gtg gct aca aca gca	5032
Glu Lys Tyr Glu Val Leu Val Leu Asn Pro Ser Val Ala Thr Thr Ala	
1515 1520 1525	

tca atg cca aag tac atg cac gcg acg tac ggc gtg aat cca aat tgc	5080
Ser Met Pro Lys Tyr Met His Ala Thr Tyr Gly Val Asn Pro Asn Cys	
1530 1535 1540 1545	
tat ttt aat ggc aaa tgt acc aac aca ggg gct tca ctt acg tac agc	5128
Tyr Phe Asn Gly Lys Cys Thr Asn Thr Gly Ala Ser Leu Thr Tyr Ser	
1550 1555 1560	
aca tat ggc atg tac ctg acc gga gca tgt tcc cgg aac tat gat gta	5176
Thr Tyr Gly Met Tyr Leu Thr Gly Ala Cys Ser Arg Asn Tyr Asp Val	
1565 1570 1575	
atc att tgt gac gaa tgc cat gct acc gat gca acc acc gtg ttg ggc	5224
Ile Ile Cys Asp Glu Cys His Ala Thr Asp Ala Thr Thr Val Leu Gly	
1580 1585 1590	
att gga aag gtc cta acc gaa gct cca tcc aaa aat gtt agg cta gtg	5272
Ile Gly Lys Val Leu Thr Glu Ala Pro Ser Lys Asn Val Arg Leu Val	
1595 1600 1605	
gtt ctt gcc acg gct acc ccc cct gga gta atc cct aca cca cat gcc	5320
Val Leu Ala Thr Ala Thr Pro Pro Gly Val Ile Pro Thr Pro His Ala	
1610 1615 1620 1625	
aac ata act gag att caa tta acc gat gaa ggc act atc ccc ttt cat	5368
Asn Ile Thr Glu Ile Gln Leu Thr Asp Glu Gly Thr Ile Pro Phe His	
1630 1635 1640	
gga aaa aag att aag gag gaa aat ctg aag aaa ggg aga cac ctt atc	5416
Gly Lys Lys Ile Lys Glu Glu Asn Leu Lys Lys Gly Arg His Leu Ile	
1645 1650 1655	
ttt gag gct acc aaa aaa cac tgt gat gag ctt gct aac gag tta gct	5464
Phe Glu Ala Thr Lys Lys His Cys Asp Glu Leu Ala Asn Glu Leu Ala	
1660 1665 1670	
cga aag gga ata aca gct gtc tct tac tat agg gga tgt gac atc tca	5512
Arg Lys Gly Ile Thr Ala Val Ser Tyr Tyr Arg Gly Cys Asp Ile Ser	
1675 1680 1685	
aaa atc cct gag ggc gac tgt gta gta gtt gcc act gat gcc ttg tgt	5560
Lys Ile Pro Glu Gly Asp Cys Val Val Val Ala Thr Asp Ala Leu Cys	
1690 1695 1700 1705	
aca ggg tac act ggt gac ttt gat tcc gtg tat gac tgc agc ctc atg	5608
Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Tyr Asp Cys Ser Leu Met	
1710 1715 1720	
gta gaa ggc aca tgc cat gtt gac ctt gac cct act ttc acc atg ggt	5656
Val Glu Gly Thr Cys His Val Asp Leu Asp Pro Thr Phe Thr Met Gly	
1725 1730 1735	
gtt cgt gtg tgc ggg gtt tca gca ata gtt aaa ggc cag cgt agg ggc	5704
Val Arg Val Cys Gly Val Ser Ala Ile Val Lys Gly Gln Arg Arg Gly	
1740 1745 1750	
cgc aca ggc cgt ggg aga gct ggc ata tac tac tat gta gac ggg agt	5752
Arg Thr Gly Arg Gly Arg Ala Gly Ile Tyr Tyr Val Asp Gly Ser	
1755 1760 1765	
tgt acc cct tcg ggt atg gtt cct gaa tgc aac att gtt gaa gcc ttc	5800
Cys Thr Pro Ser Gly Met Val Pro Glu Cys Asn Ile Val Glu Ala Phe	

1770	1775	1780	1785	
gac gca gcc aag gca tgg tat ggt ttg tca tca aca gaa gct caa act				5848
Asp Ala Ala Lys Ala Trp Tyr Gly Leu Ser Ser Thr Glu Ala Gln Thr				
	1790	1795	1800	
att ctg gac acc tat cgc acc caa cct ggg tta cct gcg ata gga gca				5896
Ile Leu Asp Thr Tyr Arg Thr Gln Pro Gly Leu Pro Ala Ile Gly Ala				
	1805	1810	1815	
aat ttg gac gag tgg gct gat ctc ttt tct atg gtc aac ccc gaa cct				5944
Asn Leu Asp Glu Trp Ala Asp Leu Phe Ser Met Val Asn Pro Glu Pro				
	1820	1825	1830	
tca ttt gtc aat act gca aaa aga act gct gac aat tat gtt ttg ttg				5992
Ser Phe Val Asn Thr Ala Lys Arg Thr Ala Asp Asn Tyr Val Leu Leu				
	1835	1840	1845	
act gca gcc caa cta caa ctg tgt cat cag tat ggc tat gct gct ccc				6040
Thr Ala Ala Gln Leu Gln Leu Cys His Gln Tyr Gly Tyr Ala Ala Pro				
	1850	1855	1860	1865
aat gac gca cca cgg tgg cag gga gcc cgg ctt ggg aaa aaa cct tgt				6088
Asn Asp Ala Pro Arg Trp Gln Gly Ala Arg Leu Gly Lys Lys Pro Cys				
	1870	1875	1880	
ggg gtt ctg tgg cgc ttg gac ggc gct gac gcc tgt cct ggc cca gag				6136
Gly Val Leu Trp Arg Leu Asp Gly Ala Asp Ala Cys Pro Gly Pro Glu				
	1885	1890	1895	
ccc agc gag gtg acc aga tac caa atg tgc ttc act gaa gtc aat act				6184
Pro Ser Glu Val Thr Arg Tyr Gln Met Cys Phe Thr Glu Val Asn Thr				
	1900	1905	1910	
tct ggg aca gcc gca ctc gct gtt ggc gtt gga gtg gct atg gct tat				6232
Ser Gly Thr Ala Ala Leu Ala Val Gly Val Gly Val Ala Met Ala Tyr				
	1915	1920	1925	
cta gcc att gac act ttt ggc gcc act tgt gtg cgg cgt tgc tgg tct				6280
Leu Ala Ile Asp Thr Phe Gly Ala Thr Cys Val Arg Arg Cys Trp Ser				
	1930	1935	1940	1945
att gca tca gtc cct acc ggt gct act gtc gcc cca gtg gtt gac gaa				6328
Ile Ala Ser Val Pro Thr Gly Ala Thr Val Ala Pro Val Val Asp Glu				
	1950	1955	1960	
gaa gaa atc gtg gag gag tgt gca tca ttc att ccc ttg gag gcc atg				6376
Glu Glu Ile Val Glu Glu Cys Ala Ser Phe Ile Pro Leu Glu Ala Met				
	1965	1970	1975	
gtt gct gca atc gat aag ctg aag agt aca ata acc aca act agt cct				6424
Val Ala Ala Ile Asp Lys Leu Lys Ser Thr Ile Thr Thr Thr Ser Pro				
	1980	1985	1990	
ttc aca ttg gaa acc gcc ctt gaa aaa ctt aac acc ttt ctt ggg cct				6472
Phe Thr Leu Glu Thr Ala Leu Glu Lys Leu Asn Thr Phe Leu Gly Pro				
	1995	2000	2005	
cat gca gct aca atc ctt gct atc ata gag tat tgc tgt ggt tta gtc				6520
His Ala Ala Thr Ile Leu Ala Ile Ile Glu Tyr Cys Cys Gly Leu Val				
	2010	2015	2020	2025
act tta cct gac aat ccc ttt gca tca tgc gtg ttt gct ttc att gcg				6568

Thr	Leu	Pro	Asp	Asn	Pro	Phe	Ala	Ser	Cys	Val	Phe	Ala	Phe	Ile	Ala		
				2030					2035					2040			
ggt	att	act	acc	cca	cta	cct	cac	aag	atc	aaa	atg	ttc	ctg	tca	tta	6616	
Gly	Ile	Thr	Thr	Pro	Leu	Pro	His	Lys	Ile	Lys	Met	Phe	Leu	Ser	Leu		
			2045					2050					2055				
ttt	gga	ggc	gca	att	gcg	tcc	aag	ctt	aca	gac	gct	aga	ggc	gca	ctg	6664	
Phe	Gly	Gly	Ala	Ile	Ala	Ser	Lys	Leu	Thr	Asp	Ala	Arg	Gly	Ala	Leu		
			2060					2065				2070					
gcg	ttc	atg	atg	gcc	ggg	gct	gcg	gga	aca	gct	ctt	ggt	aca	tgg	aca	6712	
Ala	Phe	Met	Met	Ala	Gly	Ala	Ala	Gly	Thr	Ala	Leu	Gly	Thr	Trp	Thr		
			2075					2080				2085					
tcg	gtg	ggt	ttt	gtc	ttt	gac	atg	cta	ggc	ggc	tat	gct	gcc	gcc	tca	6760	
Ser	Val	Gly	Phe	Val	Phe	Asp	Met	Leu	Gly	Gly	Tyr	Ala	Ala	Ala	Ser		
2090						2095					2100				2105		
tcc	act	gct	tgc	ttg	aca	ttt	aaa	tgc	ttg	atg	ggt	gag	tgg	ccc	act	6808	
Ser	Thr	Ala	Cys	Leu	Thr	Phe	Lys	Cys	Leu	Met	Gly	Glu	Trp	Pro	Thr		
				2110						2115				2120			
atg	gat	cag	ctt	gct	ggt	tta	gtc	tac	tcc	gcg	ttc	aat	ccg	gcc	gca	6856	
Met	Asp	Gln	Leu	Ala	Gly	Leu	Val	Tyr	Ser	Ala	Phe	Asn	Pro	Ala	Ala		
			2125					2130					2135				
gga	gtt	gtg	ggc	gtc	ttg	tca	gct	tgt	gca	atg	ttt	gct	ttg	aca	aca	6904	
Gly	Val	Val	Gly	Val	Leu	Ser	Ala	Cys	Ala	Met	Phe	Ala	Leu	Thr	Thr		
			2140					2145				2150					
gca	ggg	cca	gat	cac	tgg	ccc	aac	aga	ctt	ctt	act	atg	ctt	gct	agg	6952	
Ala	Gly	Pro	Asp	His	Trp	Pro	Asn	Arg	Leu	Leu	Thr	Met	Leu	Ala	Arg		
			2155					2160				2165					
agc	aac	act	gta	tgt	aat	gag	tac	ttt	att	gcc	act	cgt	gac	atc	cgc	7000	
Ser	Asn	Thr	Val	Cys	Asn	Glu	Tyr	Phe	Ile	Ala	Thr	Arg	Asp	Ile	Arg		
2170						2175					2180				2185		
agg	aag	ata	ctg	ggc	att	ctg	gag	gca	tct	acc	ccc	tgg	agt	gtc	ata	7048	
Arg	Lys	Ile	Leu	Gly	Ile	Leu	Glu	Ala	Ser	Thr	Pro	Trp	Ser	Val	Ile		
			2190						2195					2200			
tca	gct	tgc	atc	cgt	tgg	ctc	cac	acc	ccg	acg	gag	gat	gat	tgc	ggc	7096	
Ser	Ala	Cys	Ile	Arg	Trp	Leu	His	Thr	Pro	Thr	Glu	Asp	Asp	Cys	Gly		
			2205					2210					2215				
ctc	att	gct	tgg	ggt	cta	gag	att	tgg	cag	tat	gtg	tgc	aat	ttc	ttt	7144	
Leu	Ile	Ala	Trp	Gly	Leu	Glu	Ile	Trp	Gln	Tyr	Val	Cys	Asn	Phe	Phe		
			2220					2225				2230					
gtg	att	tgc	ttt	aat	gtc	ctt	aaa	gct	gga	gtt	cag	agc	atg	gtt	aac	7192	
Val	Ile	Cys	Phe	Asn	Val	Leu	Lys	Ala	Gly	Val	Gln	Ser	Met	Val	Asn		
			2235					2240				2245					
att	cct	ggt	tgt	cct	ttc	tac	agc	tgc	cag	aag	ggg	tac	aag	ggc	ccc	7240	
Ile	Pro	Gly	Cys	Pro	Phe	Tyr	Ser	Cys	Gln	Lys	Gly	Tyr	Lys	Gly	Pro		
2250						2255					2260				2265		
tgg	att	gga	tca	ggt	atg	ctc	caa	gca	cgc	tgt	cca	tgc	ggt	gct	gaa	7288	
Trp	Ile	Gly	Ser	Gly	Met	Leu	Gln	Ala	Arg	Cys	Pro	Cys	Gly	Ala	Glu		
				2270						2275					2280		

ctc atc ttt tct gtt gag aat ggt ttt gca aaa ctt tac aaa gga ccc	7336
Leu Ile Phe Ser Val Glu Asn Gly Phe Ala Lys Leu Tyr Lys Gly Pro	
2285 2290 2295	
aga act tgt tca aat tac tgg aga ggg gct gtt cca gtc aac gct agg	7384
Arg Thr Cys Ser Asn Tyr Trp Arg Gly Ala Val Pro Val Asn Ala Arg	
2300 2305 2310	
ctg tgt ggg tcg gct aga ccg gac cca act gat tgg act agt ctt gtc	7432
Leu Cys Gly Ser Ala Arg Pro Asp Pro Thr Asp Trp Thr Ser Leu Val	
2315 2320 2325	
gtc aat tat ggc gtt agg gac tac tgt aaa tat gag aaa atg gga gat	7480
Val Asn Tyr Gly Val Arg Asp Tyr Cys Lys Tyr Glu Lys Met Gly Asp	
2330 2335 2340 2345	
cac att ttt gtt aca gca gta tcc tct cca aat gtc tgt ttc acc cag	7528
His Ile Phe Val Thr Ala Val Ser Ser Pro Asn Val Cys Phe Thr Gln	
2350 2355 2360	
gtg ccc cca acc ttg aga gct gca gtg gcc gtg gac ggc gta cag gtt	7576
Val Pro Pro Thr Leu Arg Ala Ala Val Ala Val Asp Gly Val Gln Val	
2365 2370 2375	
cag tgt tat cta ggt gag ccc aaa act cct tgg acg aca tct gct tgc	7624
Gln Cys Tyr Leu Gly Glu Pro Lys Thr Pro Trp Thr Thr Ser Ala Cys	
2380 2385 2390	
tgt tac ggt cct gac ggt aag ggt aaa act gtt aag ctt ccc ttc cgc	7672
Cys Tyr Gly Pro Asp Gly Lys Gly Lys Thr Val Lys Leu Pro Phe Arg	
2395 2400 2405	
gtt gac ggt cac aca cct ggt gtg cgc atg caa ctt aat ttg cgt gat	7720
Val Asp Gly His Thr Pro Gly Val Arg Met Gln Leu Asn Leu Arg Asp	
2410 2415 2420 2425	
gca ctt gag aca aat gac tgt aat tcc aca aac aac act cct agt gat	7768
Ala Leu Glu Thr Asn Asp Cys Asn Ser Thr Asn Asn Thr Pro Ser Asp	
2430 2435 2440	
gaa gcc gca gtg tcc gct ctt gtt ttc aaa cag gag ttg cgg cgt aca	7816
Glu Ala Ala Val Ser Ala Leu Val Phe Lys Gln Glu Leu Arg Arg Thr	
2445 2450 2455	
aac caa ttg ctt gag gca att tca gct ggc gtt gac acc acc aaa ctg	7864
Asn Gln Leu Leu Glu Ala Ile Ser Ala Gly Val Asp Thr Thr Lys Leu	
2460 2465 2470	
cca gcc ccc tcc atc gaa gag gta gtg gta aga aag cgc cag ttc cgg	7912
Pro Ala Pro Ser Ile Glu Glu Val Val Val Arg Lys Arg Gln Phe Arg	
2475 2480 2485	
gca aga act ggt tcg ctt acc ttg cct ccc cct ccg aga tcc gtc cca	7960
Ala Arg Thr Gly Ser Leu Thr Leu Pro Pro Pro Pro Arg Ser Val Pro	
2490 2495 2500 2505	
gga gtg tca tgt cct gaa agc ctg caa cga agt gac ccg tta gaa ggt	8008
Gly Val Ser Cys Pro Glu Ser Leu Gln Arg Ser Asp Pro Leu Glu Gly	
2510 2515 2520	
cct tca aac ctc cct cct tca cca cct gtt cta cag ttg gcc atg ccg	8056
Pro Ser Asn Leu Pro Pro Ser Pro Pro Val Leu Gln Leu Ala Met Pro	
2525 2530 2535	

atg ccc ctg ttg gga gcg ggt gag tgt aac cct ttc act gca att gga	8104
Met Pro Leu Leu Gly Ala Gly Glu Cys Asn Pro Phe Thr Ala Ile Gly	
2540 2545 2550	
tgt gca atg acc gaa aca ggc gga ggc cct gat gat tta ccc agt tac	8152
Cys Ala Met Thr Glu Thr Gly Gly Gly Pro Asp Asp Leu Pro Ser Tyr	
2555 2560 2565	
cct ccc aaa aag gag gtc tct gaa tgg tca gac gaa agt tgg tca acg	8200
Pro Pro Lys Lys Glu Val Ser Glu Trp Ser Asp Glu Ser Trp Ser Thr	
2570 2575 2580 2585	
gct aca acc gtt tcc agc tac gtt act ggc ccc ccg tac cct aag ata	8248
Ala Thr Thr Val Ser Ser Tyr Val Thr Gly Pro Pro Tyr Pro Lys Ile	
2590 2595 2600	
cgg gga aag gat tcc act cag tca gcc ccc gcc aaa cgg cct aca aaa	8296
Arg Gly Lys Asp Ser Thr Gln Ser Ala Pro Ala Lys Arg Pro Thr Lys	
2605 2610 2615	
aag aag ttg gga aag agt gag ttt tcg tgc agc atg agc tac acc tgg	8344
Lys Lys Leu Gly Lys Ser Glu Phe Ser Cys Ser Met Ser Tyr Thr Trp	
2620 2625 2630	
acc gac gtg att agc ttc aaa act gct tct aaa gtt ctg tct gca act	8392
Thr Asp Val Ile Ser Phe Lys Thr Ala Ser Lys Val Leu Ser Ala Thr	
2635 2640 2645	
cgg gcc atc act agt ggt ttc ctc aaa caa aga tca ttg gtg tat gtg	8440
Arg Ala Ile Thr Ser Gly Phe Leu Lys Gln Arg Ser Leu Val Tyr Val	
2650 2655 2660 2665	
act gag ccg cgg gat gcg gag ctt aga aaa caa aaa gtc act att aat	8488
Thr Glu Pro Arg Asp Ala Glu Leu Arg Lys Gln Lys Val Thr Ile Asn	
2670 2675 2680	
aga caa cct ctg ttc ccc cca tca tac cac aag caa gtg aga ttg gct	8536
Arg Gln Pro Leu Phe Pro Pro Ser Tyr His Lys Gln Val Arg Leu Ala	
2685 2690 2695	
aag gaa aaa gct tca aaa gtt gtc ggt gtc atg tgg gac tat gat gaa	8584
Lys Glu Lys Ala Ser Lys Val Val Gly Val Met Trp Asp Tyr Asp Glu	
2700 2705 2710	
gta gca gct cac acg ccc tct aag tct gct aag tcc cac atc act ggc	8632
Val Ala Ala His Thr Pro Ser Lys Ser Ala Lys Ser His Ile Thr Gly	
2715 2720 2725	
ctt cgg ggc act gat gtt cgt tct gga gca gcc cgc aag gct gtt ctg	8680
Leu Arg Gly Thr Asp Val Arg Ser Gly Ala Ala Arg Lys Ala Val Leu	
2730 2735 2740 2745	
gac ttg cag aag tgt gtc gag gca ggt gag ata ccg agt cat tat cgg	8728
Asp Leu Gln Lys Cys Val Glu Ala Gly Glu Ile Pro Ser His Tyr Arg	
2750 2755 2760	
caa act gtg ata gtt cca aag gag gag gtc ttc gtg aag acc ccc cag	8776
Gln Thr Val Ile Val Pro Lys Glu Glu Val Phe Val Lys Thr Pro Gln	
2765 2770 2775	
aaa cca aca aag aaa ccc cca agg ctt atc tcg tac ccc cac ctt gaa	8824
Lys Pro Thr Lys Lys Pro Pro Arg Leu Ile Ser Tyr Pro His Leu Glu	

2780	2785	2790	
atg aga tgt gtt gag aag atg tac tac ggt cag gtt gct cct gac gta			8872
Met Arg Cys Val Glu Lys Met Tyr Tyr Gly Gln Val Ala Pro Asp Val			
2795	2800	2805	
gtt aaa gct gtc atg gga gat gcg tac ggg ttt gtc gac cca cgt acc			8920
Val Lys Ala Val Met Gly Asp Ala Tyr Gly Phe Val Asp Pro Arg Thr			
2810	2815	2820 2825	
cgt gtc aag cgt ctg ttg tgg atg tgg tca ccc gat gca gtc gga gcc			8968
Arg Val Lys Arg Leu Leu Ser Met Trp Ser Pro Asp Ala Val Gly Ala			
2830	2835	2840	
aca tgc gat aca gtg tgt ttt gac agt acc atc aca ccc gag gat atc			9016
Thr Cys Asp Thr Val Cys Phe Asp Ser Thr Ile Thr Pro Glu Asp Ile			
2845	2850	2855	
atg gtg gag aca gac atc tac tca gca gct aaa ctg agt gac caa cac			9064
Met Val Glu Thr Asp Ile Tyr Ser Ala Ala Lys Leu Ser Asp Gln His			
2860	2865	2870	
cga gct ggc att cac acc att gcg agg cag tta tac gct gga gga ccg			9112
Arg Ala Gly Ile His Thr Ile Ala Arg Gln Leu Tyr Ala Gly Gly Pro			
2875	2880	2885	
atg atc gct tat gat ggc cga gag atc gga tat cgt agg tgt agg tct			9160
Met Ile Ala Tyr Asp Gly Arg Glu Ile Gly Tyr Arg Arg Cys Arg Ser			
2890	2895	2900 2905	
tcc ggc gtc tat act acc tca agt tcc aac agt ttg acc tgc tgg ctg			9208
Ser Gly Val Tyr Thr Thr Ser Ser Ser Asn Ser Leu Thr Cys Trp Leu			
2910	2915	2920	
aag gta aat gct gca gcc gaa cag gct ggc atg aag aac cct cgc ttc			9256
Lys Val Asn Ala Ala Ala Glu Gln Ala Gly Met Lys Asn Pro Arg Phe			
2925	2930	2935	
ctt att tgc ggc gat gat tgc acc gta att tgg aag agc gcc gga gca			9304
Leu Ile Cys Gly Asp Asp Cys Thr Val Ile Trp Lys Ser Ala Gly Ala			
2940	2945	2950	
gat gca gac aaa caa gca atg cgt gtc ttt gct agc tgg atg aag gtg			9352
Asp Ala Asp Lys Gln Ala Met Arg Val Phe Ala Ser Trp Met Lys Val			
2955	2960	2965	
atg ggt gca cca caa gat tgt gtg cct caa ccc aaa tac agt ttg gaa			9400
Met Gly Ala Pro Gln Asp Cys Val Pro Gln Pro Lys Tyr Ser Leu Glu			
2970	2975	2980 2985	
gaa tta aca tca tgc tca tca aat gtt acc tct gga att acc aaa agt			9448
Glu Leu Thr Ser Cys Ser Ser Asn Val Thr Ser Gly Ile Thr Lys Ser			
2990	2995	3000	
ggc aag cct tac tac ttt ctt aca aga gat cct cgt atc ccc ctt ggc			9496
Gly Lys Pro Tyr Tyr Phe Leu Thr Arg Asp Pro Arg Ile Pro Leu Gly			
3005	3010	3015	
agg tgc tct gcc gag ggt ctg gga tac aac ccc agt gct gcg tgg att			9544
Arg Cys Ser Ala Glu Gly Leu Gly Tyr Asn Pro Ser Ala Ala Trp Ile			
3020	3025	3030	
ggg tat cta ata cat cac tac cca tgt ttg tgg gtt agc cgt gtg ttg			9592

Gly Tyr Leu Ile His His Tyr Pro Cys Leu Trp Val Ser Arg Val Leu
 3035 3040 3045
 gct gtc cat ttc atg gag cag atg ctc ttt gag gac aaa ctt ccc gag 9640
 Ala Val His Phe Met Glu Gln Met Leu Phe Glu Asp Lys Leu Pro Glu
 3050 3055 3060 3065
 acg gtg acc ttt gac tgg tat ggg aaa aat tat acg gtg cct gta gaa 9688
 Thr Val Thr Phe Asp Trp Tyr Gly Lys Asn Tyr Thr Val Pro Val Glu
 3070 3075 3080
 gat ctg ccc agc atc att gct ggt gtg cac ggt att gag gct ttc tcg 9736
 Asp Leu Pro Ser Ile Ile Ala Gly Val His Gly Ile Glu Ala Phe Ser
 3085 3090 3095
 gtg gtg cgc tac acc aac gct gag atc ctc aga gtt tcc caa tca cta 9784
 Val Val Arg Tyr Thr Asn Ala Glu Ile Leu Arg Val Ser Gln Ser Leu
 3100 3105 3110
 aca gac atg acc atg ccc ccc ctg cga gcc tgg cga aag aaa gcc agg 9832
 Thr Asp Met Thr Met Pro Pro Leu Arg Ala Trp Arg Lys Lys Ala Arg
 3115 3120 3125
 gcg gtc ctc gcc agc gcc aag agg cgt ggc gga gca cac gca aaa ttg 9880
 Ala Val Leu Ala Ser Ala Lys Arg Arg Gly Gly Ala His Ala Lys Leu
 3130 3135 3140 3145
 gct cgc ttc ctt ctc tgg cat gct aca tct aga cct cta cca gat ttg 9928
 Ala Arg Phe Leu Leu Trp His Ala Thr Ser Arg Pro Leu Pro Asp Leu
 3150 3155 3160
 gat aag acg agc gtg gct cgg tac acc act ttc aat tat tgt gat gtt 9976
 Asp Lys Thr Ser Val Ala Arg Tyr Thr Thr Phe Asn Tyr Cys Asp Val
 3165 3170 3175
 tac tcc ccg gag ggg gat gtg ttt att aca cca cag aga aga ttg cag 10024
 Tyr Ser Pro Glu Gly Asp Val Phe Ile Thr Pro Gln Arg Arg Leu Gln
 3180 3185 3190
 aag ttt ctt gtg aag tat ttg gct gtc att gtt ttt gcc cta ggg ctc 10072
 Lys Phe Leu Val Lys Tyr Leu Ala Val Ile Val Phe Ala Leu Gly Leu
 3195 3200 3205
 att gct gtt gga tta gcc atc agc tgaacccccca aattcaaaat taactaacag 10126
 Ile Ala Val Gly Leu Ala Ile Ser
 3210 3215
 tttttttttt tttttttttt tttagggcag cggcaacagg ggagaccccg ggcttaacga 10186
 ccccgccgat gtgagtttgg cgaccatggt ggatcagaac cgtttcgggt gaagccatgg 10246
 totgaagggg atgacgtccc ttctgggtca tccacaaaaa ccgtctcggg tgggtgagga 10306
 gtcttggtg tgtgggaagc agtcagtata attcccgtcg tgtgtggtga cgcctcacga 10366
 cgtatttgtc cgctgtgcag agcgtagtag caagggctgc accccggttt ttgttccaag 10426
 cggagggcaa ccccgcttg gaattaaaaa ctcgagcaag acgtttcccg ttgaatatgg 10486
 ctcataacac cccttgatt actgtttatg taagcagaca gttttattgt tcatgatgat 10546
 atatttttat cttgtgcaat gtaacatcag agattttgag acacaacgtg gctttgttga 10606

ataaatcgaa cttttgctga gttgaaggat cagatcacgc atcttcccga caacgcagac 10666
cgttccgtgg caaagcaaaa gttcaaaatc accaactggc ccacctacaa caaagctctc 10726
atcaaccgtg gctccctcac tttctggctg gatgatgggg cgattcaggc ctgggatgag 10786
tcagcaacac cttcttcacg aggcagacct cagcgctagc ggagtgtata ctggcttact 10846
atgttggcac tgatgagggg gtcagtgaag tgcttcatgt ggcaggagaa aaaaggctgc 10906
accggtgogt cagcagaata tgtgatacag gatataattc gcttcctcgc tctactgactc 10966
gctacgctcg gtcgttcgac tgcggcgagc ggaaatggct tacgaacggg gcggagattt 11026
cctggaagat gccaggaaga tacttaacag ggaagtgaga gggccgcggc aaagccgttt 11086
ttccataggg tccgcccccc tgacaagcat cacgaaatct gacgctcaa tctagtgggtg 11146
cgaaacccga caggactata aagataccag gcgtttcccc tggcggctcc ctctgctgc 11206
ctcctgttcc tgcctttcgg tttaccgggtg tcattccgct gttatggccg cgtttgtctc 11266
attccacgoc tgacactcag ttccgggtag gcagttcgt ccaagctgga ctgtatgcac 11326
gaaccccccg ttcagtcgga ccgctgcgcc ttatccggta actatcgtct tgagtccaac 11386
ccggaaagac atgcaaaagc accactggca gcagccactg gtaattgatt tagaggagtt 11446
agtcttgaag tcatgcgccg gttaaggcta aactgaaagg acaagttttg gtgactgcgc 11506
tcctccaagc cagttacctc ggttcaaaga gttggtagct cagagaacct tcgaaaaacc 11566
gccctgcaag gcgggtttttt cgttttcaga gcaagagatt acgcgagac caaacgatc 11626
tcaagaagat catcttatta aggggtctga cgctcagtgg aacgaaaact cacgttaagg 11686
gatttttggtc atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg 11746
aagtttttaa tcaatctaaa gtatatatga gtaaacttgg tctgacagtt accaatgctt 11806
aatcagtgag gcacctatct cagcgatctg tctatttcgt tcatccatag ttgcctgact 11866
ccccgtcgtg tagataacta cgatacggga gggcttacca tctggcccca gtgctgcaat 11926
gataccgga gacccacgct caccggctcc agatttatca gcaataaacc agccagccgg 11986
aagggccgag cgcagaagtg gtctgcaac tttatccgcc tccatccagt ctattaattg 12046
ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat 12106
tgctgcaggc atcgtgggtg caogctcgtc gtttggtatg gcttcattca gctccgggtc 12166
ccaacgatca aggcgagtta catgatcccc catgttgtgc aaaaaagcgg ttagctcctt 12226
cggctcctcg atcgttgtca gaagtaagtt ggccgcagtg ttatcactca tggttatggc 12286
agcactgcat aattctctta ctgtcatgcc atccgtaaga tgcttttctg tgactgggtga 12346
gtactcaacc aagtcattct gagaatagt tatgcggcga ccgagttgct cttgcccggc 12406
gtcaacaagg gataataccg cgccacatag cagaacttta aaagtgtca tcattggaaa 12466
acgttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca gttcagtgta 12526

<213> Artificial Sequence

<223> Description of Artificial Sequence: Synthetic
Primer

 $\langle 222 \rangle \quad (446) \dots (9268)$

accacaaaaca	ctccagtttg	ttacactccg	ctaggaatgc	tcttgaggca	ccccccctag	60
cagggcggtgg	gggatttccc	ctgcccgctc	gcagaagggg	ggagccaacc	accttagtat	120
gtaggcgggcg	ggactcatga	cgctcgcggt	atgacaagcg	ccaagcttga	cttggatggc	180
cctgatgggc	gttcatgggt	tcggtgggtg	tggcgcttta	ggcagcctcc	acgcccacca	240
cctcccagat	agagcggcgg	cactgtaggg	aagaccgggg	accggtcact	accaaggacg	300
cagacctctt	tttgagtatc	acgcctccgg	aagtagttgg	gcaagcccac	ctatatgtgt	360
tgggatggtt	ggggttagcc	atccataccg	tactgcctga	tagggtcctt	gcgaggggat	420

cct cgt ggt aga cgt cag cct atc ccc aag gca cgt cgg ccc gag ggc 664
Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg Pro Glu Gly
60 65 70

agg acc tgg gct cag ccc ggg tac cct tgg ccc ctc tat ggc aat gag	712
Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu Tyr Gly Asn Glu	
75 80 85	
ggt tgc ggg tgg gcg gga tgg ctc ctg tct ccc cgt ggc tct cgg cct	760
Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg Gly Ser Arg Pro	
90 95 100 105	
agc tgg ggc ccc aca gac ccc cgg cgt agg tcg cgc aat ttg ggt aag	808
Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg Asn Leu Gly Lys	
110 115 120	
gtc atc gat acc ctt acg tgc ggc ttc gcc gac ctc atg ggg tac ata	856
Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile	
125 130 135	
ccg ctc gtc ggc gcc cct ctt gga ggc gct gcc agg gcc ctg gcg cat	904
Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His	
140 145 150	
ggc gtc cgg gtt ctg gaa gac ggc gtg aac tat gca aca ggg aac ctt	952
Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn Leu	
155 160 165	
cct ggt tgc tct ttc tct atc ttc ctt ctg gcc ctg ctc tct tgc ctg	1000
Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys Leu	
170 175 180 185	
act gtg ccc gct tca gcc tac caa gtg cgc aat tcc tcg ggg ctt tac	1048
Thr Val Pro Ala Ser Ala Tyr Gln Val Arg Asn Ser Ser Gly Leu Tyr	
190 195 200	
cat gtc acc aat gat tgc cct aac tca agt att gtg tac gag gcg gcc	1096
His Val Thr Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala	
205 210 215	
gat gcc atc ctg cac act ccg ggg tgt gtc cct tgc gtg cgc gag ggt	1144
Asp Ala Ile Leu His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly	
220 225 230	
aac gct tcg agg tgt tgg gtg gcg gtg acc ccc acg gtg gcc acc agg	1192
Asn Ala Ser Arg Cys Trp Val Ala Val Thr Pro Thr Val Ala Thr Arg	
235 240 245	
gac ggc aaa ctc ccc aca acg cag ctt cga cgt cat atc gat ctg ctt	1240
Asp Gly Lys Leu Pro Thr Thr Gln Leu Arg Arg His Ile Asp Leu Leu	
250 255 260 265	
gtc ggg agc gcc acc ctc tgc tcg gcc ctc tac gtg ggg gac ctg tgc	1288
Val Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys	
270 275 280	
ggg tct gtc ttt ctt gtt ggt caa ctg ttt acc ttc tct ccc agg cgc	1336
Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg Arg	
285 290 295	
cac tgg acg acg caa gac tgc aat tgt tct atc tat ccc ggc cat ata	1384
His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His Ile	
300 305 310	
acg ggt cat cgc atg gca tgg gat atg atg atg aac tgg tcc cct acg	1432
Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr	
315 320 325	

gca gcg ttg gtg gta gct cag ctg ctc cgg atc cca caa gcc atc atg	1480
Ala Ala Leu Val Val Ala Gln Leu Leu Arg Ile Pro Gln Ala Ile Met	
330 335 340 345	
gac atg atc gct ggt gct cac tgg gga gtc ctg gcg ggc ata gcg tat	1528
Asp Met Ile Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala Tyr	
350 355 360	
ttc tcc atg gtg ggg aac tgg gcg aag gtc ctg gta gtg ctg ctg cta	1576
Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu Leu	
365 370 375	
ttt gcc ggc gtc gac gcg gaa acc cac gtc acc ggg gga aat gcc ggc	1624
Phe Ala Gly Val Asp Ala Glu Thr His Val Thr Gly Gly Asn Ala Gly	
380 385 390	
cgc acc acg gct ggg ctt gtt ggt ctc ctt aca cca ggc gcc aag cag	1672
Arg Thr Thr Ala Gly Leu Val Gly Leu Leu Thr Pro Gly Ala Lys Gln	
395 400 405	
aac atc caa ctg atc aac acc aac ggc agt tgg cac atc aat agc acg	1720
Asn Ile Gln Leu Ile Asn Thr Asn Gly Ser Trp His Ile Asn Ser Thr	
410 415 420 425	
gcc ttg aat tgc aat gaa agc ctt aac acc ggc tgg tta gca ggg ctc	1768
Ala Leu Asn Cys Asn Glu Ser Leu Asn Thr Gly Trp Leu Ala Gly Leu	
430 435 440	
ttc tat caa cac aaa ttc aac tct tca ggc tgt cct gag agg ttg gcc	1816
Phe Tyr Gln His Lys Phe Asn Ser Ser Gly Cys Pro Glu Arg Leu Ala	
445 450 455	
agc tgc cga cgc ctt acc gat ttt gcc cag ggc tgg ggt cct atc agt	1864
Ser Cys Arg Arg Leu Thr Asp Phe Ala Gln Gly Trp Gly Pro Ile Ser	
460 465 470	
tat gcc aac gga agc ggc ctc gac gaa cgc ccc tac tgc tgg cac tac	1912
Tyr Ala Asn Gly Ser Gly Leu Asp Glu Arg Pro Tyr Cys Trp His Tyr	
475 480 485	
cct cca aga cct tgt ggc att gtg ccc gca aag agc gtg tgt ggc ccg	1960
Pro Pro Arg Pro Cys Gly Ile Val Pro Ala Lys Ser Val Cys Gly Pro	
490 495 500 505	
gta tat tgc ttc act ccc agc ccc gtg gtg gtg gga acg acc gac agg	2008
Val Tyr Cys Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg	
510 515 520	
tcg ggc gcg cct acc tac agc tgg ggt gca aat gat acg gat gtc ttc	2056
Ser Gly Ala Pro Thr Tyr Ser Trp Gly Ala Asn Asp Thr Asp Val Phe	
525 530 535	
gtc ctt aac aac acc agg cca ccg ctg ggc aat tgg ttc ggt tgt acc	2104
Val Leu Asn Asn Thr Arg Pro Pro Leu Gly Asn Trp Phe Gly Cys Thr	
540 545 550	
tgg atg aac tca act gga ttc acc aaa gtg tgc gga gcg ccc cct tgt	2152
Trp Met Asn Ser Thr Gly Phe Thr Lys Val Cys Gly Ala Pro Pro Cys	
555 560 565	
gtc atc gga ggg gtg ggc aac aac acc ttg ctc tgc ccc act gat tgc	2200
Val Ile Gly Gly Val Gly Asn Asn Thr Leu Leu Cys Pro Thr Asp Cys	

570	575	580	585	
ttc cgc aaa cat ccg gaa gcc aca tac tct cgg tgc ggc tcc ggt ccc	2248			
Phe Arg Lys His Pro Glu Ala Thr Tyr Ser Arg Cys Gly Ser Gly Pro				
590	595	600		
tgg att aca ccc agg tgc atg gtc gac tac ccg tat agg ctt tgg cac	2296			
Trp Ile Thr Pro Arg Cys Met Val Asp Tyr Pro Tyr Arg Leu Trp His				
605	610	615		
tat cct tgt acc atc aat tac acc ata ttc aaa gtc agg atg tac gtg	2344			
Tyr Pro Cys Thr Ile Asn Tyr Thr Ile Phe Lys Val Arg Met Tyr Val				
620	625	630		
gga ggg gtc gag cac agg ctg gaa gcg gcc tgc aac tgg acg cgg ggc	2392			
Gly Gly Val Glu His Arg Leu Glu Ala Ala Cys Asn Trp Thr Arg Gly				
635	640	645		
gaa cgc tgt gat ctg gaa gac agg gac agg tcc gag ctc agc ccg ttg	2440			
Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg Ser Glu Leu Ser Pro Leu				
650	655	660	665	
ctg ctg tcc acc aca cag tgg cag gtc ctt ccg tgt tct ttc acg acc	2488			
Leu Leu Ser Thr Thr Gln Trp Gln Val Leu Pro Cys Ser Phe Thr Thr				
670	675	680		
ctg cca gcc ttg tcc acc ggc ctc atc cac ctc cac cag aac att gtg	2536			
Leu Pro Ala Leu Ser Thr Gly Leu Ile His Leu His Gln Asn Ile Val				
685	690	695		
gac gtg cag tac ttg tac ggg gta ggg tca agc atc gcg tcc tgg gcc	2584			
Asp Val Gln Tyr Leu Tyr Gly Val Gly Ser Ser Ile Ala Ser Trp Ala				
700	705	710		
att aag tgg gag tac gtc gtt ctc ctg ttc ctt ctg ctt gca gac gcg	2632			
Ile Lys Trp Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala				
715	720	725		
cgc gtc tgc tcc tgc ttg tgg atg atg tta ctc ata tcc caa gcg gag	2680			
Arg Val Cys Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu				
730	735	740	745	
gcg gct ttg gag aac ctc gta ata ctc aat gca gca tcc ctg gcc ggg	2728			
Ala Ala Leu Glu Asn Leu Val Ile Leu Asn Ala Ala Ser Leu Ala Gly				
750	755	760		
acg cac ggt ctt gtg tcc ttc ctc gtg ttc ttc tgc ttt gcg tgg tat	2776			
Thr His Gly Leu Val Ser Phe Leu Val Phe Phe Cys Phe Ala Trp Tyr				
765	770	775		
ctg aag ggt agg tgg gtg ccc gga gcg gtc tac gcc ctc tac ggg atg	2824			
Leu Lys Gly Arg Trp Val Pro Gly Ala Val Tyr Ala Leu Tyr Gly Met				
780	785	790		
tgg cct ctc ctc ctg ctc ctg ctg gcg ttg cct cag cgg gca tac gca	2872			
Trp Pro Leu Leu Leu Leu Leu Leu Ala Leu Pro Gln Arg Ala Tyr Ala				
795	800	805		
ttt gat acc gag ata att gga ggg ctg aca ata cca cct gta gta gca	2920			
Phe Asp Thr Glu Ile Ile Gly Gly Leu Thr Ile Pro Pro Val Val Ala				
810	815	820	825	
tta gtt gtc atg tct cgt ttt ggc ttc ttt gct cac ttg tta cct cgc	2968			

Leu	Val	Val	Met	Ser	Arg	Phe	Gly	Phe	Phe	Ala	His	Leu	Leu	Pro	Arg	
			830						835					840		
tgt	gct	tta	gtt	aac	tcc	tat	ctt	tgg	caa	cgt	tgg	gag	aat	tgg	ttt	3016
Cys	Ala	Leu	Val	Asn	Ser	Tyr	Leu	Trp	Gln	Arg	Trp	Glu	Asn	Trp	Phe	
			845					850					855			
tgg	aac	gtt	aca	cta	aga	ccg	gag	agg	ttt	ttc	ctt	gtg	ctg	gtt	tgt	3064
Trp	Asn	Val	Thr	Leu	Arg	Pro	Glu	Arg	Phe	Phe	Leu	Val	Leu	Val	Cys	
		860					865					870				
ttc	ccc	ggg	gcg	aca	tat	gac	gcg	ctg	gtg	act	ttc	tgt	gtg	tgt	cac	3112
Phe	Pro	Gly	Ala	Thr	Tyr	Asp	Ala	Leu	Val	Thr	Phe	Cys	Val	Cys	His	
	875					880					885					
gta	gct	ctt	cta	tgt	tta	aca	tcc	agt	gca	gca	tcg	ttc	ttt	ggg	act	3160
Val	Ala	Leu	Leu	Cys	Leu	Thr	Ser	Ser	Ala	Ala	Ser	Phe	Phe	Gly	Thr	
890					895					900					905	
gac	tct	agg	gtt	agg	gcc	cat	aga	atg	ttg	gtg	cgt	ctc	gga	aag	tgt	3208
Asp	Ser	Arg	Val	Arg	Ala	His	Arg	Met	Leu	Val	Arg	Leu	Gly	Lys	Cys	
			910						915					920		
cat	gct	tgg	tat	tct	cat	tat	gtt	ctt	aag	ttt	ttc	ctc	tta	gtg	ttt	3256
His	Ala	Trp	Tyr	Ser	His	Tyr	Val	Leu	Lys	Phe	Phe	Leu	Leu	Val	Phe	
			925					930					935			
ggg	gag	aat	ggg	gtg	ttt	ttc	tat	aag	cac	ttg	cat	ggg	gat	gtc	ttg	3304
Gly	Glu	Asn	Gly	Val	Phe	Phe	Tyr	Lys	His	Leu	His	Gly	Asp	Val	Leu	
		940					945					950				
cct	aat	gat	ttt	gcc	tcg	aaa	cta	cca	ttg	caa	gag	cca	ttt	ttc	cct	3352
Pro	Asn	Asp	Phe	Ala	Ser	Lys	Leu	Pro	Leu	Gln	Glu	Pro	Phe	Phe	Pro	
	955					960					965					
ttt	gaa	ggc	aag	gca	agg	gtc	tat	agg	aat	gaa	gga	aga	cgc	ttg	gcg	3400
Phe	Glu	Gly	Lys	Ala	Arg	Val	Tyr	Arg	Asn	Glu	Gly	Arg	Arg	Leu	Ala	
970					975					980					985	
tgt	ggg	gac	acg	gtt	gat	ggg	ttg	ccc	gtt	gtt	gcg	cgt	ctc	ggc	gac	3448
Cys	Gly	Asp	Thr	Val	Asp	Gly	Leu	Pro	Val	Val	Ala	Arg	Leu	Gly	Asp	
			990					995					1000			
ctt	gtt	ttc	gca	ggg	ttg	gct	atg	ccg	cca	gat	ggg	tgg	gcc	att	acc	3496
Leu	Val	Phe	Ala	Gly	Leu	Ala	Met	Pro	Pro	Asp	Gly	Trp	Ala	Ile	Thr	
			1005					1010					1015			
gca	cct	ttt	acg	ctg	cag	tgt	ctc	tct	gaa	cgt	ggc	acg	ctg	tca	gcg	3544
Ala	Pro	Phe	Thr	Leu	Gln	Cys	Leu	Ser	Glu	Arg	Gly	Thr	Leu	Ser	Ala	
		1020				1025						1030				
atg	gca	gtg	gtc	atg	act	ggg	ata	gac	ccc	cga	act	tgg	act	gga	act	3592
Met	Ala	Val	Val	Met	Thr	Gly	Ile	Asp	Pro	Arg	Thr	Trp	Thr	Gly	Thr	
		1035				1040					1045					
atc	ttc	aga	tta	gga	tct	ctg	gcc	act	agc	tac	atg	gga	ttc	gtt	tgt	3640
Ile	Phe	Arg	Leu	Gly	Ser	Leu	Ala	Thr	Ser	Tyr	Met	Gly	Phe	Val	Cys	
1050					1055					1060					1065	
gac	aac	gtg	ttg	tat	act	gct	cac	cat	ggc	agc	aag	ggg	cgc	cgg	ttg	3688
Asp	Asn	Val	Leu	Tyr	Thr	Ala	His	His	Gly	Ser	Lys	Gly	Arg	Arg	Leu	
			1070						1075				1080			

gct cat ccc aca ggc tcc ata cac oca ata acc gtt gac gcg gct aat	3736
Ala His Pro Thr Gly Ser Ile His Pro Ile Thr Val Asp Ala Ala Asn	
1085 1090 1095	
gac cag gac atc tat caa cca cca tgt gga gct ggg tcc ctt act cgg	3784
Asp Gln Asp Ile Tyr Gln Pro Pro Cys Gly Ala Gly Ser Leu Thr Arg	
1100 1105 1110	
tgc tct tgc ggg gag acc aag ggg tat ctg gta aca cga ctg ggg tca	3832
Cys Ser Cys Gly Glu Thr Lys Gly Tyr Leu Val Thr Arg Leu Gly Ser	
1115 1120 1125	
ttg gtt gag gtc aac aaa tcc gat gac cct tat tgg tgt gtg tgc ggg	3880
Leu Val Glu Val Asn Lys Ser Asp Asp Pro Tyr Trp Cys Val Cys Gly	
1130 1135 1140 1145	
gcc ctt ccc atg gct gtt gcc aag ggt tct tca ggt gcc ccg att ctg	3928
Ala Leu Pro Met Ala Val Ala Lys Gly Ser Ser Gly Ala Pro Ile Leu	
1150 1155 1160	
tgc tcc tcc ggg cat gtt att ggg atg ttc acc gct gct aga aat tct	3976
Cys Ser Ser Gly His Val Ile Gly Met Phe Thr Ala Ala Arg Asn Ser	
1165 1170 1175	
ggc ggt tca gtc agt cag att agg gtt agg ccg ttg gtg tgt gct gga	4024
Gly Gly Ser Val Ser Gln Ile Arg Val Arg Pro Leu Val Cys Ala Gly	
1180 1185 1190	
tac cat ccc cag tac aca gca cat gcc act ctt gat aca aaa cct act	4072
Tyr His Pro Gln Tyr Thr Ala His Ala Thr Leu Asp Thr Lys Pro Thr	
1195 1200 1205	
gtg cct aac gag tat tca gtg caa att tta att gcc ccc act ggc agc	4120
Val Pro Asn Glu Tyr Ser Val Gln Ile Leu Ile Ala Pro Thr Gly Ser	
1210 1215 1220 1225	
ggc aag tca acc aaa tta cca ctt tct tac atg cag gag aag tat gag	4168
Gly Lys Ser Thr Lys Leu Pro Leu Ser Tyr Met Gln Glu Lys Tyr Glu	
1230 1235 1240	
gtc ttg gtc cta aat ccc agt gtg gct aca aca gca tca atg cca aag	4216
Val Leu Val Leu Asn Pro Ser Val Ala Thr Thr Ala Ser Met Pro Lys	
1245 1250 1255	
tac atg cac gcg acg tac ggc gtg aat cca aat tgc tat ttt aat ggc	4264
Tyr Met His Ala Thr Tyr Gly Val Asn Pro Asn Cys Tyr Phe Asn Gly	
1260 1265 1270	
aaa tgt acc aac aca ggg gct tca ctt acg tac agc aca tat ggc atg	4312
Lys Cys Thr Asn Thr Gly Ala Ser Leu Thr Tyr Ser Thr Tyr Gly Met	
1275 1280 1285	
tac ctg acc gga gca tgt tcc cgg aac tat gat gta atc att tgt gac	4360
Tyr Leu Thr Gly Ala Cys Ser Arg Asn Tyr Asp Val Ile Ile Cys Asp	
1290 1295 1300 1305	
gaa tgc cat gct acc gat gca acc acc gtg ttg ggc att gga aag gtc	4408
Glu Cys His Ala Thr Asp Ala Thr Thr Val Leu Gly Ile Gly Lys Val	
1310 1315 1320	
cta acc gaa gct cca tcc aaa aat gtt agg cta gtg gtt ctt gcc acg	4456
Leu Thr Glu Ala Pro Ser Lys Asn Val Arg Leu Val Val Leu Ala Thr	
1325 1330 1335	

gct acc ccc cct gga gta atc cct aca cca cat gcc aac ata act gag	4504
Ala Thr Pro Pro Gly Val Ile Pro Thr Pro His Ala Asn Ile Thr Glu	
1340 1345 1350	
att caa tta acc gat gaa ggc act atc ccc ttt cat gga aaa aag att	4552
Ile Gln Leu Thr Asp Glu Gly Thr Ile Pro Phe His Gly Lys Lys Ile	
1355 1360 1365	
aag gag gaa aat ctg aag aaa ggg aga cac ctt atc ttt gag gct acc	4600
Lys Glu Glu Asn Leu Lys Lys Gly Arg His Leu Ile Phe Glu Ala Thr	
1370 1375 1380 1385	
aaa aaa cac tgt gat gag ctt gct aac gag tta gct cga aag gga ata	4648
Lys Lys His Cys Asp Glu Leu Ala Asn Glu Leu Ala Arg Lys Gly Ile	
1390 1395 1400	
aca gct gtc tct tac tat agg gga tgt gac atc tca aaa atc cct gag	4696
Thr Ala Val Ser Tyr Tyr Arg Gly Cys Asp Ile Ser Lys Ile Pro Glu	
1405 1410 1415	
ggc gac tgt gta gta gtt gcc act gat gcc ttg tgt aca ggg tac act	4744
Gly Asp Cys Val Val Val Ala Thr Asp Ala Leu Cys Thr Gly Tyr Thr	
1420 1425 1430	
ggt gac ttt gat tcc gtg tat gac tgc agc ctc atg gta gaa ggc aca	4792
Gly Asp Phe Asp Ser Val Tyr Asp Cys Ser Leu Met Val Glu Gly Thr	
1435 1440 1445	
tgc cat gtt gac ctt gac cct act ttc acc atg ggt gtt cgt gtg tgc	4840
Cys His Val Asp Leu Asp Pro Thr Phe Thr Met Gly Val Arg Val Cys	
1450 1455 1460 1465	
ggg gtt tca gca ata gtt aaa ggc cag cgt agg ggc cgc aca ggc cgt	4888
Gly Val Ser Ala Ile Val Lys Gly Gln Arg Arg Gly Arg Thr Gly Arg	
1470 1475 1480	
ggg aga gct ggc ata tac tac tat gta gac ggg agt tgt acc cct tcg	4936
Gly Arg Ala Gly Ile Tyr Tyr Tyr Val Asp Gly Ser Cys Thr Pro Ser	
1485 1490 1495	
ggt atg gtt cct gaa tgc aac att gtt gaa gcc ttc gac gca gcc aag	4984
Gly Met Val Pro Glu Cys Asn Ile Val Glu Ala Phe Asp Ala Ala Lys	
1500 1505 1510	
gca tgg tat ggt ttg tca tca aca gaa gct caa act att ctg gac acc	5032
Ala Trp Tyr Gly Leu Ser Ser Thr Glu Ala Gln Thr Ile Leu Asp Thr	
1515 1520 1525	
tat cgc acc caa cct ggg tta cct gcg ata gga gca aat ttg gac gag	5080
Tyr Arg Thr Gln Pro Gly Leu Pro Ala Ile Gly Ala Asn Leu Asp Glu	
1530 1535 1540 1545	
tgg gct gat ctc ttt tct atg gtc aac ccc gaa cct tca ttt gtc aat	5128
Trp Ala Asp Leu Phe Ser Met Val Asn Pro Glu Pro Ser Phe Val Asn	
1550 1555 1560	
act gca aaa aga act gct gac aat tat gtt ttg ttg act gca gcc caa	5176
Thr Ala Lys Arg Thr Ala Asp Asn Tyr Val Leu Leu Thr Ala Ala Gln	
1565 1570 1575	
cta caa ctg tgt cat cag tat ggc tat gct gct ccc aat gac gca cca	5224
Leu Gln Leu Cys His Gln Tyr Gly Tyr Ala Ala Pro Asn Asp Ala Pro	

1580	1585	1590	
cgg ttg cag gga gcc cgg ctt ggg aaa aaa cct tgt ggg gtt ctg tgg			5272
Arg Trp Gln Gly Ala Arg Leu Gly Lys Lys Pro Cys Gly Val Leu Trp			
1595	1600	1605	
cgc ttg gac ggc gct gac gcc tgt cct ggc cca gag ccc agc gag gtg			5320
Arg Leu Asp Gly Ala Asp Ala Cys Pro Gly Pro Glu Pro Ser Glu Val			
1610	1615	1620	1625
acc aga tac caa atg tgc ttc act gaa gtc aat act tct ggg aca gcc			5368
Thr Arg Tyr Gln Met Cys Phe Thr Glu Val Asn Thr Ser Gly Thr Ala			
1630	1635	1640	
gca ctc gct gtt ggc gtt gga gtg gct atg gct tat cta gcc att gac			5416
Ala Leu Ala Val Gly Val Gly Val Ala Met Ala Tyr Leu Ala Ile Asp			
1645	1650	1655	
act ttt ggc gcc act tgt gtg cgg cgt tgc tgg tct att gca tca gtc			5464
Thr Phe Gly Ala Thr Cys Val Arg Arg Cys Trp Ser Ile Ala Ser Val			
1660	1665	1670	
cct acc ggt gct act gtc gcc cca gtg gtt gac gaa gaa gaa atc gtg			5512
Pro Thr Gly Ala Thr Val Ala Pro Val Val Asp Glu Glu Glu Ile Val			
1675	1680	1685	
gag gag tgt gca tca ttc att ccc ttg gag gcc atg gtt gct gca atc			5560
Glu Glu Cys Ala Ser Phe Ile Pro Leu Glu Ala Met Val Ala Ala Ile			
1690	1695	1700	1705
gat aag ctg aag agt aca atc acc aca act agt cct ttc aca ttg gaa			5608
Asp Lys Leu Lys Ser Thr Ile Thr Thr Ser Pro Phe Thr Leu Glu			
1710	1715	1720	
acc gcc ctt gaa aaa ctt aac acc ttt ctt ggg cct cat gca gct aca			5656
Thr Ala Leu Glu Lys Leu Asn Thr Phe Leu Gly Pro His Ala Ala Thr			
1725	1730	1735	
atc ctt gct atc ata gag tat tgc tgt ggt tta gtc act tta cct gac			5704
Ile Leu Ala Ile Ile Glu Tyr Cys Cys Gly Leu Val Thr Leu Pro Asp			
1740	1745	1750	
aat ccc ttt gca tca tgc gtg ttt gct ttc att gcg ggt att act acc			5752
Asn Pro Phe Ala Ser Cys Val Phe Ala Phe Ile Ala Gly Ile Thr Thr			
1755	1760	1765	
cca cta cct cac aag atc aaa atg ttc ctg tca tta ttt gga ggc gca			5800
Pro Leu Pro His Lys Ile Lys Met Phe Leu Ser Leu Phe Gly Gly Ala			
1770	1775	1780	1785
att gcg tcc aag ctt aca gac gct aga ggc gca ctg gcg ttc atg atg			5848
Ile Ala Ser Lys Leu Thr Asp Ala Arg Gly Ala Leu Ala Phe Met Met			
1790	1795	1800	
gcc ggg gct gcg gga aca gct ctt ggt aca tgg aca tcg gtg ggt ttt			5896
Ala Gly Ala Ala Gly Thr Ala Leu Gly Thr Trp Thr Ser Val Gly Phe			
1805	1810	1815	
gtc ttt gac atg cta ggc ggc tat gct gcc gcc tca tcc act gct tgc			5944
Val Phe Asp Met Leu Gly Gly Tyr Ala Ala Ala Ser Ser Thr Ala Cys			
1820	1825	1830	
ttg aca ttt aaa tgc ttg atg ggt gag tgg ccc act atg gat cag ctt			5992

Leu Thr Phe Lys Cys Leu Met Gly Glu Trp Pro Thr Met Asp Gln Leu	
1835 1840 1845	
gct ggt tta gtc tac tcc gcg ttc aat ccg gcc gca gga gtt gtg ggc	6040
Ala Gly Leu Val Tyr Ser Ala Phe Asn Pro Ala Ala Gly Val Val Gly	
1850 1855 1860 1865	
gtc ttg tca gct tgt gca atg ttt gct ttg aca aca gca ggg cca gat	6088
Val Leu Ser Ala Cys Ala Met Phe Ala Leu Thr Thr Ala Gly Pro Asp	
1870 1875 1880	
cac tgg ccc aac aga ctt ctt act atg ctt gct agg agc aac act gta	6136
His Trp Pro Asn Arg Leu Leu Thr Met Leu Ala Arg Ser Asn Thr Val	
1885 1890 1895	
tgt aat gag tac ttt att gcc act cgt gac atc cgc agg aag ata ctg	6184
Cys Asn Glu Tyr Phe Ile Ala Thr Arg Asp Ile Arg Arg Lys Ile Leu	
1900 1905 1910	
ggc att ctg gag gca tct acc ccc tgg agt gtc ata tca gct tgc atc	6232
Gly Ile Leu Glu Ala Ser Thr Pro Trp Ser Val Ile Ser Ala Cys Ile	
1915 1920 1925	
cgt tgg ctc cac acc ccg acg gag gat gat tgc ggc ctc att gct tgg	6280
Arg Trp Leu His Thr Pro Thr Glu Asp Asp Cys Gly Leu Ile Ala Trp	
1930 1935 1940 1945	
ggg cta gag att tgg cag tat gtg tgc aat ttc ttt gtg att tgc ttt	6328
Gly Leu Glu Ile Trp Gln Tyr Val Cys Asn Phe Phe Val Ile Cys Phe	
1950 1955 1960	
aat gtc ctt aaa gct gga gtt cag agc atg gtt aac att cct ggt tgt	6376
Asn Val Leu Lys Ala Gly Val Gln Ser Met Val Asn Ile Pro Gly Cys	
1965 1970 1975	
cct ttc tac agc tgc cag aag ggg tac aag ggc ccc tgg att gga tca	6424
Pro Phe Tyr Ser Cys Gln Lys Gly Tyr Lys Gly Pro Trp Ile Gly Ser	
1980 1985 1990	
ggg atg ctc caa gca cgc tgt cca tgc ggt gct gaa ctc atc ttt tct	6472
Gly Met Leu Gln Ala Arg Cys Pro Cys Gly Ala Glu Leu Ile Phe Ser	
1995 2000 2005	
gtt gag aat ggt ttt gca aaa ctt tac aaa gga ccc aga act tgt tca	6520
Val Glu Asn Gly Phe Ala Lys Leu Tyr Lys Gly Pro Arg Thr Cys Ser	
2010 2015 2020 2025	
aat tac tgg aga ggg gct gtt cca gtc aac gct agg ctg tgt ggg tcg	6568
Asn Tyr Trp Arg Gly Ala Val Pro Val Asn Ala Arg Leu Cys Gly Ser	
2030 2035 2040	
gct aga ccg gac cca act gat tgg act agt ctt gtc gtc aat tat ggc	6616
Ala Arg Pro Asp Pro Thr Asp Trp Thr Ser Leu Val Val Asn Tyr Gly	
2045 2050 2055	
gtt agg gac tac tgt aaa tat gag aaa atg gga gat cac att ttt gtt	6664
Val Arg Asp Tyr Cys Lys Tyr Glu Lys Met Gly Asp His Ile Phe Val	
2060 2065 2070	
aca gca gta tcc tct cca aat gtc tgt ttc acc cag gtg ccc cca acc	6712
Thr Ala Val Ser Ser Pro Asn Val Cys Phe Thr Gln Val Pro Pro Thr	
2075 2080 2085	

ttg aga gct gca gtg gcc gtg gac ggc gta cag gtt cag tgt tat cta	6760
Leu Arg Ala Ala Val Ala Val Asp Gly Val Gln Val Gln Cys Tyr Leu	
2090 2095 2100 2105	
ggg gag ccc aaa act cct tgg acg aca tct gct tgc tgt tac ggt cct	6808
Gly Glu Pro Lys Thr Pro Trp Thr Thr Ser Ala Cys Cys Tyr Gly Pro	
2110 2115 2120	
gac ggt aag ggt aaa act gtt aag ctt ccc ttc cgc gtt gac ggt cac	6856
Asp Gly Lys Gly Lys Thr Val Lys Leu Pro Phe Arg Val Asp Gly His	
2125 2130 2135	
aca cct ggt gtg cgc atg caa ctt aat ttg cgt gat gca ctt gag aca	6904
Thr Pro Gly Val Arg Met Gln Leu Asn Leu Arg Asp Ala Leu Glu Thr	
2140 2145 2150	
aat gac tgt aat tcc aca aac aac act cct agt gat gaa gcc gca gtg	6952
Asn Asp Cys Asn Ser Thr Asn Asn Thr Pro Ser Asp Glu Ala Ala Val	
2155 2160 2165	
tcc gct ctt gtt ttc aaa cag gag ttg cgg cgt aca aac caa ttg ctt	7000
Ser Ala Leu Val Phe Lys Gln Glu Leu Arg Arg Thr Asn Gln Leu Leu	
2170 2175 2180 2185	
gag gca att tca gct ggc gtt gac acc acc aaa ctg cca gcc ccc tcc	7048
Glu Ala Ile Ser Ala Gly Val Asp Thr Thr Lys Leu Pro Ala Pro Ser	
2190 2195 2200	
atc gaa gag gta gtg gta aga aag cgc cag ttc cgg gca aga act ggt	7096
Ile Glu Glu Val Val Val Arg Lys Arg Gln Phe Arg Ala Arg Thr Gly	
2205 2210 2215	
tcg ctt acc ttg cct ccc cct ccg aga tcc gtc cca gga gtg tca tgt	7144
Ser Leu Thr Leu Pro Pro Pro Arg Ser Val Pro Gly Val Ser Cys	
2220 2225 2230	
cct gaa agc ctg caa cga agt gac ccg tta gaa ggt cct tca aac ctc	7192
Pro Glu Ser Leu Gln Arg Ser Asp Pro Leu Glu Gly Pro Ser Asn Leu	
2235 2240 2245	
cct cct tca cca cct gtt cta cag ttg gcc atg ccg atg ccc ctg ttg	7240
Pro Pro Ser Pro Pro Val Leu Gln Leu Ala Met Pro Met Pro Leu Leu	
2250 2255 2260 2265	
gga gcg ggt gag tgt aac cct ttc act gca att gga tgt gca atg acc	7288
Gly Ala Gly Glu Cys Asn Pro Phe Thr Ala Ile Gly Cys Ala Met Thr	
2270 2275 2280	
gaa aca ggc gga ggc cct gat gat tta ccc agt tac cct ccc aaa aag	7336
Glu Thr Gly Gly Gly Pro Asp Asp Leu Pro Ser Tyr Pro Pro Lys Lys	
2285 2290 2295	
gag gtc tct gaa tgg tca gac gaa agt tgg tca acg gct aca acc gtt	7384
Glu Val Ser Glu Trp Ser Asp Glu Ser Trp Ser Thr Ala Thr Thr Val	
2300 2305 2310	
tcc agc tac gtt act ggc ccc ccg tac cct aag ata cgg gga aag gat	7432
Ser Ser Tyr Val Thr Gly Pro Pro Tyr Pro Lys Ile Arg Gly Lys Asp	
2315 2320 2325	
tcc act cag tca gcc ccc gcc aaa cgg cct aca aaa aag aag ttg gga	7480
Ser Thr Gln Ser Ala Pro Ala Lys Arg Pro Thr Lys Lys Lys Leu Gly	
2330 2335 2340 2345	

aag agt gag ttt tcg tgc agc atg agc tac acc tgg acc gac gtg att	7528
Lys Ser Glu Phe Ser Cys Ser Met Ser Tyr Thr Trp Thr Asp Val Ile	
2350 2355 2360	
agc ttc aaa act gct tct aaa gtt ctg tct gca act cgg gcc atc act	7576
Ser Phe Lys Thr Ala Ser Lys Val Leu Ser Ala Thr Arg Ala Ile Thr	
2365 2370 2375	
agt ggt ttc ctc aaa caa aga tca ttg gtg tat gtg act gag ccg cgg	7624
Ser Gly Phe Leu Lys Gln Arg Ser Leu Val Tyr Val Thr Glu Pro Arg	
2380 2385 2390	
gat gcg gag ctt aga aaa caa aaa gtc act att aat aga caa cct ctg	7672
Asp Ala Glu Leu Arg Lys Gln Lys Val Thr Ile Asn Arg Gln Pro Leu	
2395 2400 2405	
ttc ccc cca tca tac cac aag caa gtg aga ttg gct aag gaa aaa gct	7720
Phe Pro Pro Ser Tyr His Lys Gln Val Arg Leu Ala Lys Glu Lys Ala	
2410 2415 2420 2425	
tca aaa gtt gtc ggt gtc atg tgg gac tat gat gaa gta gca gct cac	7768
Ser Lys Val Val Gly Val Met Trp Asp Tyr Asp Glu Val Ala Ala His	
2430 2435 2440	
acg ccc tct aag tct gct aag tcc cac atc act ggc ctt cgg ggc act	7816
Thr Pro Ser Lys Ser Ala Lys Ser His Ile Thr Gly Leu Arg Gly Thr	
2445 2450 2455	
gat gtt cgt tct gga gca gcc cgc aag gct gtt ctg gac ttg cag aag	7864
Asp Val Arg Ser Gly Ala Ala Arg Lys Ala Val Leu Asp Leu Gln Lys	
2460 2465 2470	
tgt gtc gag gca ggt gag ata ccg agt cat tat cgg caa act gtg ata	7912
Cys Val Glu Ala Gly Glu Ile Pro Ser His Tyr Arg Gln Thr Val Ile	
2475 2480 2485	
gtt cca aag gag gag gtc ttc gtg aag acc ccc cag aaa cca aca aag	7960
Val Pro Lys Glu Glu Val Phe Val Lys Thr Pro Gln Lys Pro Thr Lys	
2490 2495 2500 2505	
aaa ccc cca agg ctt atc tcg tac ccc cac ctt gaa atg aga tgt gtt	8008
Lys Pro Pro Arg Leu Ile Ser Tyr Pro His Leu Glu Met Arg Cys Val	
2510 2515 2520	
gag aag atg tac tac ggt cag gtt gct cct gac gta gtt aaa gct gtc	8056
Glu Lys Met Tyr Tyr Gly Gln Val Ala Pro Asp Val Val Lys Ala Val	
2525 2530 2535	
atg gga gat gcg tac ggg ttt gtc gac cca cgt acc cgt gtc aag cgt	8104
Met Gly Asp Ala Tyr Gly Phe Val Asp Pro Arg Thr Arg Val Lys Arg	
2540 2545 2550	
ctg ttg tcg atg tgg tca ccc gat gca gtc gga gcc aca tgc gat aca	8152
Leu Leu Ser Met Trp Ser Pro Asp Ala Val Gly Ala Thr Cys Asp Thr	
2555 2560 2565	
gtg tgt ttt gac agt acc atc aca ccc gag gat atc atg gtg gag aca	8200
Val Cys Phe Asp Ser Thr Ile Thr Pro Glu Asp Ile Met Val Glu Thr	
2570 2575 2580 2585	
gac atc tac tca gca gct aaa ctc agt gac caa cac cga gct ggc att	8248
Asp Ile Tyr Ser Ala Ala Lys Leu Ser Asp Gln His Arg Ala Gly Ile	

2590	2595	2600	
cac acc att gcg agg cag tta tac gct gga gga ccg atg atc gct tat			8296
His Thr Ile Ala Arg Gln Leu Tyr Ala Gly Gly Pro Met Ile Ala Tyr			
2605	2610	2615	
gat ggc cga gag atc gga tat cgt agg tgt agg tct tcc ggc gtc tat			8344
Asp Gly Arg Glu Ile Gly Tyr Arg Arg Cys Arg Ser Ser Gly Val Tyr			
2620	2625	2630	
act acc tca agt tcc aac agt ttg acc tgc tgg ctg aag gta aat gct			8392
Thr Thr Ser Ser Ser Asn Ser Leu Thr Cys Trp Leu Lys Val Asn Ala			
2635	2640	2645	
gca gcc gaa cag gct ggc atg aag aac cct cgc ttc ctt att tgc ggc			8440
Ala Ala Glu Gln Ala Gly Met Lys Asn Pro Arg Phe Leu Ile Cys Gly			
2650	2655	2660	2665
gat gat tgc acc gta att tgg aag agc gcc gga gca gat gca gac aaa			8488
Asp Asp Cys Thr Val Ile Trp Lys Ser Ala Gly Ala Asp Ala Asp Lys			
2670	2675	2680	
caa gca atg cgt gtc ttt gct agc tgg atg aag gtg atg ggt gca cca			8536
Gln Ala Met Arg Val Phe Ala Ser Trp Met Lys Val Met Gly Ala Pro			
2685	2690	2695	
caa gat tgt gtg cct caa ccc aaa tac agt ttg gaa gaa tta aca tca			8584
Gln Asp Cys Val Pro Gln Pro Lys Tyr Ser Leu Glu Glu Leu Thr Ser			
2700	2705	2710	
tgc tca tca aat gtt acc tct gga att acc aaa agt ggc aag cct tac			8632
Cys Ser Ser Asn Val Thr Ser Gly Ile Thr Lys Ser Gly Lys Pro Tyr			
2715	2720	2725	
tac ttt ctt aca aga gat cct cgt atc ccc ctt ggc agg tgc tct gcc			8680
Tyr Phe Leu Thr Arg Asp Pro Arg Ile Pro Leu Gly Arg Cys Ser Ala			
2730	2735	2740	2745
gag ggt ctg gga tac aac ccc agt gct gcg tgg att ggg tat cta ata			8728
Glu Gly Leu Gly Tyr Asn Pro Ser Ala Ala Trp Ile Gly Tyr Leu Ile			
2750	2755	2760	
cat cac tac cca tgt ttg tgg gtt agc cgt gtg ttg gct gtc cat ttc			8776
His His Tyr Pro Cys Leu Trp Val Ser Arg Val Leu Ala Val His Phe			
2765	2770	2775	
atg gag cag atg ctc ttt gag gac aaa ctt ccc gag acg gtg acc ttt			8824
Met Glu Gln Met Leu Phe Glu Asp Lys Leu Pro Glu Thr Val Thr Phe			
2780	2785	2790	
gac tgg tat ggg aaa aat tat acg gtg cct gta gaa gat ctg ccc agc			8872
Asp Trp Tyr Gly Lys Asn Tyr Thr Val Pro Val Glu Asp Leu Pro Ser			
2795	2800	2805	
atc att gct ggt gtg cac ggt att gag gct ttc tcg gtg gtg cgc tac			8920
Ile Ile Ala Gly Val His Gly Ile Glu Ala Phe Ser Val Val Arg Tyr			
2810	2815	2820	2825
acc aac gct gag atc ctc aga gtt tcc caa tca cta aca gac atg acc			8968
Thr Asn Ala Glu Ile Leu Arg Val Ser Gln Ser Leu Thr Asp Met Thr			
2830	2835	2840	
atg ccc ccc ctg cga gcc tgg cga aag aaa gcc agg gcg gtc ctc gcc			9016

Met Pro Pro Leu Arg Ala Trp Arg Lys Lys Ala Arg Ala Val Leu Ala
 2845 2850 2855

agc gcc aag agg cgt ggc gga gca cac gca aaa ttg gct cgc ttc ctt 9064
 Ser Ala Lys Arg Arg Gly Gly Ala His Ala Lys Leu Ala Arg Phe Leu
 2860 2865 2870

ctc tgg cat gct aca tct aga cct cta cca gat ttg gat aag acg agc 9112
 Leu Trp His Ala Thr Ser Arg Pro Leu Pro Asp Leu Asp Lys Thr Ser
 2875 2880 2885

gtg gct cgg tac acc act ttc aat tat tgt gat gtt tac tcc ccg gag 9160
 Val Ala Arg Tyr Thr Thr Phe Asn Tyr Cys Asp Val Tyr Ser Pro Glu
 2890 2895 2900 2905

ggg gat gtg ttt att aca cca cag aga aga ttg cag aag ttt ctt gtg 9208
 Gly Asp Val Phe Ile Thr Pro Gln Arg Arg Leu Gln Lys Phe Leu Val
 2910 2915 2920

aag tat ttg gct gtc att gtt ttt gcc cta ggg ctc att gct gtt gga 9256
 Lys Tyr Leu Ala Val Ile Val Phe Ala Leu Gly Leu Ile Ala Val Gly
 2925 2930 2935

tta gcc atc agc tgaacccccca aattcaaaat taactaacag tttttttttt 9308
 Leu Ala Ile Ser
 2940

tttttttttt tttagggcag cggcaacagg ggagaccccg ggcttaacga ccccgccgat 9368

gtgagtttgg cgaccatggt ggatcagaac cgtttcgggt gaagccatgg tctgaagggg 9428

atgacgtccc ttctggctca tccacaaaaa cgtctcggg tgggtgagga gtcctggctg 9488

tgtgggaagc agtcagtata attcccgctg tgtgtggtga cgcctcacga cgtatttgtc 9548

cgctgtgcag agcgtagtac caagggtgc accccggttt ttgttccaag cggagggcaa 9608

ccccgcttg gaattaaaaa ctgagcaag acgtttcccg ttgaatatgg ctcataacac 9668

ccottgtatt actgtttatg taagcagaca gttttattgt tcatgatgat atatttttat 9728

cttgtgcaat gtaacatcag agattttgag acacaacgtg gctttgttga ataaatcgaa 9788

cttttctga gttgaaggat cagatcacgc atcttcccga caacgcagac cgttccgtgg 9848

caaagcaaaa gttcaaaatc accaactggt ccacctacaa caaagctctc atcaaccgtg 9908

gctccctcac tttctggctg gatgatgggg cgattcaggc ctggtatgag tcagcaacac 9968

cttcttcacg aggcagacct cagcgctagc ggagtgtata ctggcttact atgttggcac 10028

tgatgaggggt gtcagtgaag tgcttcatgt ggcaggagaa aaaaggctgc accggtgcgt 10088

cagcagaata tgtgatacag gatataattcc gcttcctcgc tactgactc gctacgctcg 10148

gtcgttcgac tgcggcgagc ggaaatggct tacgaacggg gcggagattt cctggaagat 10208

gccaggaaga tacttaacag ggaagtgaga gggccgcggc aaagccgttt ttccataggc 10268

tccgcccccc tgacaagcat caggaatct gacgctcaaa tcagtgtggt cgaaaccgga 10328

caggactata aagataccag gcgtttcccc tggcggctcc ctcgctgcgt ctctgttcc 10388

tgccttttcgg tttaccggtg tcattccgct gttatggccg cgtttgtctc attccacgcc 10448
tgacactcag ttccgggtag gcagttcgct ccaagctgga ctgtatgcac gaaccccccg 10508
ttcagtcgga ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggaagac 10568
atgcaaaagc accactggca gcagccactg gtaattgatt tagaggagtt agtcttgaag 10628
tcatgcgccg gttaaggcta aactgaaagg acaagttttg gtgactgcgc tcctccaagc 10688
cagttacctc ggttcaaaga gttggtagct cagagaacct tcgaaaaacc gccctgcaag 10748
gcggtttttt cgttttcaga gcaagagatt acgcgcagac caaacgatc tcaagaagat 10808
catcttatta aggggtctga cgctcagtgg aacgaaaact cacgttaagg gatthttggc 10868
atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg aagthtttaa 10928
tcaatctaaa gtatatatga gtaaaacttg tctgacagtt accaatgctt aatcagtga 10988
gcacctatct cagcgatctg tctatthtct tcatccatag ttgcctgact ccccgctctg 11048
tagataacta cgatacggga gggcttacca tctggcccca gtgctgcaat gataccgca 11108
gaccacgct caccggctcc agatttatca gcaataaacc agccagccgg aagggccgag 11168
cgcagaagtg gtctgcaac tttatccgcc tccatccagt ctattaattg ttgccgggaa 11228
gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat tgctgcaggc 11288
atcgtggtgt cacgctctgc gtttggtatg gcttcattca gctccggttc ccaacgatca 11348
aggcgagtta catgatcccc catgttgtgc aaaaaagcgg ttagctcctt cggctcctcg 11408
atcgttgtca gaagtaagtt ggccgcagtg ttatcactca tggttatggc agcactgcat 11468
aattctctta ctgtcatgcc atccgtaaga tgctthttctg tgactggtga gtactcaacc 11528
aagtcattct gagaatagtg tatgcggcga ccgagttgct cttgcccggc gtcaacacgg 11588
gataataccg cgccacatag cagaacttta aaagtgtca tcattggaaa acgttcttctg 11648
gggcgaaaac totcaaggat cttaccgctg ttgagatcca gttcgatgta acccactcgt 11708
gcaccaact gatcttcagc atctthttact ttcaccagcg thttctgggtg agcaaaaaca 11768
ggaaggcaaa atgccgcaaa aaagggaata agggcgacac ggaaatgttg aatactcata 11828
ctcttcttht ttcaatatta ttgaagcatt tatcagggtt attgtctcat gagcggatac 11888
atatttgaat gtatttagaa aaataaacia atagggttgc cgcgcacatt tccccgaaa 11948
gtgccacctg acgtcgacct gaggttaatta taaccggggc cctatatatg gatcccgtaa 12008
tacgactcac tatag 12023

<210> 21

<211> 12746

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<220>

<221> CDS

<222> (446)..(9991)

<400> 21

```

accacaaaca ctccagtttg ttacactccg ctaggaatgc tcctggagca cccccctag 60
cagggcgtgg gggatttccc ctgcccgtct gcagaagggt ggagccaacc accttagtat 120
gtaggcggcg ggactcatga cgctcgctg atgacaagcg ccaagcttga cttggatggc 180
cctgatgggc gttcatgggt tcggtggtgg tggcgcttta ggcagcctcc acgcccacca 240
cctcccagat agagcggcgg cactgtaggg aagaccgggg accggtcact accaaggacg 300
cagacctctt tttgagtatc acgcctccgg aagtagttgg gcaagccac ctatatgtgt 360
tgggatgggt ggggttagcc atccataccg tactgcctga tagggtcctt gcgaggggat 420
ctgggagtct cgtagaccgt agcac atg cct gtt att tct act caa aca agt 472
                Met Pro Val Ile Ser Thr Gln Thr Ser
                1                      5

cct gta cct gcg ccc aga acg cgc aag aac aag cag acg cag gct tca 520
Pro Val Pro Ala Pro Arg Thr Arg Lys Asn Lys Gln Thr Gln Ala Ser
 10                      15                      20                      25

tat cct gtg tcc att aaa aca tct gtt gaa agg gga caa cga gca aag 568
Tyr Pro Val Ser Ile Lys Thr Ser Val Glu Arg Gly Gln Arg Ala Lys
                30                      35                      40

cgc aaa gtc cag cgc gat gct cgg cct cgt aat tac aaa att gct ggt 616
Arg Lys Val Gln Arg Asp Ala Arg Pro Arg Asn Tyr Lys Ile Ala Gly
                45                      50                      55

atc cat gat ggc ttg cag aca ttg gct cag gct gct ttg cca gct cat 664
Ile His Asp Gly Leu Gln Thr Leu Ala Gln Ala Ala Leu Pro Ala His
                60                      65                      70

ggg tgg gga cgc caa gac cct cgc cat aag tct cgc aat ctt gga atc 712
Gly Trp Gly Arg Gln Asp Pro Arg His Lys Ser Arg Asn Leu Gly Ile
 75                      80                      85

ctt ctg gat tac cct ttg ggg tgg att ggt gat gtt aca act cac aca 760
Leu Leu Asp Tyr Pro Leu Gly Trp Ile Gly Asp Val Thr Thr His Thr
 90                      95                      100                      105

cct cta gta ggc ccg ctg gtg gca gga gcg gtc gtt cga cca gtc tgc 808
Pro Leu Val Gly Pro Leu Val Ala Gly Ala Val Val Arg Pro Val Cys
                110                      115                      120

cag ata gta cgc ttg ctg gag gat gga gtc aac tgg gct act ggt tgg 856
Gln Ile Val Arg Leu Leu Glu Asp Gly Val Asn Trp Ala Thr Gly Trp
                125                      130                      135

ttc ggt gtc cac ctt ttt gtg gta tgt ctg cta tct ttg gcc tgt ccc 904
Phe Gly Val His Leu Phe Val Val Cys Leu Leu Ser Leu Ala Cys Pro
 140                      145                      150

```

tgt agt ggg tac caa gtg cgc aat tcc tcg ggg ctt tac cat gtc acc	952
Cys Ser Gly Tyr Gln Val Arg Asn Ser Ser Gly Leu Tyr His Val Thr	
155 160 165	
aat gat tgc cct aac tca agt att gtg tac gag gcg gcc gat gcc atc	1000
Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile	
170 175 180 185	
ctg cac act ccg ggg tgt gtc cct tgc gtg cgc gag ggt aac gct tcg	1048
Leu His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser	
190 195 200	
agg tgt tgg gtg gcg gtg acc ccc acg gtg gcc acc agg gac ggc aaa	1096
Arg Cys Trp Val Ala Val Thr Pro Thr Val Ala Thr Arg Asp Gly Lys	
205 210 215	
ctc ccc aca acg cag ctt cga cgt cat atc gat ctg ctt gtc ggg agc	1144
Leu Pro Thr Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser	
220 225 230	
gcc acc ctc tgc tcg gcc ctc tac gtg ggg gac ctg tgc ggg tct gtc	1192
Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Ser Val	
235 240 245	
ttt ctt gtt ggt caa ctg ttt acc ttc tct ccc agg cgc cac tgg acg	1240
Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg Arg His Trp Thr	
250 255 260 265	
acg caa gac tgc aat tgt tct atc tat ccc ggc cat ata acg ggt cat	1288
Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His Ile Thr Gly His	
270 275 280	
cgc atg gca tgg gat atg atg atg aac tgg tcc cct acg gca gcg ttg	1336
Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Ala Ala Leu	
285 290 295	
gtg gta gct cag ctg ctc cgg atc cca caa gcc atc atg gac atg atc	1384
Val Val Ala Gln Leu Leu Arg Ile Pro Gln Ala Ile Met Asp Met Ile	
300 305 310	
gct ggt gct cac tgg gga gtc ctg gcg ggc ata gcg tat ttc tcc atg	1432
Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe Ser Met	
315 320 325	
gtg ggg aac tgg gcg aag gtc ctg gta gtg ctg ctg cta ttt gcc ggc	1480
Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu Leu Phe Ala Gly	
330 335 340 345	
gtc gac gcg gaa acc cac gtc acc ggg gga aat gcc ggc cgc acc acg	1528
Val Asp Ala Glu Thr His Val Thr Gly Gly Asn Ala Gly Arg Thr Thr	
350 355 360	
gct ggg ctt gtt ggt ctc ctt aca cca ggc gcc aag cag aac atc caa	1576
Ala Gly Leu Val Gly Leu Leu Thr Pro Gly Ala Lys Gln Asn Ile Gln	
365 370 375	
ctg atc aac acc aac ggc agt tgg cac atc aat agc acg gcc ttg aat	1624
Leu Ile Asn Thr Asn Gly Ser Trp His Ile Asn Ser Thr Ala Leu Asn	
380 385 390	
tgc aat gaa agc ctt aac acc ggc tgg tta gca ggg ctc ttc tat caa	1672
Cys Asn Glu Ser Leu Asn Thr Gly Trp Leu Ala Gly Leu Phe Tyr Gln	

395	400	405	
cac aaa ttc aac tct tca ggc tgt cct gag agg ttg gcc agc tgc cga His Lys Phe Asn Ser Ser Gly Cys Pro Glu Arg Leu Ala Ser Cys Arg 410 415 420 425			1720
cgc ctt acc gat ttt gcc cag ggc tgg ggt cct atc agt tat gcc aac Arg Leu Thr Asp Phe Ala Gln Gly Trp Gly Pro Ile Ser Tyr Ala Asn 430 435 440			1768
gga agc ggc ctc gac gaa cgc ccc tac tgc tgg cac tac cct cca aga Gly Ser Gly Leu Asp Glu Arg Pro Tyr Cys Trp His Tyr Pro Pro Arg 445 450 455			1816
cct tgt ggc att gtg ccc gca aag agc gtg tgt ggc ccg gta tat tgc Pro Cys Gly Ile Val Pro Ala Lys Ser Val Cys Gly Pro Val Tyr Cys 460 465 470			1864
ttc act ccc agc ccc gtg gtg gtg gga acg acc gac agg tcg ggc gcg Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg Ser Gly Ala 475 480 485			1912
cct acc tac agc tgg ggt gca aat gat acg gat gtc ttc gtc ctt aac Pro Thr Tyr Ser Trp Gly Ala Asn Asp Thr Asp Val Phe Val Leu Asn 490 495 500 505			1960
aac acc agg cca ccg ctg ggc aat tgg ttc ggt tgt acc tgg atg aac Asn Thr Arg Pro Pro Leu Gly Asn Trp Phe Gly Cys Thr Trp Met Asn 510 515 520			2008
tca act gga ttc acc aaa gtg tgc gga gcg ccc cct tgt gtc atc gga Ser Thr Gly Phe Thr Lys Val Cys Gly Ala Pro Pro Cys Val Ile Gly 525 530 535			2056
ggg gtg ggc aac aac acc ttg ctc tgc ccc act gat tgc ttc cgc aaa Gly Val Gly Asn Asn Thr Leu Leu Cys Pro Thr Asp Cys Phe Arg Lys 540 545 550			2104
cat ccg gaa gcc aca tac tct cgg tgc ggc tcc ggt ccc tgg att aca His Pro Glu Ala Thr Tyr Ser Arg Cys Gly Ser Gly Pro Trp Ile Thr 555 560 565			2152
ccc agg tgc atg gtc gac tac ccg tat agg ctt tgg cac tat cct tgt Pro Arg Cys Met Val Asp Tyr Pro Tyr Arg Leu Trp His Tyr Pro Cys 570 575 580 585			2200
acc atc aat tac acc ata ttc aaa gtc agg atg tac gtg gga ggg gtc Thr Ile Asn Tyr Thr Ile Phe Lys Val Arg Met Tyr Val Gly Gly Val 590 595 600			2248
gag cac agg ctg gaa gcg gcc tgc aac tgg acg cgg ggc gaa cgc tgt Glu His Arg Leu Glu Ala Ala Cys Asn Trp Thr Arg Gly Glu Arg Cys 605 610 615			2296
gat ctg gaa gac agg gac agg tcc gag ctc agc ccg ttg ctg ctg tcc Asp Leu Glu Asp Arg Asp Arg Ser Glu Leu Ser Pro Leu Leu Leu Ser 620 625 630			2344
acc aca cag tgg cag gtc ctt ccg tgt tct ttc acg acc ctg cca gcc Thr Thr Gln Trp Gln Val Leu Pro Cys Ser Phe Thr Thr Leu Pro Ala 635 640 645			2392
ttg tcc acc ggc ctc atc cac ctc cac cag aac att gtg gac gtg cag			2440

Leu	Ser	Thr	Gly	Leu	Ile	His	Leu	His	Gln	Asn	Ile	Val	Asp	Val	Gln	
650					655					660					665	
tac	ttg	tac	ggg	gta	ggg	tca	agc	atc	gcg	tcc	tgg	gcc	att	aag	tgg	2488
Tyr	Leu	Tyr	Gly	Val	Gly	Ser	Ser	Ile	Ala	Ser	Trp	Ala	Ile	Lys	Trp	
				670					675					680		
gag	tac	gtc	gtt	ctc	ctg	ttc	ctt	ctg	ctt	gca	gac	gcg	cgc	gtc	tgc	2536
Glu	Tyr	Val	Val	Leu	Leu	Phe	Leu	Leu	Leu	Ala	Asp	Ala	Arg	Val	Cys	
			685					690					695			
tcc	tgc	ttg	tgg	atg	atg	tta	ctc	ata	tcc	caa	gcg	gag	gcg	gct	ttg	2584
Ser	Cys	Leu	Trp	Met	Met	Leu	Leu	Ile	Ser	Gln	Ala	Glu	Ala	Ala	Leu	
		700					705					710				
gag	aac	ctc	gta	ata	ctc	aat	gca	gca	tcc	ctg	gcc	ggg	acg	cac	ggc	2632
Glu	Asn	Leu	Val	Ile	Leu	Asn	Ala	Ala	Ser	Leu	Ala	Gly	Thr	His	Gly	
	715					720					725					
ctt	gtg	tcc	ttc	ctc	gtg	ttc	ttc	tgc	ttt	gcg	tgg	tat	ctg	aag	ggc	2680
Leu	Val	Ser	Phe	Leu	Val	Phe	Phe	Cys	Phe	Ala	Trp	Tyr	Leu	Lys	Gly	
730					735					740					745	
agg	tgg	gtg	ccc	gga	gcg	gtc	tac	gcc	ctc	tac	ggg	atg	tgg	cct	ctc	2728
Arg	Trp	Val	Pro	Gly	Ala	Val	Tyr	Ala	Leu	Tyr	Gly	Met	Trp	Pro	Leu	
				750					755					760		
ctc	ctg	ctc	ctg	ctg	gcg	ttg	cct	cag	cgg	gca	tac	gca	ctg	gac	acg	2776
Leu	Leu	Leu	Leu	Leu	Ala	Leu	Pro	Gln	Arg	Ala	Tyr	Ala	Leu	Asp	Thr	
				765				770					775			
gag	gtg	gcc	gcg	tgc	tgt	ggc	ggc	gtt	gtt	ctt	gtc	ggg	tta	atg	gcg	2824
Glu	Val	Ala	Ala	Ser	Cys	Gly	Gly	Val	Val	Leu	Val	Gly	Leu	Met	Ala	
		780					785					790				
ctg	act	ctg	tgc	cca	tat	tac	aag	cgc	tat	atc	agc	tgg	tgc	atg	tgg	2872
Leu	Thr	Leu	Ser	Pro	Tyr	Tyr	Lys	Arg	Tyr	Ile	Ser	Trp	Cys	Met	Trp	
	795					800					805					
tgg	ctt	cag	tat	ttt	ctg	acc	aga	gta	gaa	gcg	caa	ctg	cac	gtg	tgg	2920
Trp	Leu	Gln	Tyr	Phe	Leu	Thr	Arg	Val	Glu	Ala	Gln	Leu	His	Val	Trp	
810					815					820					825	
gtt	ccc	ccc	ctc	aac	gtc	cgg	ggg	ggg	cgc	gat	gcc	gtc	atc	tta	ctc	2968
Val	Pro	Pro	Leu	Asn	Val	Arg	Gly	Gly	Arg	Asp	Ala	Val	Ile	Leu	Leu	
				830					835					840		
atg	tgt	gta	gta	cac	cgc	acc	ctg	gta	ttt	gac	atc	acc	aaa	cta	ctc	3016
Met	Cys	Val	Val	His	Pro	Thr	Leu	Val	Phe	Asp	Ile	Thr	Lys	Leu	Leu	
				845				850					855			
ctg	gcc	atc	ttc	gga	ccc	ctt	tgg	att	ctt	caa	gcc	agt	ttg	ctt	aaa	3064
Leu	Ala	Ile	Phe	Gly	Pro	Leu	Trp	Ile	Leu	Gln	Ala	Ser	Leu	Leu	Lys	
		860					865					870				
gtc	ccc	tac	ttc	gtg	cgc	gtt	caa	ggc	ctt	ctc	cgg	atc	tgc	gcg	cta	3112
Val	Pro	Tyr	Phe	Val	Arg	Val	Gln	Gly	Leu	Leu	Arg	Ile	Cys	Ala	Leu	
		875				880					885					
gcg	cgg	aag	ata	gcc	gga	ggc	cat	tac	gtg	caa	atg	gcc	atc	atc	aag	3160
Ala	Arg	Lys	Ile	Ala	Gly	Gly	His	Tyr	Val	Gln	Met	Ala	Ile	Ile	Lys	
890					895					900					905	

tta ggg gcg ctt act ggc acc tat gtg tat aac cat ctc acc cct ctt	3208
Leu Gly Ala Leu Thr Gly Thr Tyr Val Tyr Asn His Leu Thr Pro Leu	
910 915 920	
cga gac tgg gcg cac aac ggc ctg cga gat ctg gcc gtg gct gtg gaa	3256
Arg Asp Trp Ala His Asn Gly Leu Arg Asp Leu Ala Val Ala Val Glu	
925 930 935	
cca gtc gtc ttc tcc cga atg gag acc aag ctc atc acg tgg ggg gca	3304
Pro Val Val Phe Ser Arg Met Glu Thr Lys Leu Ile Thr Trp Gly Ala	
940 945 950	
gat acc gcc gcg tgc ggt gac atc atc aac ggc ttg ccc gtc tct gcc	3352
Asp Thr Ala Ala Cys Gly Asp Ile Ile Asn Gly Leu Pro Val Ser Ala	
955 960 965	
cgt agg ggc cag gag ata ctg ctt ggg cca gcc gac gga atg gtc tcc	3400
Arg Arg Gly Gln Glu Ile Leu Leu Gly Pro Ala Asp Gly Met Val Ser	
970 975 980 985	
aag ggg tgg agg ttg ctg gcg ccc atc acg gcg tac gcc cag cag acg	3448
Lys Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr	
990 995 1000	
aga ggc ctc cta ggg tgt ata atc acc agc ctg act ggc cgg gac aaa	3496
Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys	
1005 1010 1015	
aac caa gtg gag ggt gag gtc cag atc gtg tca act gct acc caa acc	3544
Asn Gln Val Glu Gly Glu Val Gln Ile Val Ser Thr Ala Thr Gln Thr	
1020 1025 1030	
ttc ctg gca acg tgc atc aat ggg gta tgc tgg act gtc tac cac ggg	3592
Phe Leu Ala Thr Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly	
1035 1040 1045	
gcc gga acg agg acc atc gca tca ccc aag ggt cct gtc atc cag atg	3640
Ala Gly Thr Arg Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met	
1050 1055 1060 1065	
tat acc aat gtg gac caa gac ctt gtg ggc tgg ccc gct cct caa ggt	3688
Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly	
1070 1075 1080	
tcc cgc tca ttg aca ccc tgt acc tgc ggc tcc tcg gac ctt tac ctg	3736
Ser Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu	
1085 1090 1095	
gtc acg agg cac gcc gat gtc att ccc gtg cgc cgg cga ggt gat agc	3784
Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser	
1100 1105 1110	
agg ggt agc ctg ctt tcg ccc cgg ccc att tcc tac ttg aaa ggc tcc	3832
Arg Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser	
1115 1120 1125	
tcg ggg ggt ccg ctg ttg tgc ccc gcg gga cac gcc gtg ggc cta ttc	3880
Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu Phe	
1130 1135 1140 1145	
agg gcc gcg gtg tgc acc cgt gga gtg gct aaa gcg gtg gac ttt atc	3928
Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile	
1150 1155 1160	

cct gtg gag aac cta ggg aca acc atg aga tcc ccg gtg ttc acg gac	3976
Pro Val Glu Asn Leu Gly Thr Thr Met Arg Ser Pro Val Phe Thr Asp	
1165 1170 1175	
aac tcc tct acg cgt atg cag atc ttc gtg aag acc ctg acg ggc aag	4024
Asn Ser Ser Thr Arg Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys	
1180 1185 1190	
acc atc act ctt gag gtc gag ccc agt gac acc atc gag aat gtc aag	4072
Thr Ile Thr Leu Glu Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys	
1195 1200 1205	
gcc aag atc caa gac aag gaa ggc atc cca cct gac cag cag agg ctg	4120
Ala Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu	
1210 1215 1220 1225	
ata ttc gcg ggc aaa cag ctg gag gat ggc cgc acc ctg tcc gac tac	4168
Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr	
1230 1235 1240	
aac atc cag aaa gag tcc acc ttg cac ctg gtg ctg cgt ctc cgc ggt	4216
Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly	
1245 1250 1255	
ggc gca cct ttt acg ctg cag tgt ctc tct gaa cgt ggc acg ctg tca	4264
Gly Ala Pro Phe Thr Leu Gln Cys Leu Ser Glu Arg Gly Thr Leu Ser	
1260 1265 1270	
gcg atg gca gtg gtc atg act ggt ata gac ccc cga act tgg act gga	4312
Ala Met Ala Val Val Met Thr Gly Ile Asp Pro Arg Thr Trp Thr Gly	
1275 1280 1285	
act atc ttc aga tta gga tct ctg gcc act agc tac atg gga ttc gtt	4360
Thr Ile Phe Arg Leu Gly Ser Leu Ala Thr Ser Tyr Met Gly Phe Val	
1290 1295 1300 1305	
tgt gac aac gtg ttg tat act gct cac cat ggc agc aag ggg cgc cgg	4408
Cys Asp Asn Val Leu Tyr Thr Ala His His Gly Ser Lys Gly Arg Arg	
1310 1315 1320	
ttg gct cat ccc aca ggc tcc ata cac cca ata acc gtt gac gcg gct	4456
Leu Ala His Pro Thr Gly Ser Ile His Pro Ile Thr Val Asp Ala Ala	
1325 1330 1335	
aat gac cag gac atc tat caa cca cca tgt gga gct ggg tcc ctt act	4504
Asn Asp Gln Asp Ile Tyr Gln Pro Pro Cys Gly Ala Gly Ser Leu Thr	
1340 1345 1350	
cgg tgc tct tgc ggg gag acc aag ggg tat ctg gta aca cga ctg ggg	4552
Arg Cys Ser Cys Gly Glu Thr Lys Gly Tyr Leu Val Thr Arg Leu Gly	
1355 1360 1365	
tca ttg gtt gag gtc aac aaa tcc gat gac cct tat tgg tgt gtg tgc	4600
Ser Leu Val Glu Val Asn Lys Ser Asp Asp Pro Tyr Trp Cys Val Cys	
1370 1375 1380 1385	
ggg gcc ctt ccc atg gct gtt gcc aag ggt tct tca ggt gcc ccg att	4648
Gly Ala Leu Pro Met Ala Val Ala Lys Gly Ser Ser Gly Ala Pro Ile	
1390 1395 1400	
ctg tgc tcc tcc ggg cat gtt att ggg atg ttc acc gct gct aga aat	4696
Leu Cys Ser Ser Gly His Val Ile Gly Met Phe Thr Ala Ala Arg Asn	

1405	1410	1415	
tct ggc ggt tca gtc agt cag att agg gtt agg ccg ttg gtg tgt gct Ser Gly Gly Ser Val Ser Gln Ile Arg Val Arg Pro Leu Val Cys Ala 1420 1425 1430			4744
gga tac cat ccc cag tac aca gca cat gcc act ctt gat aca aaa cct Gly Tyr His Pro Gln Tyr Thr Ala His Ala Thr Leu Asp Thr Lys Pro 1435 1440 1445			4792
act gtg cct aac gag tat tca gtg caa att tta att gcc ccc act ggc Thr Val Pro Asn Glu Tyr Ser Val Gln Ile Leu Ile Ala Pro Thr Gly 1450 1455 1460 1465			4840
agc ggc aag tca acc aaa tta cca ctt tct tac atg cag gag aag tat Ser Gly Lys Ser Thr Lys Leu Pro Leu Ser Tyr Met Gln Glu Lys Tyr 1470 1475 1480			4888
gag gtc ttg gtc cta aat ccc agt gtg gct aca aca gca tca atg cca Glu Val Leu Val Leu Asn Pro Ser Val Ala Thr Thr Ala Ser Met Pro 1485 1490 1495			4936
aag tac atg cac gcg acg tac ggc gtg aat cca aat tgc tat ttt aat Lys Tyr Met His Ala Thr Tyr Gly Val Asn Pro Asn Cys Tyr Phe Asn 1500 1505 1510			4984
ggc aaa tgt acc aac aca ggg gct tca ctt acg tac agc aca tat ggc Gly Lys Cys Thr Asn Thr Gly Ala Ser Leu Thr Tyr Ser Thr Tyr Gly 1515 1520 1525			5032
atg tac ctg acc gga gca tgt tcc cgg aac tat gat gta atc att tgt Met Tyr Leu Thr Gly Ala Cys Ser Arg Asn Tyr Asp Val Ile Ile Cys 1530 1535 1540 1545			5080
gac gaa tgc cat gct acc gat gca acc acc gtg ttg ggc att gga aag Asp Glu Cys His Ala Thr Asp Ala Thr Thr Val Leu Gly Ile Gly Lys 1550 1555 1560			5128
gtc cta acc gaa gct cca tcc aaa aat gtt agg cta gtg gtt ctt gcc Val Leu Thr Glu Ala Pro Ser Lys Asn Val Arg Leu Val Val Leu Ala 1565 1570 1575			5176
acg gct acc ccc cct gga gta atc cct aca cca cat gcc aac ata act Thr Ala Thr Pro Pro Gly Val Ile Pro Thr Pro His Ala Asn Ile Thr 1580 1585 1590			5224
gag att caa tta acc gat gaa ggc act atc ccc ttt cat gga aaa aag Glu Ile Gln Leu Thr Asp Glu Gly Thr Ile Pro Phe His Gly Lys Lys 1595 1600 1605			5272
att aag gag gaa aat ctg aag aaa ggg aga cac ctt atc ttt gag gct Ile Lys Glu Glu Asn Leu Lys Lys Gly Arg His Leu Ile Phe Glu Ala 1610 1615 1620 1625			5320
acc aaa aaa cac tgt gat gag ctt gct aac gag tta gct cga aag gga Thr Lys Lys His Cys Asp Glu Leu Ala Asn Glu Leu Ala Arg Lys Gly 1630 1635 1640			5368
ata aca gct gtc tct tac tat agg gga tgt gac atc tca aaa atc cct Ile Thr Ala Val Ser Tyr Tyr Arg Gly Cys Asp Ile Ser Lys Ile Pro 1645 1650 1655			5416
gag ggc gac tgt gta gta gtt gcc act gat gcc ttg tgt aca ggg tac			5464

Glu Gly Asp Cys Val Val Val Ala Thr Asp Ala Leu Cys Thr Gly Tyr	
1660	1665 1670
act ggt gac ttt gat tcc gtg tat gac tgc agc ctc atg gta gaa ggc	5512
Thr Gly Asp Phe Asp Ser Val Tyr Asp Cys Ser Leu Met Val Glu Gly	
1675	1680 1685
aca tgc cat gtt gac ctt gac cct act ttc acc atg ggt gtt cgt gtg	5560
Thr Cys His Val Asp Leu Asp Pro Thr Phe Thr Met Gly Val Arg Val	
1690	1695 1700 1705
tgc ggg gtt tca gca ata gtt aaa ggc cag cgt agg ggc cgc aca ggc	5608
Cys Gly Val Ser Ala Ile Val Lys Gly Gln Arg Arg Gly Arg Thr Gly	
	1710 1715 1720
cgt ggg aga gct ggc ata tac tac tat gta gac ggg agt tgt acc cct	5656
Arg Gly Arg Ala Gly Ile Tyr Tyr Tyr Val Asp Gly Ser Cys Thr Pro	
	1725 1730 1735
tcg ggt atg gtt cct gaa tgc aac att gtt gaa gcc ttc gac gca gcc	5704
Ser Gly Met Val Pro Glu Cys Asn Ile Val Glu Ala Phe Asp Ala Ala	
	1740 1745 1750
aag gca tgg tat ggt ttg tca tca aca gaa gct caa act att ctg gac	5752
Lys Ala Trp Tyr Gly Leu Ser Ser Thr Glu Ala Gln Thr Ile Leu Asp	
	1755 1760 1765
acc tat cgc acc caa cct ggg tta cct gcg ata gga gca aat ttg gac	5800
Thr Tyr Arg Thr Gln Pro Gly Leu Pro Ala Ile Gly Ala Asn Leu Asp	
1770	1775 1780 1785
gag tgg gct gat ctc ttt tct atg gtc aac ccc gaa cct tca ttt gtc	5848
Glu Trp Ala Asp Leu Phe Ser Met Val Asn Pro Glu Pro Ser Phe Val	
	1790 1795 1800
aat act gca aaa aga act gct gac aat tat gtt ttg ttg act gca gcc	5896
Asn Thr Ala Lys Arg Thr Ala Asp Asn Tyr Val Leu Leu Thr Ala Ala	
	1805 1810 1815
caa cta caa ctg tgt cat cag tat ggc tat gct gct ccc aat gac gca	5944
Gln Leu Gln Leu Cys His Gln Tyr Gly Tyr Ala Ala Pro Asn Asp Ala	
	1820 1825 1830
cca cgg tgg cag gga gcc cgg ctt ggg aaa aaa cct tgt ggg gtt ctg	5992
Pro Arg Trp Gln Gly Ala Arg Leu Gly Lys Lys Pro Cys Gly Val Leu	
	1835 1840 1845
tgg cgc ttg gac ggc gct gac gcc tgt cct ggc cca gag ccc agc gag	6040
Trp Arg Leu Asp Gly Ala Asp Ala Cys Pro Gly Pro Glu Pro Ser Glu	
1850	1855 1860 1865
gtg acc aga tac caa atg tgc ttc act gaa gtc aat act tct ggg aca	6088
Val Thr Arg Tyr Gln Met Cys Phe Thr Glu Val Asn Thr Ser Gly Thr	
	1870 1875 1880
gcc gca ctc gct gtt ggc gtt gga gtg gct atg gct tat cta gcc att	6136
Ala Ala Leu Ala Val Gly Val Gly Val Ala Met Ala Tyr Leu Ala Ile	
	1885 1890 1895
gac act ttt ggc gcc act tgt gtg cgg cgt tgc tgg tct att gca tca	6184
Asp Thr Phe Gly Ala Thr Cys Val Arg Arg Cys Trp Ser Ile Ala Ser	
	1900 1905 1910

gtc cct acc ggt gct act gtc gcc cca gtg gtt gac gaa gaa gaa atc	6232
Val Pro Thr Gly Ala Thr Val Ala Pro Val Val Asp Glu Glu Glu Ile	
1915 1920 1925	
gtg gag gag tgt gca tca ttc att ccc ttg gag gcc atg gtt gct gca	6280
Val Glu Glu Cys Ala Ser Phe Ile Pro Leu Glu Ala Met Val Ala Ala	
1930 1935 1940 1945	
atc gat aag ctg aag agt aca ata acc aca act agt cct ttc aca ttg	6328
Ile Asp Lys Leu Lys Ser Thr Ile Thr Thr Thr Ser Pro Phe Thr Leu	
1950 1955 1960	
gaa acc gcc ctt gaa aaa ctt aac acc ttt ctt ggg cct cat gca gct	6376
Glu Thr Ala Leu Glu Lys Leu Asn Thr Phe Leu Gly Pro His Ala Ala	
1965 1970 1975	
aca atc ctt gct atc ata gag tat tgc tgt ggt tta gtc act tta cct	6424
Thr Ile Leu Ala Ile Ile Glu Tyr Cys Cys Gly Leu Val Thr Leu Pro	
1980 1985 1990	
gac aat ccc ttt gca tca tgc gtg ttt gct ttc att gcg ggt att act	6472
Asp Asn Pro Phe Ala Ser Cys Val Phe Ala Phe Ile Ala Gly Ile Thr	
1995 2000 2005	
acc cca cta cct cac aag atc aaa atg ttc ctg tca tta ttt gga ggc	6520
Thr Pro Leu Pro His Lys Ile Lys Met Phe Leu Ser Leu Phe Gly Gly	
2010 2015 2020 2025	
gca att gcg tcc aag ctt aca gac gct aga ggc gca ctg gcg ttc atg	6568
Ala Ile Ala Ser Lys Leu Thr Asp Ala Arg Gly Ala Leu Ala Phe Met	
2030 2035 2040	
atg gcc ggg gct gcg gga aca gct ctt ggt aca tgg aca tcg gtg ggt	6616
Met Ala Gly Ala Ala Gly Thr Ala Leu Gly Thr Trp Thr Ser Val Gly	
2045 2050 2055	
ttt gtc ttt gac atg cta ggc ggc tat gct gcc gcc tca tcc act gct	6664
Phe Val Phe Asp Met Leu Gly Gly Tyr Ala Ala Ala Ser Ser Thr Ala	
2060 2065 2070	
tgc ttg aca ttt aaa tgc ttg atg ggt gag tgg ccc act atg gat cag	6712
Cys Leu Thr Phe Lys Cys Leu Met Gly Glu Trp Pro Thr Met Asp Gln	
2075 2080 2085	
ctt gct ggt tta gtc tac tcc gcg ttc aat ccg gcc gca gga gtt gtg	6760
Leu Ala Gly Leu Val Tyr Ser Ala Phe Asn Pro Ala Ala Gly Val Val	
2090 2095 2100 2105	
ggc gtc ttg tca gct tgt gca atg ttt gct ttg aca aca gca ggg cca	6808
Gly Val Leu Ser Ala Cys Ala Met Phe Ala Leu Thr Thr Ala Gly Pro	
2110 2115 2120	
gat cac tgg ccc aac aga ctt ctt act atg ctt gct agg agc aac act	6856
Asp His Trp Pro Asn Arg Leu Leu Thr Met Leu Ala Arg Ser Asn Thr	
2125 2130 2135	
gta tgt aat gag tac ttt att gcc act cgt gac atc cgc agg aag ata	6904
Val Cys Asn Glu Tyr Phe Ile Ala Thr Arg Asp Ile Arg Arg Lys Ile	
2140 2145 2150	
ctg ggc att ctg gag gca tct acc ccc tgg agt gtc ata tca gct tgc	6952
Leu Gly Ile Leu Glu Ala Ser Thr Pro Trp Ser Val Ile Ser Ala Cys	
2155 2160 2165	

atc cgt tgg ctc cac acc ccg acg gag gat gat tgc ggc ctc att gct Ile Arg Trp Leu His Thr Pro Thr Glu Asp Asp Cys Gly Leu Ile Ala 2170 2175 2180 2185	7000
tgg ggt cta gag att tgg cag tat gtg tgc aat ttc ttt gtg att tgc Trp Gly Leu Glu Ile Trp Gln Tyr Val Cys Asn Phe Phe Val Ile Cys 2190 2195 2200	7048
ttt aat gtc ctt aaa gct gga gtt cag agc atg gtt aac att cct ggt Phe Asn Val Leu Lys Ala Gly Val Gln Ser Met Val Asn Ile Pro Gly 2205 2210 2215	7096
tgt cct ttc tac agc tgc cag aag ggg tac aag ggc ccc tgg att gga Cys Pro Phe Tyr Ser Cys Gln Lys Gly Tyr Lys Gly Pro Trp Ile Gly 2220 2225 2230	7144
tca ggt atg ctc caa gca cgc tgt cca tgc ggt gct gaa ctc atc ttt Ser Gly Met Leu Gln Ala Arg Cys Pro Cys Gly Ala Glu Leu Ile Phe 2235 2240 2245	7192
tct gtt gag aat ggt ttt gca aaa ctt tac aaa gga ccc aga act tgt Ser Val Glu Asn Gly Phe Ala Lys Leu Tyr Lys Gly Pro Arg Thr Cys 2250 2255 2260 2265	7240
tca aat tac tgg aga ggg gct gtt cca gtc aac gct agg ctg tgt ggg Ser Asn Tyr Trp Arg Gly Ala Val Pro Val Asn Ala Arg Leu Cys Gly 2270 2275 2280	7288
tcg gct aga ccg gac cca act gat tgg act agt ctt gtc gtc aat tat Ser Ala Arg Pro Asp Pro Thr Asp Trp Thr Ser Leu Val Val Asn Tyr 2285 2290 2295	7336
ggc gtt agg gac tac tgt aaa tat gag aaa atg gga gat cac att ttt Gly Val Arg Asp Tyr Cys Lys Tyr Glu Lys Met Gly Asp His Ile Phe 2300 2305 2310	7384
gtt aca gca gta tcc tct cca aat gtc tgt ttc acc cag gtg ccc cca Val Thr Ala Val Ser Ser Pro Asn Val Cys Phe Thr Gln Val Pro Pro 2315 2320 2325	7432
acc ttg aga gct gca gtg gcc gtg gac ggc gta cag gtt cag tgt tat Thr Leu Arg Ala Ala Val Ala Val Asp Gly Val Gln Val Gln Cys Tyr 2330 2335 2340 2345	7480
cta ggt gag ccc aaa act cct tgg acg aca tct gct tgc tgt tac ggt Leu Gly Glu Pro Lys Thr Pro Trp Thr Thr Ser Ala Cys Cys Tyr Gly 2350 2355 2360	7528
cct gac ggt aag ggt aaa act gtt aag ctt ccc ttc cgc gtt gac ggt Pro Asp Gly Lys Gly Lys Thr Val Lys Leu Pro Phe Arg Val Asp Gly 2365 2370 2375	7576
cac aca cct ggt gtg cgc atg caa ctt aat ttg cgt gat gca ctt gag His Thr Pro Gly Val Arg Met Gln Leu Asn Leu Arg Asp Ala Leu Glu 2380 2385 2390	7624
aca aat gac tgt aat tcc aca aac aac act cct agt gat gaa gcc gca Thr Asn Asp Cys Asn Ser Thr Asn Asn Thr Pro Ser Asp Glu Ala Ala 2395 2400 2405	7672
gtg tcc gct ctt gtt ttc aaa cag gag ttg cgg cgt aca aac caa ttg Val Ser Ala Leu Val Phe Lys Gln Glu Leu Arg Arg Thr Asn Gln Leu	7720

2410	2415	2420	2425	
ctt gag gca att tca gct ggc gtt gac acc acc aaa ctg cca gcc ccc				7768
Leu Glu Ala Ile Ser Ala Gly Val Asp Thr Thr Lys Leu Pro Ala Pro				
2430	2435	2440		
tcc atc gaa gag gta gtg gta aga aag cgc cag ttc cgg gca aga act				7816
Ser Ile Glu Glu Val Val Val Arg Lys Arg Gln Phe Arg Ala Arg Thr				
2445	2450	2455		
ggt tcg ctt acc ttg cct ccc cct ccg aga tcc gtc cca gga gtg tca				7864
Gly Ser Leu Thr Leu Pro Pro Pro Pro Arg Ser Val Pro Gly Val Ser				
2460	2465	2470		
tgt cct gaa agc ctg caa cga agt gac ccg tta gaa ggt cct tca aac				7912
Cys Pro Glu Ser Leu Gln Arg Ser Asp Pro Leu Glu Gly Pro Ser Asn				
2475	2480	2485		
ctc cct cct tca cca cct gtt cta cag ttg gcc atg ccg atg ccc ctg				7960
Leu Pro Pro Ser Pro Pro Val Leu Gln Leu Ala Met Pro Met Pro Leu				
2490	2495	2500	2505	
ttg gga gcg ggt gag tgt aac cct ttc act gca att gga tgt gca atg				8008
Leu Gly Ala Gly Glu Cys Asn Pro Phe Thr Ala Ile Gly Cys Ala Met				
2510	2515	2520		
acc gaa aca ggc gga ggc cct gat gat tta ccc agt tac cct ccc aaa				8056
Thr Glu Thr Gly Gly Gly Pro Asp Asp Leu Pro Ser Tyr Pro Pro Lys				
2525	2530	2535		
aag gag gtc tct gaa tgg tca gac gaa agt tgg tca acg gct aca acc				8104
Lys Glu Val Ser Glu Trp Ser Asp Glu Ser Trp Ser Thr Ala Thr Thr				
2540	2545	2550		
gtt tcc agc tac gtt act ggc ccc ccg tac cct aag ata cgg gga aag				8152
Val Ser Ser Tyr Val Thr Gly Pro Pro Tyr Pro Lys Ile Arg Gly Lys				
2555	2560	2565		
gat tcc act cag tca gcc ccc gcc aaa ccg cct aca aaa aag aag ttg				8200
Asp Ser Thr Gln Ser Ala Pro Ala Lys Arg Pro Thr Lys Lys Lys Leu				
2570	2575	2580	2585	
gga aag agt gag ttt tcg tgc agc atg agc tac acc tgg acc gac gtg				8248
Gly Lys Ser Glu Phe Ser Cys Ser Met Ser Tyr Thr Trp Thr Asp Val				
2590	2595	2600		
att agc ttc aaa act gct tct aaa gtt ctg tct gca act cgg gcc atc				8296
Ile Ser Phe Lys Thr Ala Ser Lys Val Leu Ser Ala Thr Arg Ala Ile				
2605	2610	2615		
act agt ggt ttc ctc aaa caa aga tca ttg gtg tat gtg act gag ccg				8344
Thr Ser Gly Phe Leu Lys Gln Arg Ser Leu Val Tyr Val Thr Glu Pro				
2620	2625	2630		
cgg gat gcg gag ctt aga aaa caa aaa gtc act att aat aga caa cct				8392
Arg Asp Ala Glu Leu Arg Lys Gln Lys Val Thr Ile Asn Arg Gln Pro				
2635	2640	2645		
ctg ttc ccc cca tca tac cac aag caa gtg aga ttg gct aag gaa aaa				8440
Leu Phe Pro Pro Ser Tyr His Lys Gln Val Arg Leu Ala Lys Glu Lys				
2650	2655	2660	2665	
gct tca aaa gtt gtc ggt gtc atg tgg gac tat gat gaa gta gca gct				8488

Ala	Ser	Lys	Val	Val	Gly	Val	Met	Trp	Asp	Tyr	Asp	Glu	Val	Ala	Ala	
				2670					2675					2680		
cac	acg	ccc	tct	aag	tct	gct	aag	tcc	cac	atc	act	ggc	ctt	cgg	ggc	8536
His	Thr	Pro	Ser	Lys	Ser	Ala	Lys	Ser	His	Ile	Thr	Gly	Leu	Arg	Gly	
				2685				2690					2695			
act	gat	gtt	cgt	tct	gga	gca	gcc	cgc	aag	gct	gtt	ctg	gac	ttg	cag	8584
Thr	Asp	Val	Arg	Ser	Gly	Ala	Ala	Arg	Lys	Ala	Val	Leu	Asp	Leu	Gln	
				2700				2705					2710			
aag	tgt	gtc	gag	gca	ggg	gag	ata	ccg	agt	cat	tat	cgg	caa	act	gtg	8632
Lys	Cys	Val	Glu	Ala	Gly	Glu	Ile	Pro	Ser	His	Tyr	Arg	Gln	Thr	Val	
				2715				2720					2725			
ata	gtt	cca	aag	gag	gag	gtc	ttc	gtg	aag	acc	ccc	cag	aaa	cca	aca	8680
Ile	Val	Pro	Lys	Glu	Glu	Val	Phe	Val	Lys	Thr	Pro	Gln	Lys	Pro	Thr	
				2730				2735					2740		2745	
aag	aaa	ccc	cca	agg	ctt	atc	tcg	tac	ccc	cac	ctt	gaa	atg	aga	tgt	8728
Lys	Lys	Pro	Pro	Arg	Leu	Ile	Ser	Tyr	Pro	His	Leu	Glu	Met	Arg	Cys	
				2750				2755						2760		
gtt	gag	aag	atg	tac	tac	ggg	cag	gtt	gct	cct	gac	gta	gtt	aaa	gct	8776
Val	Glu	Lys	Met	Tyr	Tyr	Gly	Gln	Val	Ala	Pro	Asp	Val	Val	Lys	Ala	
				2765				2770						2775		
gtc	atg	gga	gat	gcg	tac	ggg	ttt	gtc	gac	cca	cgt	acc	cgt	gtc	aag	8824
Val	Met	Gly	Asp	Ala	Tyr	Gly	Phe	Val	Asp	Pro	Arg	Thr	Arg	Val	Lys	
				2780				2785					2790			
cgt	ctg	ttg	tcg	atg	tgg	tca	ccc	gat	gca	gtc	gga	gcc	aca	tgc	gat	8872
Arg	Leu	Leu	Ser	Met	Trp	Ser	Pro	Asp	Ala	Val	Gly	Ala	Thr	Cys	Asp	
				2795				2800					2805			
aca	gtg	tgt	ttt	gac	agt	acc	atc	aca	ccc	gag	gat	atc	atg	gtg	gag	8920
Thr	Val	Cys	Phe	Asp	Ser	Thr	Ile	Thr	Pro	Glu	Asp	Ile	Met	Val	Glu	
				2810				2815					2820		2825	
aca	gac	atc	tac	tca	gca	gct	aaa	ctc	agt	gac	caa	cac	cga	gct	ggc	8968
Thr	Asp	Ile	Tyr	Ser	Ala	Ala	Lys	Leu	Ser	Asp	Gln	His	Arg	Ala	Gly	
				2830				2835						2840		
att	cac	acc	att	gcg	agg	cag	tta	tac	gct	gga	gga	ccg	atg	atc	gct	9016
Ile	His	Thr	Ile	Ala	Arg	Gln	Leu	Tyr	Ala	Gly	Gly	Pro	Met	Ile	Ala	
				2845				2850					2855			
tat	gat	ggc	cga	gag	atc	gga	tat	cgt	agg	tgt	agg	tct	tcc	ggc	gtc	9064
Tyr	Asp	Gly	Arg	Glu	Ile	Gly	Tyr	Arg	Arg	Cys	Arg	Ser	Ser	Gly	Val	
				2860				2865					2870			
tat	act	acc	tca	agt	tcc	aac	agt	ttg	acc	tgc	tgg	ctg	aag	gta	aat	9112
Tyr	Thr	Thr	Ser	Ser	Ser	Asn	Ser	Leu	Thr	Cys	Trp	Leu	Lys	Val	Asn	
				2875				2880					2885			
gct	gca	gcc	gaa	cag	gct	ggc	atg	aag	aac	cct	cgc	ttc	ctt	att	tgc	9160
Ala	Ala	Ala	Glu	Gln	Ala	Gly	Met	Lys	Asn	Pro	Arg	Phe	Leu	Ile	Cys	
				2890				2895					2900		2905	
ggc	gat	gat	tgc	acc	gta	att	tgg	aag	agc	gcc	gga	gca	gat	gca	gac	9208
Gly	Asp	Asp	Cys	Thr	Val	Ile	Trp	Lys	Ser	Ala	Gly	Ala	Asp	Ala	Asp	
				2910				2915						2920		

aaa caa gca atg cgt gtc ttt gct agc tgg atg aag gtg atg ggt gca	9256
Lys Gln Ala Met Arg Val Phe Ala Ser Trp Met Lys Val Met Gly Ala	
2925 2930 2935	
cca caa gat tgt gtg cct caa ccc aaa tac agt ttg gaa gaa tta aca	9304
Pro Gln Asp Cys Val Pro Gln Pro Lys Tyr Ser Leu Glu Glu Leu Thr	
2940 2945 2950	
tca tgc tca tca aat gtt acc tct gga att acc aaa agt ggc aag cct	9352
Ser Cys Ser Ser Asn Val Thr Ser Gly Ile Thr Lys Ser Gly Lys Pro	
2955 2960 2965	
tac tac ttt ctt aca aga gat cct cgt atc ccc ctt ggc agg tgc tct	9400
Tyr Tyr Phe Leu Thr Arg Asp Pro Arg Ile Pro Leu Gly Arg Cys Ser	
2970 2975 2980 2985	
gcc gag ggt ctg gga tac aac ccc agt gct gcg tgg att ggg tat cta	9448
Ala Glu Gly Leu Gly Tyr Asn Pro Ser Ala Ala Trp Ile Gly Tyr Leu	
2990 2995 3000	
ata cat cac tac cca tgt ttg tgg gtt agc cgt gtg ttg gct gtc cat	9496
Ile His His Tyr Pro Cys Leu Trp Val Ser Arg Val Leu Ala Val His	
3005 3010 3015	
ttc atg gag cag atg ctc ttt gag gac aaa ctt ccc gag acg gtg acc	9544
Phe Met Glu Gln Met Leu Phe Glu Asp Lys Leu Pro Glu Thr Val Thr	
3020 3025 3030	
ttt gac tgg tat ggg aaa aat tat acg gtg cct gta gaa gat ctg ccc	9592
Phe Asp Trp Tyr Gly Lys Asn Tyr Thr Val Pro Val Glu Asp Leu Pro	
3035 3040 3045	
agc atc att gct ggt gtg cac ggt att gag gct ttc tcg gtg gtg cgc	9640
Ser Ile Ile Ala Gly Val His Gly Ile Glu Ala Phe Ser Val Val Arg	
3050 3055 3060 3065	
tac acc aac gct gag atc ctc aga gtt tcc caa tca cta aca gac atg	9688
Tyr Thr Asn Ala Glu Ile Leu Arg Val Ser Gln Ser Leu Thr Asp Met	
3070 3075 3080	
acc atg ccc ccc ctg cga gcc tgg cga aag aaa gcc agg gcg gtc ctc	9736
Thr Met Pro Pro Leu Arg Ala Trp Arg Lys Lys Ala Arg Ala Val Leu	
3085 3090 3095	
gcc agc gcc aag agg cgt ggc gga gca cac gca aaa ttg gct cgc ttc	9784
Ala Ser Ala Lys Arg Arg Gly Gly Ala His Ala Lys Leu Ala Arg Phe	
3100 3105 3110	
ctt ctc tgg cat gct aca tct aga cct cta cca gat ttg gat aag acg	9832
Leu Leu Trp His Ala Thr Ser Arg Pro Leu Pro Asp Leu Asp Lys Thr	
3115 3120 3125	
agc gtg gct cgg tac acc act ttc aat tat tgt gat gtt tac tcc ccg	9880
Ser Val Ala Arg Tyr Thr Thr Phe Asn Tyr Cys Asp Val Tyr Ser Pro	
3130 3135 3140 3145	
gag ggg gat gtg ttt att aca cca cag aga aga ttg cag aag ttt ctt	9928
Glu Gly Asp Val Phe Ile Thr Pro Gln Arg Arg Leu Gln Lys Phe Leu	
3150 3155 3160	
gtg aag tat ttg gct gtc att gtt ttt gcc cta ggg ctc att gct gtt	9976
Val Lys Tyr Leu Ala Val Ile Val Phe Ala Leu Gly Leu Ile Ala Val	
3165 3170 3175	

gga tta gcc atc agc tgaacccccca aattcaaaat taactaacag tttttttttt 10031
Gly Leu Ala Ile Ser
3180

tttttttttt ttttagggcag cggcaacagg ggagaccccg ggottaacga ccccgccgat 10091
gtgagtttgg cgaccatggg ggatcagaac cgtttcgggt gaagccatgg tctgaagggg 10151
atgacgtccc ttctggctca tccacaaaaa ccgtctcggg tgggtgagga gtcctggctg 10211
tgtgggaagc agtcagtata attcccgctg tgtgtggtga cgcctcacga cgtatttgtc 10271
cgctgtgcag agcgtagtac caagggctgc accccggttt ttgttccaag cggagggcaa 10331
ccccgcttg gaattaaaaa ctgagcaag acgtttcccg ttgaatatgg ctcataacac 10391
cccttgatatt actgtttatg taagcagaca gttttattgt tcatgatgat atatttttat 10451
cttgtgcaat gtaacatcag agattttgag acacaacgtg gctttgttga ataaatcgaa 10511
cttttgctga gttgaaggat cagatcacgc atcttcccga caacgcagac cgttccgtgg 10571
caaagcaaaa gttcaaaatc accaactggg ccacctacaa caaagctctc atcaaccgtg 10631
gctccctcac tttctggctg gatgatgggg cgattcaggc ctggtatgag tcagcaacac 10691
cttcttcacg aggcagacct cagcgctagc ggagtgtata ctggcttact atgttggcac 10751
tgatgagggg gtcagtgaag tgcttcatgt ggcaggagaa aaaaggctgc accggtgcgt 10811
cagcagaata tgtgatacag gatataattcc gcttcctcgc tcaactgactc gctacgctcg 10871
gtcgttcgac tgcggcgagc ggaaatggct tacgaacggg gcggagattt cctggaagat 10931
gccaggaaga tacttaacag ggaagtgaag gggccgcggc aaagccgttt ttccataggc 10991
tccgcccccc tgacaagcat cagcaaatct gacgctcaaa tcagtgggtg cgaaacccga 11051
caggactata aagataccag gcgtttcccc tggcggctcc ctctgtcgct ctctgttcc 11111
tgcccttcgg tttaccggtg tcattccgct gttatggccg cgtttgtctc attccacgcc 11171
tgacactcag ttccgggtag gcagttcgct ccaagctgga ctgtatgcac gaaccccccg 11231
ttcagtcgga ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggaagac 11291
atgcaaaagc accactggca gcagccactg gtaattgatt tagaggagt agtcttgaag 11351
tcatgcgccg gttaaggcta aactgaaagg acaagttttg gtgactgcgc tcctccaagc 11411
cagttacctc ggttcaaaga gttggtagct cagagaacct tcgaaaaacc gccctgcaag 11471
gcgggttttt cgttttcaga gcaagagatt acgcgcagac caaaacgatc tcaagaagat 11531
catcttatta aggggtctga cgctcagtgg aacgaaaact cacgttaagg gatthttggtc 11591
atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttaaa 11651
tcaatctaaa gtatatatga gtaaaccttg tctgacagtt accaatgctt aatcagtgag 11711
gcacctatct cagcgatctg tctatttcgt tcatccatag ttgocctgact ccccgctcgtg 11771

tagataacta cgatacggga gggcttacca tctggcccca gtgctgcaat gataccgcga 11831
gaccacgct caccggctcc agatttatca gcaataaacc agccagccgg aagggccgag 11891
cgcagaagtg gtcctgcaac tttatccgcc tccatccagt ctattaattg ttgccgggaa 11951
gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat tgctgcaggc 12011
atcgtggtgt cacgctcgtc gtttggtatg gcttcattca gtcgggttc ccaacgatca 12071
aggcgagtta catgatcccc catgttggtgc aaaaaagcgg ttagctcctt cggtcctccg 12131
atcgttgtca gaagtaagtt ggccgcagtg ttatcactca tggttatggc agcactgcat 12191
aattctctta ctgtcatgcc atccgtaaga tgcttttctg tgactggtga gtactcaacc 12251
aagtcattct gagaatagtg tatgcggcga ccgagttgct cttgcccggc gtcaacacgg 12311
gataataccg cgccacatag cagaacttta aaagtgtca tcattggaaa acgttcttcg 12371
ggcgaaaaac tctcaaggat cttaccgctg ttgagatcca gttcgatgta acccactcgt 12431
gcaccaact gatcttcagc atcttttact ttcaccagcg tttctgggtg agcaaaaaca 12491
ggaaggcaaa atgcgcgaaa aaagggaata agggcgacac ggaaatgttg aatactcata 12551
ctcttccttt ttcaatatta ttgaagcatt tatcaggggt attgtctcat gagcggatac 12611
atatattgaat gtatttagaa aaataaaca ataggggttc cgcgcacatt tccccgaaaa 12671
gtgccacctg acgtcgacct gaggtaatta taaccggggc cctatatatg gatcccgtaa 12731
tacgactcac tatag 12746

<210> 22

<211> 11918

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<220>

<221> CDS

<222> (446)..(9163)

<400> 22

accacaaaca ctccagtttg ttacactccg ctaggaatgc tcctggagca cccccctag 60
cagggcggtgg gggatttccc ctgcccgctc gcagaagggt ggagccaacc accttagtat 120
gtaggcggcg ggactcatga cgctcgctg atgacaagcg ccaagcttga cttggatggc 180
cctgatgggc gttcatgggt tcgggtggtg tggcgcttta ggcagcctcc acgcccacca 240
cctcccagat agagcggcgg cactgtaggg aagaccggg accggtcact accaaggacg 300
cagacctctt tttgagtatc acgcctccgg aagtagttgg gcaagccac ctatatgtgt 360

```

tgggatgggtt ggggttagcc atccataccg tactgcctga tagggtcctt gcgaggggat 420

ctgggagtcct cgtagaccgt agcac atg cct gtt att tct act caa aca agt 472
                               Met Pro Val Ile Ser Thr Gln Thr Ser
                               1                               5

cct gta cct gcg ccc aga acg cgc aag aac aag cag acg cag gct tca 520
Pro Val Pro Ala Pro Arg Thr Arg Lys Asn Lys Gln Thr Gln Ala Ser
10                               15                               20                               25

tat cct gtg tcc att aaa aca tct gtt gaa agg gga caa cga gca aag 568
Tyr Pro Val Ser Ile Lys Thr Ser Val Glu Arg Gly Gln Arg Ala Lys
                               30                               35                               40

cgc aaa gtc cag cgc gat gct cgg cct cgt aat tac aaa att gct ggt 616
Arg Lys Val Gln Arg Asp Ala Arg Pro Arg Asn Tyr Lys Ile Ala Gly
                               45                               50                               55

atc cat gat ggc ttg cag aca ttg gct cag gct gct ttg cca gct cat 664
Ile His Asp Gly Leu Gln Thr Leu Ala Gln Ala Ala Leu Pro Ala His
                               60                               65                               70

ggg tgg gga cgc caa gac cct cgc cat aag tct cgc aat ctt gga atc 712
Gly Trp Gly Arg Gln Asp Pro Arg His Lys Ser Arg Asn Leu Gly Ile
75                               80                               85

ctt ctg gat tac cct ttg ggg tgg att ggt gat gtt aca act cac aca 760
Leu Leu Asp Tyr Pro Leu Gly Trp Ile Gly Asp Val Thr Thr His Thr
90                               95                               100                               105

cct cta gta ggc ccg ctg gtg gca gga gcg gtc gtt cga cca gtc tgc 808
Pro Leu Val Gly Pro Leu Val Ala Gly Ala Val Val Arg Pro Val Cys
110                               115                               120

cag ata gta cgc ttg ctg gag gat gga gtc aac tgg gct act ggt tgg 856
Gln Ile Val Arg Leu Leu Glu Asp Gly Val Asn Trp Ala Thr Gly Trp
125                               130                               135

ttc ggt gtc cac ctt ttt gtg gta tgt ctg cta tct ttg gcc tgt ccc 904
Phe Gly Val His Leu Phe Val Val Cys Leu Leu Ser Leu Ala Cys Pro
140                               145                               150

tgt agt ggg tac caa gtg cgc aat tcc tcg ggg ctt tac cat gtc acc 952
Cys Ser Gly Tyr Gln Val Arg Asn Ser Ser Gly Leu Tyr His Val Thr
155                               160                               165

aat gat tgc cct aac tca agt att gtg tac gag gcg gcc gat gcc atc 1000
Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile
170                               175                               180                               185

ctg cac act ccg ggg tgt gtc cct tgc gtg cgc gag ggt aac gct tcg 1048
Leu His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser
190                               195                               200

agg tgt tgg gtg gcg gtg acc ccc acg gtg gcc acc agg gac ggc aaa 1096
Arg Cys Trp Val Ala Val Thr Pro Thr Val Ala Thr Arg Asp Gly Lys
205                               210                               215

ctc ccc aca acg cag ctt cga cgt cat atc gat ctg ctt gtc ggg agc 1144
Leu Pro Thr Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser
220                               225                               230

```

gcc acc ctc tgc tgc gcc ctc tac gtg ggg gac ctg tgc ggg tct gtc	1192
Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Ser Val	
235 240 245	
ttt ctt gtt ggt caa ctg ttt acc ttc tct ccc agg cgc cac tgg acg	1240
Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg Arg His Trp Thr	
250 255 260 265	
acg caa gac tgc aat tgt tct atc tat ccc ggc cat ata acg ggt cat	1288
Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His Ile Thr Gly His	
270 275 280	
cgc atg gca tgg gat atg atg atg aac tgg tcc cct acg gca gcg ttg	1336
Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Ala Ala Leu	
285 290 295	
gtg gta gct cag ctg ctc cgg atc cca caa gcc atc atg gac atg atc	1384
Val Val Ala Gln Leu Leu Arg Ile Pro Gln Ala Ile Met Asp Met Ile	
300 305 310	
gct ggt gct cac tgg gga gtc ctg gcg ggc ata gcg tat ttc tcc atg	1432
Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe Ser Met	
315 320 325	
gtg ggg aac tgg gcg aag gtc ctg gta gtg ctg ctg cta ttt gcc ggc	1480
Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu Leu Phe Ala Gly	
330 335 340 345	
gtc gac gcg gaa acc cac gtc acc ggg gga aat gcc ggc cgc acc acg	1528
Val Asp Ala Glu Thr His Val Thr Gly Gly Asn Ala Gly Arg Thr Thr	
350 355 360	
gct ggg ctt gtt ggt ctc ctt aca cca ggc gcc aag cag aac atc caa	1576
Ala Gly Leu Val Gly Leu Leu Thr Pro Gly Ala Lys Gln Asn Ile Gln	
365 370 375	
ctg atc aac acc aac ggc agt tgg cac atc aat agc acg gcc ttg aat	1624
Leu Ile Asn Thr Asn Gly Ser Trp His Ile Asn Ser Thr Ala Leu Asn	
380 385 390	
tgc aat gaa agc ctt aac acc ggc tgg tta gca ggg ctc ttc tat caa	1672
Cys Asn Glu Ser Leu Asn Thr Gly Trp Leu Ala Gly Leu Phe Tyr Gln	
395 400 405	
cac aaa ttc aac tct tca ggc tgt cct gag agg ttg gcc agc tgc cga	1720
His Lys Phe Asn Ser Ser Gly Cys Pro Glu Arg Leu Ala Ser Cys Arg	
410 415 420 425	
cgc ctt acc gat ttt gcc cag ggc tgg ggt cct atc agt tat gcc aac	1768
Arg Leu Thr Asp Phe Ala Gln Gly Trp Gly Pro Ile Ser Tyr Ala Asn	
430 435 440	
gga agc ggc ctc gac gaa cgc ccc tac tgc tgg cac tac cct cca aga	1816
Gly Ser Gly Leu Asp Glu Arg Pro Tyr Cys Trp His Tyr Pro Pro Arg	
445 450 455	
cct tgt ggc att gtg ccc gca aag agc gtg tgt ggc ccg gta tat tgc	1864
Pro Cys Gly Ile Val Pro Ala Lys Ser Val Cys Gly Pro Val Tyr Cys	
460 465 470	
ttc act ccc agc ccc gtg gtg gtg gga acg acc gac agg tcg ggc gcg	1912
Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg Ser Gly Ala	
475 480 485	

cct acc tac agc tgg ggt gca aat gat acg gat gtc ttc gtc ctt aac	1960
Pro Thr Tyr Ser Trp Gly Ala Asn Asp Thr Asp Val Phe Val Leu Asn	
490 495 500 505	
aac acc agg cca ccg ctg ggc aat tgg ttc ggt tgt acc tgg atg aac	2008
Asn Thr Arg Pro Pro Leu Gly Asn Trp Phe Gly Cys Thr Trp Met Asn	
510 515 520	
tca act gga ttc acc aaa gtg tgc gga gcg ccc cct tgt gtc atc gga	2056
Ser Thr Gly Phe Thr Lys Val Cys Gly Ala Pro Pro Cys Val Ile Gly	
525 530 535	
ggg gtg ggc aac aac acc ttg ctc tgc ccc act gat tgc ttc cgc aaa	2104
Gly Val Gly Asn Asn Thr Leu Cys Pro Thr Asp Cys Phe Arg Lys	
540 545 550	
cat ccg gaa gcc aca tac tct cgg tgc ggc tcc ggt ccc tgg att aca	2152
His Pro Glu Ala Thr Tyr Ser Arg Cys Gly Ser Gly Pro Trp Ile Thr	
555 560 565	
ccc agg tgc atg gtc gac tac ccg tat agg ctt tgg cac tat cct tgt	2200
Pro Arg Cys Met Val Asp Tyr Pro Tyr Arg Leu Trp His Tyr Pro Cys	
570 575 580 585	
acc atc aat tac acc ata ttc aaa gtc agg atg tac gtg gga ggg gtc	2248
Thr Ile Asn Tyr Thr Ile Phe Lys Val Arg Met Tyr Val Gly Gly Val	
590 595 600	
gag cac agg ctg gaa gcg gcc tgc aac tgg acg ccg ggc gaa cgc tgt	2296
Glu His Arg Leu Glu Ala Ala Cys Asn Trp Thr Arg Gly Glu Arg Cys	
605 610 615	
gat ctg gaa gac agg gac agg tcc gag ctc agc ccg ttg ctg ctg tcc	2344
Asp Leu Glu Asp Arg Asp Arg Ser Glu Leu Ser Pro Leu Leu Leu Ser	
620 625 630	
acc aca cag tgg cag gtc ctt ccg tgt tct ttc acg acc ctg cca gcc	2392
Thr Thr Gln Trp Gln Val Leu Pro Cys Ser Phe Thr Thr Leu Pro Ala	
635 640 645	
ttg tcc acc ggc ctc atc cac ctc cac cag aac att gtg gac gtg cag	2440
Leu Ser Thr Gly Leu Ile His Leu His Gln Asn Ile Val Asp Val Gln	
650 655 660 665	
tac ttg tac ggg gta ggg tca agc atc gcg tcc tgg gcc att aag tgg	2488
Tyr Leu Tyr Gly Val Gly Ser Ser Ile Ala Ser Trp Ala Ile Lys Trp	
670 675 680	
gag tac gtc gtt ctc ctg ttc ctt ctg ctt gca gac gcg cgc gtc tgc	2536
Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val Cys	
685 690 695	
tcc tgc ttg tgg atg atg tta ctc ata tcc caa gcg gag gcg gct ttg	2584
Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu Ala Ala Leu	
700 705 710	
gag aac ctc gta ata ctc aat gca gca tcc ctg gcc ggg acg cac ggt	2632
Glu Asn Leu Val Ile Leu Asn Ala Ala Ser Leu Ala Gly Thr His Gly	
715 720 725	
ctt gtg tcc ttc ctc gtg ttc ttc tgc ttt gcg tgg tat ctg aag ggt	2680
Leu Val Ser Phe Leu Val Phe Phe Cys Phe Ala Trp Tyr Leu Lys Gly	

730	735	740	745	
agg tgg gtg ccc gga gcg gtc tac gcc ctc tac ggg atg tgg cct ctc				2728
Arg Trp Val Pro Gly Ala Val Tyr Ala Leu Tyr Gly Met Trp Pro Leu				
	750	755	760	
ctc ctg ctc ctg ctg gcg ttg cct cag cgg gca tac gca ttt gat acc				2776
Leu Leu Leu Leu Leu Ala Leu Pro Gln Arg Ala Tyr Ala Phe Asp Thr				
	765	770	775	
gag ata att gga ggg ctg aca ata cca cct gta gta gca tta gtt gtc				2824
Glu Ile Ile Gly Gly Leu Thr Ile Pro Pro Val Val Ala Leu Val Val				
	780	785	790	
atg tct cgt ttt ggc ttc ttt gct cac ttg tta cct cgc tgt gct tta				2872
Met Ser Arg Phe Gly Phe Phe Ala His Leu Leu Pro Arg Cys Ala Leu				
	795	800	805	
gtt aac tcc tat ctt tgg caa cgt tgg gag aat tgg ttt tgg aac gtt				2920
Val Asn Ser Tyr Leu Trp Gln Arg Trp Glu Asn Trp Phe Trp Asn Val				
	810	815	820	825
aca cta aga ccg gag agg ttt ttc ctt gtg ctg gtt tgt ttc ccc ggt				2968
Thr Leu Arg Pro Glu Arg Phe Phe Leu Val Leu Val Cys Phe Pro Gly				
	830	835	840	
gcg aca tat gac gcg ctg gtg act ttc tgt gtg tgt cac gta gct ctt				3016
Ala Thr Tyr Asp Ala Leu Val Thr Phe Cys Val Cys His Val Ala Leu				
	845	850	855	
cta tgt tta aca tcc agt gca gca tcg ttc ttt ggg act gac tct agg				3064
Leu Cys Leu Thr Ser Ser Ala Ala Ser Phe Phe Gly Thr Asp Ser Arg				
	860	865	870	
gtt agg gcc cat aga atg ttg gtg cgt ctc gga aag tgt cat gct tgg				3112
Val Arg Ala His Arg Met Leu Val Arg Leu Gly Lys Cys His Ala Trp				
	875	880	885	
tat tct cat tat gtt ctt aag ttt ttc ctc tta gtg ttt ggt gag aat				3160
Tyr Ser His Tyr Val Leu Lys Phe Phe Leu Leu Val Phe Gly Glu Asn				
	890	895	900	905
ggt gtg ttt ttc tat aag cac ttg cat ggt gat gtc ttg cct aat gat				3208
Gly Val Phe Phe Tyr Lys His Leu His Gly Asp Val Leu Pro Asn Asp				
	910	915	920	
ttt gcc tcg aaa cta cca ttg caa gag cca ttt ttc cct ttt gaa ggc				3256
Phe Ala Ser Lys Leu Pro Leu Gln Glu Pro Phe Phe Pro Phe Glu Gly				
	925	930	935	
aag gca agg gtc tat agg aat gaa gga aga cgc ttg gcg tgt ggg gac				3304
Lys Ala Arg Val Tyr Arg Asn Glu Gly Arg Arg Leu Ala Cys Gly Asp				
	940	945	950	
acg gtt gat ggt ttg ccc gtt gtt gcg cgt ctc ggc gac ctt gtt ttc				3352
Thr Val Asp Gly Leu Pro Val Val Ala Arg Leu Gly Asp Leu Val Phe				
	955	960	965	
gca ggg ttg gct atg ccg cca gat ggg tgg gcc att acc gca cct ttt				3400
Ala Gly Leu Ala Met Pro Pro Asp Gly Trp Ala Ile Thr Ala Pro Phe				
	970	975	980	985
acg ctg cag tgt ctc tct gaa cgt ggc acg ctg tca gcg atg gca gtg				3448

Thr Leu Gln Cys Leu Ser Glu Arg Gly Thr Leu Ser Ala Met Ala Val	
990 995 1000	
gtc atg act ggt ata gac ccc cga act tgg act gga act atc ttc aga	3496
Val Met Thr Gly Ile Asp Pro Arg Thr Trp Thr Gly Thr Ile Phe Arg	
1005 1010 1015	
tta gga tct ctg gcc act agc tac atg gga ttc gtt tgt gac aac gtg	3544
Leu Gly Ser Leu Ala Thr Ser Tyr Met Gly Phe Val Cys Asp Asn Val	
1020 1025 1030	
ttg tat act gct cac cat ggc agc aag ggg cgc cgg ttg gct cat ccc	3592
Leu Tyr Thr Ala His His Gly Ser Lys Gly Arg Arg Leu Ala His Pro	
1035 1040 1045	
aca ggc tcc ata cac cca ata acc gtt gac gcg gct aat gac cag gac	3640
Thr Gly Ser Ile His Pro Ile Thr Val Asp Ala Ala Asn Asp Gln Asp	
1050 1055 1060 1065	
atc tat caa cca cca tgt gga gct ggg tcc ctt act cgg tgc tct tgc	3688
Ile Tyr Gln Pro Pro Cys Gly Ala Gly Ser Leu Thr Arg Cys Ser Cys	
1070 1075 1080	
ggg gag acc aag ggg tat ctg gta aca cga ctg ggg tca ttg gtt gag	3736
Gly Glu Thr Lys Gly Tyr Leu Val Thr Arg Leu Gly Ser Leu Val Glu	
1085 1090 1095	
gtc aac aaa tcc gat gac cct tat tgg tgt gtg tgc ggg gcc ctt ccc	3784
Val Asn Lys Ser Asp Asp Pro Tyr Trp Cys Val Cys Gly Ala Leu Pro	
1100 1105 1110	
atg gct gtt gcc aag ggt tct tca ggt gcc ccg att ctg tgc tcc tcc	3832
Met Ala Val Ala Lys Gly Ser Ser Gly Ala Pro Ile Leu Cys Ser Ser	
1115 1120 1125	
ggg cat gtt att ggg atg ttc acc gct gct aga aat tct ggc ggt tca	3880
Gly His Val Ile Gly Met Phe Thr Ala Ala Arg Asn Ser Gly Gly Ser	
1130 1135 1140 1145	
gtc agt cag att agg gtt agg ccg ttg gtg tgt gct gga tac cat ccc	3928
Val Ser Gln Ile Arg Val Arg Pro Leu Val Cys Ala Gly Tyr His Pro	
1150 1155 1160	
cag tac aca gca cat gcc act ctt gat aca aaa cct act gtg cct aac	3976
Gln Tyr Thr Ala His Ala Thr Leu Asp Thr Lys Pro Thr Val Pro Asn	
1165 1170 1175	
gag tat tca gtg caa att tta att gcc ccc act ggc agc ggc aag tca	4024
Glu Tyr Ser Val Gln Ile Leu Ile Ala Pro Thr Gly Ser Gly Lys Ser	
1180 1185 1190	
acc aaa tta cca ctt tct tac atg cag gag aag tat gag gtc ttg gtc	4072
Thr Lys Leu Pro Leu Ser Tyr Met Gln Glu Lys Tyr Glu Val Leu Val	
1195 1200 1205	
cta aat ccc agt gtg gct aca aca gca tca atg cca aag tac atg cac	4120
Leu Asn Pro Ser Val Ala Thr Thr Ala Ser Met Pro Lys Tyr Met His	
1210 1215 1220 1225	
gcg acg tac ggc gtg aat cca aat tgc tat ttt aat ggc aaa tgt acc	4168
Ala Thr Tyr Gly Val Asn Pro Asn Cys Tyr Phe Asn Gly Lys Cys Thr	
1230 1235 1240	

aac aca ggg gct tca ctt acg tac agc aca tat ggc atg tac ctg acc	4216
Asn Thr Gly Ala Ser Leu Thr Tyr Ser Thr Tyr Gly Met Tyr Leu Thr	
1245 1250 1255	
gga gca tgt tcc cgg aac tat gat gta atc att tgt gac gaa tgc cat	4264
Gly Ala Cys Ser Arg Asn Tyr Asp Val Ile Ile Cys Asp Glu Cys His	
1260 1265 1270	
gct acc gat gca acc acc gtg ttg ggc att gga aag gtc cta acc gaa	4312
Ala Thr Asp Ala Thr Thr Val Leu Gly Ile Gly Lys Val Leu Thr Glu	
1275 1280 1285	
gct cca tcc aaa aat gtt agg cta gtg gtt ctt gcc acg gct acc ccc	4360
Ala Pro Ser Lys Asn Val Arg Leu Val Val Leu Ala Thr Ala Thr Pro	
1290 1295 1300 1305	
cct gga gta atc cct aca cca cat gcc aac ata act gag att caa tta	4408
Pro Gly Val Ile Pro Thr Pro His Ala Asn Ile Thr Glu Ile Gln Leu	
1310 1315 1320	
acc gat gaa ggc act atc ccc ttt cat gga aaa aag att aag gag gaa	4456
Thr Asp Glu Gly Thr Ile Pro Phe His Gly Lys Lys Ile Lys Glu Glu	
1325 1330 1335	
aat ctg aag aaa ggg aga cac ctt atc ttt gag gct acc aaa aaa cac	4504
Asn Leu Lys Lys Gly Arg His Leu Ile Phe Glu Ala Thr Lys Lys His	
1340 1345 1350	
tgt gat gag ctt gct aac gag tta gct cga aag gga ata aca gct gtc	4552
Cys Asp Glu Leu Ala Asn Glu Leu Ala Arg Lys Gly Ile Thr Ala Val	
1355 1360 1365	
tct tac tat agg gga tgt gac atc tca aaa atc cct gag ggc gac tgt	4600
Ser Tyr Tyr Arg Gly Cys Asp Ile Ser Lys Ile Pro Glu Gly Asp Cys	
1370 1375 1380 1385	
gta gta gtt gcc act gat gcc ttg tgt aca ggg tac act ggt gac ttt	4648
Val Val Val Ala Thr Asp Ala Leu Cys Thr Gly Tyr Thr Gly Asp Phe	
1390 1395 1400	
gat tcc gtg tat gac tgc agc ctc atg gta gaa ggc aca tgc cat gtt	4696
Asp Ser Val Tyr Asp Cys Ser Leu Met Val Glu Gly Thr Cys His Val	
1405 1410 1415	
gac ctt gac cct act ttc acc atg ggt gtt cgt gtg tgc ggg gtt tca	4744
Asp Leu Asp Pro Thr Phe Thr Met Gly Val Arg Val Cys Gly Val Ser	
1420 1425 1430	
gca ata gtt aaa ggc cag cgt agg ggc cgc aca ggc cgt ggg aga gct	4792
Ala Ile Val Lys Gly Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Ala	
1435 1440 1445	
ggc ata tac tac tat gta gac ggg agt tgt acc cct tcg ggt atg gtt	4840
Gly Ile Tyr Tyr Tyr Val Asp Gly Ser Cys Thr Pro Ser Gly Met Val	
1450 1455 1460 1465	
cct gaa tgc aac att gtt gaa gcc ttc gac gca gcc aag gca tgg tat	4888
Pro Glu Cys Asn Ile Val Glu Ala Phe Asp Ala Ala Lys Ala Trp Tyr	
1470 1475 1480	
ggt ttg tca tca aca gaa gct caa act att ctg gac acc tat cgc acc	4936
Gly Leu Ser Ser Thr Glu Ala Gln Thr Ile Leu Asp Thr Tyr Arg Thr	
1485 1490 1495	

caa cct ggg tta cct gcg ata gga gca aat ttg gac gag tgg gct gat	4984
Gln Pro Gly Leu Pro Ala Ile Gly Ala Asn Leu Asp Glu Trp Ala Asp	
1500 1505 1510	
ctc ttt tct atg gtc aac ccc gaa cct tca ttt gtc aat act gca aaa	5032
Leu Phe Ser Met Val Asn Pro Glu Pro Ser Phe Val Asn Thr Ala Lys	
1515 1520 1525	
aga act gct gac aat tat gtt ttg ttg act gca gcc caa cta caa ctg	5080
Arg Thr Ala Asp Asn Tyr Val Leu Leu Thr Ala Ala Gln Leu Gln Leu	
1530 1535 1540 1545	
tgt cat cag tat ggc tat gct gct ccc aat gac gca cca cgg tgg cag	5128
Cys His Gln Tyr Gly Tyr Ala Ala Pro Asn Asp Ala Pro Arg Trp Gln	
1550 1555 1560	
gga gcc cgg ctt ggg aaa aaa cct tgt ggg gtt ctg tgg cgc ttg gac	5176
Gly Ala Arg Leu Gly Lys Lys Pro Cys Gly Val Leu Trp Arg Leu Asp	
1565 1570 1575	
ggc gct gac gcc tgt cct ggc cca gag ccc agc gag gtg acc aga tac	5224
Gly Ala Asp Ala Cys Pro Gly Pro Glu Pro Ser Glu Val Thr Arg Tyr	
1580 1585 1590	
caa atg tgc ttc act gaa gtc aat act tct ggg aca gcc gca ctc gct	5272
Gln Met Cys Phe Thr Glu Val Asn Thr Ser Gly Thr Ala Ala Leu Ala	
1595 1600 1605	
gtt ggc gtt gga gtg gct atg gct tat cta gcc att gac act ttt ggc	5320
Val Gly Val Gly Val Ala Met Ala Tyr Leu Ala Ile Asp Thr Phe Gly	
1610 1615 1620 1625	
gcc act tgt gtg cgg cgt tgc tgg tct att gca tca gtc cct acc ggt	5368
Ala Thr Cys Val Arg Arg Cys Trp Ser Ile Ala Ser Val Pro Thr Gly	
1630 1635 1640	
gct act gtc gcc cca gtg gtt gac gaa gaa gaa atc gtg gag gag tgt	5416
Ala Thr Val Ala Pro Val Val Asp Glu Glu Glu Ile Val Glu Glu Cys	
1645 1650 1655	
gca tca ttc att ccc ttg gag gcc atg gtt gct gca atc gat aag ctg	5464
Ala Ser Phe Ile Pro Leu Glu Ala Met Val Ala Ala Ile Asp Lys Leu	
1660 1665 1670	
aag agt aca atc acc aca act agt cct ttc aca ttg gaa acc gcc ctt	5512
Lys Ser Thr Ile Thr Thr Thr Ser Pro Phe Thr Leu Glu Thr Ala Leu	
1675 1680 1685	
gaa aaa ctt aac acc ttt ctt ggg cct cat gca gct aca atc ctt gct	5560
Glu Lys Leu Asn Thr Phe Leu Gly Pro His Ala Ala Thr Ile Leu Ala	
1690 1695 1700 1705	
atc ata gag tat tgc tgt ggt tta gtc act tta cct gac aat ccc ttt	5608
Ile Ile Glu Tyr Cys Cys Gly Leu Val Thr Leu Pro Asp Asn Pro Phe	
1710 1715 1720	
gca tca tgc gtg ttt gct ttc att gcg ggt att act acc cca cta cct	5656
Ala Ser Cys Val Phe Ala Phe Ile Ala Gly Ile Thr Thr Pro Leu Pro	
1725 1730 1735	
cac aag atc aaa atg ttc ctg tca tta ttt gga ggc gca att gcg tcc	5704
His Lys Ile Lys Met Phe Leu Ser Leu Phe Gly Gly Ala Ile Ala Ser	

1740	1745	1750	
aag ctt aca gac gct aga ggc gca ctg gcg ttc atg atg gcc ggg gct			5752
Lys Leu Thr Asp Ala Arg Gly Ala Leu Ala Phe Met Met Ala Gly Ala			
1755	1760	1765	
gcg gga aca gct ctt ggt aca tgg aca tcg gtg ggt ttt gtc ttt gac			5800
Ala Gly Thr Ala Leu Gly Thr Trp Thr Ser Val Gly Phe Val Phe Asp			
1770	1775	1780 1785	
atg cta ggc ggc tat gct gcc gcc tca tcc act gct tgc ttg aca ttt			5848
Met Leu Gly Gly Tyr Ala Ala Ala Ser Ser Thr Ala Cys Leu Thr Phe			
1790	1795	1800	
aaa tgc ttg atg ggt gag tgg ccc act atg gat cag ctt gct ggt tta			5896
Lys Cys Leu Met Gly Glu Trp Pro Thr Met Asp Gln Leu Ala Gly Leu			
1805	1810	1815	
gtc tac tcc gcg ttc aat ccg gcc gca gga gtt gtg ggc gtc ttg tca			5944
Val Tyr Ser Ala Phe Asn Pro Ala Ala Gly Val Val Gly Val Leu Ser			
1820	1825	1830	
gct tgt gca atg ttt gct ttg aca aca gca ggg cca gat cac tgg ccc			5992
Ala Cys Ala Met Phe Ala Leu Thr Thr Ala Gly Pro Asp His Trp Pro			
1835	1840	1845	
aac aga ctt ctt act atg ctt gct agg agc aac act gta tgt aat gag			6040
Asn Arg Leu Leu Thr Met Leu Ala Arg Ser Asn Thr Val Cys Asn Glu			
1850	1855	1860 1865	
tac ttt att gcc act cgt gac atc cgc agg aag ata ctg ggc att ctg			6088
Tyr Phe Ile Ala Thr Arg Asp Ile Arg Arg Lys Ile Leu Gly Ile Leu			
1870	1875	1880	
gag gca tct acc ccc tgg agt gtc ata tca gct tgc atc cgt tgg ctc			6136
Glu Ala Ser Thr Pro Trp Ser Val Ile Ser Ala Cys Ile Arg Trp Leu			
1885	1890	1895	
cac acc ccg acg gag gat gat tgc ggc ctc att gct tgg ggt cta gag			6184
His Thr Pro Thr Glu Asp Asp Cys Gly Leu Ile Ala Trp Gly Leu Glu			
1900	1905	1910	
att tgg cag tat gtg tgc aat ttc ttt gtg att tgc ttt aat gtc ctt			6232
Ile Trp Gln Tyr Val Cys Asn Phe Phe Val Ile Cys Phe Asn Val Leu			
1915	1920	1925	
aaa gct gga gtt cag agc atg gtt aac att cct ggt tgt cct ttc tac			6280
Lys Ala Gly Val Gln Ser Met Val Asn Ile Pro Gly Cys Pro Phe Tyr			
1930	1935	1940 1945	
agc tgc cag aag ggg tac aag ggc ccc tgg att gga tca ggt atg ctc			6328
Ser Cys Gln Lys Gly Tyr Lys Gly Pro Trp Ile Gly Ser Gly Met Leu			
1950	1955	1960	
caa gca cgc tgt cca tgc ggt gct gaa ctc atc ttt tot gtt gag aat			6376
Gln Ala Arg Cys Pro Cys Gly Ala Glu Leu Ile Phe Ser Val Glu Asn			
1965	1970	1975	
ggt ttt gca aaa ctt tac aaa gga ccc aga act tgt tca aat tac tgg			6424
Gly Phe Ala Lys Leu Tyr Lys Gly Pro Arg Thr Cys Ser Asn Tyr Trp			
1980	1985	1990	
aga ggg gct gtt cca gtc aac gct agg ctg tgt ggg tcg gct aga ccg			6472

Arg Gly Ala Val Pro Val Asn Ala Arg Leu Cys Gly Ser Ala Arg Pro	
1995	2000
2005	
gac cca act gat tgg act agt ctt gtc gtc aat tat ggc gtt agg gac	6520
Asp Pro Thr Asp Trp Thr Ser Leu Val Val Asn Tyr Gly Val Arg Asp	
2010	2015
2020	2025
tac tgt aaa tat gag aaa atg gga gat cac att ttt gtt aca gca gta	6568
Tyr Cys Lys Tyr Glu Lys Met Gly Asp His Ile Phe Val Thr Ala Val	
2030	2035
2040	
tcc tct cca aat gtc tgt ttc acc cag gtg ccc cca acc ttg aga gct	6616
Ser Ser Pro Asn Val Cys Phe Thr Gln Val Pro Pro Thr Leu Arg Ala	
2045	2050
2055	
gca gtg gcc gtg gac ggc gta cag gtt cag tgt tat cta ggt gag ccc	6664
Ala Val Ala Val Asp Gly Val Gln Val Gln Cys Tyr Leu Gly Glu Pro	
2060	2065
2070	
aaa act cct tgg acg aca tct gct tgc tgt tac ggt cct gac ggt aag	6712
Lys Thr Pro Trp Thr Thr Ser Ala Cys Cys Tyr Gly Pro Asp Gly Lys	
2075	2080
2085	
ggt aaa act gtt aag ctt ccc ttc cgc gtt gac ggt cac aca cct ggt	6760
Gly Lys Thr Val Lys Leu Pro Phe Arg Val Asp Gly His Thr Pro Gly	
2090	2095
2100	2105
gtg cgc atg caa ctt aat ttg cgt gat gca ctt gag aca aat gac tgt	6808
Val Arg Met Gln Leu Asn Leu Arg Asp Ala Leu Glu Thr Asn Asp Cys	
2110	2115
2120	
aat tcc aca aac aac act cct agt gat gaa gcc gca gtg tcc gct ctt	6856
Asn Ser Thr Asn Asn Thr Pro Ser Asp Glu Ala Ala Val Ser Ala Leu	
2125	2130
2135	
gtt ttc aaa cag gag ttg cgg cgt aca aac caa ttg ctt gag gca att	6904
Val Phe Lys Gln Glu Leu Arg Arg Thr Asn Gln Leu Leu Glu Ala Ile	
2140	2145
2150	
tca gct ggc gtt gac acc acc aaa ctg cca gcc ccc tcc atc gaa gag	6952
Ser Ala Gly Val Asp Thr Thr Lys Leu Pro Ala Pro Ser Ile Glu Glu	
2155	2160
2165	
gta gtg gta aga aag cgc cag ttc cgg gca aga act ggt tcg ctt acc	7000
Val Val Val Arg Lys Arg Gln Phe Arg Ala Arg Thr Gly Ser Leu Thr	
2170	2175
2180	2185
ttg cct ccc cct ccg aga tcc gtc cca gga gtg tca tgt cct gaa agc	7048
Leu Pro Pro Pro Pro Arg Ser Val Pro Gly Val Ser Cys Pro Glu Ser	
2190	2195
2200	
ctg caa cga agt gac ccg tta gaa ggt cct tca aac ctc cct cct tca	7096
Leu Gln Arg Ser Asp Pro Leu Glu Gly Pro Ser Asn Leu Pro Pro Ser	
2205	2210
2215	
cca cct gtt cta cag ttg gcc atg ccg atg ccc ctg ttg gga gcg ggt	7144
Pro Pro Val Leu Gln Leu Ala Met Pro Met Pro Leu Leu Gly Ala Gly	
2220	2225
2230	
gag tgt aac cct ttc act gca att gga tgt gca atg acc gaa aca ggc	7192
Glu Cys Asn Pro Phe Thr Ala Ile Gly Cys Ala Met Thr Glu Thr Gly	
2235	2240
2245	

85/146

gga ggc cct gat gat tta ccc agt tac cct ccc aaa aag gag gtc tct Gly Gly Pro Asp Asp Leu Pro Ser Tyr Pro Pro Lys Lys Glu Val Ser 2250 2255 2260 2265	7240
gaa tgg tca gac gaa agt tgg tca acg gct aca acc gtt tcc agc tac Glu Trp Ser Asp Glu Ser Trp Ser Thr Ala Thr Thr Val Ser Ser Tyr 2270 2275 2280	7288
gtt act ggc ccc ccg tac cct aag ata cgg gga aag gat tcc act cag Val Thr Gly Pro Pro Tyr Pro Lys Ile Arg Gly Lys Asp Ser Thr Gln 2285 2290 2295	7336
tca gcc ccc gcc aaa cgg cct aca aaa aag aag ttg gga aag agt gag Ser Ala Pro Ala Lys Arg Pro Thr Lys Lys Lys Leu Gly Lys Ser Glu 2300 2305 2310	7384
ttt tcg tgc agc atg agc tac acc tgg acc gac gtg att agc ttc aaa Phe Ser Cys Ser Met Ser Tyr Thr Trp Thr Asp Val Ile Ser Phe Lys 2315 2320 2325	7432
act gct tct aaa gtt ctg tct gca act cgg gcc atc act agt ggt ttc Thr Ala Ser Lys Val Leu Ser Ala Thr Arg Ala Ile Thr Ser Gly Phe 2330 2335 2340 2345	7480
ctc aaa caa aga tca ttg gtg tat gtg act gag ccg cgg gat gcg gag Leu Lys Gln Arg Ser Leu Val Tyr Val Thr Glu Pro Arg Asp Ala Glu 2350 2355 2360	7528
ott aga aaa caa aaa gtc act att aat aga caa cct ctg ttc ccc cca Leu Arg Lys Gln Lys Val Thr Ile Asn Arg Gln Pro Leu Phe Pro Pro 2365 2370 2375	7576
tca tac cac aag caa gtg aga ttg gct aag gaa aaa gct tca aaa gtt Ser Tyr His Lys Gln Val Arg Leu Ala Lys Glu Lys Ala Ser Lys Val 2380 2385 2390	7624
gtc ggt gtc atg tgg gac tat gat gaa gta gca gct cac acg ccc tct Val Gly Val Met Trp Asp Tyr Asp Glu Val Ala Ala His Thr Pro Ser 2395 2400 2405	7672
aag tct gct aag tcc cac atc act ggc ctt cgg ggc act gat gtt cgt Lys Ser Ala Lys Ser His Ile Thr Gly Leu Arg Gly Thr Asp Val Arg 2410 2415 2420 2425	7720
tct gga gca gcc cgc aag gct gtt ctg gac ttg cag aag tgt gtc gag Ser Gly Ala Ala Arg Lys Ala Val Leu Asp Leu Gln Lys Cys Val Glu 2430 2435 2440	7768
gca ggt gag ata ccg agt cat tat cgg caa act gtg ata gtt cca aag Ala Gly Glu Ile Pro Ser His Tyr Arg Gln Thr Val Ile Val Pro Lys 2445 2450 2455	7816
gag gag gtc ttc gtg aag acc ccc cag aaa cca aca aag aaa ccc cca Glu Glu Val Phe Val Lys Thr Pro Gln Lys Pro Thr Lys Lys Pro Pro 2460 2465 2470	7864
agg ctt atc tcg tac ccc cac ctt gaa atg aga tgt gtt gag aag atg Arg Leu Ile Ser Tyr Pro His Leu Glu Met Arg Cys Val Glu Lys Met 2475 2480 2485	7912
tac tac ggt cag gtt gct cct gac gta gtt aaa gct gtc atg gga gat Tyr Tyr Gly Gln Val Ala Pro Asp Val Val Lys Ala Val Met Gly Asp 2490 2495 2500 2505	7960

gcg tac ggg ttt gtc gac cca cgt acc cgt gtc aag cgt ctg ttg tcg	8008
Ala Tyr Gly Phe Val Asp Pro Arg Thr Arg Val Lys Arg Leu Leu Ser	
2510 2515 2520	
atg tgg tca ccc gat gca gtc gga gcc aca tgc gat aca gtg tgt ttt	8056
Met Trp Ser Pro Asp Ala Val Gly Ala Thr Cys Asp Thr Val Cys Phe	
2525 2530 2535	
gac agt acc atc aca ccc gag gat atc atg gtg gag aca gac atc tac	8104
Asp Ser Thr Ile Thr Pro Glu Asp Ile Met Val Glu Thr Asp Ile Tyr	
2540 2545 2550	
tca gca gct aaa ctc agt gac caa cac cga gct ggc att cac acc att	8152
Ser Ala Ala Lys Leu Ser Asp Gln His Arg Ala Gly Ile His Thr Ile	
2555 2560 2565	
gcg agg cag tta tac gct gga gga ccg atg atc gct tat gat ggc cga	8200
Ala Arg Gln Leu Tyr Ala Gly Gly Pro Met Ile Ala Tyr Asp Gly Arg	
2570 2575 2580 2585	
gag atc gga tat cgt agg tgt agg tct tcc ggc gtc tat act acc tca	8248
Glu Ile Gly Tyr Arg Arg Cys Arg Ser Ser Gly Val Tyr Thr Thr Ser	
2590 2595 2600	
agt tcc aac agt ttg acc tgc tgg ctg aag gta aat gct gca gcc gaa	8296
Ser Ser Asn Ser Leu Thr Cys Trp Leu Lys Val Asn Ala Ala Ala Glu	
2605 2610 2615	
cag gct ggc atg aag aac cct cgc ttc ctt att tgc ggc gat gat tgc	8344
Gln Ala Gly Met Lys Asn Pro Arg Phe Leu Ile Cys Gly Asp Asp Cys	
2620 2625 2630	
acc gta att tgg aag agc gcc gga gca gat gca gac aaa caa gca atg	8392
Thr Val Ile Trp Lys Ser Ala Gly Ala Asp Ala Asp Lys Gln Ala Met	
2635 2640 2645	
cgt gtc ttt gct agc tgg atg aag gtg atg ggt gca cca caa gat tgt	8440
Arg Val Phe Ala Ser Trp Met Lys Val Met Gly Ala Pro Gln Asp Cys	
2650 2655 2660 2665	
gtg cct caa ccc aaa tac agt ttg gaa gaa tta aca tca tgc tca tca	8488
Val Pro Gln Pro Lys Tyr Ser Leu Glu Glu Leu Thr Ser Cys Ser Ser	
2670 2675 2680	
aat gtt acc tct gga att acc aaa agt ggc aag cct tac tac ttt ctt	8536
Asn Val Thr Ser Gly Ile Thr Lys Ser Gly Lys Pro Tyr Tyr Phe Leu	
2685 2690 2695	
aca aga gat cct cgt atc ccc ctt ggc agg tgc tct gcc gag ggt ctg	8584
Thr Arg Asp Pro Arg Ile Pro Leu Gly Arg Cys Ser Ala Glu Gly Leu	
2700 2705 2710	
gga tac aac ccc agt gct gcg tgg att ggg tat cta ata cat cac tac	8632
Gly Tyr Asn Pro Ser Ala Ala Trp Ile Gly Tyr Leu Ile His His Tyr	
2715 2720 2725	
cca tgt ttg tgg gtt agc cgt gtg ttg gct gtc cat ttc atg gag cag	8680
Pro Cys Leu Trp Val Ser Arg Val Leu Ala Val His Phe Met Glu Gln	
2730 2735 2740 2745	
atg ctc ttt gag gac aaa ctt ccc gag acg gtg acc ttt gac tgg tat	8728
Met Leu Phe Glu Asp Lys Leu Pro Glu Thr Val Thr Phe Asp Trp Tyr	

2750	2755	2760	
ggg aaa aat tat acg gtg cct gta gaa gat ctg ccc agc atc att gct			8776
Gly Lys Asn Tyr Thr Val Pro Val Glu Asp Leu Pro Ser Ile Ile Ala			
2765	2770	2775	
ggt gtg cac ggt att gag gct ttc tcg gtg gtg cgc tac acc aac gct			8824
Gly Val His Gly Ile Glu Ala Phe Ser Val Val Arg Tyr Thr Asn Ala			.
2780	2785	2790	
gag atc ctc aga gtt tcc caa tca cta aca gac atg acc atg ccc ccc			8872
Glu Ile Leu Arg Val Ser Gln Ser Leu Thr Asp Met Thr Met Pro Pro			
2795	2800	2805	
ctg cga gcc tgg cga aag aaa gcc agg gcg gtc ctc gcc agc gcc aag			8920
Leu Arg Ala Trp Arg Lys Lys Ala Arg Ala Val Leu Ala Ser Ala Lys			
2810	2815	2820	2825
agg cgt ggc gga gca cac gca aaa ttg gct cgc ttc ctt ctc tgg cat			8968
Arg Arg Gly Gly Ala His Ala Lys Leu Ala Arg Phe Leu Leu Trp His			
2830	2835	2840	
gct aca tct aga cct cta cca gat ttg gat aag acg agc gtg gct cgg			9016
Ala Thr Ser Arg Pro Leu Pro Asp Leu Asp Lys Thr Ser Val Ala Arg			
2845	2850	2855	
tac acc act ttc aat tat tgt gat gtt tac tcc ccg gag ggg gat gtg			9064
Tyr Thr Thr Phe Asn Tyr Cys Asp Val Tyr Ser Pro Glu Gly Asp Val			
2860	2865	2870	
ttt att aca cca cag aga aga ttg cag aag ttt ctt gtg aag tat ttg			9112
Phe Ile Thr Pro Gln Arg Arg Leu Gln Lys Phe Leu Val Lys Tyr Leu			
2875	2880	2885	
gct gtc att gtt ttt gcc cta ggg ctc att gct gtt gga tta gcc atc			9160
Ala Val Ile Val Phe Ala Leu Gly Leu Ile Ala Val Gly Leu Ala Ile			
2890	2895	2900	2905
agc tgaacccccca aattcaaaat taactaacag tttttttttt tttttttttt			9213
Ser			
tttagggcag cggcaacagg ggagaccccg ggcttaacga ccccgccgat gtgagtttgg			9273
cgaccatggt ggatcagaac cgtttcgggt gaagccatgg tctgaagggg atgacgtccc			9333
ttctggctca tccacaaaaa ccgtctcggg tgggtgagga gtcctggctg tgtgggaagc			9393
agtcagtata attcccgtcg tgtgtggtga cgcctcacga cgtatttgtc cgctgtgcag			9453
agcgtagtac caagggctgc accccggttt ttgttccaag cggagggcaa ccccgcttg			9513
gaattaaaaa ctcgagcaag acgtttcccg ttgaatatgg ctcataacac cccttgtatt			9573
actgtttatg taagcagaca gttttattgt tcatgatgat atatttttat cttgtgcaat			9633
gtaacatcag agattttgag acacaacgtg gctttgttga ataaatcgaa cttttgctga			9693
gttgaaggat cagatcacgc atcttcccgca caacgcagac cgttccgtgg caaagcaaaa			9753
gttcaaaatc accaaactggt ccacctacaa caaagctctc atcaacogtg gctccctcac			9813
tttctggctg gatgatgggg cgattcaggc ctggtatgag tcagcaaacac cttcttcacg			9873

aggcagacct cagcgctagc ggagtgtata ctggcttact atgttggcac tgatgaggggt 9933
gtcagtgaag tgcttcatgt ggcaggagaa aaaaggctgc accggtgcgt cagcagaata 9993
tgtgatacag gatataattcc gcttcctcgc tcaactgactc gctacgctcg gtcgttcgac 10053
tgcggcgagc ggaaatggct tacgaacggg gcggagattt cctggaagat gccaggaaga 10113
tacttaacag ggaagtgaga gggccgcggc aaagccgttt ttccataggc tccgcccccc 10173
tgacaagcat cacgaaatct gacgctcaaa tcagtgggtgg cgaaaccgga caggactata 10233
aagataccag gcgtttcccc tggcggtccc ctctgtcgct ctctgtttcc tgcctttcgg 10293
tttaccggtg tcattccgct gttatggccg cgtttgtctc attccacgcc tgacactcag 10353
ttccgggtag gcagttcgtt ccaagctgga ctgtatgcac gaaccccccg ttcagtccga 10413
ccgctgcgcc ttatccggtg actatcgtct tgagtccaac ccggaagac atgcaaaagc 10473
accactggca gcagccactg gtaattgatt tagaggagtt agtcttgaag tcatgcgccg 10533
gttaaggcta aactgaaagg acaagttttg gtgactgcgc tcctccaagc cagttacctc 10593
ggttcaaaga gttggtagct cagagaacct tcgaaaaacc gccctgcaag gcggtttttt 10653
cgttttcaga gcaagagatt acgcgagac caaacgata tcaagaagat catcttatta 10713
aggggtctga cgctcagtgg aacgaaaact cacgttaagg gatthttgtc atgagattat 10773
caaaaaggat cttcacctag atccttttaa attaaaaatg aagtttttaa tcaatctaaa 10833
gtatatatga gtaaacttgg tctgacagtt accaatgctt aatcagttag gcacctatct 10893
cagcgatctg tctatttcgt tcatccatag ttgcctgact ccccgctcgt tagataacta 10953
cgatacggga gggcttacca tctggcccca gtgctgcaat gataccgga gaccacgct 11013
caccggctcc agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg 11073
gtcctgcaac tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa 11133
gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat tgctgcaggc atcgtggtgt 11193
cacgctcgtc gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta 11253
catgatcccc catgttggtc aaaaaagcgg ttagctcctt cggctcctccg atogttgtca 11313
gaagtaagtt ggccgcagtg ttatcactca tggttatggc agcaactgat aattctotta 11373
ctgtcatgcc atccgtaaga tgcttttctg tgactgggtga gtactcaacc aagtcattct 11433
gagaatagtg tatgcggcga ccgagttgct cttgccgggc gtcaacacgg gataataccg 11493
cgccacatag cagaacttta aaagtgtca tcattggaaa acgtttcttcg gggcgaaaac 11553
tctcaaggat cttaccgctg ttgagatcca gttcgatgta acccactcgt gcaccaact 11613
gatcttcagc atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa 11673
atgccgcaaa aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt 11733
ttcaatatta ttgaagcatt tatcaggggt attgtctcat gagcggatac atatttgaat 11793

```
<210> 23
<211> 13042
<212> DNA
<213> Artificial Sequence
```

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

```
<220>
<221> CDS
<222> (342)..(9920)
```

<400> 23																										
gccagccccc	tgatgggggc	gacactccac	catgaatcac	tcccctgtga	ggaactactg	60																				
tcttcacgca	gaaagcgtct	agccatggcg	ttagtatgag	tgctcgtgcag	cctccaggac	120																				
ccccctccc	gggagagcca	tagtgggtctg	cggaaccggg	gagtacaccg	gaattgccag	180																				
gacgaccggg	tcctttcttg	gataaacccg	ctcaatgcct	ggagatttgg	gcgtgcccc	240																				
gcaagactgc	tagccgagta	gtgttgggtc	gcgaaaggcc	ttgtgggtact	gcctgatagg	300																				
gtgcttgcca	gtgccccggg	aggtctcgtg	gaccgtgcac	c atg cct gtt att tct		356																				
				Met	Pro Val Ile Ser																					
				1			5																			
act caa aca agt cct gta cct gcg ccc aga acg cgc aag aac aag cag																404										
Thr Gln Thr Ser Pro Val Pro Ala Pro Arg Thr Arg Lys Asn Lys Gln																										
	10										15					20										
acg cag gct tca tat cct gtg tcc att aaa aca tct gtt gaa agg gga																452										
Thr Gln Ala Ser Tyr Pro Val Ser Ile Lys Thr Ser Val Glu Arg Gly																										
	25										30					35										
caa cga gca aag cgc aaa gtc cag cgc gat gct cgg cct cgt aat tac																500										
Gln Arg Ala Lys Arg Lys Val Gln Arg Asp Ala Arg Pro Arg Asn Tyr																										
	40										45					50										
aaa att gct ggt atc cat gat ggc ttg cag aca ttg gct cag gct gct																548										
Lys Ile Ala Gly Ile His Asp Gly Leu Gln Thr Leu Ala Gln Ala Ala																										
	55										60					65										
ttg cca gct cat ggt tgg gga cgc caa gac cct cgc cat aag tct cgc																596										
Leu Pro Ala His Gly Trp Gly Arg Gln Asp Pro Arg His Lys Ser Arg																										
	70										75					80					85					
aat ctt gga atc ctt ctg gat tac cct ttg ggg tgg att ggt gat gtt																644										
Asn Leu Gly Ile Leu Leu Asp Tyr Pro Leu Gly Trp Ile Gly Asp Val																										
	90										95					100										
aca act cac aca cct cta gta ggc ccg ctg gtg gca gga gcg gtc gtt																692										

Thr	Thr	His	Thr	Pro	Leu	Val	Gly	Pro	Leu	Val	Ala	Gly	Ala	Val	Val	
			105					110					115			
cga	cca	gtc	tgc	cag	ata	gta	cgc	ttg	ctg	gag	gat	gga	gtc	aac	tgg	740
Arg	Pro	Val	Cys	Gln	Ile	Val	Arg	Leu	Leu	Glu	Asp	Gly	Val	Asn	Trp	
		120					125					130				
gct	act	ggg	tgg	ttc	ggg	gtc	cac	ctt	ttt	gtg	gta	tgt	ctg	cta	tct	788
Ala	Thr	Gly	Trp	Phe	Gly	Val	His	Leu	Phe	Val	Val	Cys	Leu	Leu	Ser	
	135					140					145					
ttg	gcc	tgt	ccc	tgt	agt	ggg	gcg	cgg	gtc	act	gac	cca	gac	aca	aat	836
Leu	Ala	Cys	Pro	Cys	Ser	Gly	Ala	Arg	Val	Thr	Asp	Pro	Asp	Thr	Asn	
150					155					160					165	
acc	aca	atc	ctg	acc	aat	tgc	tgc	cag	cgt	aat	cag	gtt	atc	tat	tgt	884
Thr	Thr	Ile	Leu	Thr	Asn	Cys	Cys	Gln	Arg	Asn	Gln	Val	Ile	Tyr	Cys	
				170					175					180		
tct	cct	tcc	act	tgc	cta	cac	gag	cct	ggg	tgt	gtg	atc	tgt	gcg	gac	932
Ser	Pro	Ser	Thr	Cys	Leu	His	Glu	Pro	Gly	Cys	Val	Ile	Cys	Ala	Asp	
			185					190					195			
gag	tgc	tgg	gtt	ccc	gcc	aat	ccg	tac	atc	tca	cac	cct	tcc	aat	tgg	980
Glu	Cys	Trp	Val	Pro	Ala	Asn	Pro	Tyr	Ile	Ser	His	Pro	Ser	Asn	Trp	
		200					205					210				
act	ggc	acg	gac	tcc	ttc	ttg	gct	gac	cac	att	gat	ttt	gtt	atg	ggc	1028
Thr	Gly	Thr	Asp	Ser	Phe	Leu	Ala	Asp	His	Ile	Asp	Phe	Val	Met	Gly	
	215					220					225					
gct	ctt	gtg	acc	tgt	gac	gcc	ctt	gac	att	ggg	gag	ttg	tgt	ggg	gcg	1076
Ala	Leu	Val	Thr	Cys	Asp	Ala	Leu	Asp	Ile	Gly	Glu	Leu	Cys	Gly	Ala	
230					235					240					245	
tgt	gta	tta	gtc	ggg	gac	tgg	ctt	gtc	agg	cac	tgg	ctt	att	cac	ata	1124
Cys	Val	Leu	Val	Gly	Asp	Trp	Leu	Val	Arg	His	Trp	Leu	Ile	His	Ile	
				250					255					260		
gac	ctc	aat	gaa	act	ggg	act	tgt	tac	ctg	gaa	gtg	ccc	act	gga	ata	1172
Asp	Leu	Asn	Glu	Thr	Gly	Thr	Cys	Tyr	Leu	Glu	Val	Pro	Thr	Gly	Ile	
			265					270					275			
gat	cct	ggg	ttc	cta	ggg	ttt	atc	ggg	tgg	atg	gcc	ggc	aag	gtc	gag	1220
Asp	Pro	Gly	Phe	Leu	Gly	Phe	Ile	Gly	Trp	Met	Ala	Gly	Lys	Val	Glu	
		280					285					290				
gct	gtc	atc	ttc	ttg	acc	aaa	ctg	gct	tca	caa	gta	cca	tac	gct	att	1268
Ala	Val	Ile	Phe	Leu	Thr	Lys	Leu	Ala	Ser	Gln	Val	Pro	Tyr	Ala	Ile	
	295					300					305					
gcg	act	atg	ttt	agc	agt	gta	cac	tac	ctg	gcg	gtt	ggc	gct	ctg	atc	1316
Ala	Thr	Met	Phe	Ser	Ser	Val	His	Tyr	Leu	Ala	Val	Gly	Ala	Leu	Ile	
310					315					320				325		
tac	tat	gcc	tct	cgg	ggc	aag	tgg	tat	cag	ttg	ctc	cta	gcg	ctt	atg	1364
Tyr	Tyr	Ala	Ser	Arg	Gly	Lys	Trp	Tyr	Gln	Leu	Leu	Leu	Ala	Leu	Met	
				330					335					340		
ctt	tac	ata	gaa	gcg	acc	tct	gga	aac	cct	atc	agg	gtg	ccc	act	gga	1412
Leu	Tyr	Ile	Glu	Ala	Thr	Ser	Gly	Asn	Pro	Ile	Arg	Val	Pro	Thr	Gly	
			345					350					355			

tgc tca ata gct gag ttt tgc tcg cct ttg atg ata cca tgt cct tgc	1460
Cys Ser Ile Ala Glu Phe Cys Ser Pro Leu Met Ile Pro Cys Pro Cys	
360 365 370	
cac tct tat ttg agt gag aat gtg tca gaa gtc att tgt tac agt cca	1508
His Ser Tyr Leu Ser Glu Asn Val Ser Glu Val Ile Cys Tyr Ser Pro	
375 380 385	
aag tgg acc agg cct atc act ctg gag tat aac aac tcc ata tct tgg	1556
Lys Trp Thr Arg Pro Ile Thr Leu Glu Tyr Asn Asn Ser Ile Ser Trp	
390 395 400 405	
tac ccc tat aca atc cct ggt gcg agg gga tgt atg gtt aaa ttc aaa	1604
Tyr Pro Tyr Thr Ile Pro Gly Ala Arg Gly Cys Met Val Lys Phe Lys	
410 415 420	
aat aac aca tgg ggt tgc tgc cgt att cgc aat gtg cca tcg tac tgc	1652
Asn Asn Thr Trp Gly Cys Cys Arg Ile Arg Asn Val Pro Ser Tyr Cys	
425 430 435	
act atg ggc act gat gca gtg tgg aac gac act cgc aac act tac gaa	1700
Thr Met Gly Thr Asp Ala Val Trp Asn Asp Thr Arg Asn Thr Tyr Glu	
440 445 450	
gta tgc ggt gta aca cca tgg cta aca acc gca tgg cac aac ggc tca	1748
Val Cys Gly Val Thr Pro Trp Leu Thr Thr Ala Trp His Asn Gly Ser	
455 460 465	
gcc ctg aaa ttg gct ata tta caa tac cct ggg tct aaa gaa atg ttt	1796
Ala Leu Lys Leu Ala Ile Leu Gln Tyr Pro Gly Ser Lys Glu Met Phe	
470 475 480 485	
aaa cct cat aat tgg atg tca ggc cat ttg tat ttt gag gga tca gat	1844
Lys Pro His Asn Trp Met Ser Gly His Leu Tyr Phe Glu Gly Ser Asp	
490 495 500	
acc cct ata gtt tac ttt tat gac cct gtg aat tcc act ctc cta cca	1892
Thr Pro Ile Val Tyr Phe Tyr Asp Pro Val Asn Ser Thr Leu Leu Pro	
505 510 515	
ccg gag agg tgg gct agg ttg ccc ggt acc cca cct gtg gta cgt ggt	1940
Pro Glu Arg Trp Ala Arg Leu Pro Gly Thr Pro Pro Val Val Arg Gly	
520 525 530	
tct tgg tta cag gtt ccg caa ggg ttt tac agt gat gtg aaa gac cta	1988
Ser Trp Leu Gln Val Pro Gln Gly Phe Tyr Ser Asp Val Lys Asp Leu	
535 540 545	
gcc aca gga ttg atc acc aaa gac aaa gcc tgg aaa aat tat cag gtc	2036
Ala Thr Gly Leu Ile Thr Lys Asp Lys Ala Trp Lys Asn Tyr Gln Val	
550 555 560 565	
tta tat tcc gcc acg ggt gct ttg tct ctt acg gga gtt acc acc aag	2084
Leu Tyr Ser Ala Thr Gly Ala Leu Ser Leu Thr Gly Val Thr Thr Lys	
570 575 580	
gcc gtg gtg cta att ctg ttg ggg ttg tgt ggc agc aag tat ctt att	2132
Ala Val Val Leu Ile Leu Leu Gly Leu Cys Gly Ser Lys Tyr Leu Ile	
585 590 595	
tta gcc tac ctc tgt tac ttg tcc ctt tgt ttt ggg cgc gct tct ggt	2180
Leu Ala Tyr Leu Cys Tyr Leu Ser Leu Cys Phe Gly Arg Ala Ser Gly	
600 605 610	

tac cct ttg cgt cct gtg ctc cca tcc cag tcg tat ctc caa gct ggc	2228
Tyr Pro Leu Arg Pro Val Leu Pro Ser Gln Ser Tyr Leu Gln Ala Gly	
615 620 625	
tgg gat gtt ttg tct aaa gct caa gta gct cct ttt gct ttg att ttc	2276
Trp Asp Val Leu Ser Lys Ala Gln Val Ala Pro Phe Ala Leu Ile Phe	
630 635 640 645	
ttc atc tgt tgc tat ctc cgc tgc agg cta cgt tat gct gcc ctt tta	2324
Phe Ile Cys Cys Tyr Leu Arg Cys Arg Leu Arg Tyr Ala Ala Leu Leu	
650 655 660	
ggg ttt gtg ccc atg gct gcg ggc ttg ccc cta act ttc ttt gtt gca	2372
Gly Phe Val Pro Met Ala Ala Gly Leu Pro Leu Thr Phe Phe Val Ala	
665 670 675	
gca gct gct gcc caa cca gat tat gac tgg tgg gtg cga ctg cta gtg	2420
Ala Ala Ala Ala Gln Pro Asp Tyr Asp Trp Trp Val Arg Leu Leu Val	
680 685 690	
gca ggg tta gtt ttg tgg gcc ggc cgt aac cgt ggt cac cgc ata gct	2468
Ala Gly Leu Val Leu Trp Ala Gly Arg Asn Arg Gly His Arg Ile Ala	
695 700 705	
ctg ctt gta ggt cct tgg cct ctg gta gcg ctt tta acc ctc ttg cat	2516
Leu Leu Val Gly Pro Trp Pro Leu Val Ala Leu Leu Thr Leu Leu His	
710 715 720 725	
ttg gtt acg cct gct tca gct ttt gat acc gag ata att gga ggg ctg	2564
Leu Val Thr Pro Ala Ser Ala Phe Asp Thr Glu Ile Ile Gly Gly Leu	
730 735 740	
aca ata cca cct gta gta gca tta gtt gtc atg tct cgt ttt ggc ttc	2612
Thr Ile Pro Pro Val Val Ala Leu Val Val Met Ser Arg Phe Gly Phe	
745 750 755	
ttt gct cac ttg tta cct cgc tgt gct tta gtt aac tcc tat ctt tgg	2660
Phe Ala His Leu Leu Pro Arg Cys Ala Leu Val Asn Ser Tyr Leu Trp	
760 765 770	
caa cgt tgg gag aat tgg ttt tgg aac gtt aca cta aga ccg gag agg	2708
Gln Arg Trp Glu Asn Trp Phe Trp Asn Val Thr Leu Arg Pro Glu Arg	
775 780 785	
ttt ttc ctt gtg ctg gtt tgt ttc ccc ggt gcg aca tat gac gcg ctg	2756
Phe Phe Leu Val Leu Val Cys Phe Pro Gly Ala Thr Tyr Asp Ala Leu	
790 795 800 805	
gtg act ttc tgt gtg tgt cac gta gct ctt cta tgt tta aca tcc agt	2804
Val Thr Phe Cys Val Cys His Val Ala Leu Leu Cys Leu Thr Ser Ser	
810 815 820	
gca gca tcg ttc ttt ggg act gac tct agg gtt agg gcc cat aga atg	2852
Ala Ala Ser Phe Phe Gly Thr Asp Ser Arg Val Arg Ala His Arg Met	
825 830 835	
ttg gtg cgt ctc gga aag tgt cat gct tgg tat tct cat tat gtt ctt	2900
Leu Val Arg Leu Gly Lys Cys His Ala Trp Tyr Ser His Tyr Val Leu	
840 845 850	
aag ttt ttc ctc tta gtg ttt ggt gag aat ggt gtg ttt ttc tat aag	2948
Lys Phe Phe Leu Leu Val Phe Gly Glu Asn Gly Val Phe Phe Tyr Lys	

855	860	865	
cac ttg cat ggt gat gtc ttg cct aat gat ttt gcc tcg aaa cta cca His Leu His Gly Asp Val Leu Pro Asn Asp Phe Ala Ser Lys Leu Pro 870 875 880 885			2996
ttg caa gag cca ttt ttc cct ttt gaa ggc aag gca agg gtc tat agg Leu Gln Glu Pro Phe Phe Pro Phe Glu Gly Lys Ala Arg Val Tyr Arg 890 895 900			3044
aat gaa gga aga cgc ttg gcg tgt ggg gac acg gtt gat ggt ttg ccc Asn Glu Gly Arg Arg Leu Ala Cys Gly Asp Thr Val Asp Gly Leu Pro 905 910 915			3092
gtt gtt gcg cgt ctc ggc gac ctt gtt ttc gca ggg ttg gct atg ccg Val Val Ala Arg Leu Gly Asp Leu Val Phe Ala Gly Leu Ala Met Pro 920 925 930			3140
cca gat ggg tgg gcc att acc gca cct ttt acg ctg cag tgt ctc tct Pro Asp Gly Trp Ala Ile Thr Ala Pro Phe Thr Leu Gln Cys Leu Ser 935 940 945			3188
gaa cgt ggc acg ctg tca gcg atg gca gtg gtc atg act ggt ata gac Glu Arg Gly Thr Leu Ser Ala Met Ala Val Val Met Thr Gly Ile Asp 950 955 960 965			3236
ccc cga act tgg act gga act atc ttc aga tta gga tct ctg gcc act Pro Arg Thr Trp Thr Gly Thr Ile Phe Arg Leu Gly Ser Leu Ala Thr 970 975 980			3284
agc tac atg gga ttc gtt tgt gac aac gtg ttg tat act gct cac cat Ser Tyr Met Gly Phe Val Cys Asp Asn Val Leu Tyr Thr Ala His His 985 990 995			3332
ggc agc aag ggg cgc cgg ttg gct cat ccc aca ggc tcc ata cac cca Gly Ser Lys Gly Arg Arg Leu Ala His Pro Thr Gly Ser Ile His Pro 1000 1005 1010			3380
ata acc gtt gac gcg gct aat gac cag gac atc tat caa cca cca tgt Ile Thr Val Asp Ala Ala Asn Asp Gln Asp Ile Tyr Gln Pro Pro Cys 1015 1020 1025			3428
gga gct ggg tcc ctt act cgg tgc tct tgc ggg gag acc aag ggg tat Gly Ala Gly Ser Leu Thr Arg Cys Ser Cys Gly Glu Thr Lys Gly Tyr 1030 1035 1040 1045			3476
ctg gta aca cga ctg ggg tca ttg gtt gag gtc aac aaa tcc gat gac Leu Val Thr Arg Leu Gly Ser Leu Val Glu Val Asn Lys Ser Asp Asp 1050 1055 1060			3524
cct tat tgg tgt gtg tgc ggg gcc ctt ccc atg gct gtt gcc aag ggt Pro Tyr Trp Cys Val Cys Gly Ala Leu Pro Met Ala Val Ala Lys Gly 1065 1070 1075			3572
tct tca ggt gcc ccg att ctg tgc tcc tcc ggg cat gtt att ggg atg Ser Ser Gly Ala Pro Ile Leu Cys Ser Ser Gly His Val Ile Gly Met 1080 1085 1090			3620
ttc acc gct gct aga aat tct ggc ggt tca gtc agt cag att agg gtt Phe Thr Ala Ala Arg Asn Ser Gly Gly Ser Val Ser Gln Ile Arg Val 1095 1100 1105			3668
agg ccg ttg gtg tgt gct gga tac cat ccc cag tac aca gca cat gcc			3716

Arg Pro Leu Val Cys Ala Gly Tyr His Pro Gln Tyr Thr Ala His Ala	
1110 1115 1120 1125	
act ctt gat aca aaa acg cgt atg cag atc ttc gtg aag acc ctg acg	3764
Thr Leu Asp Thr Lys Thr Arg Met Gln Ile Phe Val Lys Thr Leu Thr	
1130 1135 1140	
ggc aag acc atc act ctt gag gtc gag ccc agt gac acc atc gag aat	3812
Gly Lys Thr Ile Thr Leu Glu Val Glu Pro Ser Asp Thr Ile Glu Asn	
1145 1150 1155	
gtc aag gcc aag atc caa gac aag gaa ggc atc cca cct gac cag cag	3860
Val Lys Ala Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln	
1160 1165 1170	
agg ctg ata ttc gcg ggc aaa cag ctg gag gat ggc cgc acc ctg tcc	3908
Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser	
1175 1180 1185	
gac tac aac atc cag aaa gag tcc acc ttg cac ctg gtg ctg cgt ctc	3956
Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Leu	
1190 1195 1200 1205	
cgc ggt ggc gcg ccc atc acg gcg tac gcc cag cag acg aga ggc ctc	4004
Arg Gly Gly Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu	
1210 1215 1220	
cta ggg tgt ata atc acc agc ctg act ggc cgg gac aaa aac caa gtg	4052
Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val	
1225 1230 1235	
gag ggt gag gtc cag atc gtg tca act gct acc caa acc ttc ctg gca	4100
Glu Gly Glu Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala	
1240 1245 1250	
acg tgc atc aat ggg gta tgc tgg act gtc tac cac ggg gcc gga acg	4148
Thr Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr	
1255 1260 1265	
agg acc atc gca tca ccc aag ggt cct gtc atc cag atg tat acc aat	4196
Arg Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn	
1270 1275 1280 1285	
gtg gac caa gac ctt gtg ggc tgg ccc gct cct caa ggt tcc cgc tca	4244
Val Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser	
1290 1295 1300	
ttg aca ccc tgt acc tgc ggc tcc tcg gac ctt tac ctg gtc acg agg	4292
Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg	
1305 1310 1315	
cac gcc gat gtc att ccc gtg cgc cgg cga ggt gat agc agg ggt agc	4340
His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser	
1320 1325 1330	
ctg ctt tcg ccc cgg ccc att tcc tac ttg aaa ggc tcc tcg ggg ggt	4388
Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly	
1335 1340 1345	
ccg ctg ttg tgc ccc gcg gga cac gcc gtg ggc cta ttc agg gcc gcg	4436
Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu Phe Arg Ala Ala	
1350 1355 1360 1365	

gtg tgc acc cgt gga gtg gct aaa gcg gtg gac ttt atc cct gtg gag	4484
Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu	
1370 1375 1380	
aac cta ggg aca acc atg aga tcc ccg gtg ttc acg gac aac tcc tct	4532
Asn Leu Gly Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser	
1385 1390 1395	
cca cca gca gtg ccc cag agc ttc cag gtg gcc cac ctg cat gct ccc	4580
Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro	
1400 1405 1410	
acc ggc agc ggt aag agc acc aag gtc ccg gct gcg tac gca gcc cag	4628
Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln	
1415 1420 1425	
ggc tac aag gtg ttg gtg ctc aac ccc tct gtt gct gca acg ctg ggc	4676
Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly	
1430 1435 1440 1445	
ttt ggt gct tac atg tcc aag gcc cat ggg gtt gat cct aat atc agg	4724
Phe Gly Ala Tyr Met Ser Lys Ala His Gly Val Asp Pro Asn Ile Arg	
1450 1455 1460	
acc ggg gtg aga aca att acc act ggc agc ccc atc acg tac tcc acc	4772
Thr Gly Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr	
1465 1470 1475	
tac ggc aag ttc ctt gcc gac ggc ggg tgc tca gga ggt gct tat gac	4820
Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp	
1480 1485 1490	
ata ata att tgt gac gag tgc cac tcc acg gat gcc aca tcc atc ttg	4868
Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu	
1495 1500 1505	
ggc atc ggc act gtc ctt gac caa gca gag act gcg ggg gcg aga ctg	4916
Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu	
1510 1515 1520 1525	
gtt gtg ctc gcc act gct acc cct ccg ggc tcc gtc act gtg tcc cat	4964
Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Ser His	
1530 1535 1540	
cct aac atc gag gag gtt gct ctg tcc acc acc gga gag atc ccc ttt	5012
Pro Asn Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe	
1545 1550 1555	
tac ggc aag gct atc ccc ctc gag gtg atc aag ggg gga aga cat ctc	5060
Tyr Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu	
1560 1565 1570	
atc ttc tgc cac tca aag aag aag tgc gac gag ctc gcc gcg aag ctg	5108
Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu	
1575 1580 1585	
gtc gca ttg ggc atc aat gcc gtg gcc tac tac cgc ggt ctt gac gtg	5156
Val Ala Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val	
1590 1595 1600 1605	
tct gtc atc ccg acc agc ggc gat gtt gtc gtc gtg tcg acc gat gct	5204
Ser Val Ile Pro Thr Ser Gly Asp Val Val Val Val Ser Thr Asp Ala	
1610 1615 1620	

ctc atg act ggc ttt acc ggc gac ttc gac tct gtg ata gac tgc aac	5252
Leu Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn	
1625 1630 1635	
acg tgt gtc act cag aca gtc gat ttc agc ctt gac cct acc ttt acc	5300
Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr	
1640 1645 1650	
att gag aca acc acg ctc ccc cag gat gct gtc tcc agg act caa cgc	5348
Ile Glu Thr Thr Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg	
1655 1660 1665	
cgg ggc agg act ggc agg ggg aag cca ggc atc tat aga ttt gtg gca	5396
Arg Gly Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala	
1670 1675 1680 1685	
ccg ggg gag cgc ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag	5444
Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu	
1690 1695 1700	
tgc tat gac gcg ggc tgt gct tgg tat gag ctc acg ccc gcc gag act	5492
Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr	
1705 1710 1715	
aca gtt agg cta cga gcg tac atg aac acc ccg ggg ctt ccc gtg tgc	5540
Thr Val Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys	
1720 1725 1730	
cag gac cat ctt gaa ttt tgg gag ggc gtc ttt acg ggc ctc act cat	5588
Gln Asp His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His	
1735 1740 1745	
ata gat gcc cac ttt tta tcc cag aca aag cag agt ggg gag aac ttt	5636
Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe	
1750 1755 1760 1765	
cct tac ctg gta gcg tac caa gcc acc gtg tgc gct agg gct caa gcc	5684
Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala	
1770 1775 1780	
cct ccc cca tcg tgg gac cag atg tgg aag tgt ttg atc cgc ctt aaa	5732
Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys	
1785 1790 1795	
ccc acc ctc cat ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt	5780
Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val	
1800 1805 1810	
cag aat gaa gtc acc ctg acg cac cca atc acc aaa tac atc atg aca	5828
Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Thr	
1815 1820 1825	
tgc atg tcg gcc gac ctg gag gtc gtc acg agc acc tgg gtg ctc gtt	5876
Cys Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val	
1830 1835 1840 1845	
ggc ggc gtc ctg gct gct ctg gcc gcg tat tgc ctg tca aca ggc tgc	5924
Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys	
1850 1855 1860	
gtg gtc ata gtg ggc agg atc gtc ttg tcc ggg aag ccg gca att ata	5972
Val Val Ile Val Gly Arg Ile Val Leu Ser Gly Lys Pro Ala Ile Ile	

1865	1870	1875	
cct gac agg gag gtt ctc tac cag gag ttc gat gag atg gaa gag tgc Pro Asp Arg Glu Val Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys 1880 1885 1890			6020
tct cag cac tta ccg tac atc gag caa ggg atg atg ctc gct gag cag Ser Gln His Leu Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln 1895 1900 1905			6068
ttc aag cag aag gcc ctc ggc ctc ctg cag acc gcg tcc cgc cat gca Phe Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg His Ala 1910 1915 1920 1925			6116
gag gtt atc acc cct gct gtc cag acc aac tgg cag aaa ctc gag gtc Glu Val Ile Thr Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Val 1930 1935 1940			6164
ttt tgg gcg aag cac atg tgg aat ttc atc agt ggg ata caa tac ttg Phe Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu 1945 1950 1955			6212
gcg ggc ctg tca acg ctg cct ggt aac ccc gcc att gct tca ttg atg Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met 1960 1965 1970			6260
gct ttt aca gct gcc gtc acc agc cca cta acc act ggc caa acc ctc Ala Phe Thr Ala Ala Val Thr Ser Pro Leu Thr Thr Gly Gln Thr Leu 1975 1980 1985			6308
ctc ttc aac ata ttg ggg ggg tgg gtg gct gcc cag ctc gcc gcc ccc Leu Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro 1990 1995 2000 2005			6356
ggt gcc gct act gcc ttt gtg ggt gct ggc cta gct ggc gcc gcc atc Gly Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile 2010 2015 2020			6404
ggc agc gtt gga ctg ggg aag gtc ctc gtg gac att ctt gca ggg tat Gly Ser Val Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr 2025 2030 2035			6452
ggc gcg ggc gtg gcg gga gct ctt gta gca ttc aag atc atg agc ggt Gly Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly 2040 2045 2050			6500
gag gtc ccc tcc acg gag gac ctg gtc aat ctg ctg ccc gcc atc ctc Glu Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu 2055 2060 2065			6548
tcg cct gga gcc ctt gta gtc ggt gtg gtc tgc gca gca ata ctg cgc Ser Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg 2070 2075 2080 2085			6596
cgg cac gtt ggc ccg ggc gag ggg gca gtg caa tgg atg aac cgg cta Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu 2090 2095 2100			6644
ata gcc ttc gcc tcc cgg ggg aac cat gtt tcc ccc acg cac tac gtg Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val 2105 2110 2115			6692
ccg gag agc gat gca gcc gcc cgc gtc act gcc ata ctc agc agc ctc			6740

Pro Glu Ser Asp Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu	
2120	2125 2130
act gta acc cag ctc ctg agg cga ctg cat cag tgg ata agc tcg gag	6788
Thr Val Thr Gln Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu	
2135	2140 2145
tgt acc act cca tgc tcc ggt tcc tgg cta agg gac atc tgg gac tgg	6836
Cys Thr Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp	
2150	2155 2160 2165
ata tgc gag gtg ctg agc gac ttt aag acc tgg ctg aaa gcc aag ctc	6884
Ile Cys Glu Val Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu	
	2170 2175 2180
atg cca caa ctg cct ggg att ccc ttt gtg tcc tgc cag cgc ggg tat	6932
Met Pro Gln Leu Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr	
	2185 2190 2195
agg ggg gtc tgg cga gga gac ggc att atg cac act cgc tgc cac tgt	6980
Arg Gly Val Trp Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys	
	2200 2205 2210
gga gct gag atc act gga cat gtc aaa aac ggg acg atg agg atc gtc	7028
Gly Ala Glu Ile Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val	
	2215 2220 2225
ggt cct agg acc tgc agg aac atg tgg agt ggg acg ttc ccc att aac	7076
Gly Pro Arg Thr Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn	
	2230 2235 2240 2245
gcc tac acc acg ggc ccc tgt act ccc ctt cct gcg ccg aac tat aag	7124
Ala Tyr Thr Thr Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Lys	
	2250 2255 2260
ttc gcg ctg tgg agg gtg tct gca gag gaa tac gtg gag ata agg cgg	7172
Phe Ala Leu Trp Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Arg	
	2265 2270 2275
gtg ggg gac ttc cac tac gta tgc ggt atg act act gac aat ctt aaa	7220
Val Gly Asp Phe His Tyr Val Ser Gly Met Thr Thr Asp Asn Leu Lys	
	2280 2285 2290
tgc ccg tgc cag atc cca tgc ccc gaa ttt ttc aca gaa ttg gac ggg	7268
Cys Pro Cys Gln Ile Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly	
	2295 2300 2305
gtg cgc cta cac agg ttt gcg ccc cct tgc aag ccc ttg ctg cgg gag	7316
Val Arg Leu His Arg Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu	
	2310 2315 2320 2325
gag gta tca ttc aga gta gga ctc cac gag tac ccg gtg ggg tgc caa	7364
Glu Val Ser Phe Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln	
	2330 2335 2340
tta cct tgc gag ccc gaa ccg gac gta gcc gtg ttg acg tcc atg ctc	7412
Leu Pro Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu	
	2345 2350 2355
act gat ccc tcc cat ata aca gca gag gcg gcc ggg aga agg ttg gcg	7460
Thr Asp Pro Ser His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala	
	2360 2365 2370

aga ggg tca ccc cct tct atg gcc agc tcc tcg gct agc cag ctg tcc Arg Gly Ser Pro Pro Ser Met Ala Ser Ser Ser Ala Ser Gln Leu Ser 2375 2380 2385	7508
gct cca tct ctc aag gca act tgc acc gcc aac cat gac tcc cct gac Ala Pro Ser Leu Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp 2390 2395 2400 2405	7556
gcc gag ctc ata gag gct aac ctc ctg tgg agg cag gag atg ggc ggc Ala Glu Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly 2410 2415 2420	7604
aac atc acc agg gtt gag tca gag aac aaa gtg gtg att ctg gac tcc Asn Ile Thr Arg Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser 2425 2430 2435	7652
ttc gat ccg ctt gtg gca gag gag gat gag cgg gag gtc tcc gta cct Phe Asp Pro Leu Val Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro 2440 2445 2450	7700
gca gaa att ctg cgg aag tct cgg aga ttc gcc cgg gcc ctg ccc gtc Ala Glu Ile Leu Arg Lys Ser Arg Arg Phe Ala Arg Ala Leu Pro Val 2455 2460 2465	7748
tgg gcg cgg ccg gac tac aac ccc ccg cta gta gag acg tgg aaa aag Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys 2470 2475 2480 2485	7796
cct gac tac gaa cca cct gtg gtc cat ggc tgc ccg cta cca cct cca Pro Asp Tyr Glu Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro 2490 2495 2500	7844
cgg tcc cct cct gtg cct ccg cct cgg aaa aag cgt acg gtg gtc ctc Arg Ser Pro Pro Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu 2505 2510 2515	7892
acc gaa tca acc cta tct act gcc ttg gcc gag ctt gcc acc aaa agt Thr Glu Ser Thr Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Lys Ser 2520 2525 2530	7940
ttt ggc agc tcc tca act tcc ggc att acg ggc gac aat acg aca aca Phe Gly Ser Ser Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr 2535 2540 2545	7988
tcc tct gag ccc gcc cct tct ggc tgc ccc ccc gac tcc gac gtt gag Ser Ser Glu Pro Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Val Glu 2550 2555 2560 2565	8036
tcc tat tct tcc atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp 2570 2575 2580	8084
ctc agc gac ggg tca tgg tcg acg gtc agt agt ggg gcc gac acg gaa Leu Ser Asp Gly Ser Trp Ser Thr Val Ser Ser Gly Ala Asp Thr Glu 2585 2590 2595	8132
gat gtc gtg tgc tgc tca atg tct tat tcc tgg aca ggc gca ctc gtc Asp Val Val Cys Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val 2600 2605 2610	8180
acc ccg tgc gct gcg gaa gaa caa aaa ctg ccc atc aac gca ctg agc Thr Pro Cys Ala Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser 2615 2620 2625	8228

aac tcg ttg cta cgc cat cac aat ctg gtg tat tcc acc act tca cgc	8276
Asn Ser Leu Leu Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg	
2630 2635 2640 2645	
agt gct tgc caa agg cag aag aaa gtc aca ttt gac aga ctg caa gtt	8324
Ser Ala Cys Gln Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val	
2650 2655 2660	
ctg gac agc cat tac cag gac gtg ctc aag gag gtc aaa gca gcg gcg	8372
Leu Asp Ser His Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ala	
2665 2670 2675	
tca aaa gtg aag gct aac ttg cta tcc gta gag gaa gct tgc agc ctg	8420
Ser Lys Val Lys Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu	
2680 2685 2690	
acg ccc cca cat tca gcc aaa tcc aag ttt ggc tat ggg gca aaa gac	8468
Thr Pro Pro His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp	
2695 2700 2705	
gtc cgt tgc cat gcc aga aag gcc gta gcc cac atc aac tcc gtg tgg	8516
Val Arg Cys His Ala Arg Lys Ala Val Ala His Ile Asn Ser Val Trp	
2710 2715 2720 2725	
aaa gac ctt ctg gaa gac agt gta aca cca ata gac act acc atc atg	8564
Lys Asp Leu Leu Glu Asp Ser Val Thr Pro Ile Asp Thr Thr Ile Met	
2730 2735 2740	
gcc aag aac gag gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag	8612
Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys	
2745 2750 2755	
cca gct cgt ctc atc gtg ttc ccc gac ctg ggc gtg cgc gtg tgc gag	8660
Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu	
2760 2765 2770	
aag atg gcc ctg tac gac gtg gtt agc aag ctc ccc ctg gcc gtg atg	8708
Lys Met Ala Leu Tyr Asp Val Val Ser Lys Leu Pro Leu Ala Val Met	
2775 2780 2785	
gga agc tcc tac gga ttc caa tac tca cca gga cag cgg gtt gaa ttc	8756
Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe	
2790 2795 2800 2805	
ctc gtg caa gcg tgg aag tcc aag aag acc ccg atg ggg ttc tcg tat	8804
Leu Val Gln Ala Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr	
2810 2815 2820	
gat acc cgc tgt ttt gac tcc aca gtc act gag agc gac atc cgt acg	8852
Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr	
2825 2830 2835	
gag gag gca att tac caa tgt tgt gac ctg gac ccc caa gcc cgc gtg	8900
Glu Glu Ala Ile Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val	
2840 2845 2850	
gcc atc aag tcc ctc act gag agg ctt tat gtt ggg ggc cct ctt acc	8948
Ala Ile Lys Ser Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr	
2855 2860 2865	
aat tca agg ggg gaa aac tgc ggc tac cgc agg tgc cgc gcg agc ggc	8996
Asn Ser Arg Gly Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly	

2870	2875	2880	2885	
gta ctg aca act agc tgt ggt aac acc ctc act tgc tac atc aag gcc				9044
Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala				
2890		2895	2900	
cgg gca gcc tgt cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg				9092
Arg Ala Ala Cys Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val				
2905		2910	2915	
tgt ggc gac gac tta gtc gtt atc tgt gaa agt gcg ggg gtc cag gag				9140
Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu				
2920		2925	2930	
gac gcg gcg agc ctg aga gcc ttc acg gag gct atg acc agg tac tcc				9188
Asp Ala Ala Ser Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser				
2935		2940	2945	
gcc ccc ccc ggg gac ccc cca caa cca gaa tac gac ttg gag ctt ata				9236
Ala Pro Pro Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile				
2950		2955	2960	2965
aca tca tgc tcc tcc aac gtg tca gtc gcc cac gac ggc gct gga aag				9284
Thr Ser Cys Ser Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys				
2970		2975	2980	
agg gtc tac tac ctt acc cgt gac cct aca acc ccc ctc gcg aga gcc				9332
Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala				
2985		2990	2995	
gcg tgg gag aca gca aga cac act cca gtc aat tcc tgg cta ggc aac				9380
Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn				
3000		3005	3010	
ata atc atg ttt gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc				9428
Ile Ile Met Phe Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr				
3015		3020	3025	
cat ttc ttt agc gtc ctc ata gcc agg gat cag ctt gaa cag gct ctt				9476
His Phe Phe Ser Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu				
3030		3035	3040	3045
aac tgt gag atc tac gga gcc tgc tac tcc ata gaa cca ctg gat cta				9524
Asn Cys Glu Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu				
3050		3055	3060	
cct cca atc att caa aga ctc cat ggc ctc agc gca ttt tca ctc cac				9572
Pro Pro Ile Ile Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His				
3065		3070	3075	
agt tac tct cca ggt gaa atc aat agg gtg gcc gca tgc ctc aga aaa				9620
Ser Tyr Ser Ser Pro Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys				
3080		3085	3090	
ctt ggg gtc ccg ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc				9668
Leu Gly Val Pro Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val				
3095		3100	3105	
cgc gct agg ctt ctg tcc aga gga ggc agg gct gcc ata tgt ggc aag				9716
Arg Ala Arg Leu Leu Ser Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys				
3110		3115	3120	3125
tac ctc ttc aac tgg gca gta aga aca aag ctc aaa ctc act cca ata				9764

Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile
 3130 3135 3140
 gcg gcc gct ggc cgg ctg gac ttg tcc ggt tgg ttc acg gct ggc tac 9812
 Ala Ala Ala Gly Arg Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr
 3145 3150 3155
 agc ggg gga gac att tat cac agc gtg tct cat gcc cgg ccc cgc tgg 9860
 Ser Gly Gly Asp Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp
 3160 3165 3170
 ttc tgg ttt tgc cta ctc ctg ctc gct gca ggg gta ggc atc tac ctc 9908
 Phe Trp Phe Cys Leu Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu
 3175 3180 3185
 ctc ccc aac cga tgaaggttgg ggtaaact cggcctctt aagccatttc 9960
 Leu Pro Asn Arg
 3190
 ctgtttttttt tttttttttt tttttttttt tctttttttt tttctttcct ttccttcttt 10020
 ttttcttttc tttttccctt ctttaatggt ggctccatct tagccctagt cacggctagc 10080
 tgtgaaagggt ccgtgagcgc catgactgca gagagtgtcg atactggcct ctctgcagat 10140
 catgttctag actggaattc gtgcagagc tccctatagt gagtcgtatt agaggccgac 10200
 ttggccaaat tcgtaatcat ggtcatagct gtttctgtg tgaaattggt atccgctcac 10260
 aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctgggggtg cctaattgagt 10320
 gagctaactc acattaattg cgttgcgctc actgccgct ttcagtcgg gaaacctgtc 10380
 gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 10440
 ctcttcogct tctcgtctca ctgactcgct gcgctcggtc gttcggtgc ggcgagcgg 10500
 atcagctcac tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa 10560
 gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc 10620
 gtttttccat aggctccgcc ccctgacga gcatcacaaa aatcgacgct caagtacagag 10680
 gtggcgaaac ccgacaggac tataaagata ccaggcggtt cccctggaa gctccctcgt 10740
 gcgctctcct gttccgacct tgccgcttac cggatacctg tccgcctttc tcccttcggg 10800
 aagcgtggcg ctttctcata gtcacgctg taggtatctc agttcgggtg aggtcggtcg 10860
 ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg 10920
 taactatcgt cttgagtcca acccgtaag acacgactta tcgccactgg cagcagccac 10980
 tggtaacagg attagcagag cgaggatatg aggcggtgct acagagttct tgaagtgggtg 11040
 gcctaactac ggctacacta gaaggacagt atttggtatc tgcgctctgc tgaagccagt 11100
 taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg 11160
 tggttttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc 11220
 tttgatcttt tctacgggggt ctgacgctca gtggaacgaa aactcacgtt aagggatttt 11280

ggatcatgaga ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt 11340
taaatacaatc taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag 11400
tgaggcacct atctcagcga tctgtctatt tcgttcatcc atagttgcct gactccccgt 11460
cgtgtagata actacgatac gggagggcctt accatctggc cccagtgcctg caatgatacc 11520
gcgagaccca cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc 11580
cgagcgcaga agtggctctg caactttatc cgctccatc cagtctatta attgttgccg 11640
ggaagctaga gtaagtagtt cgccagttaa tagtttgccg aacgttggtg ccattgctac 11700
aggcatcgtg gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg 11760
atcaaggcga gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ccttcggtcc 11820
tccgatcgtt gtcagaagta agttggccgc agtggttatca ctcatggta tggcagcact 11880
gcataattct cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc 11940
aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat 12000
acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc 12060
ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac 12120
tcgtgcaccc aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa 12180
aacaggaagg caaatgcg caaaaaagg aataaggcg acacggaaat gttgaatact 12240
catactcttc ctttttcaat attattgaag cttttatcag ggttattgtc tcatgagcgg 12300
atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca ctttccccg 12360
aaaagtgcc cctgacgcgc cctgtagcgg cgcattaagc gcggcgggtg tgggtggttac 12420
gcgcagcgtg accgctacac ttgccagcgc cctagcgcgc gtcctttcg ctttcttccc 12480
ttcctttctc gccacgttcg ccggctttcc ccgtcaagct ctaaactcggg gcatcccttt 12540
agggttcga tttagtgtt tacggcacct cgacccaaa aaacttgatt aggggtgatg 12600
ttcacgtagt gggccatcgc cctgatagac ggtttttcgc ctttgacgt tggagtccac 12660
gttctttaat agtggactct tgttccaaac tggaacaaca ctcaacccta tctcgggtcta 12720
ttcttttgat ttataaggga ttttgccgat ttgcgcctat tggttaaaaa atgagctgat 12780
ttaacaaaaa ttaacgcga attttaacaa aatattaaca aaatattaac gtttacaatt 12840
tcccattcgc cattcaggct gcgcaactgt tgggaagggc gatcgggtgcg ggcctcttcg 12900
ctattacgcc agctggcgaa agggggatgt gctgcaaggc gattaagttg ggtaacgcca 12960
gggttttccc agtcacgacg ttgtaaaacg acggccagtg ccaagctgac ttggtcagcg 13020
gccgctaata cgactcacta ta 13042

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

```
<220>
<221> CDS
<222> (342)..(9143)
```

<400> 24																	
gccagccccc	tgatgggggc	gacactccac	catgaatcac	tcccctgtga	ggaactactg											60	
tcttcacgca	gaaagcgtct	agccatggcg	ttagtatgag	tgctcgtgcag	cctccaggac											120	
ccccctccc	gggagagcca	tagtgggtctg	cggaaccggt	gagtacaccg	gaattgccag											180	
gacgaccggg	tcctttcttg	gataaacccg	ctcaatgcct	ggagatttgg	gcgtgcccc											240	
gcaagactgc	tagccgagta	gtgttgggtc	gcgaaaggcc	ttgtgggtact	gcctgatagg											300	
gtgcttgcca	gtgccccggg	aggtctcgtg	gaccgtgcac	c atg	cct gtt att tct											356	
					Met Pro Val Ile Ser												
					1											5	
act caa aca agt cct gta cct gcg ccc aga acg cgc aag aac aag cag																404	
Thr Gln Thr Ser Pro Val Pro Ala Pro Arg Thr Arg Lys Asn Lys Gln																	
					10						15						20
acg cag gct tca tat cct gtg tcc att aaa aca tct gtt gaa agg gga																452	
Thr Gln Ala Ser Tyr Pro Val Ser Ile Lys Thr Ser Val Glu Arg Gly																	
					25						30						35
caa cga gca aag cgc aaa gtc cag cgc gat gct cgg cct cgt aat tac																500	
Gln Arg Ala Lys Arg Lys Val Gln Arg Asp Ala Arg Pro Arg Asn Tyr																	
					40						45						50
aaa att gct ggt atc cat gat ggc ttg cag aca ttg gct cag gct gct																548	
Lys Ile Ala Gly Ile His Asp Gly Leu Gln Thr Leu Ala Gln Ala Ala																	
					55						60						65
ttg cca gct cat ggt tgg gga cgc caa gac cct cgc cat aag tct cgc																596	
Leu Pro Ala His Gly Trp Gly Arg Gln Asp Pro Arg His Lys Ser Arg																	
					70						75						80
															85		
aat ctt gga atc ctt ctg gat tac cct ttg ggg tgg att ggt gat gtt																644	
Asn Leu Gly Ile Leu Leu Asp Tyr Pro Leu Gly Trp Ile Gly Asp Val																	
					90						95						100
aca act cac aca cct cta gta ggc ccg ctg gtg gca gga gcg gtc gtt																692	
Thr Thr His Thr Pro Leu Val Gly Pro Leu Val Ala Gly Ala Val Val																	
					105						110						115
cga cca gtc tgc cag ata gta cgc ttg ctg gag gat gga gtc aac tgg																740	
Arg Pro Val Cys Gln Ile Val Arg Arg Leu Leu Glu Asp Gly Val Asn Trp																	
					120						125						130
gct act ggt tgg ttc ggt gtc cac ctt ttt gtg gta tgt ctg cta tct																788	
Ala Thr Gly Trp Phe Gly Val His Leu Phe Val Val Cys Leu Leu Ser																	

135	140	145	
ttg gcc tgt ccc tgt agt ggg gcg cgg gtc act gac cca gac aca aat Leu Ala Cys Pro Cys Ser Gly Ala Arg Val Thr Asp Pro Asp Thr Asn 150 155 160 165			836
acc aca atc ctg acc aat tgc tgc cag cgt aat cag gtt atc tat tgt Thr Thr Ile Leu Thr Asn Cys Cys Gln Arg Asn Gln Val Ile Tyr Cys 170 175 180			884
tct cct tcc act tgc cta cac gag cct ggt tgt gtg atc tgt gcg gac Ser Pro Ser Thr Cys Leu His Glu Pro Gly Cys Val Ile Cys Ala Asp 185 190 195			932
gag tgc tgg gtt ccc gcc aat ccg tac atc tca cac cct tcc aat tgg Glu Cys Trp Val Pro Ala Asn Pro Tyr Ile Ser His Pro Ser Asn Trp 200 205 210			980
act ggc acg gac tcc ttc ttg gct gac cac att gat ttt gtt atg ggc Thr Gly Thr Asp Ser Phe Leu Ala Asp His Ile Asp Phe Val Met Gly 215 220 225			1028
gct ctt gtg acc tgt gac gcc ctt gac att ggt gag ttg tgt ggt gcg Ala Leu Val Thr Cys Asp Ala Leu Asp Ile Gly Glu Leu Cys Gly Ala 230 235 240 245			1076
tgt gta tta gtc ggt gac tgg ctt gtc agg cac tgg ctt att cac ata Cys Val Leu Val Gly Asp Trp Leu Val Arg His Trp Leu Ile His Ile 250 255 260			1124
gac ctc aat gaa act ggt act tgt tac ctg gaa gtg ccc act gga ata Asp Leu Asn Glu Thr Gly Thr Cys Tyr Leu Glu Val Pro Thr Gly Ile 265 270 275			1172
gat cct ggg ttc cta ggg ttt atc ggg tgg atg gcc ggc aag gtc gag Asp Pro Gly Phe Leu Gly Phe Ile Gly Trp Met Ala Gly Lys Val Glu 280 285 290			1220
gct gtc atc ttc ttg acc aaa ctg gct tca caa gta cca tac gct att Ala Val Ile Phe Leu Thr Lys Leu Ala Ser Gln Val Pro Tyr Ala Ile 295 300 305			1268
gcg act atg ttt agc agt gta cac tac ctg gcg gtt ggc gct ctg atc Ala Thr Met Phe Ser Ser Val His Tyr Leu Ala Val Gly Ala Leu Ile 310 315 320 325			1316
tac tat gcc tct cgg ggc aag tgg tat cag ttg ctc cta gcg ctt atg Tyr Tyr Ala Ser Arg Gly Lys Trp Tyr Gln Leu Leu Leu Ala Leu Met 330 335 340			1364
ctt tac ata gaa gcg acc tct gga aac cct atc agg gtg ccc act gga Leu Tyr Ile Glu Ala Thr Ser Gly Asn Pro Ile Arg Val Pro Thr Gly 345 350 355			1412
tgc tca ata gct gag ttt tgc tgc cct ttg atg ata cca tgt cct tgc Cys Ser Ile Ala Glu Phe Cys Ser Pro Leu Met Ile Pro Cys Pro Cys 360 365 370			1460
cac tct tat ttg agt gag aat gtg tca gaa gtc att tgt tac agt cca His Ser Tyr Leu Ser Glu Asn Val Ser Glu Val Ile Cys Tyr Ser Pro 375 380 385			1508
aag tgg acc agg cct atc act ctg gag tat aac aac tcc ata tct tgg			1556

Lys Trp Thr Arg Pro Ile Thr Leu Glu Tyr Asn Asn Ser Ile Ser Trp 390 395 400 405	
tac ccc tat aca atc cct ggt gcg agg gga tgt atg gtt aaa ttc aaa Tyr Pro Tyr Thr Ile Pro Gly Ala Arg Gly Cys Met Val Lys Phe Lys 410 415 420	1604
aat aac aca tgg ggt tgc tgc cgt att cgc aat gtg cca tcg tac tgc Asn Asn Thr Trp Gly Cys Cys Arg Ile Arg Asn Val Pro Ser Tyr Cys 425 430 435	1652
act atg ggc act gat gca gtg tgg aac gac act cgc aac act tac gaa Thr Met Gly Thr Asp Ala Val Trp Asn Asp Thr Arg Asn Thr Tyr Glu 440 445 450	1700
gta tgc ggt gta aca cca tgg cta aca acc gca tgg cac aac ggc tca Val Cys Gly Val Thr Pro Trp Leu Thr Thr Ala Trp His Asn Gly Ser 455 460 465	1748
gcc ctg aaa ttg gct ata tta caa tac cct ggg tct aaa gaa atg ttt Ala Leu Lys Leu Ala Ile Leu Gln Tyr Pro Gly Ser Lys Glu Met Phe 470 475 480 485	1796
aaa cct cat aat tgg atg tca ggc cat ttg tat ttt gag gga tca gat Lys Pro His Asn Trp Met Ser Gly His Leu Tyr Phe Glu Gly Ser Asp 490 495 500	1844
acc cct ata gtt tac ttt tat gac cct gtg aat tcc act ctc cta cca Thr Pro Ile Val Tyr Phe Tyr Asp Pro Val Asn Ser Thr Leu Leu Pro 505 510 515	1892
ccg gag agg tgg gct agg ttg ccc ggt acc cca cct gtg gta cgt ggt Pro Glu Arg Trp Ala Arg Leu Pro Gly Thr Pro Pro Val Val Arg Gly 520 525 530	1940
tct tgg tta cag gtt ccg caa ggg ttt tac agt gat gtg aaa gac cta Ser Trp Leu Gln Val Pro Gln Gly Phe Tyr Ser Asp Val Lys Asp Leu 535 540 545	1988
gcc aca gga ttg atc acc aaa gac aaa gcc tgg aaa aat tat cag gtc Ala Thr Gly Leu Ile Thr Lys Asp Lys Ala Trp Lys Asn Tyr Gln Val 550 555 560 565	2036
tta tat tcc gcc acg ggt gct ttg tct ctt acg gga gtt acc acc aag Leu Tyr Ser Ala Thr Gly Ala Leu Ser Leu Thr Gly Val Thr Thr Lys 570 575 580	2084
gcc gtg gtg cta att ctg ttg ggg ttg tgt ggc agc aag tat ctt att Ala Val Val Leu Ile Leu Leu Gly Leu Cys Gly Ser Lys Tyr Leu Ile 585 590 595	2132
tta gcc tac ctc tgt tac ttg tcc ctt tgt ttt ggg cgc gct tct ggt Leu Ala Tyr Leu Cys Tyr Leu Ser Leu Cys Phe Gly Arg Ala Ser Gly 600 605 610	2180
tac cct ttg cgt cct gtg ctc cca tcc cag tcg tat ctc caa gct ggc Tyr Pro Leu Arg Pro Val Leu Pro Ser Gln Ser Tyr Leu Gln Ala Gly 615 620 625	2228
tgg gat gtt ttg tct aaa gct caa gta gct cct ttt gct ttg att ttc Trp Asp Val Leu Ser Lys Ala Gln Val Ala Pro Phe Ala Leu Ile Phe 630 635 640 645	2276

ttc atc tgt tgc tat ctc cgc tgc agg cta cgt tat gct gcc ctt tta	2324
Phe Ile Cys Cys Tyr Leu Arg Cys Arg Leu Arg Tyr Ala Ala Leu Leu	
650 655 660	
ggg ttt gtg ccc atg gct gcg ggc ttg ccc cta act ttc ttt gtt gca	2372
Gly Phe Val Pro Met Ala Ala Gly Leu Pro Leu Thr Phe Phe Val Ala	
665 670 675	
gca gct gct gcc caa cca gat tat gac tgg tgg gtg cga ctg cta gtg	2420
Ala Ala Ala Ala Gln Pro Asp Tyr Asp Trp Trp Val Arg Leu Leu Val	
680 685 690	
gca ggg tta gtt ttg tgg gcc ggc cgt aac cgt ggt cac cgc ata gct	2468
Ala Gly Leu Val Leu Trp Ala Gly Arg Asn Arg Gly His Arg Ile Ala	
695 700 705	
ctg ctt gta ggt cct tgg cct ctg gta gcg ctt tta acc ctc ttg cat	2516
Leu Leu Val Gly Pro Trp Pro Leu Val Ala Leu Leu Thr Leu Leu His	
710 715 720 725	
ttg gtt acg cct gct tca gct ctg gac acg gag gtg gcc gcg tcg tgt	2564
Leu Val Thr Pro Ala Ser Ala Leu Asp Thr Glu Val Ala Ala Ser Cys	
730 735 740	
ggc ggc gtt gtt ctt gtc ggg tta atg gcg ctg act ctg tcg cca tat	2612
Gly Gly Val Val Leu Val Gly Leu Met Ala Leu Thr Leu Ser Pro Tyr	
745 750 755	
tac aag cgc tat atc agc tgg tgc atg tgg tgg ctt cag tat ttt ctg	2660
Tyr Lys Arg Tyr Ile Ser Trp Cys Met Trp Trp Leu Gln Tyr Phe Leu	
760 765 770	
acc aga gta gaa gcg caa ctg cac gtg tgg gtt ccc ccc ctc aac gtc	2708
Thr Arg Val Glu Ala Gln Leu His Val Trp Val Pro Pro Leu Asn Val	
775 780 785	
cgg ggg ggg cgc gat gcc gtc atc tta ctc atg tgt gta gta cac ccg	2756
Arg Gly Gly Arg Asp Ala Val Ile Leu Leu Met Cys Val Val His Pro	
790 795 800 805	
acc ctg gta ttt gac atc acc aaa cta ctc ctg gcc atc ttc gga ccc	2804
Thr Leu Val Phe Asp Ile Thr Lys Leu Leu Ala Ile Phe Gly Pro	
810 815 820	
ctt tgg att ctt caa gcc agt ttg ctt aaa gtc ccc tac ttc gtg cgc	2852
Leu Trp Ile Leu Gln Ala Ser Leu Leu Lys Val Pro Tyr Phe Val Arg	
825 830 835	
gtt caa ggc ctt ctc cgg atc tgc gcg cta gcg cgg aag ata gcc gga	2900
Val Gln Gly Leu Leu Arg Ile Cys Ala Leu Ala Arg Lys Ile Ala Gly	
840 845 850	
ggt cat tac gtg caa atg gcc atc atc aag tta ggg gcg ctt act ggc	2948
Gly His Tyr Val Gln Met Ala Ile Ile Lys Leu Gly Ala Leu Thr Gly	
855 860 865	
acc tat gtg tat aac cat ctc acc cct ctt cga gac tgg gcg cac aac	2996
Thr Tyr Val Tyr Asn His Leu Thr Pro Leu Arg Asp Trp Ala His Asn	
870 875 880 885	
ggc ctg cga gat ctg gcc gtg gct gtg gaa cca gtc gtc ttc tcc cga	3044
Gly Leu Arg Asp Leu Ala Val Ala Val Glu Pro Val Val Phe Ser Arg	
890 895 900	

atg gag acc aag ctc atc acg tgg ggg gca gat acc gcc gcg tgc ggt Met Glu Thr Lys Leu Ile Thr Trp Gly Ala Asp Thr Ala Ala Cys Gly 905 910 915	3092
gac atc atc aac ggc ttg ccc gtc tct gcc cgt agg ggc cag gag ata Asp Ile Ile Asn Gly Leu Pro Val Ser Ala Arg Arg Gly Gln Glu Ile 920 925 930	3140
ctg ctt ggg cca gcc gac gga atg gtc tcc aag ggg tgg agg ttg ctg Leu Leu Gly Pro Ala Asp Gly Met Val Ser Lys Gly Trp Arg Leu Leu 935 940 945	3188
gcg ccc atc acg gcg tac gcc cag cag acg aga ggc ctc cta ggg tgt Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly Cys 950 955 960 965	3236
ata atc acc agc ctg act ggc cgg gac aaa aac caa gtg gag ggt gag Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu 970 975 980	3284
gtc cag atc gtg tca act gct acc caa acc ttc ctg gca acg tgc atc Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala Thr Cys Ile 985 990 995	3332
aat ggg gta tgc tgg act gtc tac cac ggg gcc gga acg agg acc atc Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr Ile 1000 1005 1010	3380
gca tca ccc aag ggt cct gtc atc cag atg tat acc aat gtg gac caa Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val Asp Gln 1015 1020 1025	3428
gac ctt gtg ggc tgg ccc gct cct caa ggt tcc cgc tca ttg aca ccc Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser Leu Thr Pro 1030 1035 1040 1045	3476
tgt acc tgc ggc tcc tcg gac ctt tac ctg gtc acg agg cac gcc gat Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp 1050 1055 1060	3524
gtc att ccc gtg cgc cgg cga ggt gat agc agg ggt agc ctg ctt tcg Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser 1065 1070 1075	3572
ccc cgg ccc att tcc tac ttg aaa ggc tcc tcg ggg ggt ccg ctg ttg Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu 1080 1085 1090	3620
tgc ccc gcg gga cac gcc gtg ggc cta ttc agg gcc gcg gtg tgc acc Cys Pro Ala Gly His Ala Val Gly Leu Phe Arg Ala Ala Val Cys Thr 1095 1100 1105	3668
cgt gga gtg gct aaa gcg gtg gac ttt atc cct gtg gag aac cta ggg Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn Leu Gly 1110 1115 1120 1125	3716
aca acc atg aga tcc ccg gtg ttc acg gac aac tcc tct cca cca gca Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala 1130 1135 1140	3764
gtg ccc cag agc ttc cag gtg gcc cac ctg cat gct ccc acc ggc agc Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser	3812

1145	1150	1155	
ggt aag agc acc aag gtc ccg gct gcg tac gca gcc cag ggc tac aag			3860
Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys			
1160	1165	1170	
gtg ttg gtg ctc aac ccc tct gtt gct gca acg ctg ggc ttt ggt gct			3908
Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala			
1175	1180	1185	
tac atg tcc aag gcc cat ggg gtt gat cct aat atc agg acc ggg gtg			3956
Tyr Met Ser Lys Ala His Gly Val Asp Pro Asn Ile Arg Thr Gly Val			
1190	1195	1200	1205
aga aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc aag			4004
Arg Thr Ile Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys			
1210	1215	1220	
ttc ctt gcc gac ggc ggg tgc tca gga ggt gct tat gac ata ata att			4052
Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile			
1225	1230	1235	
tgt gac gag tgc cac tcc acg gat gcc aca tcc atc ttg ggc atc ggc			4100
Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly			
1240	1245	1250	
act gtc ctt gac caa gca gag act gcg ggg gcg aga ctg gtt gtg ctc			4148
Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu			
1255	1260	1265	
gcc act gct acc cct ccg ggc tcc gtc act gtg tcc cat cct aac atc			4196
Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Ser His Pro Asn Ile			
1270	1275	1280	1285
gag gag gtt gct ctg tcc acc acc gga gag atc ccc ttt tac ggc aag			4244
Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys			
1290	1295	1300	
gct atc ccc ctc gag gtg atc aag ggg gga aga cat ctc atc ttc tgc			4292
Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys			
1305	1310	1315	
cac tca aag aag aag tgc gac gag ctc gcc gcg aag ctg gtc gca ttg			4340
His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu			
1320	1325	1330	
ggc atc aat gcc gtg gcc tac tac cgc ggt ctt gac gtg tct gtc atc			4388
Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile			
1335	1340	1345	
ccg acc agc ggc gat gtt gtc gtc gtg tcg acc gat gct ctc atg act			4436
Pro Thr Ser Gly Asp Val Val Val Val Ser Thr Asp Ala Leu Met Thr			
1350	1355	1360	1365
ggc ttt acc ggc gac ttc gac tct gtg ata gac tgc aac acg tgt gtc			4484
Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val			
1370	1375	1380	
act cag aca gtc gat ttc agc ctt gac cct acc ttt acc att gag aca			4532
Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr			
1385	1390	1395	
acc acg ctc ccc cag gat gct gtc tcc agg act caa cgc cgg ggc agg			4580

Thr Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg	
1400	1405 1410
act ggc agg ggg aag cca ggc atc tat aga ttt gtg gca ccg ggg gag	4628
Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu	
1415	1420 1425
cgc ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag tgc tat gac	4676
Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp	
1430	1435 1440 1445
gcg ggc tgt gct tgg tat gag ctc acg ccc gcc gag act aca gtt agg	4724
Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg	
1450	1455 1460
cta cga gcg tac atg aac acc ccg ggg ctt ccc gtg tgc cag gac cat	4772
Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His	
1465	1470 1475
ctt gaa ttt tgg gag ggc gtc ttt acg ggc ctc act cat ata gat gcc	4820
Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala	
1480	1485 1490
cac ttt tta tcc cag aca aag cag agt ggg gag aac ttt cct tac ctg	4868
His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Pro Tyr Leu	
1495	1500 1505
gta gcg tac caa gcc acc gtg tgc gct agg gct caa gcc cct ccc cca	4916
Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro	
1510	1515 1520 1525
tcg tgg gac cag atg tgg aag tgt ttg atc cgc ctt aaa ccc acc ctc	4964
Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu	
1530	1535 1540
cat ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt cag aat gaa	5012
His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu	
1545	1550 1555
gtc acc ctg acg cac cca atc acc aaa tac atc atg aca tgc atg tcg	5060
Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Thr Cys Met Ser	
1560	1565 1570
gcc gac ctg gag gtc gtc acg agc acc tgg gtg ctc gtt ggc ggc gtc	5108
Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val	
1575	1580 1585
ctg gct gct ctg gcc gcg tat tgc ctg tca aca ggc tgc gtg gtc ata	5156
Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile	
1590	1595 1600 1605
gtg ggc agg atc gtc ttg tcc ggg aag ccg gca att ata cct gac agg	5204
Val Gly Arg Ile Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg	
1610	1615 1620
gag gtt ctc tac cag gag ttc gat gag atg gaa gag tgc tct cag cac	5252
Glu Val Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His	
1625	1630 1635
tta ccg tac atc gag caa ggg atg atg ctc gct gag cag ttc aag cag	5300
Leu Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln	
1640	1645 1650

aag gcc ctc ggc ctc ctg cag acc gcg tcc cgc cat gca gag gtt atc Lys Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg His Ala Glu Val Ile 1655 1660 1665	5348
acc cct gct gtc cag acc aac tgg cag aaa ctc gag gtc ttt tgg gcg Thr Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Val Phe Trp Ala 1670 1675 1680 1685	5396
aag cac atg tgg aat ttc atc agt ggg ata caa tac ttg gcg ggc ctg Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu 1690 1695 1700	5444
tca acg ctg cct ggt aac ccc gcc att gct tca ttg atg gct ttt aca Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr 1705 1710 1715	5492
gct gcc gtc acc agc cca cta acc act ggc caa acc ctc ctc ttc aac Ala Ala Val Thr Ser Pro Leu Thr Thr Gly Gln Thr Leu Leu Phe Asn 1720 1725 1730	5540
ata ttg ggg ggg tgg gtg gct gcc cag ctc gcc gcc ccc ggt gcc gct Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala 1735 1740 1745	5588
act gcc ttt gtg ggt gct ggc cta gct ggc gcc gcc atc ggc agc gtt Thr Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile Gly Ser Val 1750 1755 1760 1765	5636
gga ctg ggg aag gtc ctc gtg gac att ctt gca ggg tat ggc gcg ggc Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala Gly 1770 1775 1780	5684
gtg gcg gga gct ctt gta gca ttc aag atc atg agc ggt gag gtc ccc Val Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro 1785 1790 1795	5732
tcc acg gag gac ctg gtc aat ctg ctg ccc gcc atc ctc tcg cct gga Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly 1800 1805 1810	5780
gcc ctt gta gtc ggt gtg gtc tgc gca gca ata ctg cgc cgg cac gtt Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His Val 1815 1820 1825	5828
ggc ccg ggc gag ggg gca gtg caa tgg atg aac cgg cta ata gcc ttc Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe 1830 1835 1840 1845	5876
gcc tcc cgg ggg aac cat gtt tcc ccc acg cac tac gtg ccg gag agc Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser 1850 1855 1860	5924
gat gca gcc gcc cgc gtc act gcc ata ctc agc agc ctc act gta acc Asp Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr 1865 1870 1875	5972
cag ctc ctg agg cga ctg cat cag tgg ata agc tcg gag tgt acc act Gln Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr 1880 1885 1890	6020
cca tgc tcc ggt tcc tgg cta agg gac atc tgg gac tgg ata tgc gag Pro Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu 1895 1900 1905	6068

gtg ctg agc gac ttt aag acc tgg ctg aaa gcc aag ctc atg cca caa 6116
 Val Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln
 1910 1915 1920 1925

ctg cct ggg att ccc ttt gtg tcc tgc cag cgc ggg tat agg ggg gtc 6164
 Leu Pro Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Arg Gly Val
 1930 1935 1940

tgg cga gga gac ggc att atg cac act cgc tgc cac tgt gga gct gag 6212
 Trp Arg Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu
 1945 1950 1955

atc act gga cat gtc aaa aac ggg acg atg agg atc gtc ggt cct agg 6260
 Ile Thr Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg
 1960 1965 1970

acc tgc agg aac atg tgg agt ggg acg ttc ccc att aac gcc tac acc 6308
 Thr Cys Arg Asn Met Trp Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr
 1975 1980 1985

acg ggc ccc tgt act ccc ctt cct gcg ccg aac tat aag ttc gcg ctg 6356
 Thr Gly Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Lys Phe Ala Leu
 1990 1995 2000 2005

tgg agg gtg tct gca gag gaa tac gtg gag ata agg cgg gtg ggg gac 6404
 Trp Arg Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Arg Val Gly Asp
 2010 2015 2020

ttc cac tac gta tcg ggt atg act act gac aat ctt aaa tgc ccg tgc 6452
 Phe His Tyr Val Ser Gly Met Thr Thr Asp Asn Leu Lys Cys Pro Cys
 2025 2030 2035

cag atc cca tcg ccc gaa ttt ttc aca gaa ttg gac ggg gtg cgc cta 6500
 Gln Ile Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu
 2040 2045 2050

cac agg ttt gcg ccc cct tgc aag ccc ttg ctg cgg gag gag gta tca 6548
 His Arg Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser
 2055 2060 2065

ttc aga gta gga ctc cac gag tac ccg gtg ggg tcg caa tta cct tgc 6596
 Phe Arg Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys
 2070 2075 2080 2085

gag ccc gaa ccg gac gta gcc gtg ttg acg tcc atg ctc act gat ccc 6644
 Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro
 2090 2095 2100

tcc cat ata aca gca gag gcg gcc ggg aga agg ttg gcg aga ggg tca 6692
 Ser His Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser
 2105 2110 2115

ccc cct tct atg gcc agc tcc tcg gct agc cag ctg tcc gct cca tct 6740
 Pro Pro Ser Met Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser
 2120 2125 2130

ctc aag gca act tgc acc gcc aac cat gac tcc cct gac gcc gag ctc 6788
 Leu Lys Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu
 2135 2140 2145

ata gag gct aac ctc ctg tgg agg cag gag atg ggc ggc aac atc acc 6836
 Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr

2150	2155	2160	2165	
agg gtt gag tca gag aac aaa gtg gtg att ctg gac tcc ttc gat ccg Arg Val Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro 2170 2175 2180				6884
ctt gtg gca gag gag gat gag cgg gag gtc tcc gta cct gca gaa att Leu Val Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu Ile 2185 2190 2195				6932
ctg cgg aag tct cgg aga ttc gcc cgg gcc ctg ccc gtc tgg gcg cgg Leu Arg Lys Ser Arg Arg Phe Ala Arg Ala Leu Pro Val Trp Ala Arg 2200 2205 2210				6980
ccg gac tac aac ccc ccg cta gta gag acg tgg aaa aag cct gac tac Pro Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr 2215 2220 2225				7028
gaa cca cct gtg gtc cat ggc tgc ccg cta cca cct cca cgg tcc cct Glu Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro Arg Ser Pro 2230 2235 2240 2245				7076
cct gtg cct ccg cct cgg aaa aag cgt acg gtg gtc ctc acc gaa tca Pro Val Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser 2250 2255 2260				7124
acc cta tct act gcc ttg gcc gag ctt gcc acc aaa agt ttt ggc agc Thr Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Lys Ser Phe Gly Ser 2265 2270 2275				7172
tcc tca act tcc ggc att acg ggc gac aat acg aca aca tcc tct gag Ser Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu 2280 2285 2290				7220
ccc gcc cct tct ggc tgc ccc ccc gac tcc gac gtt gag tcc tat tct Pro Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Val Glu Ser Tyr Ser 2295 2300 2305				7268
tcc atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat ctc agc gac Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp 2310 2315 2320 2325				7316
ggg tca tgg tcg acg gtc agt agt ggg gcc gac acg gaa gat gtc gtg Gly Ser Trp Ser Thr Val Ser Ser Gly Ala Asp Thr Glu Asp Val Val 2330 2335 2340				7364
tgc tgc tca atg tct tat tcc tgg aca ggc gca ctc gtc acc ccg tgc Cys Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys 2345 2350 2355				7412
gct gcg gaa gaa caa aaa ctg ccc atc aac gca ctg agc aac tcg ttg Ala Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu 2360 2365 2370				7460
cta cgc cat cac aat ctg gtg tat tcc acc act tca cgc agt gct tgc Leu Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys 2375 2380 2385				7508
caa agg cag aag aaa gtc aca ttt gac aga ctg caa gtt ctg gac agc Gln Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser 2390 2395 2400 2405				7556
cat tac cag gac gtg ctc aag gag gtc aaa gca gcg gcg tca aaa gtg				7604

His Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ala Ser Lys Val	
2410	2415 2420
aag gct aac ttg cta tcc gta gag gaa gct tgc agc ctg acg ccc cca	7652
Lys Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Thr Pro Pro	
2425	2430 2435
cat tca gcc aaa tcc aag ttt ggc tat ggg gca aaa gac gtc cgt tgc	7700
His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys	
2440	2445 2450
cat gcc aga aag gcc gta gcc cac atc aac tcc gtg tgg aaa gac ctt	7748
His Ala Arg Lys Ala Val Ala His Ile Asn Ser Val Trp Lys Asp Leu	
2455	2460 2465
ctg gaa gac agt gta aca cca ata gac act acc atc atg gcc aag aac	7796
Leu Glu Asp Ser Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn	
2470	2475 2480 2485
gag gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag cca gct cgt	7844
Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg	
2490	2495 2500
ctc atc gtg ttc ccc gac ctg ggc gtg cgc gtg tgc gag aag atg gcc	7892
Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala	
2505	2510 2515
ctg tac gac gtg gtt agc aag ctc ccc ctg gcc gtg atg gga agc tcc	7940
Leu Tyr Asp Val Val Ser Lys Leu Pro Leu Ala Val Met Gly Ser Ser	
2520	2525 2530
tac gga ttc caa tac tca cca gga cag cgg gtt gaa ttc ctc gtg caa	7988
Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln	
2535	2540 2545
gcg tgg aag tcc aag aag acc ccg atg ggg ttc tcg tat gat acc cgc	8036
Ala Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg	
2550	2555 2560 2565
tgt ttt gac tcc aca gtc act gag agc gac atc cgt acg gag gag gca	8084
Cys Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala	
2570	2575 2580
att tac caa tgt tgt gac ctg gac ccc caa gcc cgc gtg gcc atc aag	8132
Ile Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys	
2585	2590 2595
tcc ctc act gag agg ctt tat gtt ggg ggc cct ctt acc aat tca agg	8180
Ser Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg	
2600	2605 2610
ggg gaa aac tgc ggc tac cgc agg tgc cgc gcg agc ggc gta ctg aca	8228
Gly Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr	
2615	2620 2625
act agc tgt ggt aac acc ctc act tgc tac atc aag gcc cgg gca gcc	8276
Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala	
2630	2635 2640 2645
tgt cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg tgt ggc gac	8324
Cys Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp	
2650	2655 2660

gac tta gtc gtt atc tgt gaa agt gcg ggg gtc cag gag gac gcg gcg Asp Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala 2665 2670 2675	8372
agc ctg aga gcc ttc acg gag gct atg acc agg tac tcc gcc ccc ccc Ser Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro 2680 2685 2690	8420
ggg gac ccc cca caa cca gaa tac gac ttg gag ctt ata aca tca tgc Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys 2695 2700 2705	8468
tcc tcc aac gtg tca gtc gcc cac gac ggc gct gga aag agg gtc tac Ser Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr 2710 2715 2720 2725	8516
tac ctt acc cgt gac cct aca acc ccc ctc gcg aga gcc gcg tgg gag Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu 2730 2735 2740	8564
aca gca aga cac act cca gtc aat tcc tgg cta ggc aac ata atc atg Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met 2745 2750 2755	8612
ttt gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc cat ttc ttt Phe Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe 2760 2765 2770	8660
agc gtc ctc ata gcc agg gat cag ctt gaa cag gct ctt aac tgt gag Ser Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asn Cys Glu 2775 2780 2785	8708
atc tac gga gcc tgc tac tcc ata gaa cca ctg gat cta cct cca atc Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile 2790 2795 2800 2805	8756
att caa aga ctc cat ggc ctc agc gca ttt tca ctc cac agt tac tct Ile Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser 2810 2815 2820	8804
cca ggt gaa atc aat agg gtg gcc gca tgc ctc aga aaa ctt ggg gtc Pro Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val 2825 2830 2835	8852
ccg ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc cgc gct agg Pro Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg 2840 2845 2850	8900
ctt ctg tcc aga gga ggc agg gct gcc ata tgt ggc aag tac ctc ttc Leu Leu Ser Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe 2855 2860 2865	8948
aac tgg gca gta aga aca aag ctc aaa ctc act cca ata gcg gcc gct Asn Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala 2870 2875 2880 2885	8996
ggc cgg ctg gac ttg tcc ggt tgg ttc acg gct ggc tac agc ggg gga Gly Arg Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly 2890 2895 2900	9044
gac att tat cac agc gtg tct cat gcc cgg ccc cgc tgg ttc tgg ttt Asp Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Phe Trp Phe 2905 2910 2915	9092

tgc cta ctc ctg ctc gct gca ggg gta ggc atc tac ctc ctc ccc aac 9140
Cys Leu Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn
2920 . 2925 2930

cga tgaaggttg ggtaaact cgcgcctctt aagccatttc ctgttttttt 9193
Arg

tttttttttt tttttttttt tttttttttt ttttttttct ttcttttttt ttttcttttc 9253
tttttccctt ctttaattgg ggctccatct tagccctagt cacggctagc tgtgaaaggt 9313
ccgtgagccg catgactgca gagagtgtgt atactggcct ctctgcagat catgttctag 9373
actggaattc gtogacgagc tccctatagt gagtcgtatt agaggccgac ttggccaaat 9433
tcgtaatcat ggtcatagct gtttcctgtg tgaaattggt atccgctcac aattccacac 9493
aacatacgag ccggaagcat aaagtgtaaa gcctgggggt cctaattgagt gagctaactc 9553
acattaattg cgttgcgctc actgccgctt ttccagtcgg gaaacctgtc gtgccagctg 9613
cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggag ctcttccgct 9673
tcttcgctca ctgactcgct gcgctcggtc gttcggctgc ggcgagcgg atcagctcac 9733
tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga 9793
gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 9853
aggctccgcc ccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac 9913
ccgacaggac tataaagata ccaggcgttt cccctggaa gctccctcgt gcgctctcct 9973
gttccgacct tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg 10033
ctttctcata gctcacgctg taggtatctc agttcgggtg aggtcgttcg ctccaagctg 10093
ggctgtgtgc acgaaccccc cgttcagccc gaccgctgag ccttatccgg taactatcgt 10153
cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg 10213
attagcagag cgaggatagt aggcgggtgt acagagttct tgaagtgggt gcctaactac 10273
ggctacacta gaaggacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga 10333
aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 10393
gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 10453
tctacgggggt ctgacgctca gtggaacgaa aactcacgtt aagggatatt ggtcatgaga 10513
ttatcaaaaa ggatcttcac ctgatacctt ttaaattaaa aatgaagttt taaatcaatc 10573
taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct 10633
atctcagcga tctgtctatt tcgttcaccc atagttgcct gactccccgt cgtgtagata 10693
actacgatac gggagggcct accatctggc ccagtgctg caatgatacc gcgagacca 10753
cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga 10813
agtggctctg caactttatc cgcctccatc cagtctatta attggtgccg ggaagctaga 10873

gtaagtagtt cgccagttaa tagtttgcgc aacgttggtg ccattgctac aggcacgtg 10933
gtgtcacgct cgtcgttttg tatggcttca ttcagctccg gttcccaacg atcaaggcga 10993
gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt 11053
gtcagaagta agttggccgc agtggttatca ctcatgggta tggcagcact gcataattct 11113
cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca 11173
ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat 11233
accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga 11293
aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac tctgtcaccc 11353
aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 11413
caaaatgccg caaaaaaggg aataaggcg acacggaaat gttgaatact catactcttc 11473
ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg atacatat 11533
gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcc 11593
cctgacgcgc cctgtagcgg cgcattaagc gcggcgggtg tgggtggttac gcgcagcgtg 11653
accgctacac ttgccagcgc ctagcgccc gctcctttcg ctttcttccc ttcctttctc 11713
gccacgttcg ccggctttcc cgtcaagct ctaaatcggg gcaccccttt aggggtccga 11773
tttagtgctt tacggcacct cgaccccaaa aaacttgatt aggggtgatg ttcacgtagt 11833
gggccatcgc cctgatagac ggtttttcgc cctttgacgt tggagtccac gttctttaat 11893
agtggactct tggtccaaac tggaacaaca ctcaacccta tctcgggtcta ttcttttgat 11953
ttataaggga ttttgccgat ttcggcctat tgggttaaaaa atgagctgat ttaacaaaaa 12013
tttaacgcga attttaacaa aatattaaca aaatattaac gtttacaatt tccattcgc 12073
cattcaggct gcgcaactgt tgggaagggc gatcgggtgc ggctcttcg ctattacgcc 12133
agctggcgaa agggggatgt gctgcaaggc gattaagttg ggtaacgcca gggttttccc 12193
agtcacgacg ttgtaaaacg acggccagt ccaagctgac ttgggtcagc gccgctaata 12253
cgactcacta ta 12265

<210> 25

<211> 13147

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<220>

<221> CDS

<222> (342)..(10025)

<400> 25

```

gccagccccc tgatgggggc gacactccac catgaatcac tcccctgtga ggaactactg 60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgtcgtgcag cctccaggac 120
ccccctccc gggagagcca tagtgggtctg cggaaccggg gagtacaccg gaattgccag 180
gacgaccggg tcctttcttg gataaaccgg ctcaatgcct ggagatttgg gcgtgcccc 240
gcaagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtgggtact gcctgatagg 300
gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac c atg agc acg aat cct 356
                                         Met Ser Thr Asn Pro
                                         1           5

aaa cct caa aga aaa acc aaa cgt aac acc aac cgt cgc cca cag gac 404
Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp
                        10                        15                        20

gtc aag ttc ccg ggt ggc ggt cag atc gtt ggt gga gtt tac ttg ttg 452
Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu
                        25                        30                        35

ccg cgc agg ggc cct aga ttg ggt gtg cgc gcg acg agg aag act tcc 500
Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser
                        40                        45                        50

gag cgg tcg caa cct cga ggt aga cgt cag cct atc ccc aag gca cgt 548
Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg
                        55                        60                        65

cgg ccc gag ggc agg acc tgg gct cag ccc ggg tac cct tgg ccc ctc 596
Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu
                        70                        75                        80                        85

tat ggc aat gag ggt tgc ggg tgg gcg gga tgg ctc ctg tct ccc cgt 644
Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg
                        90                        95                        100

ggc tct cgg cct agc tgg ggc ccc aca gac ccc cgg cgt agg tcg cgc 692
Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg
                        105                        110                        115

aat ttg ggt aag gtc atc gat acc ctt acg tgc ggc ttc gcc gac ctc 740
Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
                        120                        125                        130

atg ggg tac ata ccg ctc gtc ggc gcc cct ctt gga ggc gct gcc agg 788
Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg
                        135                        140                        145

gcc ctg gcg cat ggc gtc cgg gtt ctg gaa gac ggc gtg aac tat gca 836
Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala
                        150                        155                        160                        165

aca ggg aac ctt cct ggt tgc tct ttc tct atc ttc ctt ctg gcc ctg 884
Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu
                        170                        175                        180

ctc tct tgc ctg act gtg ccc gct tca gcc gcg cgg gtc act gac cca 932

```

Leu Ser Cys	Leu Thr Val	Pro Ala	Ser Ala Ala	Arg Val	Thr Asp Pro	
185			190		195	
gac aca aat acc aca atc ctg acc aat tgc tgc cag cgt aat cag gtt	980					
Asp Thr Asn Thr Thr Ile Leu Thr Asn Cys Cys Gln Arg Asn Gln Val			205		210	
200						
atc tat tgt tct cct tcc act tgc cta cac gag cct ggt tgt gtg atc	1028					
Ile Tyr Cys Ser Pro Ser Thr Cys Leu His Glu Pro Gly Cys Val Ile			220		225	
215						
tgt gcg gac gag tgc tgg gtt ccc gcc aat ccg tac atc tca cac cct	1076					
Cys Ala Asp Glu Cys Trp Val Pro Ala Asn Pro Tyr Ile Ser His Pro			235		240	245
230						
tcc aat tgg act ggc acg gac tcc ttc ttg gct gac cac att gat ttt	1124					
Ser Asn Trp Thr Gly Thr Asp Ser Phe Leu Ala Asp His Ile Asp Phe			250		255	260
gtt atg ggc gct ctt gtg acc tgt gac gcc ctt gac att ggt gag ttg	1172					
Val Met Gly Ala Leu Val Thr Cys Asp Ala Leu Asp Ile Gly Glu Leu			265		270	275
tgt ggt gcg tgt gta tta gtc ggt gac tgg ctt gtc agg cac tgg ctt	1220					
Cys Gly Ala Cys Val Leu Val Gly Asp Trp Leu Val Arg His Trp Leu			280		285	290
att cac ata gac ctc aat gaa act ggt act tgt tac ctg gaa gtg ccc	1268					
Ile His Ile Asp Leu Asn Glu Thr Gly Thr Cys Tyr Leu Glu Val Pro			295		300	305
act gga ata gat cct ggg ttc cta ggg ttt atc ggg tgg atg gcc ggc	1316					
Thr Gly Ile Asp Pro Gly Phe Leu Gly Phe Ile Gly Trp Met Ala Gly			310		315	320
						325
aag gtc gag gct gtc atc ttc ttg acc aaa ctg gct tca caa gta cca	1364					
Lys Val Glu Ala Val Ile Phe Leu Thr Lys Leu Ala Ser Gln Val Pro			330		335	340
tac gct att gcg act atg ttt agc agt gta cac tac ctg gcg gtt ggc	1412					
Tyr Ala Ile Ala Thr Met Phe Ser Ser Val His Tyr Leu Ala Val Gly			345		350	355
gct ctg atc tac tat gcc tct cgg ggc aag tgg tat cag ttg ctc cta	1460					
Ala Leu Ile Tyr Tyr Ala Ser Arg Gly Lys Trp Tyr Gln Leu Leu Leu			360		365	370
gcg ctt atg ctt tac ata gaa gcg acc tct gga aac cct atc agg gtg	1508					
Ala Leu Met Leu Tyr Ile Glu Ala Thr Ser Gly Asn Pro Ile Arg Val			375		380	385
ccc act gga tgc tca ata gct gag ttt tgc tcg cct ttg atg ata cca	1556					
Pro Thr Gly Cys Ser Ile Ala Glu Phe Cys Ser Pro Leu Met Ile Pro			390		395	400
						405
tgt cct tgc cac tct tat ttg agt gag aat gtg tca gaa gtc att tgt	1604					
Cys Pro Cys His Ser Tyr Leu Ser Glu Asn Val Ser Glu Val Ile Cys			410		415	420
tac agt cca aag tgg acc agg cct atc act ctg gag tat aac aac tcc	1652					
Tyr Ser Pro Lys Trp Thr Arg Pro Ile Thr Leu Glu Tyr Asn Asn Ser			425		430	435

ata tct tgg tac ccc tat aca atc cct ggt gcg agg gga tgt atg gtt Ile Ser Trp Tyr Pro Tyr Thr Ile Pro Gly Ala Arg Gly Cys Met Val 440 445 450	1700
aaa ttc aaa aat aac aca tgg ggt tgc tgc cgt att cgc aat gtg cca Lys Phe Lys Asn Asn Thr Trp Gly Cys Cys Arg Ile Arg Asn Val Pro 455 460 465	1748
tcg tac tgc act atg ggc act gat gca gtg tgg aac gac act cgc aac Ser Tyr Cys Thr Met Gly Thr Asp Ala Val Trp Asn Asp Thr Arg Asn 470 475 480 485	1796
act tac gaa gta tgc ggt gta aca cca tgg cta aca acc gca tgg cac Thr Tyr Glu Val Cys Gly Val Thr Pro Trp Leu Thr Thr Ala Trp His 490 495 500	1844
aac ggc tca gcc ctg aaa ttg gct ata tta caa tac cct ggg tct aaa Asn Gly Ser Ala Leu Lys Leu Ala Ile Leu Gln Tyr Pro Gly Ser Lys 505 510 515	1892
gaa atg ttt aaa cct cat aat tgg atg tca ggc cat ttg tat ttt gag Glu Met Phe Lys Pro His Asn Trp Met Ser Gly His Leu Tyr Phe Glu 520 525 530	1940
gga tca gat acc cct ata gtt tac ttt tat gac cct gtg aat tcc act Gly Ser Asp Thr Pro Ile Val Tyr Phe Tyr Asp Pro Val Asn Ser Thr 535 540 545	1988
ctc cta cca ccg gag agg tgg gct agg ttg ccc ggt acc cca cct gtg Leu Leu Pro Pro Glu Arg Trp Ala Arg Leu Pro Gly Thr Pro Pro Val 550 555 560 565	2036
gta cgt ggt tct tgg tta cag gtt ccg caa ggg ttt tac agt gat gtg Val Arg Gly Ser Trp Leu Gln Val Pro Gln Gly Phe Tyr Ser Asp Val 570 575 580	2084
aaa gac cta gcc aca gga ttg atc acc aaa gac aaa gcc tgg aaa aat Lys Asp Leu Ala Thr Gly Leu Ile Thr Lys Asp Lys Ala Trp Lys Asn 585 590 595	2132
tat cag gtc tta tat tcc gcc acg ggt gct ttg tct ctt acg gga gtt Tyr Gln Val Leu Tyr Ser Ala Thr Gly Ala Leu Ser Leu Thr Gly Val 600 605 610	2180
acc acc aag gcc gtg gtg cta att ctg ttg ggg ttg tgt ggc agc aag Thr Thr Lys Ala Val Val Leu Ile Leu Leu Gly Leu Cys Gly Ser Lys 615 620 625	2228
tat ctt att tta gcc tac ctc tgt tac ttg tcc ctt tgt ttt ggg cgc Tyr Leu Ile Leu Ala Tyr Leu Cys Tyr Leu Ser Leu Cys Phe Gly Arg 630 635 640 645	2276
gct tct ggt tac cct ttg cgt cct gtg ctc cca tcc cag tcg tat ctc Ala Ser Gly Tyr Pro Leu Arg Pro Val Leu Pro Ser Gln Ser Tyr Leu 650 655 660	2324
caa gct ggc tgg gat gtt ttg tct aaa gct caa gta gct cct ttt gct Gln Ala Gly Trp Asp Val Leu Ser Lys Ala Gln Val Ala Pro Phe Ala 665 670 675	2372
ttg att ttc ttc atc tgt tgc tat ctc cgc tgc agg cta cgt tat gct Leu Ile Phe Phe Ile Cys Cys Tyr Leu Arg Cys Arg Leu Arg Tyr Ala 680 685 690	2420

gcc ctt tta ggg ttt gtg ccc atg gct gcg ggc ttg ccc cta act ttc	2468
Ala Leu Leu Gly Phe Val Pro Met Ala Ala Gly Leu Pro Leu Thr Phe	
695 700 705	
ttt gtt gca gca gct gct gcc caa cca gat tat gac tgg tgg gtg cga	2516
Phe Val Ala Ala Ala Ala Ala Gln Pro Asp Tyr Asp Trp Trp Val Arg	
710 715 720 725	
ctg cta gtg gca ggg tta gtt ttg tgg gcc ggc cgt aac cgt ggt cac	2564
Leu Leu Val Ala Gly Leu Val Leu Trp Ala Gly Arg Asn Arg Gly His	
730 735 740	
cgc ata gct ctg ctt gta ggt cct tgg cct ctg gta gcg ctt tta acc	2612
Arg Ile Ala Leu Leu Val Gly Pro Trp Pro Leu Val Ala Leu Leu Thr	
745 750 755	
ctc ttg cat ttg gtt acg cct gct tca gct ttt gat acc gag ata att	2660
Leu Leu His Leu Val Thr Pro Ala Ser Ala Phe Asp Thr Glu Ile Ile	
760 765 770	
gga ggg ctg aca ata cca cct gta gta gca tta gtt gtc atg tct cgt	2708
Gly Gly Leu Thr Ile Pro Pro Val Val Ala Leu Val Val Met Ser Arg	
775 780 785	
ttt ggc ttc ttt gct cac ttg tta cct cgc tgt gct tta gtt aac tcc	2756
Phe Gly Phe Phe Ala His Leu Leu Pro Arg Cys Ala Leu Val Asn Ser	
790 795 800 805	
tat ctt tgg caa cgt tgg gag aat tgg ttt tgg aac gtt aca cta aga	2804
Tyr Leu Trp Gln Arg Trp Glu Asn Trp Phe Trp Asn Val Thr Leu Arg	
810 815 820	
ccg gag agg ttt ttc ctt gtg ctg gtt tgt ttc ccc ggt gcg aca tat	2852
Pro Glu Arg Phe Phe Leu Val Leu Val Cys Phe Pro Gly Ala Thr Tyr	
825 830 835	
gac gcg ctg gtg act ttc tgt gtg tgt cac gta gct ctt cta tgt tta	2900
Asp Ala Leu Val Thr Phe Cys Val Cys His Val Ala Leu Leu Cys Leu	
840 845 850	
aca tcc agt gca gca tcg ttc ttt ggg act gac tct agg gtt agg gcc	2948
Thr Ser Ser Ala Ala Ser Phe Phe Gly Thr Asp Ser Arg Val Arg Ala	
855 860 865	
cat aga atg ttg gtg cgt ctc gga aag tgt cat gct tgg tat tct cat	2996
His Arg Met Leu Val Arg Leu Gly Lys Cys His Ala Trp Tyr Ser His	
870 875 880 885	
tat gtt ctt aag ttt ttc ctc tta gtg ttt ggt gag aat ggt gtg ttt	3044
Tyr Val Leu Lys Phe Phe Leu Leu Val Phe Gly Glu Asn Gly Val Phe	
890 895 900	
ttc tat aag cac ttg cat ggt gat gtc ttg cct aat gat ttt gcc tcg	3092
Phe Tyr Lys His Leu His Gly Asp Val Leu Pro Asn Asp Phe Ala Ser	
905 910 915	
aaa cta cca ttg caa gag cca ttt ttc cct ttt gaa ggc aag gca agg	3140
Lys Leu Pro Leu Gln Glu Pro Phe Phe Pro Phe Glu Gly Lys Ala Arg	
920 925 930	
gtc tat agg aat gaa gga aga cgc ttg gcg tgt ggg gac acg gtt gat	3188
Val Tyr Arg Asn Glu Gly Arg Arg Leu Ala Cys Gly Asp Thr Val Asp	

935	940	945	
ggt ttg ccc gtt gtt gcg cgt ctc ggc gac ctt gtt ttc gca ggg ttg Gly Leu Pro Val Val Ala Arg Leu Gly Asp Leu Val Phe Ala Gly Leu 950 955 960 965			3236
gct atg ccg cca gat ggg tgg gcc att acc gca cct ttt acg ctg cag Ala Met Pro Pro Asp Gly Trp Ala Ile Thr Ala Pro Phe Thr Leu Gln 970 975 980			3284
tgt ctc tct gaa cgt ggc acg ctg tca gcg atg gca gtg gtc atg act Cys Leu Ser Glu Arg Gly Thr Leu Ser Ala Met Ala Val Val Met Thr 985 990 995			3332
ggt ata gac ccc cga act tgg act gga act atc ttc aga tta gga tct Gly Ile Asp Pro Arg Thr Trp Thr Gly Thr Ile Phe Arg Leu Gly Ser 1000 1005 1010			3380
ctg gcc act agc tac atg gga ttc gtt tgt gac aac gtg ttg tat act Leu Ala Thr Ser Tyr Met Gly Phe Val Cys Asp Asn Val Leu Tyr Thr 1015 1020 1025			3428
gct cac cat ggc agc aag ggg cgc cgg ttg gct cat ccc aca ggc tcc Ala His His Gly Ser Lys Gly Arg Arg Leu Ala His Pro Thr Gly Ser 1030 1035 1040 1045			3476
ata cac cca ata acc gtt gac gcg gct aat gac cag gac atc tat caa Ile His Pro Ile Thr Val Asp Ala Ala Asn Asp Gln Asp Ile Tyr Gln 1050 1055 1060			3524
cca cca tgt gga gct ggg tcc ctt act cgg tgc tct tgc ggg gag acc Pro Pro Cys Gly Ala Gly Ser Leu Thr Arg Cys Ser Cys Gly Glu Thr 1065 1070 1075			3572
aag ggg tat ctg gta aca cga ctg ggg tca ttg gtt gag gtc aac aaa Lys Gly Tyr Leu Val Thr Arg Leu Gly Ser Leu Val Glu Val Asn Lys 1080 1085 1090			3620
tcc gat gac cct tat tgg tgt gtg tgc ggg gcc ctt ccc atg gct gtt Ser Asp Asp Pro Tyr Trp Cys Val Cys Gly Ala Leu Pro Met Ala Val 1095 1100 1105			3668
gcc aag ggt tct tca ggt gcc ccg att ctg tgc tcc tcc ggg cat gtt Ala Lys Gly Ser Ser Gly Ala Pro Ile Leu Cys Ser Ser Gly His Val 1110 1115 1120 1125			3716
att ggg atg ttc acc gct gct aga aat tct ggc ggt tca gtc agt cag Ile Gly Met Phe Thr Ala Ala Arg Asn Ser Gly Gly Ser Val Ser Gln 1130 1135 1140			3764
att agg gtt agg ccg ttg gtg tgt gct gga tac cat ccc cag tac aca Ile Arg Val Arg Pro Leu Val Cys Ala Gly Tyr His Pro Gln Tyr Thr 1145 1150 1155			3812
gca cat gcc act ctt gat aca aaa acg cgt atg cag atc ttc gtg aag Ala His Ala Thr Leu Asp Thr Lys Thr Arg Met Gln Ile Phe Val Lys 1160 1165 1170			3860
acc ctg acg ggc aag acc atc act ctt gag gtc gag ccc agt gac acc Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Pro Ser Asp Thr 1175 1180 1185			3908
atc gag aat gtc aag gcc aag atc caa gac aag gaa ggc atc cca cct			3956

123/146

Ile Glu Asn Val Lys Ala Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro 1190 1195 1200 1205	
gac cag cag agg ctg ata ttc gcg ggc aaa cag ctg gag gat ggc cgc Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg 1210 1215 1220	4004
acc ctg tcc gac tac aac atc cag aaa gag tcc acc ttg cac ctg gtg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val 1225 1230 1235	4052
ctg cgt ctc cgc ggt ggc gcg ccc atc acg gcg tac gcc cag cag acg Leu Arg Leu Arg Gly Gly Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr 1240 1245 1250	4100
aga ggc ctc cta ggg tgt ata atc acc agc ctg act ggc cgg gac aaa Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys 1255 1260 1265	4148
aac caa gtg gag ggt gag gtc cag atc gtg tca act gct acc caa acc Asn Gln Val Glu Gly Glu Val Gln Ile Val Ser Thr Ala Thr Gln Thr 1270 1275 1280 1285	4196
ttc ctg gca acg tgc atc aat ggg gta tgc tgg act gtc tac cac ggg Phe Leu Ala Thr Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly 1290 1295 1300	4244
gcc gga acg agg acc atc gca tca ccc aag ggt cct gtc atc cag atg Ala Gly Thr Arg Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met 1305 1310 1315	4292
tat acc aat gtg gac caa gac ctt gtg ggc tgg ccc gct cct caa ggt Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly 1320 1325 1330	4340
tcc cgc tca ttg aca ccc tgt acc tgc ggc tcc tcg gac ctt tac ctg Ser Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu 1335 1340 1345	4388
gtc acg agg cac gcc gat gtc att ccc gtg cgc cgg cga ggt gat agc Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser 1350 1355 1360 1365	4436
agg ggt agc ctg ctt tcg ccc cgg ccc att tcc tac ttg aaa ggc tcc Arg Gly Ser Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser 1370 1375 1380	4484
tcg ggg ggt ccg ctg ttg tgc ccc gcg gga cac gcc gtg ggc cta ttc Ser Gly Gly Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu Phe 1385 1390 1395	4532
agg gcc gcg gtg tgc acc cgt gga gtg gct aaa gcg gtg gac ttt atc Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile 1400 1405 1410	4580
cct gtg gag aac cta ggg aca acc atg aga tcc ccg gtg ttc acg gac Pro Val Glu Asn Leu Gly Thr Thr Met Arg Ser Pro Val Phe Thr Asp 1415 1420 1425	4628
aac tcc tct cca cca gca gtg ccc cag agc ttc cag gtg gcc cac ctg Asn Ser Ser Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu 1430 1435 1440 1445	4676

cat gct ccc acc ggc agc ggt aag agc acc aag gtc ccg gct gcg tac	4724
His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr	
1450 1455 1460	
gca gcc cag ggc tac aag gtg ttg gtg ctc aac ccc tct gtt gct gca	4772
Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala	
1465 1470 1475	
acg ctg ggc ttt ggt gct tac atg tcc aag gcc cat ggg gtt gat cct	4820
Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly Val Asp Pro	
1480 1485 1490	
aat atc agg acc ggg gtg aga aca att acc act ggc agc ccc atc acg	4868
Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr	
1495 1500 1505	
tac tcc acc tac ggc aag ttc ctt gcc gac ggc ggg tgc tca gga ggt	4916
Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly	
1510 1515 1520 1525	
gct tat gac ata ata att tgt gac gag tgc cac tcc acg gat gcc aca	4964
Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr	
1530 1535 1540	
tcc atc ttg ggc atc ggc act gtc ctt gac caa gca gag act gcg ggg	5012
Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly	
1545 1550 1555	
gcg aga ctg gtt gtg ctc gcc act gct acc cct ccg ggc tcc gtc act	5060
Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr	
1560 1565 1570	
gtg tcc cat cct aac atc gag gag gtt gct ctg tcc acc acc gga gag	5108
Val Ser His Pro Asn Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu	
1575 1580 1585	
atc ccc ttt tac ggc aag gct atc ccc ctc gag gtg atc aag ggg gga	5156
Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly	
1590 1595 1600 1605	
aga cat ctc atc ttc tgc cac tca aag aag aag tgc gac gag ctc gcc	5204
Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala	
1610 1615 1620	
gcg aag ctg gtc gca ttg ggc atc aat gcc gtg gcc tac tac cgc ggt	5252
Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly	
1625 1630 1635	
ctt gac gtg tct gtc atc ccg acc agc ggc gat gtt gtc gtc gtg tcg	5300
Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Val Val Val Ser	
1640 1645 1650	
acc gat gct ctc atg act ggc ttt acc ggc gac ttc gac tct gtg ata	5348
Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val Ile	
1655 1660 1665	
gac tgc aac acg tgt gtc act cag aca gtc gat ttc agc ctt gac cct	5396
Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro	
1670 1675 1680 1685	
acc ttt acc att gag aca acc acg ctc ccc cag gat gct gtc tcc agg	5444
Thr Phe Thr Ile Glu Thr Thr Thr Leu Pro Gln Asp Ala Val Ser Arg	
1690 1695 1700	

act caa cgc cgg ggc agg act ggc agg ggg aag cca ggc atc tat aga	5492
Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg	
1705 1710 1715	
ttt gtg gca ccg ggg gag cgc ccc tcc ggc atg ttc gac tcg tcc gtc	5540
Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val	
1720 1725 1730	
ctc tgt gag tgc tat gac gcg ggc tgt gct tgg tat gag ctc acg ccc	5588
Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro	
1735 1740 1745	
gcc gag act aca gtt agg cta cga gcg tac atg aac acc ccg ggg ctt	5636
Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu	
1750 1755 1760 1765	
ccc gtg tgc cag gac cat ctt gaa ttt tgg gag ggc gtc ttt acg ggc	5684
Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly	
1770 1775 1780	
ctc act cat ata gat gcc cac ttt tta tcc cag aca aag cag agt ggg	5732
Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly	
1785 1790 1795	
gag aac ttt cct tac ctg gta gcg tac caa gcc acc gtg tgc gct agg	5780
Glu Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg	
1800 1805 1810	
gct caa gcc cct ccc cca tcg tgg gac cag atg tgg aag tgt ttg atc	5828
Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile	
1815 1820 1825	
cgc ctt aaa ccc acc ctc cat ggg cca aca ccc ctg cta tac aga ctg	5876
Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu	
1830 1835 1840 1845	
ggc gct gtt cag aat gaa gtc acc ctg acg cac cca atc acc aaa tac	5924
Gly Ala Val Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr	
1850 1855 1860	
atc atg aca tgc atg tcg gcc gac ctg gag gtc gtc acg agc acc tgg	5972
Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp	
1865 1870 1875	
gtg ctc gtt ggc ggc gtc ctg gct gct ctg gcc gcg tat tgc ctg tca	6020
Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser	
1880 1885 1890	
aca ggc tgc gtg gtc ata gtg ggc agg atc gtc ttg tcc ggg aag ccg	6068
Thr Gly Cys Val Val Ile Val Gly Arg Ile Val Leu Ser Gly Lys Pro	
1895 1900 1905	
gca att ata cct gac agg gag gtt ctc tac cag gag ttc gat gag atg	6116
Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Gln Glu Phe Asp Glu Met	
1910 1915 1920 1925	
gaa gag tgc tct cag cac tta ccg tac atc gag caa ggg atg atg ctc	6164
Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Val Glu Gln Gly Met Met Leu	
1930 1935 1940	
gct gag cag ttc aag cag aag gcc ctc ggc ctc ctg cag acc gcg tcc	6212
Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Ser	

1945	1950	1955	
cgc cat gca gag gtt atc acc cct gct gtc cag acc aac tgg cag aaa			6260
Arg His Ala Glu Val Ile Thr Pro Ala Val Gln Thr Asn Trp Gln Lys			
1960	1965	1970	
ctc gag gtc ttt tgg gcg aag cac atg tgg aat ttc atc agt ggg ata			6308
Leu Glu Val Phe Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile			
1975	1980	1985	
caa tac ttg gcg ggc ctg tca acg ctg cct ggt aac ccc gcc att gct			6356
Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala			
1990	1995	2000	2005
tca ttg atg gct ttt aca gct gcc gtc acc agc cca cta acc act ggc			6404
Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro Leu Thr Thr Gly			
2010	2015	2020	
caa acc ctc ctc ttc aac ata ttg ggg ggg tgg gtg gct gcc cag ctc			6452
Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu			
2025	2030	2035	
gcc gcc ccc ggt gcc gct act gcc ttt gtg ggt gct ggc cta gct ggc			6500
Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala Gly			
2040	2045	2050	
gcc gcc atc ggc agc gtt gga ctg ggg aag gtc ctc gtg gac att ctt			6548
Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu Val Asp Ile Leu			
2055	2060	2065	
gca ggg tat ggc gcg ggc gtg gcg gga gct ctt gta gca ttc aag atc			6596
Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Ile			
2070	2075	2080	2085
atg agc ggt gag gtc ccc tcc acg gag gac ctg gtc aat ctg ctg ccc			6644
Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro			
2090	2095	2100	
gcc atc ctc tcg cct gga gcc ctt gta gtc ggt gtg gtc tgc gca gca			6692
Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala			
2105	2110	2115	
ata ctg cgc cgg cac gtt ggc ccg ggc gag ggg gca gtg caa tgg atg			6740
Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met			
2120	2125	2130	
aac cgg cta ata gcc ttc gcc tcc cgg ggg aac cat gtt tcc ccc acg			6788
Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr			
2135	2140	2145	
cac tac gtg ccg gag agc gat gca gcc gcc cgc gtc act gcc ata ctc			6836
His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val Thr Ala Ile Leu			
2150	2155	2160	2165
agc agc ctc act gta acc cag ctc ctg agg cga ctg cat cag tgg ata			6884
Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu His Gln Trp Ile			
2170	2175	2180	
agc tcg gag tgt acc act cca tgc tcc ggt tcc tgg cta agg gac atc			6932
Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Ile			
2185	2190	2195	
tgg gac tgg ata tgc gag gtg ctg agc gac ttt aag acc tgg ctg aaa			6980

Trp	Asp	Trp	Ile	Cys	Glu	Val	Leu	Ser	Asp	Phe	Lys	Thr	Trp	Leu	Lys		
		2200					2205					2210					
gcc	aag	ctc	atg	cca	caa	ctg	cct	ggg	att	ccc	ttt	gtg	tcc	tgc	cag	7028	
Ala	Lys	Leu	Met	Pro	Gln	Leu	Pro	Gly	Ile	Pro	Phe	Val	Ser	Cys	Gln		
	2215					2220					2225						
cgc	ggg	tat	agg	ggg	gtc	tgg	cga	gga	gac	ggc	att	atg	cac	act	cgc	7076	
Arg	Gly	Tyr	Arg	Gly	Val	Trp	Arg	Gly	Asp	Gly	Ile	Met	His	Thr	Arg		
	2230				2235					2240					2245		
tgc	cac	tgt	gga	gct	gag	atc	act	gga	cat	gtc	aaa	aac	ggg	acg	atg	7124	
Cys	His	Cys	Gly	Ala	Glu	Ile	Thr	Gly	His	Val	Lys	Asn	Gly	Thr	Met		
				2250					2255					2260			
agg	atc	gtc	ggg	cct	agg	acc	tgc	agg	aac	atg	tgg	agt	ggg	acg	ttc	7172	
Arg	Ile	Val	Gly	Pro	Arg	Thr	Cys	Arg	Asn	Met	Trp	Ser	Gly	Thr	Phe		
		2265						2270					2275				
ccc	att	aac	gcc	tac	acc	acg	ggc	ccc	tgt	act	ccc	ctt	cct	gcg	ccg	7220	
Pro	Ile	Asn	Ala	Tyr	Thr	Thr	Gly	Pro	Cys	Thr	Pro	Leu	Pro	Ala	Pro		
	2280						2285					2290					
aac	tat	aag	ttc	gcg	ctg	tgg	agg	gtg	tct	gca	gag	gaa	tac	gtg	gag	7268	
Asn	Tyr	Lys	Phe	Ala	Leu	Trp	Arg	Val	Ser	Ala	Glu	Glu	Tyr	Val	Glu		
	2295					2300				2305							
ata	agg	cgg	gtg	ggg	gac	ttc	cac	tac	gta	tcg	ggg	atg	act	act	gac	7316	
Ile	Arg	Arg	Val	Gly	Asp	Phe	His	Tyr	Val	Ser	Gly	Met	Thr	Thr	Asp		
	2310				2315				2320						2325		
aat	ctt	aaa	tgc	ccg	tgc	cag	atc	cca	tcg	ccc	gaa	ttt	ttc	aca	gaa	7364	
Asn	Leu	Lys	Cys	Pro	Cys	Gln	Ile	Pro	Ser	Pro	Glu	Phe	Phe	Thr	Glu		
				2330					2335					2340			
ttg	gac	ggg	gtg	cgc	cta	cac	agg	ttt	gcg	ccc	cct	tgc	aag	ccc	ttg	7412	
Leu	Asp	Gly	Val	Arg	Leu	His	Arg	Phe	Ala	Pro	Pro	Cys	Lys	Pro	Leu		
			2345					2350					2355				
ctg	cgg	gag	gag	gta	tca	ttc	aga	gta	gga	ctc	cac	gag	tac	ccg	gtg	7460	
Leu	Arg	Glu	Glu	Val	Ser	Phe	Arg	Val	Gly	Leu	His	Glu	Tyr	Pro	Val		
		2360					2365					2370					
ggg	tcg	caa	tta	cct	tgc	gag	ccc	gaa	ccg	gac	gta	gcc	gtg	ttg	acg	7508	
Gly	Ser	Gln	Leu	Pro	Cys	Glu	Pro	Glu	Pro	Asp	Val	Ala	Val	Leu	Thr		
	2375					2380					2385						
tcc	atg	ctc	act	gat	ccc	tcc	cat	ata	aca	gca	gag	gcg	gcc	ggg	aga	7556	
Ser	Met	Leu	Thr	Asp	Pro	Ser	His	Ile	Thr	Ala	Glu	Ala	Ala	Gly	Arg		
	2390				2395					2400					2405		
agg	ttg	gcg	aga	ggg	tca	ccc	cct	tct	atg	gcc	agc	tcc	tcg	gct	agc	7604	
Arg	Leu	Ala	Arg	Gly	Ser	Pro	Pro	Ser	Met	Ala	Ser	Ser	Ser	Ala	Ser		
			2410						2415					2420			
cag	ctg	tcc	gct	cca	tct	ctc	aag	gca	act	tgc	acc	gcc	aac	cat	gac	7652	
Gln	Leu	Ser	Ala	Pro	Ser	Leu	Lys	Ala	Thr	Cys	Thr	Ala	Asn	His	Asp		
			2425					2430					2435				
tcc	cct	gac	gcc	gag	ctc	ata	gag	gct	aac	ctc	ctg	tgg	agg	cag	gag	7700	
Ser	Pro	Asp	Ala	Glu	Leu	Ile	Glu	Ala	Asn	Leu	Leu	Trp	Arg	Gln	Glu		
		2440					2445					2450					

atg ggc ggc aac atc acc agg gtt gag tca gag aac aaa gtg gtg att	7748
Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn Lys Val Val Ile	
2455 2460 2465	
ctg gac tcc ttc gat ccg ctt gtg gca gag gag gat gag cgg gag gtc	7796
Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp Glu Arg Glu Val	
2470 2475 2480 2485	
tcc gta cct gca gaa att ctg cgg aag tct cgg aga ttc gcc cgg gcc	7844
Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg Phe Ala Arg Ala	
2490 2495 2500	
ctg ccc gtc tgg gcg cgg ccg gac tac aac ccc ccg cta gta gag acg	7892
Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu Val Glu Thr	
2505 2510 2515	
tgg aaa aag cct gac tac gaa cca cct gtg gtc cat ggc tgc ccg cta	7940
Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His Gly Cys Pro Leu	
2520 2525 2530	
cca cct cca cgg tcc cct cct gtg cct ccg cct cgg aaa aag cgt acg	7988
Pro Pro Pro Arg Ser Pro Pro Val Pro Pro Pro Arg Lys Lys Arg Thr	
2535 2540 2545	
gtg gtc ctc acc gaa tca acc cta tct act gcc ttg gcc gag ctt gcc	8036
Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu Ala Glu Leu Ala	
2550 2555 2560 2565	
acc aaa agt ttt ggc agc tcc tca act tcc ggc att acg ggc gac aat	8084
Thr Lys Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile Thr Gly Asp Asn	
2570 2575 2580	
acg aca aca tcc tct gag ccc gcc cct tct ggc tgc ccc ccc gac tcc	8132
Thr Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys Pro Pro Asp Ser	
2585 2590 2595	
gac gtt gag tcc tat tct tcc atg ccc ccc ctg gag ggg gag cct ggg	8180
Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly	
2600 2605 2610	
gat ccg gat ctc agc gac ggg tca tgg tcg acg gtc agt agt ggg gcc	8228
Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser Ser Gly Ala	
2615 2620 2625	
gac acg gaa gat gtc gtg tgc tgc tca atg tct tat tcc tgg aca ggc	8276
Asp Thr Glu Asp Val Val Cys Cys Ser Met Ser Tyr Ser Trp Thr Gly	
2630 2635 2640 2645	
gca ctc gtc acc ccg tgc gct gcg gaa gaa caa aaa ctg ccc atc aac	8324
Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys Leu Pro Ile Asn	
2650 2655 2660	
gca ctg agc aac tcg ttg cta cgc cat cac aat ctg gtg tat tcc acc	8372
Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr Ser Thr	
2665 2670 2675	
act tca cgc agt gct tgc caa agg cag aag aaa gtc aca ttt gac aga	8420
Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val Thr Phe Asp Arg	
2680 2685 2690	
ctg caa gtt ctg gac agc cat tac cag gac gtg ctc aag gag gtc aaa	8468
Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu Lys Glu Val Lys	
2695 2700 2705	

gca gcg gcg tca aaa gtg aag gct aac ttg cta tcc gta gag gaa gct	8516
Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser Val Glu Glu Ala	
2710 2715 2720 2725	
tgc agc ctg acg ccc cca cat tca gcc aaa tcc aag ttt ggc tat ggg	8564
Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys Phe Gly Tyr Gly	
2730 2735 2740	
gca aaa gac gtc cgt tgc cat gcc aga aag gcc gta gcc cac atc aac	8612
Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val Ala His Ile Asn	
2745 2750 2755	
tcc gtg tgg aaa gac ctt ctg gaa gac agt gta aca cca ata gac act	8660
Ser Val Trp Lys Asp Leu Leu Glu Asp Ser Val Thr Pro Ile Asp Thr	
2760 2765 2770	
acc atc atg gcc aag aac gag gtt ttc tgc gtt cag cct gag aag ggg	8708
Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu Lys Gly	
2775 2780 2785	
ggt cgt aag cca gct cgt ctc atc gtg ttc ccc gac ctg ggc gtg cgc	8756
Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly Val Arg	
2790 2795 2800 2805	
gtg tgc gag aag atg gcc ctg tac gac gtg gtt agc aag ctc ccc ctg	8804
Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Lys Leu Pro Leu	
2810 2815 2820	
gcc gtg atg gga agc tcc tac gga ttc caa tac tca cca gga cag cgg	8852
Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg	
2825 2830 2835	
gtt gaa ttc ctc gtg caa gcg tgg aag tcc aag aag acc ccg atg ggg	8900
Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys Thr Pro Met Gly	
2840 2845 2850	
ttc tcg tat gat acc cgc tgt ttt gac tcc aca gtc act gag agc gac	8948
Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu Ser Asp	
2855 2860 2865	
atc cgt acg gag gag gca att tac caa tgt tgt gac ctg gac ccc caa	8996
Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Asp Leu Asp Pro Gln	
2870 2875 2880 2885	
gcc cgc gtg gcc atc aag tcc ctc act gag agg ctt tat gtt ggg ggc	9044
Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu Tyr Val Gly Gly	
2890 2895 2900	
cct ctt acc aat tca agg ggg gaa aac tgc ggc tac cgc agg tgc cgc	9092
Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr Arg Arg Cys Arg	
2905 2910 2915	
gcg agc ggc gta ctg aca act agc tgt ggt aac acc ctc act tgc tac	9140
Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr	
2920 2925 2930	
atc aag gcc cgg gca gcc tgt cga gcc gca ggg ctc cag gac tgc acc	9188
Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu Gln Asp Cys Thr	
2935 2940 2945	
atg ctc gtg tgt ggc gac gac tta gtc gtt atc tgt gaa agt gcg ggg	9236
Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser Ala Gly	

2950	2955	2960	2965	
gtc cag gag gac gcg gcg agc ctg aga gcc ttc acg gag gct atg acc				9284
Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr Glu Ala Met Thr				
2970	2975	2980		
agg tac tcc gcc ccc ccc ggg gac ccc cca caa cca gaa tac gac ttg				9332
Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu				
2985	2990	2995		
gag ctt ata aca tca tgc tcc tcc aac gtg tca gtc gcc cac gac ggc				9380
Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His Asp Gly				
3000	3005	3010		
gct gga aag agg gtc tac tac ctt acc cgt gac cct aca acc ccc ctc				9428
Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu				
3015	3020	3025		
gcg aga gcc gcg tgg gag aca gca aga cac act cca gtc aat tcc tgg				9476
Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn Ser Trp				
3030	3035	3040	3045	
cta ggc aac ata atc atg ttt gcc ccc aca ctg tgg gcg agg atg ata				9524
Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp Ala Arg Met Ile				
3050	3055	3060		
ctg atg acc cat ttc ttt agc gtc ctc ata gcc agg gat cag ctt gaa				9572
Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg Asp Gln Leu Glu				
3065	3070	3075		
cag gct ctt aac tgt gag atc tac gga gcc tgc tac tcc ata gaa cca				9620
Gln Ala Leu Asn Cys Glu Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro				
3080	3085	3090		
ctg gat cta cct cca atc att caa aga ctc cat ggc ctc agc gca ttt				9668
Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly Leu Ser Ala Phe				
3095	3100	3105		
tca ctc cac agt tac tct cca ggt gaa atc aat agg gtg gcc gca tgc				9716
Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala Ala Cys				
3110	3115	3120	3125	
ctc aga aaa ctt ggg gtc ccg ccc ttg cga gct tgg aga cac cgg gcc				9764
Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp Arg His Arg Ala				
3130	3135	3140		
cgg agc gtc cgc gct agg ctt ctg tcc aga gga ggc agg gct gcc ata				9812
Arg Ser Val Arg Ala Arg Leu Leu Ser Arg Gly Gly Arg Ala Ala Ile				
3145	3150	3155		
tgt ggc aag tac ctc ttc aac tgg gca gta aga aca aag ctc aaa ctc				9860
Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu Lys Leu				
3160	3165	3170		
act cca ata gcg gcc gct ggc cgg ctg gac ttg tcc ggt tgg ttc acg				9908
Thr Pro Ile Ala Ala Ala Gly Arg Leu Asp Leu Ser Gly Trp Phe Thr				
3175	3180	3185		
gct ggc tac agc ggg gga gac att tat cac agc gtg tct cat gcc cgg				9956
Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val Ser His Ala Arg				
3190	3195	3200	3205	
ccc cgc tgg ttc tgg ttt tgc cta ctc ctg ctc gct gca ggg gta ggc				10004

Pro Arg Trp Phe Trp Phe Cys Leu Leu Leu Leu Ala Ala Gly Val Gly
 3210 3215 3220

atc tac ctc ctc ccc aac cga tgaagggttg ggtaaact ccggcctctt 10055
 Ile Tyr Leu Leu Pro Asn Arg
 3225

aagccatttc ctgttttttt tttttttttt tttttttttt tctttttttt tttcttttct 10115
 ttctttcttt ttttctttt tttttccctt ctttaatggg ggctccatct tagccctagt 10175
 cacggctagc tgtgaaaggc ccgtgagccg catgactgca gagagtgcctg atactggcct 10235
 ctctgcagat catgttctag actggaattc gtcgacgagc tccctatagt gagtcgtatt 10295
 agaggccgac ttggccaaat tcgtaatcat ggtcatagct gtttcctgtg tgaaattgtt 10355
 atccgctcac aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctgggggtg 10415
 cctaattgagt gagctaactc acattaattg cgttgcgctc actgcccgct ttccagtcgg 10475
 gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc 10535
 gtattggggc ctcttcgct tctcgcctca ctgactcgct gcgctcggtc gttcggctgc 10595
 ggcgagcggg atcagctcac tcaaaggcgg taatacgggt atccacagaa tcaggggata 10655
 acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg 10715
 cgttgctggc gtttttccat aggctccgcc ccctgacga gcatcacaaa aatcgacgct 10775
 caagtcagag gtggcgaaac ccgacaggac tataaagata ccaggcggtt cccctggaa 10835
 gctccctcgt gcgctctcct gttccgacct tgccgcttac cggatacctg tccgcctttc 10895
 tcccttcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcgggtg 10955
 aggtcggttc ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg 11015
 ccttatccgg taactatcgt cttgagtcca acccggttaag acacgactta tcgccactgg 11075
 cagcagccac tggtaacagg attagcagag cgaggatatgt aggcggtgct acagagttct 11135
 tgaagtgggt gcctaactac ggctacacta gaaggacagt atttggtatc tgcgctctgc 11195
 tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg 11255
 ctggtagcgg tggttttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc 11315
 aagaagatcc ttgatcttt tctacggggt ctgacgctca gtggaacgaa aactcacgtt 11375
 aagggatttt ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa 11435
 aatgaagttt taaatcaatc taaagtatat atgagtaaac ttgggtctgac agttaccaat 11495
 gcttaatcag tgaggcacct atctcagcga tctgtctatt tcgttcaccc atagttgcct 11555
 gactccccgt cgtgtagata actacgatac gggagggcct accatctggc ccagtgctg 11615
 caatgatacc gcgagacca cgctcaccgg ctccagattt atcagcaata aaccagccag 11675
 ccggaagggc cgagcgaga agtggctctg caactttatc cgcctccatc cagtctatta 11735

attgttgccg ggaagctaga gtaagtagtt cgccagttaa tagtttgccg aacgttggtt 11795
ccattgctac aggcacgtg gtgtcacgct cgtcgtttg tatggcttca ttcagctccg 11855
gttcccaacg atcaaggcga gttacatgat ccccatgtt gtgcaaaaaa gcggttagct 11915
ccttcgggtcc tccgatcggt gtcagaagta agttggccgc agtgttatca ctcatgggta 11975
tggcagcaact gcataattct cttactgtca tgccatccgt aagatgcttt tctgtgactg 12035
gtgagtactc aaccaagtca ttctgagaat agtgatgacg gcgaccgagt tgctcttgcc 12095
cggcgtaaat acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg 12155
gaaaacgttc ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga 12215
tgtaaccacac tcgtgcaccc aactgatctt cagcatcttt tactttcacc agcgtttctg 12275
ggtgagcaaa aacaggaagg caaatgccg caaaaaaggg aataaggcg acacggaaat 12335
gttgaatact catactcttc ctttttcaat attattgaag catttatcag ggttattgtc 12395
tcatgagcgg atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca 12455
catttccccg aaaagtgcc cctgacgcgc cctgtagcgg cgcattaagc gcggcggtg 12515
tggtgggttac gcgcagcgtg accgctacac ttgccagcgc cctagcgccc gctcctttcg 12575
ctttcttccc ttcctttctc gccacgttcg ccggctttcc ccgtcaagct ctaaactcgg 12635
gcatcccttt agggttccga tttagtgtt tacggcacct cgaccccaa aaacttgatt 12695
aggggtgatgg ttcacgtagt gggccatcgc cctgatagac ggtttttcgc cctttgacgt 12755
tggaagtcac gttctttaat agtggactct tgttccaaac tggaacaaca ctcaacccta 12815
tctcgggtcta ttcttttgat ttataaggga ttttgccgat ttcggcctat tggttaaaaa 12875
atgagctgat ttaacaaaaa tttaacgcga attttaacaa aatattaaca aaatattaac 12935
gtttacaatt tccatttcgc cattcaggct gcgcaactgt tgggaagggc gatcgggtgcg 12995
ggcctcttcg ctattacgcc agctggcgaa agggggatgt gctgcaaggc gattaagttg 13055
ggtaacgccg gggttttccc agtcacgacg ttgtaaaacg acggccagt ccaagctgac 13115
ttggtcagcg gccgctaata cgactcacta ta 13147

<210> 26

<211> 12370

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<220>

<221> CDS

<222> (342) .. (9248)

<400> 26

```

gccagccccc tgatgggggc gacactccac catgaatcac tcccctgtga ggaactactg 60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgtcgtgcag cctccaggac 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg tcctttcttg gataaaccgc ctcaatgcct ggagatttgg gcgtgcccc 240
gcaagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcca gtgccccggg aggtctcgta gaccgtgcac c atg agc acg aat cct 356
                                         Met Ser Thr Asn Pro
                                         1             5

aaa cct caa aga aaa acc aaa cgt aac acc aac cgt cgc cca cag gac 404
Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp
                        10                      15                      20

gtc aag ttc ccg ggt ggc ggt cag atc gtt ggt gga gtt tac ttg ttg 452
Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu
                        25                      30                      35

ccg cgc agg ggc cct aga ttg ggt gtg cgc gcg acg agg aag act tcc 500
Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser
                        40                      45                      50

gag cgg tcg caa cct cga ggt aga cgt cag cct atc ccc aag gca cgt 548
Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg
                        55                      60                      65

cgg ccc gag ggc agg acc tgg gct cag ccc ggg tac cct tgg ccc ctc 596
Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu
                        70                      75                      80                      85

tat ggc aat gag ggt tgc ggg tgg gcg gga tgg ctc ctg tct ccc cgt 644
Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg
                        90                      95                      100

ggc tct cgg cct agc tgg ggc ccc aca gac ccc cgg cgt agg tcg cgc 692
Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg
                        105                      110                      115

aat ttg ggt aag gtc atc gat acc ctt acg tgc ggc ttc gcc gac ctc 740
Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
                        120                      125                      130

atg ggg tac ata ccg ctc gtc ggc gcc cct ctt gga ggc gct gcc agg 788
Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg
                        135                      140                      145

gcc ctg gcg cat ggc gtc ccg gtt ctg gaa gac ggc gtg aac tat gca 836
Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala
                        150                      155                      160                      165

aca ggg aac ctt cct ggt tgc tct ttc tct atc ttc ctt ctg gcc ctg 884
Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu
                        170                      175                      180

ctc tct tgc ctg act gtg ccc gct tca gcc gcg cgg gtc act gac cca 932
Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Ala Arg Val Thr Asp Pro
                        185                      190                      195

```

gac aca aat acc aca atc ctg acc aat tgc tgc cag cgt aat cag gtt	980
Asp Thr Asn Thr Thr Ile Leu Thr Asn Cys Cys Gln Arg Asn Gln Val	
200 205 210	
atc tat tgt tct cct tcc act tgc cta cac gag cct ggt tgt gtg atc	1028
Ile Tyr Cys Ser Pro Ser Thr Cys Leu His Glu Pro Gly Cys Val Ile	
215 220 225	
tgt gcg gac gag tgc tgg gtt ccc gcc aat ccg tac atc tca cac cct	1076
Cys Ala Asp Glu Cys Trp Val Pro Ala Asn Pro Tyr Ile Ser His Pro	
230 235 240 245	
tcc aat tgg act ggc acg gac tcc ttc ttg gct gac cac att gat ttt	1124
Ser Asn Trp Thr Gly Thr Asp Ser Phe Leu Ala Asp His Ile Asp Phe	
250 255 260	
gtt atg ggc gct ctt gtg acc tgt gac gcc ctt gac att ggt gag ttg	1172
Val Met Gly Ala Leu Val Thr Cys Asp Ala Leu Asp Ile Gly Glu Leu	
265 270 275	
tgt ggt gcg tgt gta tta gtc ggt gac tgg ctt gtc agg cac tgg ctt	1220
Cys Gly Ala Cys Val Leu Val Gly Asp Trp Leu Val Arg His Trp Leu	
280 285 290	
att cac ata gac ctc aat gaa act ggt act tgt tac ctg gaa gtg ccc	1268
Ile His Ile Asp Leu Asn Glu Thr Gly Thr Cys Tyr Leu Glu Val Pro	
295 300 305	
act gga ata gat cct ggg ttc cta ggg ttt atc ggg tgg atg gcc ggc	1316
Thr Gly Ile Asp Pro Gly Phe Leu Gly Phe Ile Gly Trp Met Ala Gly	
310 315 320 325	
aag gtc gag gct gtc atc ttc ttg acc aaa ctg gct tca caa gta cca	1364
Lys Val Glu Ala Val Ile Phe Leu Thr Lys Leu Ala Ser Gln Val Pro	
330 335 340	
tac gct att gcg act atg ttt agc agt gta cac tac ctg gcg gtt ggc	1412
Tyr Ala Ile Ala Thr Met Phe Ser Ser Val His Tyr Leu Ala Val Gly	
345 350 355	
gct ctg atc tac tat gcc tct cgg ggc aag tgg tat cag ttg ctc cta	1460
Ala Leu Ile Tyr Tyr Ala Ser Arg Gly Lys Trp Tyr Gln Leu Leu Leu	
360 365 370	
gcg ctt atg ctt tac ata gaa gcg acc tct gga aac cct atc agg gtg	1508
Ala Leu Met Leu Tyr Ile Glu Ala Thr Ser Gly Asn Pro Ile Arg Val	
375 380 385	
ccc act gga tgc tca ata gct gag ttt tgc tgc cct ttg atg ata cca	1556
Pro Thr Gly Cys Ser Ile Ala Glu Phe Cys Ser Pro Leu Met Ile Pro	
390 395 400 405	
tgt cct tgc cac tct tat ttg agt gag aat gtg tca gaa gtc att tgt	1604
Cys Pro Cys His Ser Tyr Leu Ser Glu Asn Val Ser Glu Val Ile Cys	
410 415 420	
tac agt cca aag tgg acc agg cct atc act ctg gag tat aac aac tcc	1652
Tyr Ser Pro Lys Trp Thr Arg Pro Ile Thr Leu Glu Tyr Asn Asn Ser	
425 430 435	
ata tct tgg tac ccc tat aca atc cct ggt gcg agg gga tgt atg gtt	1700
Ile Ser Trp Tyr Pro Tyr Thr Ile Pro Gly Ala Arg Gly Cys Met Val	
440 445 450	

aaa ttc aaa aat aac aca tgg ggt tgc tgc cgt att cgc aat gtg cca	1748
Lys Phe Lys Asn Asn Thr Trp Gly Cys Cys Arg Ile Arg Asn Val Pro	
455 460 465	
tcg tac tgc act atg ggc act gat gca gtg tgg aac gac act cgc aac	1796
Ser Tyr Cys Thr Met Gly Thr Asp Ala Val Trp Asn Asp Thr Arg Asn	
470 475 480 485	
act tac gaa gta tgc ggt gta aca cca tgg cta aca acc gca tgg cac	1844
Thr Tyr Glu Val Cys Gly Val Thr Pro Trp Leu Thr Thr Ala Trp His	
490 495 500	
aac ggc tca gcc ctg aaa ttg gct ata tta caa tac cct ggg tct aaa	1892
Asn Gly Ser Ala Leu Lys Leu Ala Ile Leu Gln Tyr Pro Gly Ser Lys	
505 510 515	
gaa atg ttt aaa cct cat aat tgg atg tca ggc cat ttg tat ttt gag	1940
Glu Met Phe Lys Pro His Asn Trp Met Ser Gly His Leu Tyr Phe Glu	
520 525 530	
gga tca gat acc cct ata gtt tac ttt tat gac cct gtg aat tcc act	1988
Gly Ser Asp Thr Pro Ile Val Tyr Phe Tyr Asp Pro Val Asn Ser Thr	
535 540 545	
ctc cta cca ccg gag agg tgg gct agg ttg ccc ggt acc cca cct gtg	2036
Leu Leu Pro Pro Glu Arg Trp Ala Arg Leu Pro Gly Thr Pro Pro Val	
550 555 560 565	
gta cgt ggt tct tgg tta cag gtt ccg caa ggg ttt tac agt gat gtg	2084
Val Arg Gly Ser Trp Leu Gln Val Pro Gln Gly Phe Tyr Ser Asp Val	
570 575 580	
aaa gac cta gcc aca gga ttg atc acc aaa gac aaa gcc tgg aaa aat	2132
Lys Asp Leu Ala Thr Gly Leu Ile Thr Lys Asp Lys Ala Trp Lys Asn	
585 590 595	
tat cag gtc tta tat tcc gcc acg ggt gct ttg tct ctt acg gga gtt	2180
Tyr Gln Val Leu Tyr Ser Ala Thr Gly Ala Leu Ser Leu Thr Gly Val	
600 605 610	
acc acc aag gcc gtg gtg cta att ctg ttg ggg ttg tgt ggc agc aag	2228
Thr Thr Lys Ala Val Val Leu Ile Leu Leu Gly Leu Cys Gly Ser Lys	
615 620 625	
tat ctt att tta gcc tac ctc tgt tac ttg tcc ctt tgt ttt ggg cgc	2276
Tyr Leu Ile Leu Ala Tyr Leu Cys Tyr Leu Ser Leu Cys Phe Gly Arg	
630 635 640 645	
gct tct ggt tac cct ttg cgt cct gtg ctc cca tcc cag tcg tat ctc	2324
Ala Ser Gly Tyr Pro Leu Arg Pro Val Leu Pro Ser Gln Ser Tyr Leu	
650 655 660	
caa gct ggc tgg gat gtt ttg tct aaa gct caa gta gct cct ttt gct	2372
Gln Ala Gly Trp Asp Val Leu Ser Lys Ala Gln Val Ala Pro Phe Ala	
665 670 675	
ttg att ttc ttc atc tgt tgc tat ctc cgc tgc agg cta cgt tat gct	2420
Leu Ile Phe Phe Ile Cys Cys Tyr Leu Arg Cys Arg Leu Arg Tyr Ala	
680 685 690	
gcc ctt tta ggg ttt gtg ccc atg gct gcg ggc ttg ccc cta act ttc	2468
Ala Leu Leu Gly Phe Val Pro Met Ala Ala Gly Leu Pro Leu Thr Phe	

695	700	705	
ttt gtt gca gca gct gct gcc caa cca gat tat gac tgg tgg gtg cga Phe Val Ala Ala Ala Ala Ala Gln Pro Asp Tyr Asp Trp Trp Val Arg 710 715 720 725			2516
ctg cta gtg gca ggg tta gtt ttg tgg gcc ggc cgt aac cgt ggt cac Leu Leu Val Ala Gly Leu Val Leu Trp Ala Gly Arg Asn Arg Gly His 730 735 740			2564
cgc ata gct ctg ctt gta ggt cct tgg cct ctg gta gcg ctt tta acc Arg Ile Ala Leu Leu Val Gly Pro Trp Pro Leu Val Ala Leu Leu Thr 745 750 755			2612
ctc ttg cat ttg gtt acg cct gct tca gct ctg gac acg gag gtg gcc Leu Leu His Leu Val Thr Pro Ala Ser Ala Leu Asp Thr Glu Val Ala 760 765 770			2660
gcg tcg tgt ggc ggc gtt gtt ctt gtc ggg tta atg gcg ctg act ctg Ala Ser Cys Gly Gly Val Val Leu Val Gly Leu Met Ala Leu Thr Leu 775 780 785			2708
tcg cca tat tac aag cgc tat atc agc tgg tgc atg tgg tgg ctt cag Ser Pro Tyr Tyr Lys Arg Tyr Ile Ser Trp Cys Met Trp Trp Leu Gln 790 795 800 805			2756
tat ttt ctg acc aga gta gaa gcg caa ctg cac gtg tgg gtt ccc ccc Tyr Phe Leu Thr Arg Val Glu Ala Gln Leu His Val Trp Val Pro Pro 810 815 820			2804
ctc aac gtc cgg ggg ggg cgc gat gcc gtc atc tta ctc atg tgt gta Leu Asn Val Arg Gly Gly Arg Asp Ala Val Ile Leu Leu Met Cys Val 825 830 835			2852
gta cac ccg acc ctg gta ttt gac atc acc aaa cta ctc ctg gcc atc Val His Pro Thr Leu Val Phe Asp Ile Thr Lys Leu Leu Leu Ala Ile 840 845 850			2900
ttc gga ccc ctt tgg att ctt caa gcc agt ttg ctt aaa gtc ccc tac Phe Gly Pro Leu Trp Ile Leu Gln Ala Ser Leu Leu Lys Val Pro Tyr 855 860 865			2948
ttc gtg cgc gtt caa ggc ctt ctc cgg atc tgc gcg cta gcg cgg aag Phe Val Arg Val Gln Gly Leu Leu Arg Ile Cys Ala Leu Ala Arg Lys 870 875 880 885			2996
ata gcc gga ggt cat tac gtg caa atg gcc atc atc aag tta ggg gcg Ile Ala Gly Gly His Tyr Val Gln Met Ala Ile Ile Lys Leu Gly Ala 890 895 900			3044
ctt act ggc acc tat gtg tat aac cat ctc acc cct ctt cga gac tgg Leu Thr Gly Thr Tyr Val Tyr Asn His Leu Thr Pro Leu Arg Asp Trp 905 910 915			3092
gcg cac aac ggc ctg cga gat ctg gcc gtg gct gtg gaa cca gtc gtc Ala His Asn Gly Leu Arg Asp Leu Ala Val Ala Val Glu Pro Val Val 920 925 930			3140
ttc tcc cga atg gag acc aag ctc atc acg tgg ggg gca gat acc gcc Phe Ser Arg Met Glu Thr Lys Leu Ile Thr Trp Gly Ala Asp Thr Ala 935 940 945			3188
gcg tgc ggt gac atc atc aac ggc ttg ccc gtc tct gcc cgt agg ggc			3236

Ala Cys Gly Asp Ile Ile Asn Gly Leu Pro Val Ser Ala Arg Arg Gly 950 955 960 965	
cag gag ata ctg ctt ggg cca gcc gac gga atg gtc tcc aag ggg tgg Gln Glu Ile Leu Leu Gly Pro Ala Asp Gly Met Val Ser Lys Gly Trp 970 975 980	3284
agg ttg ctg gcg ccc atc acg gcg tac gcc cag cag acg aga ggc ctc Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu 985 990 995	3332
cta ggg tgt ata atc acc agc ctg act ggc cgg gac aaa aac caa gtg Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val 1000 1005 1010	3380
gag ggt gag gtc cag atc gtg tca act gct acc caa acc ttc ctg gca Glu Gly Glu Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala 1015 1020 1025	3428
acg tgc atc aat ggg gta tgc tgg act gtc tac cac ggg gcc gga acg Thr Cys Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr 1030 1035 1040 1045	3476
agg acc atc gca tca ccc aag ggt cct gtc atc cag atg tat acc aat Arg Thr Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn 1050 1055 1060	3524
gtg gac caa gac ctt gtg ggc tgg ccc gct cct caa ggt tcc cgc tca Val Asp Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser 1065 1070 1075	3572
ttg aca ccc tgt acc tgc ggc tcc tcg gac ctt tac ctg gtc acg agg Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg 1080 1085 1090	3620
cac gcc gat gtc att ccc gtg cgc cgg cga ggt gat agc agg ggt agc His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser 1095 1100 1105	3668
ctg ctt tcg ccc cgg ccc att tcc tac ttg aaa ggc tcc tcg ggg ggt Leu Leu Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly 1110 1115 1120 1125	3716
ccg ctg ttg tgc ccc gcg gga cac gcc gtg ggc cta ttc agg gcc gcg Pro Leu Leu Cys Pro Ala Gly His Ala Val Gly Leu Phe Arg Ala Ala 1130 1135 1140	3764
gtg tgc acc cgt gga gtg gct aaa gcg gtg gac ttt atc cct gtg gag Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu 1145 1150 1155	3812
aac cta ggg aca acc atg aga tcc ccg gtg ttc acg gac aac tcc tct Asn Leu Gly Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser 1160 1165 1170	3860
cca cca gca gtg ccc cag agc ttc cag gtg gcc cac ctg cat gct ccc Pro Pro Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro 1175 1180 1185	3908
acc ggc agc ggt aag agc acc aag gtc ccg gct gcg tac gca gcc cag Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln 1190 1195 1200 1205	3956

ggc tac aag gtg ttg gtg ctc aac ccc tct gtt gct gca acg ctg ggc	4004
Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly	
1210 1215 1220	
ttt ggt gct tac atg tcc aag gcc cat ggg gtt gat cct aat atc agg	4052
Phe Gly Ala Tyr Met Ser Lys Ala His Gly Val Asp Pro Asn Ile Arg	
1225 1230 1235	
acc ggg gtg aga aca att acc act ggc agc ccc atc acg tac tcc acc	4100
Thr Gly Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr	
1240 1245 1250	
tac ggc aag ttc ctt gcc gac ggc ggg tgc tca gga ggt gct tat gac	4148
Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp	
1255 1260 1265	
ata ata att tgt gac gag tgc cac tcc acg gat gcc aca tcc atc ttg	4196
Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu	
1270 1275 1280 1285	
ggc atc ggc act gtc ctt gac caa gca gag act gcg ggg gcg aga ctg	4244
Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu	
1290 1295 1300	
gtt gtg ctc gcc act gct acc cct ccg ggc tcc gtc act gtg tcc cat	4292
Val Val Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Ser His	
1305 1310 1315	
cct aac atc gag gag gtt gct ctg tcc acc acc gga gag atc ccc ttt	4340
Pro Asn Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe	
1320 1325 1330	
tac ggc aag gct atc ccc ctc gag gtg atc aag ggg gga aga cat ctc	4388
Tyr Gly Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu	
1335 1340 1345	
atc ttc tgc cac tca aag aag aag tgc gac gag ctc gcc gcg aag ctg	4436
Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu	
1350 1355 1360 1365	
gtc gca ttg ggc atc aat gcc gtg gcc tac tac cgc ggt ctt gac gtg	4484
Val Ala Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val	
1370 1375 1380	
tct gtc atc ccg acc agc ggc gat gtt gtc gtc gtg tcg acc gat gct	4532
Ser Val Ile Pro Thr Ser Gly Asp Val Val Val Val Ser Thr Asp Ala	
1385 1390 1395	
ctc atg act ggc ttt acc ggc gac ttc gac tct gtg ata gac tgc aac	4580
Leu Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn	
1400 1405 1410	
acg tgt gtc act cag aca gtc gat ttc agc ctt gac cct acc ttt acc	4628
Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr	
1415 1420 1425	
att gag aca acc acg ctc ccc cag gat gct gtc tcc agg act caa cgc	4676
Ile Glu Thr Thr Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg	
1430 1435 1440 1445	
cgg ggc agg act ggc agg ggg aag cca ggc atc tat aga ttt gtg gca	4724
Arg Gly Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala	
1450 1455 1460	

ccg ggg gag cgc ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag	4772
Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu	
1465 1470 1475	
tgc tat gac gcg ggc tgt gct tgg tat gag ctc acg ccc gcc gag act	4820
Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr	
1480 1485 1490	
aca gtt agg cta cga gcg tac atg aac acc ccg ggg ctt ccc gtg tgc	4868
Thr Val Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys	
1495 1500 1505	
cag gac cat ctt gaa ttt tgg gag ggc gtc ttt acg ggc ctc act cat	4916
Gln Asp His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His	
1510 1515 1520 1525	
ata gat gcc cac ttt tta tcc cag aca aag cag agt ggg gag aac ttt	4964
Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe	
1530 1535 1540	
cct tac ctg gta gcg tac caa gcc acc gtg tgc gct agg gct caa gcc	5012
Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala	
1545 1550 1555	
cct ccc cca tcg tgg gac cag atg tgg aag tgt ttg atc cgc ctt aaa	5060
Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys	
1560 1565 1570	
ccc acc ctc cat ggg cca aca ccc ctg cta tac aga ctg ggc gct gtt	5108
Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val	
1575 1580 1585	
cag aat gaa gtc acc ctg acg cac cca atc acc aaa tac atc atg aca	5156
Gln Asn Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Thr	
1590 1595 1600 1605	
tgc atg tcg gcc gac ctg gag gtc gtc acg agc acc tgg gtg ctc gtt	5204
Cys Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val	
1610 1615 1620	
ggc ggc gtc ctg gct gct ctg gcc gcg tat tgc ctg tca aca ggc tgc	5252
Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys	
1625 1630 1635	
gtg gtc ata gtg ggc agg atc gtc ttg tcc ggg aag ccg gca att ata	5300
Val Val Ile Val Gly Arg Ile Val Leu Ser Gly Lys Pro Ala Ile Ile	
1640 1645 1650	
cct gac agg gag gtt ctc tac cag gag ttc gat gag atg gaa gag tgc	5348
Pro Asp Arg Glu Val Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys	
1655 1660 1665	
tct cag cac tta ccg tac atc gag caa ggg atg atg ctc gct gag cag	5396
Ser Gln His Leu Pro Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln	
1670 1675 1680 1685	
ttc aag cag aag gcc ctc ggc ctc ctg cag acc gcg tcc cgc cat gca	5444
Phe Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Ser Arg His Ala	
1690 1695 1700	
gag gtt atc acc cct gct gtc cag acc aac tgg cag aaa ctc gag gtc	5492
Glu Val Ile Thr Pro Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Val	

1705	1710	1715	
ttt tgg gcg aag cac atg tgg aat ttc atc agt ggg ata caa tac ttg			5540
Phe Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu			
1720	1725	1730	
gcg ggc ctg tca acg ctg cct ggt aac ccc gcc att gct tca ttg atg			5588
Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met			
1735	1740	1745	
gct ttt aca gct gcc gtc acc agc cca cta acc act ggc caa acc ctc			5636
Ala Phe Thr Ala Ala Val Thr Ser Pro Leu Thr Thr Gly Gln Thr Leu			
1750	1755	1760	1765
ctc ttc aac ata ttg ggg ggg tgg gtg gct gcc cag ctc gcc gcc ccc			5684
Leu Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro			
1770	1775	1780	
ggt gcc gct act gcc ttt gtg ggt gct ggc cta gct ggc gcc gcc atc			5732
Gly Ala Ala Thr Ala Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Ile			
1785	1790	1795	
ggc agc gtt gga ctg ggg aag gtc ctc gtg gac att ctt gca ggg tat			5780
Gly Ser Val Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr			
1800	1805	1810	
ggc gcg ggc gtg gcg gga gct ctt gta gca ttc aag atc atg agc ggt			5828
Gly Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly			
1815	1820	1825	
gag gtc ccc tcc acg gag gac ctg gtc aat ctg ctg ccc gcc atc ctc			5876
Glu Val Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu			
1830	1835	1840	1845
tcg cct gga gcc ctt gta gtc ggt gtg gtc tgc gca gca ata ctg cgc			5924
Ser Pro Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg			
1850	1855	1860	
cgg cac gtt ggc ccg ggc gag ggg gca gtg caa tgg atg aac cgg cta			5972
Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu			
1865	1870	1875	
ata gcc ttc gcc tcc cgg ggg aac cat gtt tcc ccc acg cac tac gtg			6020
Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val			
1880	1885	1890	
ccg gag agc gat gca gcc gcc cgc gtc act gcc ata ctc agc agc ctc			6068
Pro Glu Ser Asp Ala Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu			
1895	1900	1905	
act gta acc cag ctc ctg agg cga ctg cat cag tgg ata agc tcg gag			6116
Thr Val Thr Gln Leu Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu			
1910	1915	1920	1925
tgt acc act cca tgc tcc ggt tcc tgg cta agg gac atc tgg gac tgg			6164
Cys Thr Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp			
1930	1935	1940	
ata tgc gag gtg ctg agc gac ttt aag acc tgg ctg aaa gcc aag ctc			6212
Ile Cys Glu Val Leu Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu			
1945	1950	1955	
atg cca caa ctg cct ggg att ccc ttt gtg tcc tgc cag cgc ggg tat			6260

ttc gat ccg ctt gtg gca gag gag gat gag cgg gag gtc tcc gta cct	7028
Phe Asp Pro Leu Val Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro	
2215 2220 2225	
gca gaa att ctg cgg aag tct cgg aga ttc gcc cgg gcc ctg ccc gtc	7076
Ala Glu Ile Leu Arg Lys Ser Arg Arg Phe Ala Arg Ala Leu Pro Val	
2230 2235 2240 2245	
tgg gcg cgg ccg gac tac aac ccc ccg cta gta gag acg tgg aaa aag	7124
Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu Val Glu Thr Trp Lys Lys	
2250 2255 2260	
cct gac tac gaa cca cct gtg gtc cat ggc tgc ccg cta cca cct cca	7172
Pro Asp Tyr Glu Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Pro	
2265 2270 2275	
cgg tcc cct cct gtg cct ccg cct cgg aaa aag cgt acg gtg gtc ctc	7220
Arg Ser Pro Pro Val Pro Pro Arg Lys Lys Arg Thr Val Val Leu	
2280 2285 2290	
acc gaa tca acc cta tct act gcc ttg gcc gag ctt gcc acc aaa agt	7268
Thr Glu Ser Thr Leu Ser Thr Ala Leu Ala Glu Leu Ala Thr Lys Ser	
2295 2300 2305	
ttt ggc agc tcc tca act tcc ggc att acg ggc gac aat acg aca aca	7316
Phe Gly Ser Ser Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr	
2310 2315 2320 2325	
tcc tct gag ccc gcc cct tct ggc tgc ccc ccc gac tcc gac gtt gag	7364
Ser Ser Glu Pro Ala Pro Ser Gly Cys Pro Pro Asp Ser Asp Val Glu	
2330 2335 2340	
tcc tat tct tcc atg ccc ccc ctg gag ggg gag cct ggg gat ccg gat	7412
Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp	
2345 2350 2355	
ctc agc gac ggg tca tgg tcg acg gtc agt agt ggg gcc gac acg gaa	7460
Leu Ser Asp Gly Ser Trp Ser Thr Val Ser Ser Gly Ala Asp Thr Glu	
2360 2365 2370	
gat gtc gtg tgc tgc tca atg tct tat tcc tgg aca ggc gca ctc gtc	7508
Asp Val Val Cys Cys Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val	
2375 2380 2385	
acc ccg tgc gct gcg gaa gaa caa aaa ctg ccc atc aac gca ctg agc	7556
Thr Pro Cys Ala Ala Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser	
2390 2395 2400 2405	
aac tcg ttg cta cgc cat cac aat ctg gtg tat tcc acc act tca cgc	7604
Asn Ser Leu Leu Arg His His Asn Leu Val Tyr Ser Thr Thr Ser Arg	
2410 2415 2420	
agt gct tgc caa agg cag aag aaa gtc aca ttt gac aga ctg caa gtt	7652
Ser Ala Cys Gln Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val	
2425 2430 2435	
ctg gac agc cat tac cag gac gtg ctc aag gag gtc aaa gca gcg gcg	7700
Leu Asp Ser His Tyr Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ala	
2440 2445 2450	
tca aaa gtg aag gct aac ttg cta tcc gta gag gaa gct tgc agc ctg	7748
Ser Lys Val Lys Ala Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu	
2455 2460 2465	

acg ccc cca cat tca gcc aaa tcc aag ttt ggc tat ggg gca aaa gac	7796
Thr Pro Pro His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp	
2470 2475 2480 2485	
gtc cgt tgc cat gcc aga aag gcc gta gcc cac atc aac tcc gtg tgg	7844
Val Arg Cys His Ala Arg Lys Ala Val Ala His Ile Asn Ser Val Trp	
2490 2495 2500	
aaa gac ctt ctg gaa gac agt gta aca cca ata gac act acc atc atg	7892
Lys Asp Leu Leu Glu Asp Ser Val Thr Pro Ile Asp Thr Thr Ile Met	
2505 2510 2515	
gcc aag aac gag gtt ttc tgc gtt cag cct gag aag ggg ggt cgt aag	7940
Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys	
2520 2525 2530	
cca gct cgt ctc atc gtg ttc ccc gac ctg ggc gtg cgc gtg tgc gag	7988
Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu	
2535 2540 2545	
aag atg gcc ctg tac gac gtg gtt agc aag ctc ccc ctg gcc gtg atg	8036
Lys Met Ala Leu Tyr Asp Val Val Ser Lys Leu Pro Leu Ala Val Met	
2550 2555 2560 2565	
gga agc tcc tac gga ttc caa tac tca cca gga cag cgg gtt gaa ttc	8084
Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe	
2570 2575 2580	
ctc gtg caa gcg tgg aag tcc aag aag acc ccg atg ggg ttc tcg tat	8132
Leu Val Gln Ala Trp Lys Ser Lys Lys Thr Pro Met Gly Phe Ser Tyr	
2585 2590 2595	
gat acc cgc tgt ttt gac tcc aca gtc act gag agc gac atc cgt acg	8180
Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr	
2600 2605 2610	
gag gag gca att tac caa tgt tgt gac ctg gac ccc caa gcc cgc gtg	8228
Glu Glu Ala Ile Tyr Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val	
2615 2620 2625	
gcc atc aag tcc ctc act gag agg ctt tat gtt ggg ggc cct ctt acc	8276
Ala Ile Lys Ser Leu Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr	
2630 2635 2640 2645	
aat tca agg ggg gaa aac tgc ggc tac cgc agg tgc cgc gcg agc ggc	8324
Asn Ser Arg Gly Glu Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly	
2650 2655 2660	
gta ctg aca act agc tgt ggt aac acc ctc act tgc tac atc aag gcc	8372
Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Ile Lys Ala	
2665 2670 2675	
cgg gca gcc tgt cga gcc gca ggg ctc cag gac tgc acc atg ctc gtg	8420
Arg Ala Ala Cys Arg Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val	
2680 2685 2690	
tgt ggc gac gac tta gtc gtt atc tgt gaa agt gcg ggg gtc cag gag	8468
Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu	
2695 2700 2705	
gac gcg gcg agc ctg aga gcc ttc acg gag gct atg acc agg tac tcc	8516
Asp Ala Ala Ser Leu Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser	

2710	2715	2720	2725	
gcc ccc ccc ggg gac ccc cca caa cca gaa tac gac ttg gag ctt ata				8564
Ala Pro Pro Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile				
	2730	2735	2740	
aca tca tgc tcc tcc aac gtg tca gtc gcc cac gac ggc gct gga aag				8612
Thr Ser Cys Ser Ser Asn Val Ser Val Ala His Asp Gly Ala Gly Lys				
	2745	2750	2755	
agg gtc tac tac ctt acc cgt gac cct aca acc ccc ctc gcg aga gcc				8660
Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala				
	2760	2765	2770	
gcg tgg gag aca gca aga cac act cca gtc aat tcc tgg cta ggc aac				8708
Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn				
	2775	2780	2785	
ata atc atg ttt gcc ccc aca ctg tgg gcg agg atg ata ctg atg acc				8756
Ile Ile Met Phe Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr				
	2790	2795	2800	2805
cat ttc ttt agc gtc ctc ata gcc agg gat cag ctt gaa cag gct ctt				8804
His Phe Phe Ser Val Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu				
	2810	2815	2820	
aac tgt gag atc tac gga gcc tgc tac tcc ata gaa cca ctg gat cta				8852
Asn Cys Glu Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu				
	2825	2830	2835	
cct cca atc att caa aga ctc cat ggc ctc agc gca ttt tca ctc cac				8900
Pro Pro Ile Ile Gln Arg Leu His Gly Leu Ser Ala Phe Ser Leu His				
	2840	2845	2850	
agt tac tct cca ggt gaa atc aat agg gtg gcc gca tgc ctc aga aaa				8948
Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys				
	2855	2860	2865	
ctt ggg gtc ccg ccc ttg cga gct tgg aga cac cgg gcc cgg agc gtc				8996
Leu Gly Val Pro Pro Leu Arg Ala Trp Arg His Arg Ala Arg Ser Val				
	2870	2875	2880	2885
cgc gct agg ctt ctg tcc aga gga ggc agg gct gcc ata tgt ggc aag				9044
Arg Ala Arg Leu Leu Ser Arg Gly Gly Arg Ala Ala Ile Cys Gly Lys				
	2890	2895	2900	
tac ctc ttc aac tgg gca gta aga aca aag ctc aaa ctc act cca ata				9092
Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile				
	2905	2910	2915	
gcg gcc gct ggc cgg ctg gac ttg tcc ggt tgg ttc acg gct ggc tac				9140
Ala Ala Ala Gly Arg Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr				
	2920	2925	2930	
agc ggg gga gac att tat cac agc gtg tct cat gcc cgg ccc cgc tgg				9188
Ser Gly Gly Asp Ile Tyr His Ser Val Ser His Ala Arg Pro Arg Trp				
	2935	2940	2945	
ttc tgg ttt tgc cta ctc ctg ctc gct gca ggg gta ggc atc tac ctc				9236
Phe Trp Phe Cys Leu Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu				
	2950	2955	2960	2965
ctc ccc aac cga tgaagggttg ggtaaact ccggcctctt aagccatttc				9288

Leu Pro Asn Arg

ctgtttttttt tttttttttt tttttttttt tctttttttt tttcttttctt ttcctttcttt 9348
ttttcctttc tttttccctt ctttaaatggt ggctccatct tagccctagt cacggctagc 9408
tgtgaaagggt ccgtgagccg catgactgca gagagtgtg atactggcct ctctgcagat 9468
catgtttctag actggaattc gtcgacgagc tccctatagt gagtctgtatt agaggccgac 9528
ttggccaaat tcgtaatcat ggctcatagct gtttcctgtg tgaaattggt atccgctcac 9588
aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctggggtg cctaattgagt 9648
gagctaactc acattaattg cgttgcgctc actgcccgtt ttccagtcgg gaaacctgtc 9708
gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcy 9768
ctcttccgct tcctcgctca ctgactcgct gcgctcggtc gttcggtgc ggcgagcgg 9828
atcagctcac tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa 9888
gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc 9948
gtttttccat aggctccgcc cccctgacga gcatcacaaa aatcgacgct caagtacag 10008
gtggcgaaac ccgacaggac tataaagata ccaggcggtt cccctggaa gctccctcgt 10068
gcgctctcct gttccgacct tgccgcttac cggatacctg tccgcctttc tcccttcggg 10128
aagcgtggcg ctttctcata gctcacgctg taggtatctc agttcggtgt aggtcgttcg 10188
ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccg 10248
taactatcgt cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac 10308
tggtaacagg attagcagag cgaggtatgt aggcggtgct acagagttct tgaagtgtg 10368
gcctaactac ggctacacta gaaggacagt atttggtatc tgcgctctgc tgaagccagt 10428
taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgy 10488
tggttttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc 10548
tttgatcttt tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggatttt 10608
ggctcatgaga ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt 10668
taaataaata taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag 10728
tgaggcacct atctcagcga tctgtctatt tcgttcatcc atagttgcct gactccccgt 10788
cgtgtagata actacgatac gggagggcct accatctggc ccagtgctg caatgatacc 10848
gcgagacca cgtcacccg ctccagattt atcagcaata aaccagccag ccggaagggc 10908
cgagcgcaga agtggctcct caactttatc cgctccatc cagtctatta attgttgccg 10968
ggaagctaga gtaagtagtt cgccagttaa tagtttgccg aacgttggtg ccattgctac 11028
aggcatcgtg gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg 11088
atcaaggcga gttacatgat ccccatggt gtgcaaaaaa gcggttagct ccttcgggtc 11148

tccgatcggtt gtcagaagta agttggccgc agtggttatca ctcatgggta tggcagcact 11208
gcataaattct cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc 11268
aaccaagtca ttctgagaat agtgatatgcg gcgaccgagt tgctcttgcc cggcgtcaat 11328
acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc 11388
ttcggggcgca aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac 11448
tcgtgcaccc aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa 11508
aacaggaagg caaaatgccg caaaaaaggg aataagggcg acacggaaat gttgaatact 11568
catactcttc ctttttcaat attattgaag catthtatcag ggthattgtc tcatgagcgg 11628
atacatattt gaatgtattt agaaaaataa acaaataaggg gttccgcgca catttccccg 11688
aaaagtgcc cctgacgcgc cctgtagcgg cgcattaagc gcggcgggtg tgggtgggttac 11748
gcgcagcgtg accgctacac ttgccagcgc cctagcgccc gctcctttcg ctttcttccc 11808
ttcctttctc gccacgttcg ccggctttcc ccgtcaagct ctaaactcggg gcaccccttt 11868
agggttccga tttagtgtt tacggcacct cgaccccaaa aaacttgatt aggggtgatgg 11928
ttcacgtagt gggccatcgc cctgatagac ggtttttcgc cctttgacgt tggagtccac 11988
gttctttaat agtggactct tgttccaaac tggaacaaca ctcaacccta tctcgggtcta 12048
ttcttttgat ttataaggga ttttgccgat ttcggcctat tggttaaaaa atgagctgat 12108
ttaacaaaaa tttaacgcga attttaacaa aatattaaca aaatattaac gtttacaatt 12168
tcccatcgc cattcaggct gcgcaactgt tgggaagggc gatcgggtgcg ggcctcttcg 12228
ctattacgcc agctggcgaa agggggatgt gctgcaaggc gattaagttg ggtaacgcca 12288
gggttttccc agtcacgacg ttgtaaaacg acggccagtg ccaagctgac ttggtcagcg 12348
gccgctaata cgactcacta ta 12370

<210> 27

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
Primer

<400> 27

aattcggccc tgcaggccac aacagtc

27