Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3081362 C 2024/04/09

(11)(21) 3 081 362

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(22) Date de dépôt/Filing Date: 2007/07/24

(41) Mise à la disp. pub./Open to Public Insp.: 2008/08/07

(45) Date de délivrance/Issue Date: 2024/04/09

(62) Demande originale/Original Application: 2 993 381

(30) Priorité/Priority: 2006/07/24 (US60/832,780)

(51) Cl.Int./Int.Cl. *C12Q 1/6883* (2018.01), *C12Q 1/68* (2018.01), *C12Q 1/6827* (2018.01), *C12Q 1/6844* (2018.01), *C12Q 1/6858* (2018.01)

(72) Inventeurs/Inventors:
WATNICK, TERRY J., US;
GARCIA-GONZALEZ, MIGUEL, US;
GERMINO, GREGORY G., US;
JONES, JEFFREY G., US

(73) Propriétaires/Owners:
ATHENA DIAGNOSTICS, INC., US;
THE JOHNS HOPKINS UNIVERSITY, US

(74) Agent: MARKS & CLERK

(54) Titre : PROCEDE DE DETECTION DE L'APPARITION DE LA MALADIE POLYKYSTIQUE RENALE AUTOSOMALE DOMINANTE (MPRAD) EN FONCTION DE LA PRESENCE D'UNE MODIFICATION DE SEQUENCE NUCLEOTIDIQUE DANS LAPOLYKYSTOSE RENALE TYPE DOMINANT (PKD) 1

(54) Title: METHOD OF DETECTING THE OCCURRENCE OF ADPKD BASED ON PRESENCE OF NUCLEOTIDE SEQUENCE ALTERATION IN PKD1

(57) Abrégé/Abstract:
A method of detecting or predicting the occurrence of autosomal dominant polycystic kidney disease (ADPKD) in an individual is provided. The method involves detecting the presence of one or more nucleotide sequence alterations in a PKD 1 gene having the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:7 in a nucleic acid sample obtained from said individual.

Abstract

A method of detecting or predicting the occurrence of autosomal dominant polycystic kidney disease (ADPKD) in an individual is provided. The method involves detecting the presence of one or more nucleotide sequence alterations in a PKD 1 gene having the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:7 in a nucleic acid sample obtained from said individual.

Date Reçue/Date Received 2020-05-26

METHOD OF DETECTING THE OCCURRENCE OF ADPKD BASED ON PRESENCE OF NUCLEOTIDE SEQUENCE ALTERATION IN PKD1

GOVERNMENT SUPPORT

The invention was supported in part by Becas FPI de investigacion from Ministerio de Ciencia y Technologia (Spain) and grants R01DK70617, P50-DK57325 and R37DK48006 from the National Institutes of Health. The Governments have certain rights in the invention.

BACKGROUND OF THE INVENTION

Autosomal dominant polycystic kidney disease (ADPKD) is an exceptionally common inherited disorder in humans, affecting approximately one in every 600 to 1000 individuals (Gabow P.A., *N Engl J Med* 329(5):332-342, 1993). The disease is characterized by age dependent growth of renal cysts such that end-stage renal disease (ESRD) typically ensues during mid-adulthood. ADPKD may alternatively, or in addition, involve cysts in other organs including liver and spleen, as well as gastrointestinal, cardiovascular, and musculoskeletal abnormalities (Gabow P.A., *N Engl J Med* 329(5):332-342, 1993; Gabow P et al., *Adv Nephrol* 18:19-32, 1989). Both ADPKD type 1 and type 2 share the entire range of renal and extrarenal manifestations, but type 2 appears to have a delayed onset relative to type 1. The common phenotypic complications observed for ADPKD which include hypertension, hematuria and urinary tract infection, seem to be clinically milder in type 2 patients.

Approximately 85 percent of ADPKD cases are caused by mutations in the PKD1 gene [MIM 601313], which is located on chromosome 16, while the remainder are due to mutations in the PKD2 gene [MIM 173910] located on chromosome 4 (Peters et al., *Contrib Nephrol* 97:128-139, 1992; European

Polycystic Kidney Disease Consortium, *Cell*, 77(6):881-894, 1994; International Polycystic Kidney Disease Consortium, *Cell* 81(2):289-298, 1995; Hughes J. et al, *Nat Genet* 10(2):151-160, 1995; Mochizuki T. et al., *Science* 272(5266):1339-1342, 1996). However, genetic testing for ADPKD has posed a unique set of challenges in terms of DNA diagnostics. PKD1 analysis in particular has been complicated because the 5' portion of the gene (exons 1-34) is replicated in at least five highly homologous copies (with less than 2% divergence) elsewhere on chromosome 16 (Hughes J. et al, *Nat Genet* 10(2):151-160, 1995). Further complicating PKD1 mutant analysis, PKD1 has a high rate of potentially non-pathogenic DNA variation; thus the nature of each change detected must be verified. Several techniques have been used to detect mutations in the PKD1 gene including using gene-specific primers to amplify large products screened via nested PCR techniques, denaturing high-performance liquid chromatography (DHPLC) to screen nested PCR products for mutations and direct sequencing of the entire PKD1 coding sequence (Watnick TJ et al., *Hum Mol Genet* 6(9):1473-1481, 1997; Watnick TJ et al., *Mol Cell* 2(2):247-251, 1998; Watnick T. et al., *Am J Hum Genet* 65(6):1561-1571, 1999; Phakdeekitcharoen B. et al., *Kidney Int* 58(4):1400-1412, 2000; Phakdeekitcharoen B. et al., *J Am Soc Nephrol* 12:955-963, 2001; Thomas R. et al., *Am J Hum Genet* 65(1):39-49, 1999; Perichot R.A., *Hum Genet* 105(3):231-239, 1999; Perichot R. et al., *Eur J Hum Genet* 8(5):353-359, 2000; Afzal A.R. et al., *Genet* 4(4):365-370; Rossetti S. et al., *Lancet* 361(9376):2196-2201, 2003; Rossetti S. et al., *Kidney Int* 61:1588-1599, 2002; Rossetti S. et al., *Am J Hum Genet* 68(1):46-63, 2001, Inoue S. et al., *Hum Mutat* 19(6):622-628, 2002; Burtey S. et al., *J Med Genet* 39(6):422-429, 2002; Mizoguchi M. et al., *J Hum Genet* 46(9):511-517, 2001; Zhang D.Y. et al., *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 21(3):211-214, 2004). However, some of these strategies may not be cost effective for routine clinical sample analysis and/or their mutation detection rate has not been established or is inadequate. For example, direct DNA sequencing of the entire coding regions of PKD1 and PKD2 is considered necessary because no mutational hot spots have been identified in either PKD1 or PKD2. Although several pathogenic mutations in PKD1 and PKD2 have been identified, the known mutations do not account for all those individuals with

ADPKD. Thus, to accurately diagnose and treat the disease, there remains a need to identify other mutations of PKD1 or PKD2 which are linked to ADPKD.

SUMMARY OF THE INVENTION

Several novel nucleotide sequence alterations in the PKD1 and PKD2 genes have been identified that are associated with ADPKD. The mutations in PKD1 and PKD2 were found by direct sequencing of the genes and the pathogenicity of the mutations determined using a combination of various analyses and algorithms. The mutations in the PKD1 and PKD2 genes identified as pathogenic can be used to detect and/or predict the occurrence of ADPKD in an individual. This is important clinically in diagnostic and prognostic analysis of the genes for ADPKD.

Accordingly, the invention relates to methods of detecting or predicting the occurrence of ADPKD in an individual. In one aspect, the present invention relates to a method of detecting or predicting the occurrence of autosomal dominant polycystic kidney disease (ADPKD) in an individual comprising detecting the presence of one or more nucleotide sequence alterations in a PKD1 gene having the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:7 in a nucleic acid sample obtained from said individual, wherein said one or more alterations are selected from the group consisting of: a deletion of TTTAA at nucleotide positions 559 to 563 of SEQ ID NO:1, an insertion of CT at nucleotide position 1124 of SEQ ID NO:1, an insertion of an A, T, G, or C at nucleotide position 2291 of SEQ ID NO:1, an insertion of an A, T, G, or C at nucleotide position 2297 of SEQ ID NO:1, an insertion a T at nucleotide position 5365 of SEQ ID NO:1, an insertion of a G at nucleotide position 6666 of SEQ ID NO:1, an insertion of an A at nucleotide position 6881 of SEQ ID NO:1, a deletion of a T at nucleotide position 8713 of SEQ ID NO:1, an insertion of an A, T, G, or C at nucleotide position 9134 of SEQ ID NO:1, an insertion of 5 nucleotides at nucleotide position 9536 of SEQ ID NO:1, a deletion of a T at nucleotide position 10239 of SEQ ID NO:1, a change of a C to an A at nucleotide position 483 of SEQ ID NO:1, a change of a C to a T at nucleotide position 4517 of SEQ ID NO:1, a change of a C to an A at nucleotide position 7006 of SEQ ID NO:1, a change of a C to T at nucleotide position 8267 of SEQ ID NO:1, a change of a G to a T at nucleotide position 8639 of SEQ ID NO:1, a change of a G

Date Reçue/Date Received 2020-05-26

to an A at nucleotide position 20168 of SEQ ID NO:7, a change of a G to a T at nucleotide position 31025 of SEQ ID NO:7, a change of a G to a C at nucleotide position 33415 of SEQ ID NO:7, a deletion of CAA between nucleotide positions 508 to 516 of SEQ ID NO:1, a deletion of TGG at nucleotide positions 1848 to 1850 of SEQ ID NO:1, a deletion of CCAACTCCG at nucleotide positions 8892 to 8900 of SEQ ID NO:1, a deletion of AAG at nucleotide positions 9905 to 9907 of SEQ ID NO:1, a deletion of CTC at nucleotide positions 10070 to 10072 of SEQ ID NO:1, a deletion of TGG at nucleotide positions 12597 to 12599 of SEQ ID NO:1, a change of a C to an A at nucleotide position 1023 of SEQ ID NO:1, a change of a G to an A at nucleotide position 385 of SEQ ID NO:1, a change of an A to a G at nucleotide position 1470 of SEQ ID NO:1, a change of a C to a T at nucleotide position 4262 of SEQ ID NO:1, a change of a T to an A at nucleotide position 8855 of SEQ ID NO:1, a change of an A to a G at nucleotide position 1794 of SEQ ID NO:1, a change of a G to an A at nucleotide position 6036 of SEQ ID NO:1, a change of a C to a T at nucleotide position 2042 of SEQ ID NO:1, a change of a C to a T at nucleotide position 3351 of SEQ ID NO:1, a change of an A to a G at nucleotide position 6756 of SEQ ID NO:1, a change of a C to a T at nucleotide position 5793 of SEQ ID NO:1, a change of a C to a T at nucleotide position 6707 of SEQ ID NO:1, a change of a G to a C at nucleotide position 10187 of SEQ ID NO:1, a change of a C to a G at nucleotide position 7116 of SEQ ID NO:1, a change of an A to a G at nucleotide position 10311 of SEQ ID NO:1, a change of a T to a C at nucleotide position 7554 of SEQ ID NO:1, a change of a C to a T at nucleotide position 7757 of SEQ ID NO:1, a change of a T to a C at nucleotide position 8067 of SEQ ID NO:1, a change of a C to a T at nucleotide position 8138 of SEQ ID NO:1, a change of a C to a T at nucleotide position 8509 of SEQ ID NO:1, a change of a C to an A at nucleotide position 10096 of SEQ ID NO:1 and a change of a C to a T at nucleotide position 12658 of SEQ ID NO:1. The detection of one or more of the listed nucleotide sequence alterations indicates that the individual has ADPKD or will develop ADPKD. In one embodiment, at least one nucleotide sequence alteration other than the one or more nucleotide sequence alterations listed above is also detected in SEQ ID NO:1 and/or SEQ ID NO:4, wherein the at least one nucleotide sequence alteration which is also detected is associated with ADPKD. In another aspect, the

one or more nucleotide sequence alterations are detected by sequencing, polymerase chain reaction (PCR), DHPLC or combinations of the foregoing.

The present invention also relates to a method of detecting or predicting the occurrence of autosomal dominant polycystic kidney disease (ADPKD) in an individual comprising detecting the presence of one or more nucleotide sequence alterations in a PKD2 gene having the nucleotide sequence of SEQ ID NO:4 in a nucleic acid sample obtained from said individual, wherein said one or more alterations are selected from the group consisting of: an insertion of an A at nucleotide position 2226 of SEQ ID NO:4, a deletion of AG at nucleotide positions 2422 to 2423 of SEQ ID NO:4, a change of a C to a T at nucleotide position 2680 of SEQ ID NO:4, IVS7-1G>A, IVS8+5G>A, a deletion of TGG at nucleotide positions 374-376 of SEQ ID NO:4 and a deletion of TTC between nucleotide positions 1876-1881 of SEQ ID NO:4, wherein detection of the one or more nucleotide sequence alterations indicates that the individual has ADPKD or will develop ADPKD. In one embodiment, at least one nucleotide sequence alteration other than the one or more nucleotide sequence alterations listed above is also detected in SEQ ID NO:1 and/or SEQ ID NO:4, wherein the at least one nucleotide sequence alteration also detected is associated with ADPKD. In yet another embodiment, the one or more nucleotide sequence alterations are detected by sequencing, PCR, DHPLC or combinations thereof.

The present invention further relates to a method for detecting in an individual the presence or absence of a mutant PKD gene comprising obtaining a nucleic acid sample from the individual and detecting the presence or absence of one or more nucleotide sequence alterations in a PKD1 or PKD2 gene of the individual, wherein the one or more alterations are selected from the group consisting of: a deletion of TTTAA at nucleotide positions 559 to 563 of SEQ ID NO: 1, an insertion of CT at nucleotide position 1124 of SEQ ID NO:1, an insertion of an A, T, G, or C at nucleotide position 2291 of SEQ ID NO:1, an insertion of an A, T, G, or C at nucleotide position 2297 of SEQ ID NO:1, an insertion of a T at nucleotide position 5365 of SEQ ID NO:1, an insertion of a G at nucleotide position 6666 of SEQ ID NO:1, an insertion of an A at nucleotide position 6881 of SEQ ID NO:1, a deletion of a T at nucleotide position 8713 of SEQ ID NO:1, an insertion of an A, T,

G, or C at nucleotide position 9134 of SEQ ID NO:1, an insertion of 5 nucleotides at nucleotide position 9536 of SEQ ID NO:1, a deletion of a T at nucleotide position 10239 of SEQ ID NO:1, a change of a C to an A at nucleotide position 483 of SEQ ID NO:1,a change of a C to a T at nucleotide position 4517 of SEQ ID NO:1, a change of a C to an A at nucleotide position 7006 of SEQ ID NO:1, a change of a C to T at nucleotide position 8267 of SEQ ID NO:1, a change of a G to a T at nucleotide position 8639 of SEQ ID NO:1, a change of a G to an A at nucleotide position 20168 of SEQ ID NO:7, a change of a G to a T at nucleotide position 31025 of SEQ ID NO:7, a change of a G to a C at nucleotide position 33415 of SEQ ID NO:7, a deletion of CAA between nucleotide positions 508 to 516 of SEQ ID NO:1, a deletion of TGG at nucleotide positions 1848 to 1850 of SEQ ID NO:1, a deletion of CCAACTCCG at nucleotide positions 8892 to 8900 of SEQ ID NO:1, a deletion of AAG at nucleotide positions 9905 to 9907 of SEQ ID NO:1, a deletion of CTC at nucleotide positions 10070 to 10072 of SEQ ID NO:1, a deletion of TGG at nucleotide positions 12597 to 12599 of SEQ ID NO:1, a change of a C to an A at nucleotide position 1023 of SEQ ID NO:1, a change of a G to an A at nucleotide position 385 of SEQ ID NO:1, a change of an A to a G at nucleotide position 1470 of SEQ ID NO:1, a change of a C to a T at nucleotide position 4262 of SEQ ID NO:1, a change of a T to an A at nucleotide position 8855 of SEQ ID NO:1, a change of an A to a G at nucleotide position 1794 of SEQ ID NO:1, a change of a G to an A at nucleotide position 6036 of SEQ ID NO:1, a change of a C to a T at nucleotide position 2042 of SEQ ID NO:1, a change of a C to a T at nucleotide position 3351 of SEQ ID NO:1, a change of an A to a G at nucleotide position 6756 of SEQ ID NO:1, a change of a C to a T at nucleotide position 5793 of SEQ ID NO:1, a change of a C to a T at nucleotide position 6707 of SEQ ID NO:1, a change of a G to a C at nucleotide position 10187 of SEQ ID NO:1, a change of a C to a G at nucleotide position 7116 of SEQ ID NO:1, a change of an A to a G at nucleotide position 10311 of SEQ ID NO:1, a change of a T to a C at nucleotide position 7554 of SEQ ID NO:1, a change of a C to a T at nucleotide position 7757 of SEQ ID NO:1, a change of a T to a C at nucleotide position 8067 of SEQ ID NO:1, a change of a C to a T at nucleotide position 8138 of SEQ ID NO:1, a change of a C to a T at nucleotide position 8509 of SEQ ID NO:1, a change of a C to an A at nucleotide

position 10096 of SEQ ID NO:1, a change of a C to a T at nucleotide position 12658 of SEQ ID NO:1, a change of a C to an A at nucleotide position 7476 of SEQ ID NO:1, a change of a C to a G at nucleotide position 3527 of SEQ ID NO:1, a change of a C to an A at nucleotide position 1947 of SEQ ID NO:1, a change of an A to a G at nucleotide position 3312 of SEQ ID NO:1, a change of a C to a G at nucleotide position 4391 of SEQ ID NO:1, a change of a T to an A at nucleotide position 11040 of SEQ ID NO:1, a change of a G to a T at nucleotide position 840 of SEQ ID NO:1, a change of a G to an A at nucleotide position 7197 of SEQ ID NO:1, a change of a G to a C at nucleotide position 351 of SEQ ID NO:1, a change of a G to an A at nucleotide position 4757 of SEQ ID NO:1, a change of an A to a C at nucleotide position 1023 of SEQ ID NO:1, an insertion of: an A at nucleotide position 2226 of SEQ ID NO:4, a deletion of AG at nucleotide positions 2422 to 2423 of SEQ ID NO:4, a change of a C to a T at nucleotide position 2680 of SEQ ID NO:4, IVS7-1G>A, IVS8+5G>A, a deletion of TGG at nucleotide positions 374-376 of SEQ ID NO:4, a deletion of TTC between nucleotide positions 1876-1881 of SEQ ID NO:4 and a change of a G to an A at nucleotide position 634 of SEQ ID NO:4, wherein detection of the one or more nucleotide sequence alterations is indicative of a mutant PKD gene. In one embodiment, the presence or absence of the one or more nucleotide sequence alterations in the PKD1 or PKD2 gene of the individual indicates that the individual has ADPKD. In another embodiment, the presence or absence of one or more nucleotide sequence alterations in the PKD1 or PKD2 nucleic acid sequence is detected by sequencing, PCR and/or DHPLC.

The identification of mutations associated with ADPKD provides conclusive diagnostic information, allows the blood relatives of an individual to be pre-symptomatically and inexpensively evaluated for counseling and planning using targeted PKD gene analysis and allows prospective living-related kidney donors to be tested and subsequently accepted or rejected for donation with greater certainty. Pre-symptomatic testing for ADPKD may be particularly relevant not only in the evaluation of living kidney donors from ADPKD families, but also in the early detection for treatment with new agents that may be indicated for use early in the course of the disease (e.g., before cystic disease is apparent), family planning, the detection of ADPKD in young individuals (e.g., those under 30) for whom

ultrasound imaging may not be accurate and/or adequate or in those families with PKD2-associated ADPKD, a clinically milder disease. In addition, clinicians may encounter patients with atypical cystic disease in whom the diagnosis is not obvious. Thus, using the novel, pathogenic mutations identified in the PKD1 and PKD2 genes, the methods of the invention help to better assist in the diagnosis and management of existing ADPKD and/or predict the likelihood of the occurrence of ADPKD in an individual.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1E depict the PKD1 coding sequence (GenBank Accession No. L33243) (SEQ ID NO:1).

Figures 2A-2B depict the PKD2 coding sequence (GenBank Accession Nos. AF004859 - AF004873) (SEQ ID NO:4).

Figures 3A-3T depict wild-type PKD1 cDNA coding sequence according to one embodiment of the invention. Exon and PCR product junctions are depicted above the nucleotide sequence and amino acids are positioned under the center of each codon.

Figures 4A-4D depict wild-type PKD2 cDNA coding sequence according to one embodiment of the invention. Exon and PCR product junctions are depicted above the nucleotide sequence and amino acids are positioned under the center of each codon.

Figure 5A illustrates missense mutations affecting the PKD1 repeats and C-lectin domain. Changes that disrupt the consensus sequence are red (dark-shaded) those that do not are yellow (light-shaded). Consensus sequence code: l (aliphatic), a (aromatic), c (charged), s (small residue), p (polar residue), b (big residue), h (hydrophobic), capital letters represent the corresponding amino acid codon.

Figures 5B and 5C illustrate ribbon diagrams of the PKD repeat (5B) and C-lectin domain (5C) with potential pathogenic missense changes indicated.

Figure 6 illustrates a schematic of PKD1 mutant polypeptides with the location of each amino acid substitution indicated, M (missense) or P (polymorphism) and a photograph of a western blot of the full-length, flagged-tagged PKD1 constructs for each mutant protein and any cleavage products. FL:

Date Reçue/Date Received 2020-05-26

full-length, NTF: PKD1 N-terminal cleavage fragment, CTF; PKD1 C-terminal cleavage fragment.

Figures 7A-7B are schematics illustrating all the PKD1 (5A) and PKD2 (5B) mutations identified. Numbers 1-113 refer to identifiers of the mutations in Table 8.

Figures 8A and 8B are schematic representations of polycystin-1 (PC-1) (8A) and polycystin-2 (PC-2) (8B). The location of pathogenic (Class I and Class II, see Example) mutations are indicated.

DETAILED DESCRIPTION OF THE INVENTION

The PKD genes are genomic DNA sequences that map to chromosomal position 16p13.3 (PKD1) or chromosomal position 4q21-23 (PKD2) and give rise to messenger RNA molecules encoding PKD1 and PKD2 proteins. The PKD1 and PKD2 genes comprise the sequences of SEQ ID NO:1 and SEQ ID NO:4, respectively, which include introns and putative regulatory sequences. Like many other genes, PKD1 and PKD2 gene sequences, when compared among individuals, show sequence variations that do not affect gene expression or expression and/or function of the gene product.

The PKD1 gene (e.g., GenBank Accession Number L39891, SEQ ID NO:7) spans about 54 kb of genomic DNA on chromosome 16 (16p13.3) and contains a 12,906 basepair coding sequence divided into 46 exons from which a 14 kb mRNA is transcribed. The protein product of PKD1, polycystin-1 (PC-1) (GeneBank Accession No. AAC37576, SEQ ID NO:3), is a 4303 amino acid protein with a predicted mass of 460 kDa which forms multiprotein complexes at the cell membrane and is thought to function in cell-cell and cell-matrix signal regulation. (Arnould T et al., *J Biol. Chem* 273:6013-6018, 1992; Parnell S.C. et al., *J Biol Chem*, 277:19566-19572, 2002; Bhunia A.K. et al., *Cell*, 109:157-168, 2002; Nauli S.M. et al., *Nat Genet* 33:129-137, 2003). Approximately 75% of the PKD1 gene is duplicated and shares about 97% identity with its homologous copies. The reiterated region encompasses a 50 kb (5') portion of the gene containing the first 34 exons. Only the most 3', 5.7 kb of the gene, containing exons 35-46, is unique to PKD1. Another notable feature of the PKD1 gene is a polypyrimidine tract in intron 21 that is 2.5 kb long, the longest described in the human genome.

Date Reçue/Date Received 2020-05-26

The PKD2 gene (see e.g., GenBank Accession Numbers AF004859 (exon1) - AF004873 (exon 15), SEQ ID NO:4) (see also GenBank Accession Number V50928) spans 68 kb of genomic DNA and is located on chromosome 4 (4q21-23). PKD2 contains 15 exons and encodes a 5.4 kb transcript (see e.g., GenBank Accession Number NM000297) from which a 968-amino acid protein product, polycystin-2 (PC-2) of approximately 110 kDa is generated (SEQ ID NO:6) (see also GenBank Accession Number NP00288). Polycystin-2 has been shown to interact with the carboxy-terminus of PC-1 and functions as a cation channel in complex with PC-1. (Gonzalez-Perrett S. et al., *Proc Natl Acad Sci* USA 98:1182-1187, 2000; Vassilev P.M. et al, *Biochem Biophys Res Commun* 282:341-350, 2001; Koulen P. et al., *Nat Cell Biol* 4:191-197, 2002; Hanaoka K. et al., *Nature* 408:990-994, 2000). Unlike PKD1, PKD2 is a single copy gene, making its analysis much more straight-forward. See Table 1 for a summary of the PKD genes. Further discussion of PKD1 and PKD2 genes, gene and protein alterations and methods of detecting the same can be found in US 2006/0246504, US 2003/0008288, WO 2002/006529, US 2005/017399, US 7,083,915, US 6,031,088, US 6,228,591, US 2007/0166755, US 2005/0100898, US 6,916,619, US 6,656,681, US 6,485,960, US 6,380,360 and WO 1995/018225.

Table 1. PKD gene description

Gene Description	***PKD1***	***PKD2***	
Chromosome	16p13.3	4q21-23	
Genomic length	54 kb	68 kb	
Exons	46	15	
Base pairs	12,909	2,904	
Codons	4,303	968	
Protein	Polycystin-1	Polycystin-2	
Analysis:			**Total**
Long Range PCRs	8	-	8
Amplicons	54	17	71
Base Pairs evaluated (including adjacent intronic sequence)	13,830	3,204	17,034

PKD gene analysis

Genomic DNA obtained from a sample from a subject can be used as the template for generating one or more PKD-specific amplification products (e.g., long-range PKD amplification products). DNA testing is advantageous as it has the potential to provide genetic information to an isolated individual (e.g., when family members are unavailable for linkage studies. Both copies of the PKD genes in an individual should be analyzed/sequenced to identify bona fide gene mutations, as mutations have been detected on a normal haplotype and/or in combination with other amino acid truncating mutations.

A sample can be a biological material which is isolated from its natural environment containing target nucleic acid (e.g., a nucleic acid comprising a PKD gene), and may consist of purified or isolated nucleic acid, or may comprise a biological sample such as a tissue sample, a biological fluid sample, or a cell sample comprising the target nucleic acid. Collecting a tissue sample also includes *in vitro* harvest of cultured human cells derived from an individual's tissue or any means of *in vivo* sampling directly from a subject, for example, by blood draw, spinal tap, tissue smear or tissue biopsy. Optionally, tissue samples can be stored before analysis by well known storage means that preserve a sample's nucleic acid(s) in an analyzable condition, such as quick freezing, or a controlled freezing regime, in the presence of a cryoprotectant, for example, dimethyl sulfoxide (DMSO), glycerol, or propanediol-sucrose. Tissue samples can also be pooled before or after storage for purposes of amplifying them for analysis. In some embodiments, the sample contains DNA, tissue or cells from two or more different individuals. In another embodiment, the amount of sample necessary to analyze a PKD gene is dependent on the type of sample (e.g., more than 5 milliliters of blood) and this amount is best assessed by one of skill in the art. Preferably, aseptic techniques are used to obtain these samples to avoid their contamination.

Methods of isolating genomic DNA from a particular sample are well known and routine (see Sambrook et al., supra, 1989). In a particular embodiment, amplification of the genomic PKD DNA has advantages over the cDNA amplification process, including, for example, the allowance of the analysis of exons and introns of the PKD gene. As such, a target sequence of interest associated with

Date Reçue/Date Received 2020-05-26

either an intron or exon sequence of a PKD gene can be amplified and characterized. A target sequence of interest is any sequence or locus of a PKD gene that contains or is thought to contain a nucleotide sequence alteration, including those alterations that correlate with a PKD-associated disorder or disease (e.g., ADPKD).

Mutations in a PKD gene can be detected by amplification, including, for example, by polymerase chain reaction (PCR), ligase chain reaction, self sustained sequence replication, a transcriptional amplification system, Q-Beta Replicase, or any other nucleic acid amplification method, followed by the detection of the amplification products. Accordingly, in one embodiment, genomic DNA extracted from whole blood serves as a template for highly specific PKD1 gene amplification by long-range amplification of 8 segments encompassing the entire *PKD1* duplicated region. The specific long-range amplification prevents the spurious amplification of *PKD1* homologs that would otherwise confound the analysis. These PKD1 homologs are sequences which are closely related to PKD1, but which do not encode an expressed PKD1 gene product. In fact, analysis of the PKD1 gene had not been amenable to genetic analysis largely because of the presence of at least three highly homologous copies of the gene that map proximal to PKD1 along chromosome 16 (16p13.1). The sequence of these PKD1 gene homologs are contained in GenBank Accession Nos. AC002039, AC010488, AC040158, AF320593 AND AF320594.

Several examples of such homologs that map to chromosomal location 16p13.1 or 4q21-23 have been identified and sequenced. A PKD1 homologue may share more than 95% sequence identity to an authentic PKD gene.

In some embodiments of the invention, a nested amplification is performed using amplified products in a preceding amplification reaction as templates. Preferably, the nested amplification reaction is a nested PCR using PCR amplified products from a preceding PCR reaction as templates. In addition to optimizing the annealing temperature of the primers, "nested" amplification can be used to increase the specificity and sensitivity of the PKD-specific amplification assay. For example, a method comprising a nested PCR can involve two sequential PCR reactions. After multiple cycles of PCR (e.g., 10 to 40, or 10 to 30 or 10 to 20 cycles) with the first pair of primers comprising at least one PKD-specific primer (e.g., a PKD-specific

Date Reçue/Date Received 2020-05-26

primer and a control primer or two PKD-specific primers), a small amount aliquot of the first reaction (e.g., 1μl of a 50μl reaction) serves as the template for a second round comprising multiple cycles of PCR reaction (e.g., 10 to 40, or 10 to 30 or 10 to 20 cycles) with a new set of primers comprising at least one PKD-specific primer (e.g., a PKD-specific primer and a control primer or two PKD-specific primers) that anneal to sequences internal to, or nested between, the first pair.

In a particular embodiment, the 8 long range PCR products described above serve as template for 43 nested PCR reactions and cover exons 1-34 of the PKD1 gene. The unique region of the PKD1 gene (exons 35-46) and the entire PKD*2* gene are amplified from genomic DNA as 28 additional gene segments. Using the nested PCR procedure, the template that is successfully amplified is selected twice for PKD-specificity. The use of nested PCR can also greatly enhance the yield of the species-specific product and, therefore, the sensitivity of the assay, when a single primer pair fails by itself.

Methods for designing primers and for performing PCR are known in the art (see Current Protocols in Molecular Biology, supra). The general criteria for selecting primers applies to primers for both the long-range PCR and nested PCR. With regard to primer for the nested PCR, both nested primers should anneal to sequences internal to (e.g., within) the first pair of primers and at least one of the nested primers. Some PKD1-specific primers which eliminate unintended amplification of PKD1 homologs have been developed (see, e.g., U.S. 2003/0008288). Other such primers can be designed, where a "PKD-specific" primer would be a nucleic acid sequence which anneals to a sequence within a PKD gene (including introns and exons) under specific stringent conditions. A PKD-specific primer, anneals to a unique site present in the authentic expressed PKD1 gene, and not to PKD1 homologs or other sequences under specific stringent conditions. Thus, PKD-specific primers can be designed using these unique PKD sites. The length of a unique site may vary from several nucleotides to thousands of nucleotides. Most of unique sites that have been identified comprises less than or equal to 100 nucleotides, e.g., less than or equal to 50 nucleotides, or less than or equal to 30 nucleotides. Amplification using PKD-specific primers increases the specificity of the amplification reaction and reduces

the amount of by-products amplified from PKD homologs. The primers may be 10 to 60 nucleotides in length, for example, 18-52 nucleotides in length.

The 71 PCR products are bi-directionally sequenced to detect nucleotide sequence alterations. In a particular embodiment, all PCR primers comprise a tag (e.g., M13 forward and reverse primer sequences) to permit bi-directional sequencing of all fragments with the same primers. Methods of sequencing DNA are well-known in the art and are dependent on the primer position and/or fragment length. For example, in one embodiment, sequencing is performed using ABI Big Dye terminator chemistry followed by electrophoresis on an ABI 3730 capillary sequencer. Nucleotide alterations of the invention can be detected in a PKD sequence to assess existing or potential ADPKD. Novel alterations identified can be clinically interpreted as disease-associated mutations, for example, frameshift or nonsense mutations or invariant splice site changes. Benign polymorphisms would include silent or conservative missense mutations, intronic variants and synonymous codon changes.

Sequence alterations in a PKD gene can also be detected using denaturing high performance liquid chromatography (DHPLC). DHPLC has been used to detect sequence variants by separating a heteroduplex (resulting from the presence of a mutation) and a homoduplex having the same basepair length. This separation is based on the fact that a heteroduplex has a lower melting temperature (T_m) than a homoduplex. DHPLC can separate heteroduplexes that differ by as little as one base pair under certain conditions. The "heteroduplex site separation temperature" or "midpoint temperature" or " T_m " is defined herein to mean, the temperature at which one or more base pairs denature, i.e., separate, at the site of base pair mismatch in a heteroduplex DNA fragment. When DHPLC is carried out at a partially denaturing temperature, i.e., a temperature sufficient to denature a heteroduplex at the site of a base pair mismatch, homoduplexes can be separated from heteroduplexes having the same base pair length and detected by various methods (e.g., gel electrophoresis). DHPLC can also be used to separate duplexes having different basepairs in length.

Evaluation of Identified PKD Nucleotide Alterations

Numerous novel nucleotide alterations in PKD have been identified (see Tables 4-7). These sequence alterations were then evaluated to determine whether

they were pathogenic, this is, resulted in an altered PKD gene product (e.g., protein, polypeptide). A "nucleotide sequence alteration" or "nucleotide alteration" or "mutation" refers to a nucleotide sequence modification including one or more substitutions (transitions or transversions), deletions (including loss of locus), insertions (including duplications), translocations, inversions and/or other modifications relative to a normal PKD gene (e.g., SEQ ID NO:1, SEQ ID NO:7 or SEQ ID NO:4). Thus, a nucleotide alteration/change in a PKD1 or PKD2 nucleotide sequence (e.g., DNA or mRNA) can be a deletion, insertion, substitution or inversion, or can be silent such that there is no change in the reading frame of a polypeptide encoded by the PKD polynucleotide. Pathogenic mutations are those nucleic acid alterations that result in an amino acid change (e.g., a non-silent or non-conservative change) and/or introduces a STOP codon into the nucleotide sequence, or changes nucleotide sequence involved in transcription or translation of the PKD1 or PKD2 nucleotide sequence; for example, a change that results in altered splicing of a PKD1 or PKD2 gene transcript into an mRNA (see Figures 7A and 7B). An "amino acid alteration" refers to an amino acid modification including a substitution, a frameshift, a deletion, a truncation and an insertion, and/or other modifications relative to the normal PKD amino acid sequence (e.g., SEQ ID NO:3 or SEQ ID NO:6). Thus, a mutation in a PKD gene sequence can result in the expression of a truncated PKD polypeptide, or even a complete loss of expression of the PKD polypeptide.

In contrast, polymorphic mutations or variants are those nucleic acid alterations that do not alter and/or are not expected to alter a PKD protein/ polypeptide in the above-described manner and/or do not correlate with the signs or symptoms of a PKD-associated disorder such as ADPKD (see Tables 8 and 9). These mutations include, for example, nucleotide substitutions that do not result in a change in the encoded amino acid, i.e., silent mutations, in which the wild type (see, e.g., SEQ ID NOs:1, 7 or 4) and mutant codons both encode the same amino acid; those that do not segregate with the disease or those that are found in a panel of unaffected individuals. Nucleic acid alterations that cause conservative amino acid substitutions in which a wild-type amino acid (see, e.g., SEQ ID NOs:3 or 6) is substituted for another amino acid with similar properties, may also be non-

Date Reçue/Date Received 2020-05-26

pathogenic polymorphic mutations, as it would be expected that the secondary structure and hydropathic nature of the PKD polypeptide would be substantially unchanged by these mutations. In general, the following groups of amino acid substitutions are thought to be conservative: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. With respect to PKD mutations, polymorphisms are then defined as: (i) sequence variants not predicted to alter an amino acid; (ii) missense changes found in homozygosity in at least one individual; (iii) intronic sequences of unknown significance; or (iv) changes in the 3' UTR of unknown significance. Accordingly, polymorphic mutations would be expected to result in a PKD protein/polypeptide that is still properly expressed and/or fully functional; that is, these variants would not be expected to be associated with ADPKD.

Nucleotide sequence alterations identified in PKD1 and PKD2 genes can be evaluated for pathogenicity in a number of ways. Mutant PKD nucleotide sequence can be compared to wild-type PKD sequence (SEQ ID NOs:1 and 4) and the effect of the nucleic acid sequence alterations on amino acid codon(s) assessed. For example, a change in nucleotide sequence that produces a stop codon (e.g., UGA, UAA, UAG) or a frameshift, which generally results in a nonsensical polypeptide and/or also produces a stop codon, or that alters a consensus donor/acceptor splice site would result in a non-functional PKD protein, a truncated PKD protein, or obliterate its expression altogether. These mutations would be expected to be pathogenic and thus correlates with ADPKD.

PKD nucleic acid sequence alterations that do not result in the production of a stop codon, frameshift or splice site mutation can also be assessed by comparing the mutant PKD amino acid sequence to the wild-type PKD amino acid sequence from various species to determine if the alteration affects an amino acid residue that is conserved across several species. In particular, an amino acid change (i.e., a missense mutation) or a deletion of several adjacent nucleotide residues (e.g., a deletion of 3, 6 or 9 nucleotides) which would cause a complete deletion of one or more amino acid residues (i.e., an in-frame deletion; see also Table 5) would result in a PKD polypeptide that is still expressed. The change or loss of an amino acid residue conserved across several species (e.g., human, canine, mouse, fish, fruit fly,

nematode, etc), where a "conserved" amino acid residue is one that is identical or has similar properties (e.g., ala, pro, gly, glu, asp, gln asn, ser, thr), would strongly indicate that the amino acid residue is important/critical to PKD protein function. Accordingly, such PKD mutations might also be expected to be associated with and/or predictive of ADPKD.

Furthermore, there are also several algorithms that can be used to predict/evaluate alterations to a PKD nucleic acid sequence, particularly those that result in a missense mutation. These algorithms include, for example, the Miller/ Kumar matrix (Miller M.P. and Kumar S., *Hum Mol Genet* 10(21):2319-2328, 2001); Grantham's chemical difference matrix;

Splice Site Prediction by

Neural Network (SSPNN) (see also Reese M.G. et al., *J Comput Biol* 4(3):311-323, 1997); Automated Splice Site Analyses

(ASSA) (see also, Nalla V.K. et al., *Hum Mutat* 25(4):334-342, 2005 and Rogan P.K. et al., *Hum Mutat* 12(3)153-171, 1998);

Prediction of Protein Sorting Signals and Localization Sites in

Amino Acid Sequences II (PSORT II) (see also Krogh A. et al., *J Mol Biol* 305:567-580, 2001); and Transmembrane Helices

Prediction (TMHMM), (see alsoGrimm D.H. et al, *J Biol Chem* 278:36786-36793, 2003). By predicting mRNA and/or

protein structure, function and motifs, these and other algorithms can help determine the likelihood that a mutation (e.g., a missense mutation) represents a pathogenic change as opposed to a polymorphism.

Further assessment of PKD mutations not clearly pathogenic could also be aided with a dataset comprising complete sequence information from a population of unaffected, ethnically diverse individuals. Normal or wild-type PKD1 and PKD2 sequence information from such a population would be a useful control for comparison to novel PKD mutations identified to both evaluate the presence or absence of a sequence variant in the control population and expand the spectrum of known non-pathogenic sequence variants. Having such a dataset to compare to PKD

mutations that have been identified would be advantageous diagnostically and prognostically, especially in the analysis of individuals having less than a 50% probability of having ADPKD (e.g., individuals not the progeny and/or siblings of an individual with ADPKD).

The effect of mutations in a PKD gene on a PKD gene product can be assessed and/or confirmed by expressing a polynucleotide having or constructed (e.g., a recombinant polynucleotide) to have the identified mutation(s). The polynucleotide can comprise the mutant PKD polypeptide or a portion of a recombinant nucleic acid molecule, which, for example, can encode a fusion PKD protein (e.g., a tagged PKD protein). The mutant polynucleotide or recombinant nucleic acid molecule can be inserted into a vector, which can be an expression vector, and can be derived from a plasmid, a virus or the like. The expression vector generally contains an origin of replication, a promoter, and one or more genes that allow phenotypic selection of transformed cells containing the vector. Expression vectors suitable for use are well-known in the art e.g., a T7-based expression vector for expression in bacteria, a pMSXND expression vector for expression in mammalian cells or baculovirus-derived vectors for expression in insect cells and the like. The choice of a vector will depend on the size of the polynucleotide sequence and the host cell to be employed. Thus, the vector used in the methods of the invention can be plasmids, phages, cosmids, phagemids, viruses (e.g., retroviruses, parainfluenzavirus, herpesviruses, reoviruses, paramyxoviruses, and the like), or selected portions thereof (e.g., coat protein, spike glycoprotein, capsid protein). For example, cosmids and phagemids are typically used where the specific nucleic acid sequence to be analyzed or modified is large because these vectors are able to stably propagate large polynucleotides. Cosmids and phagemids are particularly suited for the expression or manipulation of a PKD polynucleotide (e.g., SEQ ID NO:1) or a mutant PKD1 polynucleotide.

A variety of host-expression vector systems can be utilized to express wildtype PKD polynucleotide sequence (e.g., SEQ ID NO:1 or SEQ ID NO:4), the PKD coding sequence (e.g., SEQ ID NO:2 or SEQ ID NO:5) and a variant or mutant PKD1 or PKD2 polynucleotide. In a particular embodiment, the PKD polynucleotide(s) is tagged (e.g., FLAG, Myc, biotin, streptavadin, avadin and the

Date Reçue/Date Received 2020-05-26

like) to aid in purification and/or visualization of the PKD polypeptide after it has been exposed. Such host-expression systems represent vehicles by which the nucleotide sequences of interest can be produced and subsequently purified, and also represent cells that, when transformed or transfected with the appropriate nucleotide coding sequences, can express a PKD protein, including a PKD variant or mutant polypeptide or peptide portion thereof *in situ*. Such cells include, but are not limited to, microorganisms such as bacteria (e.g., *E. coli*, *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing a PKD1 polynucleotide, or oligonucleotide portion thereof (wild type, variant or other mutant); yeast (e.g., Saccharomyces, Pichia) transformed with recombinant yeast expression vectors containing a PKD polynucleotide, or oligonucleotide portions thereof (wild type, variant or other PKD mutant); insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing a PKD polynucleotide, or oligonucleotide portion thereof (wild type, PKD variant or other mutant); plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus or tobacco mosaic virus) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing a mutant PKD polynucleotide, or oligonucleotide portion thereof; or mammalian cell systems (e.g., HEK293, COS, CHO, BHK, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5K promoter). Further discussion of vectors and expressions systems for PKD polynucleotides can be found, for example, in US 2003/0008288.

For instance, the PKD1 gene product, polycystin-1 (PC-1), which is believed to function as a cell surface signaling receptor at cell-cell and cell-matrix junctions and as a mechano-sensor in renal cells, is an 11-transmembrane glycoprotein with a long N-terminal extracellular region and short cytoplasmic tail (Boletta A. and Germino G.G. , *Trends Cell Biol* 13(9):484-492, 2003; Harris P.C. and Torres V.E., *Curr Opin Nephrol Hypentens* 15(4):456-463, 2006; Nauli S.M. et al., *Nat Genet* 33(2):129-137, 2003; Hughes J. et al., *Nat Genet* 10(2):151-160, 1995) (see also Figure 8A). PC1 has several amino acid sequence motifs of interest (e.g., receptor

for egg jelly (REJ) domain, G-protein coupled receptor proteolytic site (GPS), C-type lectin domain, leucine rich repeat (LRR), polycystic kidney disease repeat (PKD-R), transmembrane domain (TM), coiled-coil domain (CC)) (see also FIG. 4). A site useful for evaluation of PC-1 function/activity is the GPS domain, a site at which the PC-1 protein undergoes cleavage (Qian F. et al., *Proc Natl Acad Sci USA* 99(26):16981-16986, 2002). Cleavage of PC1 at this site produces an N-terminal fragment (NTF) and a C-terminal fragment (CTF) and this cleavage is critical for normal PC-1 function (Qian F. et al., *Proc Natl Acad Sci* USA 24:99(26):16981-16986). Thus, expression and cleavage of the PKD1 gene product can be used to assess the pathogenicity of identified PKD1 mutations, particularly missense mutations. PKD1 mutants can be constructed (e.g., in an expression vector) and expressed (as, e.g., a recombinantly tagged fusion protein) in the above-described manner and the cleavage of the PKD1 mutant gene products assayed (e.g., by immunoprecipitation and/or western blot, fluoresence of a tag, radioactivity or the like).

One or more of the above-described methods to assess/evaluate PKD mutations can be used to determine whether PKD1 or PKD2 gene mutations that have been identified are benign polymorphisms or pathogenic, such that the mutations can be associated with ADPKD and, subsequently used to diagnose or predict ADPKD in, for instance, the methods of the invention.

Methods of the invention

The PKD mutations identified and determined to be pathogenic are listed in Tables 4-7. These mutations are used in the methods of the invention to detect or predict the occurrence of ADPKD in an individual or detect the presence or absence of a mutant PKD gene in an individual. Specifically, ADPKD is detected or the occurrence of ADPKD is predicted by detecting the presence of one or more of the identified nucleotide sequence alterations in a PKD1 gene having the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:7 in a nucleic acid sample obtained from an individual. Similarly, ADPKD can be detected or predicted in an individual using the methods of the invention by detecting the presence of one or more of the identified nucleotide sequence alterations in a PKD2 gene having the nucleotide

Date Reçue/Date Received 2020-05-26

sequence of SEQ ID NO:4 in a nucleic acid sample obtained from an individual. As several mutations in the PKD genes that are associated with ADPKD have been detected in just a single individual/family (see e.g., Table 7), these other nucleotide sequence alterations in a PKD1 gene (e.g., SEQ ID NO:1or 7) and/or PKD2 gene (SEQ ID NO:4) not listed above (see *Summary of Invention* and Tables 4-7) can also be detected in the methods of the invention. The methods can be performed by obtaining a sample (e.g., biological fluid, tissue, cell) from an individual by one or more procedures (e.g., DNA isolation method/kit) and/or one or more methods (e.g., sequencing, PCR, DHPLC) as described above.

In addition, the invention relates to methods of detecting the presence or absence of a mutant PKD gene in an individual by obtaining a nucleic acid sample from the individual (e.g., biological fluid, tissue or cell sample), by the above-described methods (e.g., DNA isolation method/kit) and detecting the presence or absence of one or more of the identified nucleotide sequence alterations in a PKD1 or PKD2 gene, by using one or more of the above-described processes (e.g., sequencing, PCR, DHPLC or the like). In a particular embodiment, detection of one or more of the identified PKD nucleotide sequence alterations indicates that the individual has ADPKD or may develop ADPKD.

EXEMPLIFICATION

Patient Recruitment and Clinical Evaluation

Eighty-two unrelated ADPKD patients were recruited from outpatient nephrology clinics. The Johns Hopkins Institutional Review board approved the study and informed consent was obtained from each patient. A diagnosis of ADPKD was based on established ultrasound criteria described (Ravine et al., *Lancet* 2:343(8901):824-7, 1994). A detailed medical history was obtained from each participant at the time of entry into the study. A coded blood sample was collected from each proband and sent to Athena Diagnostics, Inc. for mutation analysis. In most cases routine laboratory data were obtained as part of the standard medical evaluation.

Baseline characteristics of the study population are summarized in Table 2. The average age of the study participants was 46.5 years of age. Only 22% had

Date Reçue/Date Received 2020-05-26

reached ESRD at the time that mutation analysis was performed. The average glomerular filtration rate (GFR) for those that had not reached ESRD was 68 ml/min. Family history was either unknown or was negative for ADPKD in 34% of the patients.

Mutation Analysis

DNA sequence analysis of patient samples was performed using methods described in detail previously and optimized at Athena Diagnostics, Inc (Watnick TJ et al., *Hum Mol Genet* 6(9):1473-1481, 1997; Watnick TJ et al., *Mol Cell* 2(2):247-251, 1998; Watnick T. et al., *Am J Hum Genet* 65(6):1561-1571, 1999; Phakdeekitcharoen B. et al., *Kidney Int* 58(4):1400-1412, 2000; Phakdeekitcharoen B. et al., *J Am Soc Nephrol* 12:955-963, 2001).

For example, genomic DNA is derived from whole blood using a Puregene® DNA extraction kit (Gentra Systems, Inc. Minneapolis, MN) or other suitable extraction method. Amplified DNA product served as a template for highly specific long-range PCR amplification of the 8 segments encompassing the entire PKD1 duplicated region, to prevent the amplification of PKD1 homologs that would confound the analysis. The 8 long range PCR products served as template for 43 nested PCR reactions while the unique region of the PKD1 gene and the entire PKD2 gene were amplified from genomic DNA as 28 additional gene segments. PCR primers were tagged with M13 forward and reverse primer sequences to permit bi-directional sequencing of all fragments with the same primers.

PCR products were then bi-directionally sequenced, for example, using ABI Big Dye™ terminator chemistry (versions 3.1 and 1.1 depending upon primer position and/or fragment length) followed by electrophoresis on an ABI 3730 capillary sequencer (Applera Corporation, Norwalk, CT). This process provides sequence data for the entire coding region of the PKD1 and PKD2 genes including the highly conserved exon-intron splice junctions.

Analysis of normal samples

A normal population was selected from anonymized samples, older than 65, submitted to Athena Diagnostics, Inc for ataxia testing. PCR products from a minimum of 171 individuals were sequenced to determine the frequency of certain

common variants in either PKD1 or PKD2. Complete DNA analysis was not performed for these samples.

Generation of PC-1 variant constructs for cleavage testing

Missense variants were generated, for example using the QuickChange™ Site-Directed Mutagenesis Kit (Stratagene). The full-length wild type PKD1 cDNA construct and three of the constructs have been previously described (Q3016R, F3064L, F2853S) (Hanaoka K. et al., *Nature* 408:990-994, 2000; Qian F. et al., *Proc Natl Acad Sci* USA 24:99(26):16981-16986, 2002).

Cleavage Assay

Constructs were transfected into HEK293 cells using Lipofectamine Plus™ (Life Technologies, Rockville, MD). After transfection, the cells were lysed in buffer [20 mM sodium phosphate, pH 7.2, 150 mM NaCl, 1 mM EDTA, 10% (vol/vol) glycerol, 0.5% Triton™ X-100] for 1 hr on ice in the presence of protease inhibitor (Roche Molecular Biochemicals). The cell lysates were immunoprecipitated (IP) using ANTI-FLAG® M2 beads Affinity Gel Freezer-Safe (SIGMA) and then resolved on a NuPAGE® 3–8% Tris-Acetate Gel (Invitrogen). The IP products were electro-blotted onto an Immobilon™ transfer membrane (MILLIPORE) and probed with α-Leucine-rich-repeat (LRR) and α-C-terminus (CT) antibodies for PC1. These antibodies have been previously described (Boletta A. et al., *Mol Cell* 6:1267-1273, 2000; Qian F. et al., *Proc Natl Acad Sci* USA 24:99(26):16981-16986, 2002).

Results

DNA sequence variance analysis identified three categories of variants. Class I variants were defined as those having definitive pathogenic sequence variants, including stop codons, frameshift and splice site alterations, that are diagnostic without additional information (Tables 3, 4). Class II variants included those demonstrating in-frame deletions or amino acid substitutions determined likely to be pathogenic based on various algorithms, as described in detail below. Class III variants included those where no pathogenic changes were confirmed.

Class I variants

Forty-two percent (N=34) of the study population had stop codons, frameshift or splice site alterations (Tables 3, 4). Twenty-four of these alterations occurred in PKD1 (29% of total sample) and 10 in PKD2 (12% of total sample). The mutations found in Class I variants were expected to result in premature truncation of a PKD1 or PKD2 protein and therefore segregate with ADPKD.

Class II variants

Thirty participants had either an in-frame deletion or at least one amino acid substitution deemed likely to be pathogenic (Tables 5 and 6). A total of 8 unique in-frame deletions (6 in the PKD1 gene and 2 in PKD2 gene) were detected (Table 5). In each case, the deletion affected one or more residues fully or highly conserved between *Fugu rubripes* (Fugu fish) and *Mus musculus* (mouse) polycystin proteins.

There were 10 individuals with no other truncating PKD mutations who had unique intronic variants. Two of the predicted splice site mutations did not directly affect a consensus splice donor/acceptor site; JHU573 and JHU595 had an intronic change at the 5th base pair from the intron 24 splice donor site (IVS24+5 G>C) that affected a residue that is highly conserved as a guanine in 84% of donor splice sites. Both the Neural Network Splice Site prediction program (SSPN) and Automated Splice Site Analyses (ASSA) predicted that these variants resulted in improper splicing, as such an alteration would severely disrupt the architecture of the splice donor site at the exon 24/intron 24 boundary. JHU105 had a similar alteration (IVS8+5,G>A) at the 5th basepair from the end of PKD2 exon 8 splice donor site (i.e., the 5th nucleotide base counted from left to right after nucleotide residue 1964 of SEQ ID NO:5 into the following intron (intron 8)), in which the highly conserved guanine residue was replaced by an adenine. In addition, IVS37-10C>A (JHU 604), was previously reported to segregate with ADPKD in a European family (Bogdanova, M. et al., *Hum Mutat* 16(2):166-174, 2000). JHU562 also had a PKD2 pathogenic mutation that affected a splice site, IVS7-1G>A (i.e., a change from a guanine to an adenine at the 1st nucleotide residue counted right to left from the beginning of exon 8 (e.g., nucleotide residue 1783 of SEQ ID NO:5) into the previous intron (intron 7)), which resulted in the loss of the acceptor site for exon 7.

Date Reçue/Date Received 2020-05-26

Most of the remaining participants had a combination of amino acid substitutions, primarily in PKD1. Three major criteria were used to judge the pathogenicity of each missense variant. Conservation of the altered residue between human polycystin-1 and Fugu fish and mouse proteins was examined. Amino acids that were considered "fully conserved" were those that were identical in all three species, while amino acids with similar properties (i.e. belonging to the same class) were deemed to be "highly conserved" residues. In addition, a pathogenicity score for each missense variant was assigned using the matrix of Miller and Kumar (Miller M.P. and Kumar S., *Hum Mol Genet* 12(21):2319-2328, 2001), which defines the relative likelihood that a missense change represents a pathogenic alteration versus a polymorphism. This algorithm was developed by using interspecies sequence comparisons coupled with Grantham's chemical difference matrix to determine the common attributes of amino acid replacement mutations across 7 disease genes (including tuberous sclerosis and cystic fibrosis). Other investigators have used this strategy to assist in characterizing amino acid substitutions (Sharp A.M. et al., *J Med Genet* 42(4):336-349, 2005). Finally, literature was reviewed to determine whether any of the variants had been reported by others to occur in unaffected individuals. Several amino acid substitutions (N=13, Table 9), detected in homozygosity in one or more individuals, were classified as polymorphisms. Since germ line ADPKD mutations are heterozygous, one of these changes would have to be associated with a wild type allele, presumably inherited from an unaffected parent.

Analysis of individual amino acid substitutions, grouped by patient, is summarized in Table 6. An amino acid substitution was deemed to be pathogenic, if it occurred at a fully or highly conserved amino acid residue and if it was also predicted to have a higher pathogenic potential using the matrix of Miller and Kumar (Table 6, shaded in Gray). Using these strict criteria, 24 of 30 patients had one or more pathogenic amino acid substitutions. Six of these missense changes were predicted to disrupt structural determinants of either the C-type lectin (Y420C, Y528C) or one of the PKD repeats (S1047L, R1340W, R1351W, T1861I) (Figures 5A and 5B). Three of the missense changes (Q3016R, E2771K, F2853S) were previously shown to disrupt polycystin-1 cleavage, a property that is critical for

Date Reçue/Date Received 2020-05-26

normal polycystin-1 function (see Figure 6) (Qian F. et al., *Proc Natl Acad Sci* USA 24:99(26):16981-16986, 2002).

Recurrent PKD1 variants (R2200C, Q739R, G2814R, Q2182R, G2309R, R1340W) that met the criteria for pathogenicity were observed in 7 individuals and were also present in other individuals who harbored either chain terminating mutations or other predicted pathogenic amino acid substitutions (Tables 4, 6 and 7). For example, R2200C was present in 4 patients, JHU584, JHU606, JHU111 and JHU573. The latter two individuals had a PKD1 frame shift mutation and a splice site mutation, respectively. This association suggested that these changes represented polymorphisms. To further characterize the missense mutations, 342 normal chromosomes were sequenced to identify polymorphisms and the R2200C sequence alteration was seen in a small (1.4%) fraction but greater than the polymorphism threshold of 1%. Likewise Q739R (this study 6.4%) and G2814R (Rossetti et al., 0.9%) have also been reported in a small percentage of the unaffected population and are or may be polymorphisms, respectively (Thomas T. et al., *Am J Hum Genet* 65(1):39-49, 1999 and Rossetti S. et al., *Kidney Int* 61(5):1588-1599, 2002).

If patients with only these pathogenic recurrent variants (without additional chain terminating mutations or other pathogenic amino acid substitution) were eliminated, then approximately 21% of the sample (N=17/82 patients) would be predicted to harbor a pathogenic PKD1 missense mutation.

Five participants JHU 602 (*N*=2), JHU100 (*N*=3), JHU588 (*N*=2), JHU411 (*N*=2), JHU114 (*N*=2) had more than one PKD1 amino acid variant that met the criteria for pathogenicity. This observation raises the possibility that a combination of missense changes in cis might cooperatively result in a diminished level of functional PKD1 protein (Reiterova J. et al., *Hum Mutat* 19(5):573, 2002).

In contrast with PKD1, only two PKD2 amino acid substitutions were detected among the 37 patients lacking chain-terminating mutations. One change (M800L in JHU559, Table 6), was not considered pathogenic by the criteria of the present system and did not segrate with disease in a PKD2 family. A second PKD2 substitution, A190T, was found in 3 patients and, likewise, did not meet the criteria for pathogenicity as it was identified in 3.2% of normal chromosomes (Table 6).

Date Reçue/Date Received 2020-05-26

In assessing Class II variants, detection of in-frame deletions was a useful predictor of pathogenicity. Also amino acid substitutions resulting in loss of polycystin-1 cleavage were predictive of pathogenicity.

Class I and Class II amino acid changes in the PKD-1 protein (polycystin-1) and PKD-2 protein (polycystin-2) are depicted in a schematic in Figure 8.

Class III variants

Eighteen subjects in the study lacked definitive pathogenic sequence alterations (Tables 6 and 9). Of these, 9 had clear and extensive family history of polycystic kidney disease (Table 9). The other 9 had enlarged kidneys with cysts, with 4 of these individuals suffering from significant renal dysfunction (GFR < 40) at the time of DNA testing.

Failure to detect pathogenic or potentially pathogenic changes in a subset of individuals with polycystic kidney disease may be due to several reasons. Mutational events in individuals with Class III tests could involve introns or other regulatory regions that were not assayed by the methodology that was used. Direct sequencing might also miss deletions or duplications, which would appear as an area of homozygous normal sequence. Alternatively, the stringent criteria used may have identified some missense changes as benign when they are in fact pathogenic. For example, JHU617, with an extensive family history of ADPKD, was found to have a unique leucine to valine chage in PKD repeat 4 that was judged more likely to be a polymorphism by the matrix of Miller/Kumar. Nevertheless, this change does disrupt the structure of PKD repeat 4 and could be pathogenic (see Figure 5A and 5B). In addition, as reported by Reynolds, missense variants may unexpectedly activate cryptic splice sites, thereby reducing the level of normal transcript (Reynolds D.M. et al., *J Am Soc Nephrol* 10(11):2342-2351, 1999).

Functional analysis of missense changes

To confirm that a subset of PKD1 amino acid substitutions predicted to be pathogenic disrupted the functional properties of the protein, full-length mutant constructs were generated and transiently expressed in HEK293 cells. Figure 6 demonstrates that E2771K, Q3016R and F2853S disrupt cleavage, as do three additional missense changes, R2643C, R2767C and L2619P.

Date Reçue/Date Received 2020-05-26

Polymorphism and variability in PKD genes

In addition to the sequence alterations described in Tables 4-7, a large number of polymorphisms were detected (Table 9) (see also Figures 7A and 7B). Polymorphisms are defined as: (i) sequence variants not predicted to alter an amino acid; (ii) missense changes found in homozygosity in at least one patient; (iii) intronic sequences of unknown significance; or (iv) changes in the 3' UTR of unknown significance.

Further discussion of the above example can be found in M.A. Garcia-Gonzalez et al., Evaluating the clinical utility of a molecular genetic test for polycystic kidney disease, Mol Genet Metab, 2007; 92(1-2); 160-167.

While this invention has been particularly shown and described with references to example embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein.

Date Reçue/Date Received 2020-05-26

Table 2. Cohort characteristics. *N=82 subjects. [€] ESRD defined as transplant, dialysis or MDRD GFR <10 ml/minute. [¥]N = 80 patients.

% Female*	50%
Average Age at time of Test*	46.5 (range 1-73y)
% ESRD[€, *]	20.7%
Average GFR (ml/min)[¥]	68.7 (range 14-126)
% Liver cysts*	74.3%
% Vascular complications*	9.8%
% Unknown or no Family history*	30.5%

5

Table 3. PKD mutations definitively pathogenic.

Gene	Truncation and Splicing			*Total %*
	Stop Codon	**Frameshift**	**Splicing**	
PKD1	8 (9.8%)	14 (17.1%)	4 (4.9%)	31.7%
PKD2	6 (7.3%)	3 (3.7%)	2 (2.4%)	13.4%
Total %	17.1%	20.8%	7.3%	45.1%

Date Reçue/Date Received 2020-05-26

Table 4. Truncating and Splice site mutations. Leucine Rich Repeat (LRR), Polycystic Kidney Disease Repeat (PKD-R), Receptor for Egg Jelly domain (REJ), Transmembrane (TM), Coiled Coil (CC), Novel change (N). See *Full Reference List* for mutation references.

ID	Mutation		Mutation Effect		Exon	Domain	Rate	Ref.
	cDNA	Protein	Stop Codon	Splice Site				
PKD1 gene								
Frameshift:								
JHU111	559delTTTAA	N116fsX	117		3	LRR2	1/164	N.
JHU568	1124insCT	W305 fsX	334		5	PKDR1	1/164	N.
JHU582	2291ins1	P694 fsX	713		10		1/164	N.
JHU585	2297ins1	A696 fsX	713		10		1/164	N.
JHU15	5225delAG	R1672 fsX	1721		15	PKDR11	1/164	1, 8, 23
JHU508	5365insT	V1718 fsX	1770		15	PKDR12	1/164	N.
JHU613	6666insG	D2152 fsX	2174		15	REJ	1/164	N.
JHU611	6881insA	P2224 fsX	2261		15	REJ	1/164	N.
JHU577	8713delT	F2834 fsX	2874		23		1/164	N.
JHU600	9134ins1	P2975 fsX	3068		24		1/164	N.
JHU579	9536ins5	I3109 fsX	3317		26		2/164	N.
JHU609	9536ins5	I3109 fsX	3317		26		2/164	N.
JHU599	10239delT	L3343 fsX	3395		30	TM3	1/164	N.
JHU104	11587delG	G3792 fsX	3824		40		1/164	26
Nonsense:								
JHU605	483 C>A	S91X	91		2	LRR1	1/164	N.
JHU567	4517 C>T	R1436X	1436		15	PKDR8	1/164	N.
JHU108	7006 C>A	Y2265X	2265		15	REJ	1/164	N.
JHU563	7499 C>T	R2430X	2430		18	REJ	1/164	2, 3
JHU593	7877 C>T	Q2556X	2556		19	REJ	1/164	N.

Date Recue/Date Received 2020-05-20

JHU083	8267 C>T	Q2686X	2686		22	REJ	1/164	N.
JHU574	8639 G>T	E2810X	2810		23	REJ	1/164	N.
JHU620	12243 G>A	W4011X	4011		44	TM9	1/164	N.
Splicing:								
JHU572		IVS4+1G>A.		Loss of donor site	4		1/164	N.
JHU580		IVS19+1G>T.		Loss of donor site	19	REJ	1/164	N.
JHU573		IVS24+5G>C.		Loss of donor site	24		2/164	N.
JHU595		IVS24+5G>C.		Loss of donor site	24		2/164	N.
***PKD2* gene**								
Frameshift:								
JHU586	2226insA	720fsX			11		2/164	N.
JHU116	2226insA	720fsX			11		2/164	N.
JHU591	2422delAG	786fsX			12	CC	1/164	N.
Nonsense:								
JHU578	982 C>T	R306X			4	TM1	2/164	5
JHU583	982 C>T	R306X			4	TM1	2/164	5
JHU607	2224 C>T	R742X			11		1/164	6
JHU594	2680C>T	R872X			14		3/164	N.
JHU566	2680C>T	R872X			14		3/164	N.
JHU608	2680C>T	R872X			14		3/164	N.
Splicing:								
JHU562		IVS7-1G>A		Loss of acceptor site	7		1/164	N.
JHU105[L2]		IVS8+5G>A		Loss of donor site	8		1/164	N.

Table 5. In-Frame Deletions. Leucine rich repeat-2 (LRR2), polycystic kidney disease repeat (PKD-R), receptor for egg jelly domain (REJ), Transmembrane (TM), coiled coil (CC), Novel change (N),* Disrupts the Consensus sequence for the Domain.

ID	Mutation		Exon	Domain	Conservation			Variant	Ref.
	cDNA	Protein			Fugu	Mouse	Level		
PKD1 gene									
JHU115	514-551delCAA	N101del	3	LRR2	N	N	Fully	1/164	N.
JHU107[L1]	1848-1851delTGG	V546del	8		V	V	Fully	1/164	N.
JHU560	8892-8898delCCAACTCCG	ANS2894del	23		AGA	VGS	Highly	1/164	N.
JHU592	9905-9909delAAG	K3232del	28	PLAT	I	K	Highly	1/164	N.
JHU571	10070-10074delCTC	L3287del	29	TM2	L	L	Fully	1/164	N.
JHU112	12597-12600delTGG	V4129del	45		V	V	Fully	1/164	N.
PKD2 gene									
JHU 596	374-378delTGG	V103del	1	Poly-Glu	-	V	Highly	1/164	N.
JHU416[L2]	1879-1882delTTC	F605del	8	TM5	-	F	Highly	1/164	N.

5

Date Reçue/Date Received 2020-05-26

Table 6. Families with One or More Amino Acid Changes. Families underlined are those with one or more amino acid change that meets criteria of pathogenicity (grey shadow) and not found in patients with definitive pathogenic sequence variants. Fp = do not disrupt cleavage; Fm = disrupt cleavege. See *Full Reference List* for mutation references.

ID	Mutation			Exon	Domain	Graham	Conservation			Rate	Ref.	Total # of variants.	
	Gene	cDNA	Protein				Fugu	Mouse	Level			PKD 1	PKD 2
JHU612	*PKD1*	1023C>A	A271D	5	WSC	Path.H.	A	A	Fully	1/164	N.	4	0
	PKD1	385G>A	A92T	2		Equal	F	A	Highly	1/164	N.		
JHU602	*PKD1*	1470A>G	Y420C	6	C-LECT*	Path.H.	F	Y	Highly	1/164	N.	25	1
	PKD1	4262C>T	R1351W	15	PKDR7*	Path.H.	R	R	Fully	1/164	N.		
	PKD1	8855T>A	W2882R	23		Path.H.	G	Q	No	1/164	N.		
	PKD1	9109G>C	E2966D^{Fp}	24		Poly. H.	G	E	Highly	1/164	10, 3, 24, 27		
JHU103	*PKD1*	1794A>G	Y528C	7	C-LECT*	Path.H.	Y	Y	Fully	1/164	N.	28	1
	PKD1	6036G>A	R1942H	15	PKDR14	Equal	R	R	Fully	1/164	N.		
JHU001	*PKD1*	2042C>T	R611W	9		Path.H.	R	R	Fully	1/164	N.	6	0
	PKD1	8651G>A	G2814R	23	REJ	Path.H.	A	G	Highly	6/164	8, 9		
JHU411	*PKD1*	3351C>T	S1047L	13	PKDR4*	Path.H.	M	S	Highly	1/164	N.	60	1
	PKD1	6756A>G	Q2182R	15	REJ	Path.H.	G	Q	Highly	2/164	N.		
	PKD2	634G>A	A190T	1		Equal	-	A	Highly	3/164	N.		
JHU100	*PKD1*	5793C>T	T1861I	15	PKDR13*	Path.H.	T	S	Highly	1/164	N.	8	1
	PKD1	6707C>T	R2166C	15	REJ	Path.H.	P	R	Highly	1/164	N.		
	PKD1	4229C>T	R1340W	15	PKDR6*	Path.H.	H	H	Fully	3/164	8		
JHU564	*PKD1*	10187G>C	G3326R	30	TM3	Path.H.	G	G	Fully	1/164	N.	4	1
	PKD1	7116C>G	A2302G	15	REJ	Equal	S	A	Highly	1/164	N.		
	PKD1	10311A>G	I3367V	31		Poly. H.	I	V	Highly	1/164	N.		
JHU588	*PKD1*	7554T>C	L2448P	18	REJ	Path.H.	L	L	Fully	1/164	N.	39	0
	PKD1	4229C>T	R1340W	15	PKDR6*	Path.H.	H	H	Highly	3/164	8		
JHU603	*PKD1*	7757C>T	R2516C	19	REJ	Path.H.	R	R	Fully	1/164	N.	4	0

Date Recue/Date Received 2020-05-20

JHU569	PKD1	8067T>C	L2619P	20	REJ	Path.H.	L	L	Fully	1/164	N.	22	2
	PKD1	8411C>A	P2734T	23	REJ	Equal	P	P	Fully	1/164	3		
	PKD1	8415A>T	Q2735L	23	REJ	Equal	S	Q	Highly	1/164	3		
JHU597	PKD1	8138C>T	R2643C	21	REJ	Path.H.	R	R	Fully	1/164	N.	3	1
JHU101	PKD1	8509C>T	R2767C	23	REJ	Path.H.	R	R	Fully	1/164	N.	4	2
JHU109	PKD1	8522G>A	E2771K[Fm]	23	REJ	Path.H.	E	E	Fully	1/164	3, 24, 23	3	0
JHU589	PKD1	8769T>C	F2853S[Fm]	23		Path.H.	F	F	Fully	1/164	4, 24	22	2
JHU 576	PKD1	10096C>A	N3295K	29	TM2	Path.H.	N	N	Fully	1/164	N.	3	0
JHU114	PKD1	12658C>T	R4149C	46		Path.H.	R	R	Fully	1/164	N.	22	1
	PKD1	4229C>T	R1340W	15	PKDR6*	Path.H.	H	H	Highly	3/164	8		
JHU601 B	PKD1	9258A>G	Q3016R[Fm]	25	GPS	Path.H.	Q	Q	Fully	1/164	4, 3, 27, 24	43	1
	PKD1	2427A>G	Q739R	11		Path.H.	R	Q	Highly	11/164	21		
JHU565	PKD1	7476C>A	T2422K	18	REJ	Equal	T	T	Fully	1/164	N.	24	0
	PKD1	3527C>G	L1106V	15	PKDR4	Poly. H.	S	V	No	1/164	N.		
	PKD1	3713C>T	P1168S	15	PKDR5	Equal	-	P	Highly	2/164	8		
JHU570	PKD1	1947C>A	P579Q	9		Equal	P	P	Fully	1/164	N.	5	1
	PKD1	2427A>G	Q739R	11		Path.H.	R	Q	Highly	11/164	21		
JHU575	PKD1	3312A>G	N1034S	13	PKD R4	Poly. H.	G	S	No	1/164	N.	17	1
JHU178	PKD1	3713C>T	P1168S	15	PKDR5	Equal	-	P	Highly	2/164	8	2	1
	PKD2	634G>A	A190T	1		Equal	-	A	Highly	3/164	N.		
JHU610	PKD2	634G>A	A190T	1		Equal	-	A	Highly	3/164	N.	40	2
JHU617	PKD1	4391C>G	L1394V	15	PKDR8*	Poly. H.	V	L	Highly	1/164	N.	5	1
	PKD1	11040T>A	L3730Q	39		Equal	F	L	Highly	1/164	N.		
JHU587	PKD1	840G>T	C210F	5		Equal	C	C	Fully	1/164	N.	6	2
	PKD1	7197G>A	R2329Q	16	REJ	Equal	E	R	Highly	1/164	N.		
	PKD1	2427A>G	Q739R	11		Path.H.	R	Q	Highly	11/164	21		
JHU559	PKD1	351G>C	C47S	1	LRR-N	Poly. H.	W	C	Highly	1/164	N.	24	3
	PKD2	2464A>C	M800L	13		Poly. H.	-	M	Highly	1/164	11		
JHU606	PKD1	6809C>T	R2200C	15	REJ	Path.H.	R	R	Fully	4/164	23	5	2
JHU584	PKD1	6809C>T	R2200C	15	REJ	Path.H.	R	R	Fully	4/164	23	20	1

JHU106	PKD1	3651G>A	G28KR	23	RJ	Fail.H.	A	G	Highly	6/164	3.8	4	1
JHU614	PKD1	4757G>A	A1516T	15	PKDR9	Equal	T	I	No	2/164	N.	10	0
	PKD1	1973A>C	E586D	9		Equal	A	E	Highly	1/164	N.		
	PKD1	2427A>G	Q739R	11		Fail.H.	R	Q	Highly	1/164	23		

Table 7. Families with multiple PKD mutations associated with ADPKD. Occasionally, families with a mutation associated to the disease had other change that could be classified also as associated to the disease by meeting our criteria or disrupting the consensus sequence of the Domain*.

Pedigree	Mutations Disease Associated		Amino acid changes highly pathogenic		# of Changes per patient	
	PKD1	*PKD2*	*PKD1*	*PKD2*	*PKD1*	*PKD2*
JHU605	S91X				4	0
JHU 567	R1436X				24	0
JHU108	Y2265X		Q739R		5	1
JHU563	R2430X		Q739R	R807Q	19	1
JHU593	Q2556X				3	2
JHU083	Q2686X				5	0
JHU574	E2810X		G2814R		4	0
JHU620	W4011X				22	0
JHU568	W305 fsX		W305C*		4	1
JHU582	P694 fsX				1	0
JHU585	A696 fsX				6	2
JHU508	V1718 fsX				5	2
JHU613	D2152 fsX		E624K		21	0
JHU611	P2224 fsX				29	1
JHU600	P2975 fsX		Q739R		25	2
JHU609	I3109 fsX				41	1
JHU579	I3109 fsX		G2814R		23	2
JHU577	F2834fsX		Q739R		4	1
JHU111	N116 fsX		R2200C S1619F		7	1
JHU15	R1672 fsX				5	1
JHU599	L3343 fsX		*R1312Q*		20	1
JHU104	G3792 fsX				4	2
JHU115	N101del				20	0

Date Reçue/Date Received 2020-05-26

JHU107	V546del		R1142W		25	1
JHU560	ANS2894del				21	1
JHU592	K3232del				3	1
JHU571	L3287del				10	0
JHU112	V4129del		S4053F		19	2
JHU580	IVS19+1G>T.		G2814R		5	1
JHU573	IVS24+5G>C		R2200C		5	0
JHU595	IVS24+5G>C				16	2
JHU572	IVS4+1G>A				17	1
JHU578		R306X			3	3
JHU583		R306X			5	1
JHU607		R742X			21	1
JHU594		R872X			22	3
JHU566		R872X			1	3
JHU608		R872X	G2814R		4	2
JHU596		V103del	Q2182R		35	1
JHU416		F605del			2	3
JHU591		786fsX			4	2
JHU562		IVS7-1G>A			3	2
JHU105		IVS8+5G>A	T1773I*		3	2
JHU116		720 fsX	Q739R		3	2
JHU586		720 fsX	T1773I*		22	1

Table 8. Families without disease-associated PKD mutations. *disrupts the consensus sequence. ^^predicted to generate a new splice site.

ID	Non-pathogenic missense		Intronic Changes		Family history	# of Changes	
	PKD1	PKD2	PKD1	PKD2		*PKD1*	*PKD2*
JHU565	L1106V P1168S T2422K				Yes	24	0
JHU570	Q739R P579Q				Yes	5	1
JHU575	N1034S				No	17	1
JHU178	P1168S	A190T			Yes	2	1
JHU610		A190T			No	40	2
JHU617	L1394V* L3730Q				Yes	5	1
JHU587	C210F Q739R R2329Q				No	6	2
JHU559	C47S	M800L			Yes	24	3
JHU604	Q739R		IVS37-10C>A^^		Yes	2	0
JHU606	R2200C				No	5	2
JHU584	R2200C				No	20	1
JHU590			IVS24+28G>T^^		Yes	3	1
JHU106	G2814R				Yes	4	1
JHU614	E586D Q739R A1516T				No	10	0
JHU102[L1]					Yes	21	0
JHU616					Yes	17	0
JHU615					No	0	1
JHU110[L3]					Yes	3	0
JHU113					No	2	1
JHU598					No	19	0

Date Reçue/Date Received 2020-05-26

Table 9. Polymorphisms Identified. See *Full Reference List* for mutation references.

ID#	Designation	cDNA Change (s)	Location	Domain	Frequency	Ref.
PKD1 Polymorphisms.						
-	T263S(H)	1004C>T	Exon 5		2/164	N.
-	P572S(H)	1925C>T	Exon 8		4/164	8.
-	M1092T(H)	3486T>C	Exon 14	PKD R4	30/164	8
-	W1399R(H)	4406T>G	Exon 15	PKD R8	22/164	1, 8, 16
-	V1943I(H)	6038G>A	Exon 15	PKD R14	5/164	8
-	E2548Q(H)	7853G>C	Exon 19	REJ	4/164	1
-	H2638R(H)	8124A>G	Exon 21	REJ	32/164	1
-	P2674S(H)	8231C>T	Exon 21	REJ	2/164	3, 8
-	F3066L (H)	9407T>C	Exon 25		38/164	3, 17, 34
-	V3408L(H)	10433G>C	Exon 33		5/164	N.
-	A3511V(H)	10743C>T	Exon 35		13/164	3, 8
-	I4044V(H)	12341A>G	Exon 44	TM10	42/164	3, 8, 17, 18, 14,10
-	A4058V(H)	12386C>T	Exon 45		12/164	8, 10
1		104C>T	Exon 1	5'UTR	1/164	N.
2		145C>T	Exon 1	5'UTR	2/164	N.
3		160C>T	Exon 1	5'UTR	1/164	N.
4		210C>T	Exon 1	5'UTR	1/164	N.
5	L72L	425C>T	Exon 1	LRR1	2/164	N.
6	G109G	538A>T	Exon 3	LRR2	1/164	N.
7	IVS4+1G>A(H)		Intron 4		1/164	N.
8	S196S	799C>T	Exon 5		2/164	N.
9	A341A	1234C>T	Exon 5	PKD R1	5/164	3

10	L373L(H)	1330T>C	Exon 5		36/164	3, 8, 15
11	G441G	1534G>A	Exon 6	C-LECT	1/164	N.
12	H570H	1921C>T	Exon 8		1/164	3, 8
13	IVS9+2del7		Intron 9		12/164	N.
14	IVS9+2 T>A		Intron 9		1/164	N.
15	IVS9+28del7 (H)		Intron 9		4/164	8
16	ISV9-44G>C		Intron 9		1/164	8
17	IVS9-4A>G		Intron 9		42/164	8
18	IVS10-4 G>A		Intron 10		1/164	N.
19	P738P(H)	2425C>G	Exon 11		4/164	N.
20	A745A	2448C>G	Exon 11		1/164	N.
21	A898A	2905A>C	Exon 11	PKD R2	4/164	8, 9
22	P900P	2911G>A	Exon 11	PKD R2	10/164	8, 16, 9
23	D910D	2941C>T	Exon 11	PKD R2	10/164	8, 16, 9
24	IVS11-5C>T		Intron 11		2/164	8
25	IVS11+23C>T(H)		Intron 11		4/164	N.
26	IVS12-15C>T		Intron 12		5/164	N.
27	G1021G(H)	3274T>C	Exon 13	PKD R4	30/164	8, 16, 9
28	L1037L	3392A>G	Exon 13	PKD R4	15/164	9
29	E1061E	3394G>A	Exon 14	PKD R4	1/164	N.
30	P1076P	3439G>A	Exon 14	PKD R4	1/164	N.
31	A1124A	3583C>T	Exon 15	PKD R4	25/164	8,9
32	S1125S	3586C>T	Exon 15	PKD R5	25/164	8,9
33	F1163F	3700C>T	Exon 15	PKD R5	1/164	N.
34	T1171T	3724C>G	Exon 15	PKD R5	1/164	N.

35	D1310D	4141C>T	Exon 15	PKD R7	1/164	N.
36	L1357L	4282G>T	Exon 15	PKD R7	1/164	N.
37	S1373S	4330C>T	Exon 15	PKD R7	1/164	N.
38	S1452S	4567T>C	Exon 15	PKD R8	1/164	N.
39	P1511P	4744G>A	Exon 15	PKD R9	1/164	N.
40	A1555A(H)	4876A>C	Exon 15	Extracellular	42/164	16, 1, 9
41	T1558T	4885G>A	Exon 15	Extracellular	9/164	2
42	S1603S	5020C>T	Exon 15	Extracellular	1/164	N.
43	T1724T(H)	5383C>T	Exon 15	PKD R12	40/164	8, 9, 21
44	A1818A(H)	5665G>A	Exon 15	PKD R13	5/164	8, 9
45	G1860G	5791C>A	Exon 15	PKD R13	1/164	N.
46	A1894A	5893C>T	Exon 15	PKD R14	1/164	8, 9
47	L1921L	5974G>A	Exon 15	PKD R14	2/164	8, 9
48	V2026V	6289C>T	Exon 15	PKD R15	1/164	N.
49	R2121R	6574C>T	Exon 15	PKD R16	1/164	N.
50	T2180T	6751C>T	Exon 15	REJ	1/164	N.
51	A2202A	6817G>A	Exon 15	REJ	1/164	N.
52	V2257V	6982G>A	Exon 15	REJ	1/164	N.
53	G2309G	7138C>T	Exon 16	REJ	4/164	8, 9
54	IVS16+10 G>A		Intron 16	REJ	1/164	N.
55	R2359R	7289G>C	Exon 17	REJ	3/164	N.
56	L2389L(H)	7376T>C	Exon 17	REJ	46/164	1, 2, 8, 9.
57	G2425G	7486C>T	Exon 18	REJ	1/164	N.
58	L2481L(H)	7652C>T	Exon 18	REJ	39/164	1, 8
59	IVS19-24 C>A		Intron 19	REJ	2/164	N.

60	L2570L(H)	7919T>C	Exon 20	REJ	31/164	1, 9
61	IVS20+ C>A		Intron20	REJ	1/164	N.
62	ISV20-16C>G		Intron20	REJ	2/164	N.
63	T2708M	8334C>T	Exon 22	REJ	1/164	3, 8
64	IVS22+8G>A (H)		Intron 22	REJ	1/164	1, 8
65	S2729S	8398G>A	Exon 23	REJ	2/164	N.
66	A2749A	8458G>A	Exon 23	REJ	1/164	N.
67	S2766S	8509C>T	Exon 23	REJ	1/164	13
68	D2789D	8578C>T	Exon 23	REJ	2/164	N.
69	S2813S	8650C>T	Exon 23	REJ	2/164	3, 8, 24
70	S2893S	8890C>G	Exon 23		2/164	3
71	A2971A(H)	9124T>C	Exon 24		2/164	N.
72	IVS24-20G>A (H)		Intron 24		3/164	N.
73	IVS24-17A>G(H)		Intron 24		6/164	N.
74	IVS24+17A>G		Intron 24		32/164	N.
75	S3007S	9232C>T	Exon 25		1/164	N.
76	V3065V(H)	9406G>C	Exon 25		38/164	24
77	V3090V	9481C>T	Exon 26	TM1	3/164	N.
78	P3110P(H)	9543T>C	Exon 26		37/164	6
79	IVS26+76C>A		Intron26		1/164	N.
80	IVS27-13T>C(H)		Intron27		15/164	8
81	T3223T	9880G>A	Exon 28	PLAT	2/164	6, 3, 8
82	S3265S	10006C>T	Exon 29		1/164	N.
83	IVS29-4 C>T		Intron29		1/164	N.
84	A3455A	10576C>T	Exon 34		1/164	N.

85	L3589L	10976C>T	Exon 36	TM5	5/164	N.
86	IVS37-4C>T		Intron 37		1/164	N.
87	IVS38+11G>A		Intron 38		4/164	N.
88	R3752R	11385C>A	Exon 39	Polycystin motif	1/164	N.
89	L3753L	11465G>C	Exon 39	Polycystin motif	1/164	N.
90	IVS39-25del72bp		Intron 39		1/164	7, 3
91	IVS41+C>T		Intron 41		1/164	N.
92	IVS41+5insGGG		Intron 41		2/164	8
93	IVS41-11C>T	C>T	Intron 41		2/164	N.
94	S3893S(H)	11890C>T	Exon 42		3/164	8
95	IVS43+42C>A		Intron 43		6/164	N.
96	R3971R	12124C>T	Exon 43		3/164	N.
97	L4025L	12286C>T	Exon 44		1/164	N.
98	L4035L	12316C>T	Exon 44	TM10	1/164	N.
99	IVS44+22delG		Intron44		4/164	N.
100	L4089L	12478C>G	Exon 45	TM11	1/164	N.
101	A4091A(H)	12484A>G	Exon 45	TM11	43/164	8, 3, 17, 18, 7
102	L4136L(H)	12617C>T	Exon 45		13/164	8, 14
103	V4152V	12667C>T	Exon 46		2/164	N.
104	P4161P	12696C>A	Exon 46		1/164	N.
105	S4189S	12778C>T	Exon 46		1/164	6
106	P4209P(H)	12838T>C	Exon 46		40/164	8, 6, 3
107	L4221L	12874C>T	Exon 46	COILED COIL	1/164	N.
108	A4255A	12978C>T	Exon 46		1/164	N.
109		13135G>A	3'UTR		2/164	8

Date Reçue/Date Received 2020-05-26

PKD2 Polymorphisms.						
-	R28P(H)	149C>T	Exon 1		50/164	8, 10, 22
110	R60R	246G>A	Exon 1		1/164	N.
111	G140G(H)	486G>A	Exon 1		22/164	N.
112	IVS6-4C>T		Intron 6		1/164	N.
113	L539L	1683G>C	Exon 7		1/164	N.

Full Reference List

1. Watnick T, Phakdeekitcharoen B, Johnson A, *et al.* Mutation detection of PKD1 identifies a novel mutation common to three families with aneurysms and/or very-early-onset disease. *Am J Hum Genet 65(6):1561-71, 1999.*

2. Phakdeekitcharoen B, Watnick TJ, Ahn C, *et al.* Thirteen novel mutations of the replicated region of PKD1 in an Asian population. *Kidney Int 58(4):1400-12, 2000.*

3. Rossetti S, Strmecki L, Gamble V, *et al.* Mutation analysis of the entire PKD1 gene: genetic and diagnostic implications. *Am J Hum Genet 68(1):46-63, 2001.*

4. Peral B, Gamble V, Strong C, *et al.* Identification of mutations in the duplicated region of the polycystic kidney disease 1 gene (PKD1) by a novel approach. *Am J Hum Genet 60(6):1399-410, 1997.*

5. Veldhuisen B, Saris JJ, de Haij S, *et al.* A spectrum of mutations in the second gene for autosomal dominant polycystic kidney disease (PKD2). *Am J Hum Genet 61(3):547-55, 1997.*

6. Peral B, Ong AC, San Millan JL, *et al.* A stable, nonsense mutation associated with a case of infantile onset polycystic kidney disease 1 (PKD1). *Hum Mol Genet 5(4):539-42, 1996.*

7. Peral B, San Millan JL, Ong A, *et al.* Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene reveals six novel mutations. *Am J Hum Genet 58(1):86-96, 1996.*

8. Rossetti S, Chauveau D, Walker D, *et al.* A complete mutation screen of the ADPKD genes by DHPLC. *Kidney Int 61, 1588–1599, 2002.*

9. Thomas R, McConnell R, Whittacker J, *et al.* Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR. *Am J Hum Genet 65(1):39-49, 1999.*

10. Rossetti S, Bresin E, Restagno G, *et al.* Autosomal dominant polycystic kidney disease (ADPKD) in an Italian family carrying a novel nonsense mutation and two missense changes in exons 44 and 45 of the PKD1 Gene. *Am J Med Genet 16;65(2):155-9, 1996.*

11. Reiterova J, Stekrova J, Peters DJ, *et al.* Four novel mutations of the PKD2 gene in Czech families with autosomal dominant polycystic kidney disease. *Hum Mutat 19(5):573, 2002.*

12. Hanaoka K, Qian F, Boletta A, *et al.* Co-assembly of polycystin-1 and -2 produces unique cation-permeable currents. *Nature 408, 990-994, 2000.*

13. Inoue S, Inoue K, Utsunomiya M, *et al.* Mutation analysis in PKD1 of Japanese autosomal dominant polycystic kidney disease patients. *Hum Mutat 19(6):622-8, 2002.*

14. Perrichot RA, Mercier B, Simon PM, *et al.* DGGE screening of PKD1 gene reveals novel mutations in a large cohort of 146 unrelated patients. *Hum Genet 105(3):231-9, 1999.*

15. Bogdanova N, McCluskey M, Sikmann K, *et al.* Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene in 41 Bulgarian and Australian kindreds reveals a prevalence of protein truncating mutations. *Hum Mutat 16(2):166-74, 2000.*

16. Watnick TJ, Torres VE, Gandolph MA, *et al.* Somatic mutation in individual liver cysts supports a two-hit model of cystogenesis in autosomal dominant polycystic kidney disease. *Mol Cell 2(2):247-51, 1998.*

17. Perrichot R, Mercier B, Quere I, *et al.* Novel mutations in the duplicated region of PKD1 gene. *Eur J Hum Genet 8(5):353-9, 2000.*

18. Boletta, A., Qian, F., Onuchic, L. F., *et al.* Polycystin-1, the gene product of PKD1, induces resistance to apoptosis and spontaneous tubulogenesis in MDCK cells. *Mol. Cell 6, 1267-1273, 2000.*

Date Reçue/Date Received 2020-05-26

19. Aguiari G, Savelli S, Garbo M, *et al.* Novel splicing and missense mutations in autosomal dominant polycystic kidney disease 1 (PKD1) gene: expression of mutated genes. *Hum Mutat 16(5):444-5, 2000.*

20. Bycroft M, Bateman A, Clarke J, *et al.* The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. *EMBO J. 15;18(2):297-305, 1999.*

21. Torra R, Viribay M, Telleria D, *et al.* Seven novel mutations of the PKD2 gene in families with autosomal dominant polycystic kidney disease. *Kidney Int 56(1):28-33, 1999.*

22. Rossetti S, Chauveau D, Kubly V, *et al.* Association of mutation position in polycystic kidney disease 1 (PKD1) gene and development of a vascular phenotype. *Lancet 28;361(9376):2196-201, 2003.*

23. Afzal AR, Florencio RN, Taylor R, *et al.* Novel mutations in the duplicated region of the polycystic kidney disease 1 (PKD1) gene provides supporting evidence for gene conversion. *Genet 4(4):365-70, 2000.*

24. Roelfsema JH, Spruit L, Saris JJ, *et al.* Mutation detection in the repeated part of the PKD1 gene. *Am J Hum Genet 61(5):1044-52, 1997.*

25. Bogdanova N, McCluskey M, Sikmann K, *et al.* Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene in 41 Bulgarian and Australian kindreds reveals a prevalence of protein truncating mutations. *Hum Mutat 16(2):166-74, 2000.*

26. Qian F, Boletta A, Bhunia AK, Xu H, *et al.* Cleavage of polycystin-1 requires the receptor for egg jelly domain and is disrupted by human autosomal-dominant polycystic kidney disease 1-associated mutations. *Proc Natl Acad Sci U S A 24;99(26):16981-6, 2002.*

27. Gabow PA. Autosomal dominant polycystic kidney disease. *N Engl J Med 29;329(5):332-42, 1993.*

Date Reçue/Date Received 2020-05-26

We claim:

1. A method for diagnosis or predicting occurrence of autosomal dominant polycystic kidney disease (ADPKD), comprising: a) generating a PKD1 nucleic acid by (i) performing a first amplification reaction comprising a long range amplification of a PKD1 gene in a sample from a human subject suspected of having ADPKD; and (ii) performing a second amplification reaction and a third amplification reaction which form a nested PCR; b) contacting the PKD1 nucleic acid generated in step a) with an oligonucleotide that specifically hybridizes to a mutation; wherein the mutation is a change of a T to an A at a PKD1 nucleic acid position corresponding to position 8855 of SEQ ID NO: 1, wherein the oligonucleotide is detectably labeled and comprises a sequence of less than 100 nucleotides that is fully complementary to a portion of the PKD1 nucleic acid that comprises the mutation; c) detecting hybridization of the oligonucleotide with the PKD1 nucleic acid; wherein detection of hybridization is indicative of a mutant PKD1 gene in the human subject; and d) diagnosis or predicting occurrence of ADPKD in the human subject when the mutant PKD1 gene is detected.

2. The method of claim 1, wherein the method further comprises use of a technique selected from the group consisting of sequencing, one or more additional polymerase chain reactions (PCR), denaturing high performance liquid chromatography and a combination thereof.

3. The method of claim 1, wherein the oligonucleotide is detectably labeled with a radiolabel, an enzymatic label or a fluorescent label.

4. The method of claim 1, wherein the nested PCR uses different PKD-specific primers.

5. The method of claim 4, wherein the method further comprises use of a primer that specifically amplifies a PKD1 nucleic acid, but does not specifically amplify a PKD1 homolog that shares more than 95% sequence identity to SEQ ID NO: 1 but does not encode a PKD1 gene product.

Date Reçue/Date Received 2023-02-07

6. The method of claim 5, wherein the PKD1 homolog comprises a sequence identified with a GenBank Accession Number selected from any one of AC002039, AC010488, AC040158, AF320593 and AF320594.

7. The method of claim 1, wherein the human subject suspected of having ADPKD is a blood relative of an individual with ADPKD.

Date Reçue/Date Received 2023-02-07

Figure 1

```
gcactgcagc gccagcgtcc gagcgggcgg ccgagctccc ggagcggcct ggccccgagc 60
cccgagcggg cgtcgctcag cagcaggtcg cggccgcgca gccccatcca gccccgcgcc 120
cgccatgccg tccgcgggcc ccgcctgagc tgcggtctcc gcgcgcgggc gggcctgggg 180
acggcggggc catgcgcgcg ctgccctaac gatgccgccc gccgcgcccg cccgcctggc 240
gctggccctg ggcctgggcc tgtggctcgg ggcgctggcg gggggccccg ggcgcggctg 300
cgggccctgc gagcccccct gcctctgcgg cccagcgccc ggcgccgcct gccgcgtcaa 360
ctgctcgggc cgcgggctgc ggacgctcgg tcccgcgctg cgcatccccg cggacgccac 420
agcgctagac gtctcccaca acctgctccg ggcgctggac gttgggctcc tggcgaacct 480
ctcggcgctg gcagagctgg atataagcaa caacaagatt tctacgttag aagaaggaat 540
atttgctaat ttatttaatt taagtgaaat aaacctgagt gggaacccgt ttgagtgtga 600
ctgtggcctg gcgtggctgc cgcgatgggc ggaggagcag caggtgcggg tggtgcagcc 660
cgaggcagcc acgtgtgctg ggcctggctc cctggctggc cagcctctgc ttggcatccc 720
cttgctggac agtggctgtg gtgaggagta tgtcgcctgc ctccctgaca acagctcagg 780
caccgtggca gcagtgtcct tttcagctgc ccacgaaggc ctgcttcagc cagaggcctg 840
cagcgccttc tgcttctcca ccggccaggg cctcgcagcc ctctcggagc agggctggtg 900
cctgtgtggg gcggcccagc cctccagtgc ctcctttgcc tgcctgtccc tctgctccgg 960
ccccccgcca cctcctgccc ccacctgtag gggccccacc ctcctccagc acgtcttccc 1020
tgcctcccca ggggccaccc tggtggggcc ccacggacct ctggcctctg gccagctagc 1080
agccttccac atcgctgccc cgctccctgt cactgccaca cgctgggact tcggagacgg 1140
ctccgccgag gtggatgccg ctgggccggc tgcctcgcat cgctatgtgc tgcctgggcg 1200
ctatcacgtg acggccgtgc tggccctggg ggccggctca gccctgctgg ggacagacgt 1260
gcaggtggaa gcggcacctg ccgccctgga gctcgtgtgc ccgtcctcgg tgcagagtga 1320
cgagagcctt gacctcagca tccagaaccg cggtggttca ggcctggagg ccgcctacag 1380
catcgtggcc ctgggcgagg agccggcccg agcggtgcac ccgctctgcc cctcggacac 1440
ggagatcttc cctggcaacg ggcactgcta ccgcctggtg gtggagaagg cggcctggct 1500
gcaggcgcag gagcagtgtc aggcctgggc cggggccgcc ctggcaatgg tggacagtcc 1560
cgccgtgcag cgcttcctgg tctcccgggt caccaggagc ctagacgtgt ggatcggctt 1620
ctcgactgtg caggggtgg aggtgggccc agcgccgcag ggcgaggcct tcagcctgga 1680
gagctgccag aactggctgc ccggggagcc acacccagcc acagccgagc actgcgtccg 1740
gctcgggccc accgggtggt gtaacaccga cctgtgctca gcgccgcaca gctacgtctg 1800
cgagctgcag cccggaggcc cagtgcagga tgccgagaac ctcctcgtgg gagcgcccag 1860
tggggacctg cagggacccc tgacgcctct ggcacagcag gacggcctct cagccccgca 1920
cgagcccgtg gaggtcatgg tattcccggg cctgcgtctg agccgtgaag ccttcctcac 1980
cacggccgaa tttgggaccc aggagctccg gcggcccgcc cagctgcggc tgcaggtgta 2040
ccggctcctc agcacagcag ggaccccgga gaacggcagc gagcctgaga gcaggtcccc 2100
ggacaacagg acccagctgg cccccgcgtg catgccaggg ggacgctggt gccctggagc 2160
caacatctgc ttgccgctgg acgcctcttg ccacccccag gcctgcgcca atggctgcac 2220
gtcagggcca gggctacccg gggcccccta tgcgctatgg agagagttcc tcttctccgt 2280
tgccgcgggg ccccccgcgc agtactcggt caccctccac ggccaggatg tcctcatgct 2340
ccctggtgac ctcgttggct tgcagcacga cgctggccct ggcgccctcc tgcactgctc 2400
gccggctccc ggccaccctg gtccccaggc cccgtacctc tccgccaacg cctcgtcatg 2460
gctgccccac ttgccagccc agctggaggg cacttgggcc tgccctgcct gtgccctgcg 2520
gctgcttgca gccacggaac agctcaccgt gctgctgggc ttgaggccca accctggact 2580
gcggatgcct gggcgctatg aggtccgggc agaggtgggc aatggcgtgt ccaggcacaa 2640
cctctcctgc agctttgacg tggtctcccc agtggctggg ctgcgggtca tctaccctgc 2700
cccccgcgac ggccgcctct acgtgcccac caacggctca gccttggtgc tccaggtgga 2760
ctctggtgcc aacgccacgg ccacggctcg ctggcctggg ggcagtgtca gcgcccgctt 2820
tgagaatgtc tgccctgccc tggtggccac cttcgtgccc ggctgccccc gggagaccaa 2880
cgataccctg ttctcagtgg tagcactgcc gtggctcagt gaggggagc acgtggtgga 2940
cgtggtggtg gaaaacagcg ccagccgggc caacctcagc ctgcgggtga cggcggagga 3000
gcccatctgt ggcctccgcg ccacgcccag ccccgaggcc cgtgtactgc agggagtcct 3060
agtgaggtac agccccgtgg tggaggccgg ctcggacatg gtcttccggt ggaccatcaa 3120
cgacaagcag tccctgacct tccagaacgt ggtcttcaat gtcatttatc agagcgcggc 3180
ggtcttcaag ctctcactga cggcctccaa ccacgtgagc aacgtcaccg tgaactacaa 3240
cgtaaccgtg gagcggatga acaggatgca gggtctgcag gtctccacag tgccggccgt 3300
gctgtccccc aatgccacgc tagcactgac ggcgggcgtg ctggtggact cggccgtgga 3360
ggtggccttc ctgtggaact ttggggatgg ggagcaggcc ctccaccagt tccagcctcc 3420
```

FIG. 1A

Figure 1 con.

```
gtacaacgag tccttcccgg ttccagaccc ctcggtggcc caggtgctgg tggagcacaa 3480
tgtcatgcac acctacgctg ccccaggtga gtacctcctg accgtgctgg catctaatgc 3540
cttcgagaac ctgacgcagc aggtgcctgt gagcgtgcgc gcctccctgc cctccgtggc 3600
tgtgggtgtg agtgacggcg tcctggtggc cggccggccc gtcaccttct acccgcaccc 3660
gctgccctcg cctgggggtg ttctttacac gtgggacttc ggggacggct cccctgtcct 3720
gacccagagc cagccggctg ccaaccacac ctatgcctcg aggggcacct accacgtgcg 3780
cctggaggtc aacaacacgg tgagcggtgc ggcggcccag gcggatgtgc gcgtctttga 3840
ggagctccgc ggactcagcg tggacatgag cctggccgtg gagcagggcg ccccgtggt 3900
ggtcagcgcc gcggtgcaga cgggcgacaa catcacgtgg accttcgaca tgggggacgg 3960
caccgtgctg tcgggcccgg aggcaacagt ggagcatgtg tacctgcggg cacagaactg 4020
cacagtgacc gtgggtgcgg ccagccccgc cggccacctg gcccggagcc tgcacgtgct 4080
ggtcttcgtc ctggaggtgc tgcgcgttga acccgccgcc tgcatcccca cgcagcctga 4140
cgcgcggctc acggcctacg tcaccgggaa cccggcccac tacctcttcg actggacctt 4200
cggggatggc tcctccaaca cgaccgtgcg ggggtgcccg acggtgacac acaacttcac 4260
gcggagcggc acgttccccc tggcgctggt gctgtccagc cgcgtgaaca gggcgcatta 4320
cttcaccagc atctgcgtgg agccagaggt gggcaacgtc accctgcagc cagaggcca 4380
gtttgtgcag ctcggggacg aggcctggct ggtggcatgt gcctggcccc cgttccccta 4440
ccgctacacc tgggactttg gcaccgagga agccgccccc acccgtgcca ggggccctga 4500
ggtgacgttc atctaccgag acccaggctc ctatcttgtg acagtcaccg cgtccaacaa 4560
catctctgct gccaatgact cagccctggt ggaggtgcag gagcccgtgc tggtcaccag 4620
catcaaggtc aatggctccc ttgggctgga gctgcagcag ccgtacctgt tctctgctgt 4680
gggccgtggg cgccccgcca gctacctgtg ggatctgggg gacggtgggt ggctcgaggg 4740
tccggaggtc acccacgctt acaacagcac aggtgacttc accgttaggg tggccggctg 4800
gaatgaggtg agccgcagcg aggcctggct caatgtgacg gtgaagcggc gcgtgcgggg 4860
gctcgtcgtc aatgcaagcc gcacggtggt gcccctgaat gggagcgtga gcttcagcac 4920
gtcgctggag gccggcagtg atgtgcgcta ttcctgggtg ctctgtgacc gctgcacgcc 4980
catccctggg ggtcctacca tctcttacac cttccgctcc gtgggcacct tcaatatcat 5040
cgtcacggct gagaacgagg tgggctccgc ccaggacagc atcttcgtct atgtcctgca 5100
gctcatagag gggctgcagg tggtgggcgg tggccgctac ttccccacca accacacggt 5160
acagctgcag gccgtggtta gggatggcac caacgtctcc tacagctgga ctgcctggag 5220
ggacaggggc ccggccctgg ccggcagcgg caaaggcttc tcgctcaccg tgctcgaggc 5280
cggcacctac catgtgcagc tgcgggccac caacatgctg ggcagcgcct gggccgactg 5340
caccatggac ttcgtggagc ctgtggggtg gctgatggtg accgcctccc cgaacccagc 5400
tgccgtcaac acaagcgtca ccctcagtgc cgagctggct ggtggcagtg gtgtcgtata 5460
cacttggtcc ttggaggagg ggctgagctg ggagacctcc gagccattta ccacccatag 5520
cttccccaca cccggcctgc acttggtcac catgacggca gggaacccgc tgggctcagc 5580
caacgccacc gtggaagtgg atgtgcaggt gcctgtgagt ggcctcagca tcagggccag 5640
cgagcccgga ggcagcttcg tggcggccgg gtcctctgtg ccctttgggg ggcagctggc 5700
cacgggcacc aatgtgagct ggtgctgggc tgtgccccgc ggcagcagca agcgtggccc 5760
tcatgtcacc atggtcttcc cggatgctgg caccttctcc atccggctca atgcctccaa 5820
cgcagtcagc tgggtctcag ccacgtacaa cctcacggcg gaggagccca tcgtgggcct 5880
ggtgctgtgg gccagcagca aggtggtggc gcccgggcag ctggtccatt ttcagatcct 5940
gctggctgcc ggctcagctg tcaccttccg cctgcaggtc ggcggggcca accccgaggt 6000
gctccccggg ccccgtttct cccacagctt ccccgcgtc ggagaccacg tggtgagcgt 6060
gcggggcaaa aaccacgtga gctgggccca ggcgcaggtg cgcatcgtgg tgctggaggc 6120
cgtgagtggg ctgcagatgc ccaactgctg cgagcctggc atcgccacgg gcactgagag 6180
gaacttcaca gcccgcgtgc agcgcggctc tcgggtcgcc tacgcctggt acttctcgct 6240
gcagaaggtc cagggcgact cgctggtcat cctgtcgggc cgcgacgtca cctacacgcc 6300
cgtggccgcg gggctgttgg agatccaggt gcgcgccttc aacgccctgg gcagtgagaa 6360
ccgcacgctg gtgctggagg ttcaggacgc cgtccagtat gtggccctgc agagcggccc 6420
ctgcttcacc aaccgctcgg cgcagtttga ggccgccacc agccccagcc cccggcgtgt 6480
ggcctaccac tgggactttg gggatgggtc gccagggcag gacacagatg agcccagggc 6540
cgagcactcc tacctgaggc ctggggacta ccgcgtgcag gtgaacgcct ccaacctggt 6600
gagcttcttc gtggcgcagg ccacggtgac cgtccaggtg ctggcctgcc gggagccgga 6660
ggtggacgtg gtcctgcccc tgcaggtgct gatgcggcga tcacagcgca actacttgga 6720
ggcccacgtt gacctgcgcg actgcgtcac ctaccagact gagtaccgct gggaggtgta 6780
tcgcaccgcc agctgccagc ggccggggcg cccagcgcgt gtggccctgc ccggcgtgga 6840
```

FIG. 1B

Figure 1 con.

```
cgtgagccgg cctcggctgg tgctgccgcg gctggcgctg cctgtggggc actactgctt 6900
tgtgtttgtc gtgtcatttg gggacacgcc actgacacag agcatccagg ccaatgtgac 6960
ggtggccccc gagcgcctgg tgcccatcat tgagggtggc tcataccgcg tgtggtcaga 7020
cacacgggac ctggtgctgg atgggagcga gtcctacgac cccaacctgg aggacggcga 7080
ccagacgccg ctcagtttcc actgggcctg tgtggcttcg acacagaggg aggctggcgg 7140
gtgtgcgctg aactttgggc cccgcgggag cagcacggtc accattccac gggagcggct 7200
ggcggctggc gtggagtaca ccttcagcct gaccgtgtgg aaggccggcc gcaaggagga 7260
ggccaccaac cagacggtgc tgatccggag tggccgggtg cccattgtgt ccttggagtg 7320
tgtgtcctgc aaggcacagg ccgtgtacga agtgagccgc agctcctacg tgtacttgga 7380
gggccgctgc ctcaattgca gcagcggctc caagcgaggg cggtgggctg cacgtacgtt 7440
cagcaacaag acgctggtgc tggatgagac caccacatcc acgggcagtg caggcatgcg 7500
actggtgctg cggcggggcg tgctgcggga cggcgaggga tacaccttca cgctcacggt 7560
gctgggccgc tctggcgagg aggagggctg cgcctccatc cgcctgtccc ccaaccgccc 7620
gccgctgggg ggctcttgcc gcctcttccc actgggcgct gtgcacgccc tcaccaccaa 7680
ggtgcacttc gaatgcacgg gctggcatga cgcggaggat gctggcgccc cgctggtgta 7740
cgccctgctg ctgcggcgct gtcgccaggg ccactgcgag gagttctgtg tctacaaggg 7800
cagcctctcc agctacggag ccgtgctgcc cccgggtttc aggccacact tcgaggtggg 7860
cctggccgtg gtggtgcagg accagctggg agccgctgtg gtcgccctca acaggtcttt 7920
ggccatcacc ctcccagagc ccaacggcag cgcaacgggg ctcacagtct ggctgcacgg 7980
gctcaccgct agtgtgctcc cagggctgct gcggcaggcc gatccccagc acgtcatcga 8040
gtactcgttg gccctggtca ccgtgctgaa cgagtacgag cgggccctgg acgtggcggc 8100
agagcccaag cacgagcggc agcaccgagc ccagatacgc aagaacatca cggagactct 8160
ggtgtccctg agggtccaca ctgtggatga catccagcag atcgctgctg cgctggccca 8220
gtgcatgggg cccagcaggg agctcgtatg ccgctcgtgc ctgaagcaga cgctgcacaa 8280
gctggaggcc atgatgctca tcctgcaggc agagaccacc gcgggcaccg tgacgcccac 8340
cgccatcgga gacagcatcc tcaacatcac aggagacctc atccacctgg ccagctcgga 8400
cgtgcgggca ccacagccct cagagctggg agccgagtca ccatctcgga tggtggcgtc 8460
ccaggcctac aacctgacct ctgccctcat gcgcatcctc atgcgctccc gcgtgctcaa 8520
cgaggagccc ctgacgctgg cgggcgagga gatcgtggcc cagggcaagc gctcggaccc 8580
gcggagcctg ctgtgctatg gcggcgcccc agggcctggc tgccacttct ccatccccga 8640
ggctttcagc ggggccctgg ccaacctcag tgacgtggtg cagctcatct ttctggtgga 8700
ctccaatccc tttcccttg gctatatcag caactacacc gtctccacca aggtggcctc 8760
gatggcattc cagacacagg ccggcgccca gatccccatc gagcggctgg cctcagagcg 8820
cgccatcacc gtgaaggtgc ccaacaactc ggactgggct gcccggggcc accgcagctc 8880
cgccaactcc gccaactccg ttgtggtcca gccccaggcc tccgtcggtg ctgtggtcac 8940
cctggacagc agcaaccctg cggccgggct gcatctgcag ctcaactata cgctgctgga 9000
cggccactac ctgtctgagg aacctgagcc ctacctggca gtctacctac actcggagcc 9060
ccggcccaat gagcacaact gctcggctag caggaggatc cgcccagagt cactccaggg 9120
tgctgaccac cggccctaca ccttcttcat ttccccgggg agcagagacc cagcggggag 9180
ttaccatctg aacctctcca gccacttccg ctggtcggcg ctgcaggtgt ccgtgggcct 9240
gtacacgtcc ctgtgccagt acttcagcga ggaggacatg gtgtggcgga cagaggggct 9300
gctgcccctg gaggagacct cgccccgcca ggccgtctgc ctcacccgcc acctcaccgc 9360
cttcggcgcc agcctcttcg tgcccccaag ccatgtccgc tttgtgtttc ctgagccgac 9420
agcggatgta aactacatcg tcatgctgac atgtgctgtg tgcctggtga cctacatggt 9480
catggccgcc atcctgcaca agctggacca gttggatgcc agccggggcc gcgccatccc 9540
tttctgtggg cagcggggcc gcttcaagta cgagatcctc gtcaagacag gctggggccg 9600
gggctcaggt accacggccc acgtgggcat catgctgtat ggggtggaca gccggagcgg 9660
ccaccggcac ctggacggcg acagagcctt ccaccgcaac agcctggaca tcttccggat 9720
cgccaccccg cacagcctgg gtagcgtgtg gaagatccga gtgtggcacg acaacaaagg 9780
gctcagccct gcctggttcc tgcagcacgt catcgtcagg gacctgcaga cggcacgcag 9840
cgccttcttc ctggtcaatg actggctttc ggtggagacg gaggccaacg ggggcctggt 9900
ggagaaggag gtgctggccg cgagcgacgc agccctttg cgcttccggc gcctgctggt 9960
ggctgagctg cagcgtggct tctttgacaa gcacatctgg ctctccatat gggaccggcc 10020
gcctcgtagc cgtttcactc gcatccagag ggccacctgc tgcgttctcc tcatctgcct 10080
cttcctgggc gccaacgccg tgtggtacgg ggctgttggc gactctgcct acagcacggg 10140
gcatgtgtcc aggctgagcc cgctgagcgt cgacacagtc gctgttggcc tggtgtccag 10200
cgtggttgtc tatcccgtct acctggccat ccttttctc ttccggatgt cccggagcaa 10260
```

FIG. 1C

Figure 1 con.

```
ggtggctggg agcccgagcc ccacacctgc cgggcagcag gtgctggaca tcgacagctg 10320
cctggactcg tccgtgctgg acagctcctt cctcacgttc tcaggcctcc acgctgaggc 10380
ctttgttgga cagatgaaga gtgacttgtt tctggatgat tctaagagtc tggtgtgctg 10440
gccctccggc gagggaacgc tcagttggcc ggacctgctc agtgacccgt ccattgtggg 10500
tagcaatctg cggcagctgg cacggggcca ggcgggccat gggctgggcc cagaggagga 10560
cggcttctcc ctggccagcc cctactcgcc tgccaaatcc ttctcagcat cagatgaaga 10620
cctgatccag caggtccttg ccgagggggt cagcagccca gcccctaccc aagacaccca 10680
catggaaacg gacctgctca gcagcctgtc cagcactcct ggggagaaga cagagacgct 10740
ggcgctgcag aggctggggg agctggggcc acccagccca ggcctgaact gggaacagcc 10800
ccaggcagcg aggctgtcca ggacaggact ggtggagggt ctgcggaagc gcctgctgcc 10860
ggcctggtgt gcctccctgg cccacgggct cagcctgctc ctggtggctg tggctgtggc 10920
tgtctcaggg tgggtgggtg cgagcttccc cccgggcgtg agtgttgcgt ggctcctgtc 10980
cagcagcgcc agcttcctgg cctcattcct cggctgggag ccactgaagg tcttgctgga 11040
agccctgtac ttctcactgg tggccaagcg gctgcacccg gatgaagatg acaccctggt 11100
agagagcccg gctgtgacgc ctgtgagcgc acgtgtgccc cgcgtacggc caccccacgg 11160
ctttgcactc ttcctggcca aggaagaagc ccgcaaggtc aagaggctac atggcatgct 11220
gcggagcctc ctggtgtaca tgcttttttct gctggtgacc ctgctggcca gctatgggga 11280
tgcctcatgc catgggcacg cctaccgtct gcaaagcgcc atcaagcagg agctgcacag 11340
ccgggccttc ctggccatca cgcggtctga ggagctctgg ccatggatgg cccacgtgct 11400
gctgccctac gtccacggga accagtccag cccagagctg gggcccccac ggctgcggca 11460
ggtgcggctg caggaagcac tctacccaga ccctcccggc cccagggtcc acacgtgctc 11520
ggccgcagga ggcttcagca ccagcgatta cgacgttggc tgggagagtc ctcacaatgg 11580
ctcggggacg tgggcctatt cagcgccgga tctgctgggg gcatggtcct ggggctcctg 11640
tgccgtgtat gacagcgggg gctacgtgca ggagctgggc ctgagcctgg aggagagccg 11700
cgaccggctg cgcttcctgc agctgcacaa ctggctggac aacaggagcc gcgctgtgtt 11760
cctggagctc acgcgctaca gcccggccgt ggggctgcac gccgccgtca cgctgcgcct 11820
cgagttcccg gcggccggcc gcgccctggc cgccctcagc gtccgccctt ttgcgctgcg 11880
ccgcctcagc gcgggcctct cgctgcctct gctcacctcg gtgtgcctgc tgctgttcgc 11940
cgtgcacttc gccgtggccg aggcccgtac ttggcacagg gaagggcgct ggcgcgtgct 12000
gcggctcgga gcctgggcgc ggtggctgct ggtggcgctg acggcggcca cggcactggt 12060
acgcctcgcc cagctgggtg ccgctgaccg ccagtggacc cgtttcgtgc gcggccgccc 12120
gcgccgcttc actagcttcg accaggtggc gcagctgagc tccgcagccc gtggcctggc 12180
ggcctcgctg ctcttcctgc ttttggtcaa ggctgcccag cagctacgct tcgtgcgcca 12240
gtggtccgtc tttggcaaga cattatgccg agctctgcca gagctcctgg gggtcacctt 12300
gggcctggtg gtgctcgggg tagcctacgc ccagctggcc atcctgctcg tgtcttcctg 12360
tgtggactcc ctctggagcg tggcccaggc cctgttggtg ctgtgccctg ggactgggct 12420
ctctaccctg tgtcctgccg agtcctggca cctgtcaccc ctgctgtgtg tggggctctg 12480
ggcactgcgg ctgtggggcg ccctacggct gggggctgtt attctccgct ggcgctacca 12540
cgccttgcgt ggagagctgt accggccggc ctgggagccc caggactacg agatggtgga 12600
gttgttcctg cgcaggctgc gcctctggat gggcctcagc aaggtcaagg agttccgcca 12660
caaagtccgc tttgaaggga tggagccgct gccctctcgc tcctccaggg gctccaaggt 12720
atccccggat gtgcccccac ccagcgctgg ctccgatgcc tcgcaccctt ccacctcctc 12780
cagccagctg gatgggctga gcgtgagcct gggccggctg gggacaaggt gtgagcctga 12840
gccctcccgc ctccaagccg tgttcgaggc cctgctcacc cagtttgacc gactcaacca 12900
ggccacagag gacgtctacc agctggagca gcagctgcac agcctgcaag gccgcaggag 12960
cagccgggcg cccgccggat cttcccgtgg cccatccccg ggcctgcggc cagcactgcc 13020
cagccgcctt gcccgggcca gtcggggtgt ggacctggcc actggcccca gcaggacacc 13080
ccttcgggcc aagaacaagg tccaccccag cagcacttag tcctccttcc tggcgggggt 13140
gggccgtgga gtcggagtgg acaccgctca gtattacttt ctgccgctgt caaggccgag 13200
ggccaggcag aatggctgca cgtaggttcc ccagagagca ggcaggggca tctgtctgtc 13260
tgtgggcttc agcactttaa agaggctgtg tggccaacca ggacccaggg tcccctcccc 13320
agctcccttg ggaaggacac agcagtattg gacggtttct agcctctgag atgctaattt 13380
atttccccga gtcctcaggt acagcgggct gtgcccggcc ccacccccctg ggcagatgtc 13440
ccccactgct aaggctgctg gcttcaggga gggttagcct gcaccgccgc caccctgccc 13500
ctaagttatt acctctccag ttcctaccgt actccctgca ccgtctcact gtgtgtctcg 13560
tgtcagtaat ttatatggtg ttaaaatgtg tatatttttg tatgtcacta ttttcactag 13620
ggctgagggg cctgcgccca gagctggcct cccccaacac ctgctgcgct tggtaggtgt 13680
```

FIG. 1D

Figure 1 con.

```
ggtggcgtta tggcagcccg gctgctgctt ggatgcgagc ttggccttgg gccggtgctg 13740
ggggcacagc tgtctgccag gcactctcat caccccagag gccttgtcat cctcccttgc 13800
cccaggccag gtagcaagag agcagcgccc aggcctgctg gcatcaggtc tgggcaagta 13860
gcaggactag gcatgtcaga ggaccccagg gtggttagag gaaaagactc ctcctggggg 13920
ctggctccca gggtggagga aggtgactgt gtgtgtgtgt gtgtgcgcgc gcgacgcgcg 13980
agtgtgctgt atggcccagg cagcctcaag gccctcggag ctggctgtgc ctgcttctgt 14040
gtaccacttc tgtgggcatg gccgcttcta gagcctcgac acccccccaa cccccgcacc 14100
aagcagacaa agtcaataaa agagctgtct gactgc                          14136
```

FIG. 1E

Figure 2

```
ggctcctgag gcgcacagcg ccgagcgcgg cgccgcgcac ccgcgcgccg gacgccagtg 60
accgcgatgg tgaactccag tcgcgtgcag cctcagcagc ccggggacgc caagcggccg 120
cccgcgcccc gcgcgccgga cccgggccgg ctgatggctg gctgcgcggc cgtgggcgcc 180
agcctcgccg ccccgggccg cctctgcgag cagcggggcc tggagatcga gatgcagcgc 240
atccggcagg cggccgcgcg ggaccccccg gccggagccg cggcctcccc ttctcctccg 300
ctctcgtcgt gctcccggca ggcgtggagc cgcgataacc ccggcttcga ggccgaggag 360
gaggaggagg aggtggaagg ggaagaaggc ggaatggtgg tggagatgga cgtagagtgg 420
cgcccgggca gccggaggtc ggccgcctcc tcggccgtga gctccgtggg cgcgcggagc 480
cgggggcttg ggggctacca cggcgcgggc cacccgagcg ggaggcggcg ccggcgagag 540
gaccagggcc cgccgtgccc cagcccagtc ggcggcgggg acccgctgca tcgccacctc 600
cccctggaag ggcagccgcc ccgagtggcc tgggcggaga ggctggttcg cgggctgcga 660
ggtgtaagag cgcgcgaccc gcagcggcag atgcacgaac cagaacggcc ggcgccggng 720
gcttcttaaa taaaatgata tcttttcttt tcttcattat tattttaaag gtctctgggg 780
aacaagactc atggaggaaa gcagcactaa ccgagagaaa taccttaaaa gtgttttacg 840
ggaactggtc acatacctcc tttttctcat agtcttgtgc atctgtaagt agaatatttc 900
cttgcactaa tgggaaagtt ttgaaacgat gtgaatttgt ccaaaatgtt tatccacagg 960
aacaatccct ttgtgaaggc tgctggtatg tggatgtgtg ccggttccct tggggcgttc 1020
atttggatct ttctgtgttc cagtgaccta cggcatgatg agctccaatg tgtactacta 1080
cacccggatg atgtcacagc tcttcctaga cacccccgtg tccaaaacgg agaaaactaa 1140
ctttaaaact ctgtcttcca tggaagactt ctggaaggta tttggaaata actttgaaag 1200
tacctctcta tcacaagcca atgcttggtt atgcaacgat gcaggcaggg caaagcagcg 1260
gcatgagctt gaacttnnnn agatgttnnc tttcttttag ttcacagaag gctccttatt 1320
ggatgggctg tactggaaga tgcagcccag caaccagact gaagctgaca accgaagttt 1380
catcttctat gagaacctgc tgttaggggt tccacgaata cggcaactcc gagtcagaaa 1440
tggatcctgc tctatccccc aggacttgag agatgaaatt aaagagtgct atgatgtcta 1500
ctctgtcagt agtgaagata gggctccctt tgggccccga aatggaaccg cgtaagtgtc 1560
tgtgactcat tggcactcgg tgatattcat ccttgtaatt gcctcaagtg ttccactgat 1620
tgtaactgtt tgtttttngg ttttgttttt aatcagttgg atctacacaa gtgaaaaaga 1680
cttgaatggt agtagccact ggggaatcat tgcaacttat agtggagctg gctattatct 1740
ggatttgtca agaacaagag aggaaacagc tgcacaagtt gctagcctca agaaaaatgt 1800
ctggctggac cgaggaacca gggcaacttt tattgacttc tcagtgtaca acgccaacat 1860
taacctgttc tgtgtggtca ggtgtgtgac tgaggacatg catccctcct atttctgtgt 1920
ggttgtacat acatcctatt ctagggttac ccagaaaaac ctttttntgc aggttgttat 1980
tgttttaatt gttcttattt acatgcaggt tattggttga attcccagca acaggtggtg 2040
tgattccatc ttggcaattt cagcctttaa agctgatccg atatgtcaca acttttgatt 2100
tcttcctggc agcctgtgag attatctttt gtttctttat cttttactat gtggtggaag 2160
agatattgga aattcgcatt cacaaactac actatttcag gagtttctgg aattgtctgg 2220
atgttgtgat cgttgtggta ggtccganca ncancaccaa atttcctatt ctattctaca 2280
agnatgttaa caattaatac attggtgaag aaaaatatac tagtcatatt aaggtaagtt 2340
tcatatttct aaaacactgt aataaaatat aaatattttg cttttcagct gtcagtggta 2400
gctataggaa ttaacatata cagaacatca aatgtggagg tgctactaca gtttctggaa 2460
gatcaaaata ctttccccaa ctttgagcat ctggcatatt ggcagataca gttcaacaat 2520
atagctgctg tcacagtatt ttttgtctgg attaaggtaa tttataaatt tcatgttcta 2580
cattnnaaat aatattttct ttaaaaaaaa tgagttccac aaaancatgc gaaacaatgt 2640
tttattatac acagtcacac catttggttt atccattcat ctattgatgt cttctctctc 2700
ttacagctct tcaaattcat caattttaac aggaccatga gccagctctc gacaaccatg 2760
tctcgatgtg ccaaagacct gtttggcttt gctattatgt tcttcattat tttcctagcg 2820
tatgctcagt tggcatacct tgtctttggc actcaggtcg atgacttcag tactttccaa 2880
gagtgtatgt aagtatatat gaaattaaga agaaaaattt agtcagagta gncactgttg 2940
cgtggacant ctttggtttt gtattgtggt gntttgtntt atttttatag cttcactcaa 3000
ttccgtatca ttttgggcga tatcaacttt gcagagattg aggaagctaa tcgagttttg 3060
ggaccaattt atttcactac atttgtgttc tttatgttct tcattctttt ggtatgtaca 3120
tttatattta tagtggaggt tcaatttaaa cttcgtaaat ccttgtcttc tctttttga 3180
ttgataattc caaattatgt ttcttccttt aatttttgcc ctcctttcat ttacaaacag 3240
aatatgtttt tggctatcat caatgatact tactctgaag tgaaatctga cttggcacag 3300
cagaaagctg aaatggaact ctcagatctt atcagaaagg taggaaaaac cttaattctc 3360
aaaaattctt ctgtttctga cataaaatga gcattgtttc acccanattt tagaatacnc 3420
```

FIG. 2A

Figure 2 con.

```
taaaccaagt cttttatttt ttctctctct gatagggcta ccataaagct ttggtcaaac 3480
taaaactgaa aaaaaatacc gtggatgaca tttcagagag tctgcggcaa ggaggaggca 3540
agttaaactt tgacgaactt cgacaagatc tcaaagggtg agaatcatgc ttcctgaggt 3600
tctnaaaaat tcctgcttct aaagataaat tcctggtgat aagagtattt ctagcccaag 3660
ggctcatggg aacanaggat gaatgttatc tgtatcctct ctctaatttc aggaagggcc 3720
atactgatgc agagattgag gcaatattca caaagtacga ccaagatgga gaccaagaac 3780
tgaccgaaca tgaacatcag cagatgagag acgacttgga gaaagagagg gtgggtctgg 3840
tttaggagna accggatttg atttggtacc tacaacacca cacttctgtg gggtctcagt 3900
gttctgctcc tcactcagtg acccccttgtt cttcaggagg acctggattt ggatcacagt 3960
tctttaccac gtcccatgag cagccgaagt ttccctcgaa gcctggatga ctctgaggag 4020
gatgacgatg aagatagcgg acatagctcc agaaggaggg gaagcatttc tagtggcgtt 4080
tcttacgaag agtttcaagt gtaagtataa aggaattggc agaatttgcg tngacaattt 4140
gtcccctctgt actgtgtttt ccttgcagcc tggtgagacg agtggaccgg atggagcatt 4200
ccatcggcag catagtgtcc aagattgacg ccgtgatcgt gaagctagag attatggagc 4260
gagccaaact gaagaggagg gaggtgctgg gaaggctgtt ggatggggtg gccgaggtca 4320
gtagtcatga gctgaanaca ccgctgctga gcatggtgtt attaatnnna atatatgttg 4380
ctgacagttg tatttnaagt attnactgac ccccaacacc agtttctttt tccctttta 4440
ggatgaaagg ctgggtcgtg acagtgaaat ccatagggaa cagatggaac ggctagtacg 4500
tgaagagttg gaacgctggg aatccgatga tgcagcttcc cagatcagtc atggtttagg 4560
cacgccagtg ggactaaatg gtcaacctcg ccccagaagc tcccgcccat cttcctccca 4620
atctacagaa ggcatggaag gtgcaggtgg aaatgggagt tctaatgtcc acgtatgata 4680
tgtgtgtttc agtatgtgtg tttctaataa gtgaggaagt ggctgtcctg aattgctgta 4740
acaagcacac tatttatatg ccctgaccac cataggatgc tagtctttgt gaccgattgc 4800
taatcttctg cactttaatt tattttatat aaactttacc catggttcaa agatttttt 4860
ttctttttct catataagaa atctaggtgt aaatattgag tacagaaaaa aaatcttcat 4920
gatgtgtatt gagcggtacg cccagttgcc accatgactg agtcttctca gttgacaatg 4980
aagtagcctt ttaaagctag aaaactgtca aagggcttct gagtttcatt tccagtcaca 5040
aaaatcagta ttgttatttt tttccaagag tgtgaaggaa aatggggcaa ttcctttcca 5100
ctctggcata gttcatgagc ttaatacata gctttctttt aagaaaggag cctttttttt 5160
caactagctt cctggggtaa acttttctaa aagataaaat gggaaggaac tccaaactat 5220
gatagaatct gtgtgaatgg ttaagatgaa tgttaaatac tatgcttttt tgtaagttga 5280
tcgtatctga tgtctgtggg actaactgta tcacttaatt tttaccttat tttggctcta 5340
atttgaataa gctgagtaaa accaccaaag atcagttata ggataaaatg gcatctctaa 5400
ccataacaca ggagaattgg aaggagccct aagttgtcac tcagtttaat ttcttttaat 5460
ggttagttta gcctaaagat ttatctgcat attcttttte ccatgtggct ctactcattt 5520
gcaactgaat ttaatgttat aactcatcta gtgagaccaa cttactaaat ttttagtatg 5580
cactgaaagt ttttatccaa caattatgtt cattttaagc aaaattttaa gaaagttttg 5640
aaattcataa agcatttggt tttaaactat tttaagaata tagtactcgg tcaggtatgn 5700
nncacgcctg taatcccagc actttgggag gccgaaacag gcgaatcact tgagcccagg 5760
agttcaagac caacatgggc aatgtggcga aactccatct ctacaaaaaa tgcaaaaata 5820
aaaaatatag tactcaagta ttcttgatcc tgtgtttcaa aactagaatt tgtaatgcaa 5880
atggagctca gtctaataaa aaagaggttt tggtattaaa agttcataca ttagacagta 5940
tcagccaaaa tttgagttag caacactgtt ttctttacga gagggtctca cccaaattta 6000
tggggagaaa tctatttctc aaaaaaaaaa aatcttcttt tacagaaatg ttgagtaagg 6060
tgacattttg agcgctaata agcaaaagag catgcagtgc tgttgaataa ccctcacttg 6120
gagaaccaag agaatcctgt cgtttaatgc tatattttaa tttcacaagt tgttcattta 6180
actggtagaa tgtcagtcca atctccaatg agaacatgag caaatagacc tttccaggtt 6240
gaaagtgaaa catactgggt ttctgtaagt ttttcctcat ggcttcatct ctatctttac 6300
tttctcttga atatgctaca caaagttctt tattactaca tactaaagtt tgcattccag 6360
ggatattgac tgtacatatt tatgtatatg taccatgttg ttacatgtaa acaaacttca 6420
atttgaagtg cagctattat gtggtatcca tgtgtatcga ccatgtgcca tatatcaatt 6480
atggtcacta gaaagtctct ttatgatact ttttattgta ctgtttttca tttcacttgc 6540
aaaatttttgc agaattcctc ctttctaccc ataaattaca tataattttt cttctttagt 6600
catggagaac ncccccccat catctcancc ctattanctt tcccatgtgt actggtatta 6660
ttaaaaagac atttacatac gcaagttttt cactgacaan caagaatgtt attaatgtgt 6720
aatactgagc acntttactt cttaataaa                                   6749
```

FIG. 2B

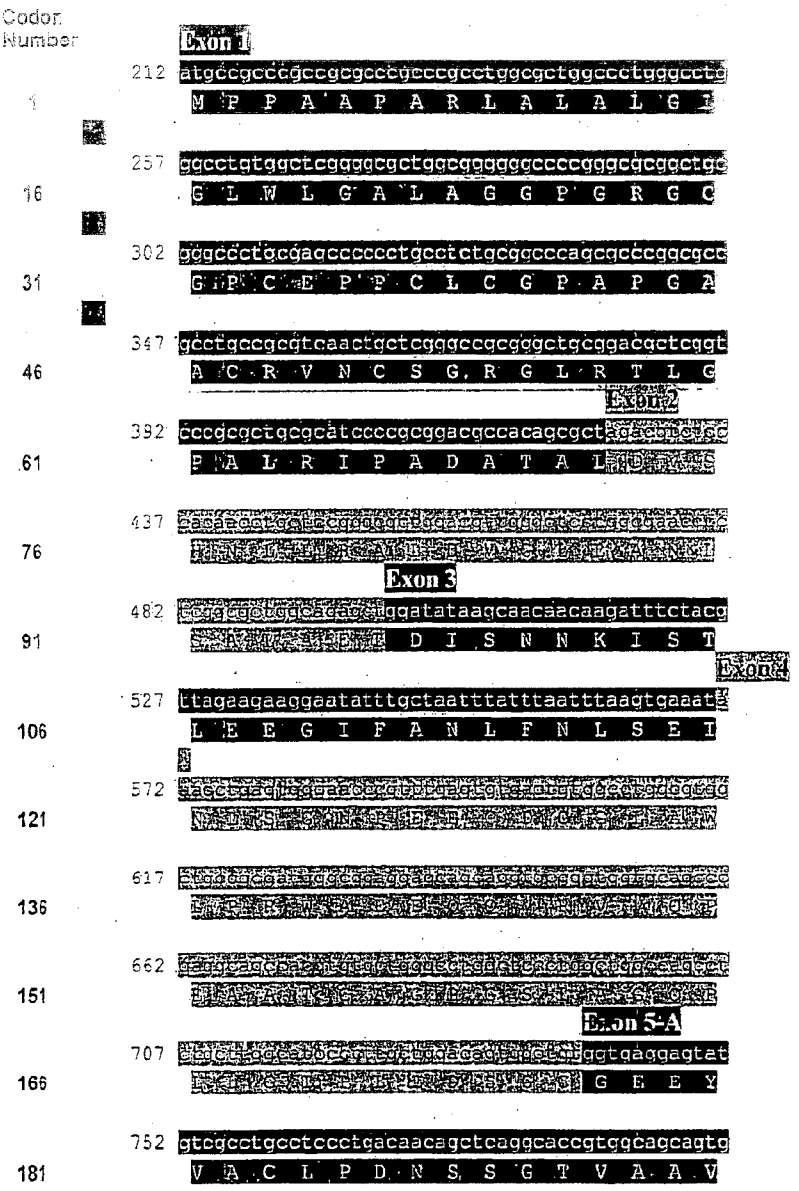


FIG. 3A

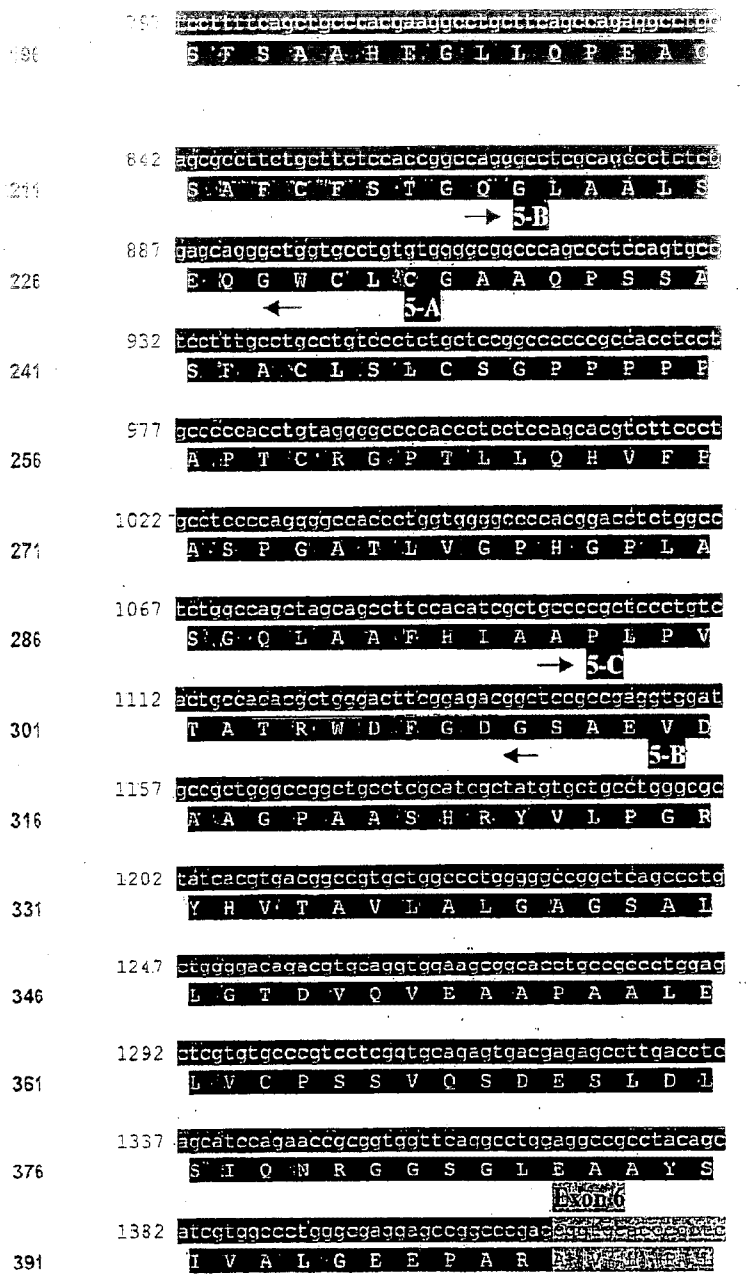


FIG. 3B

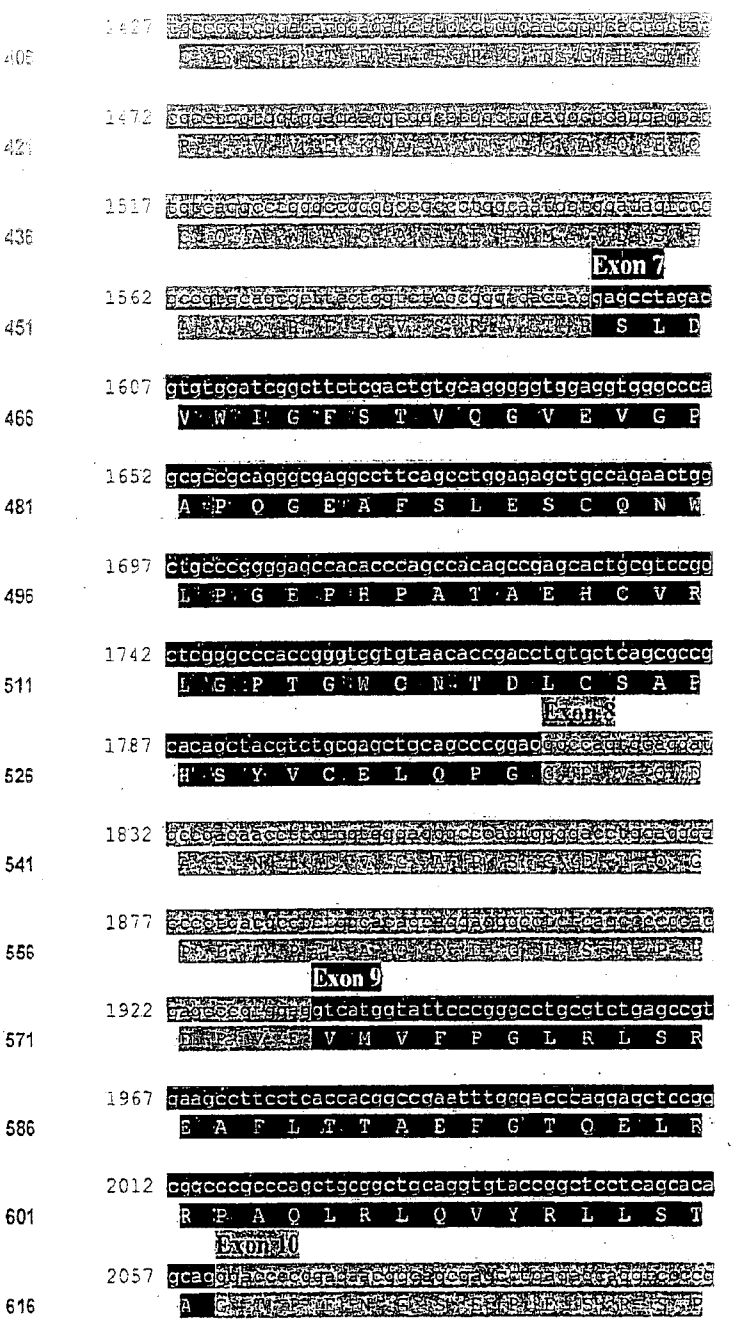


FIG. 3C

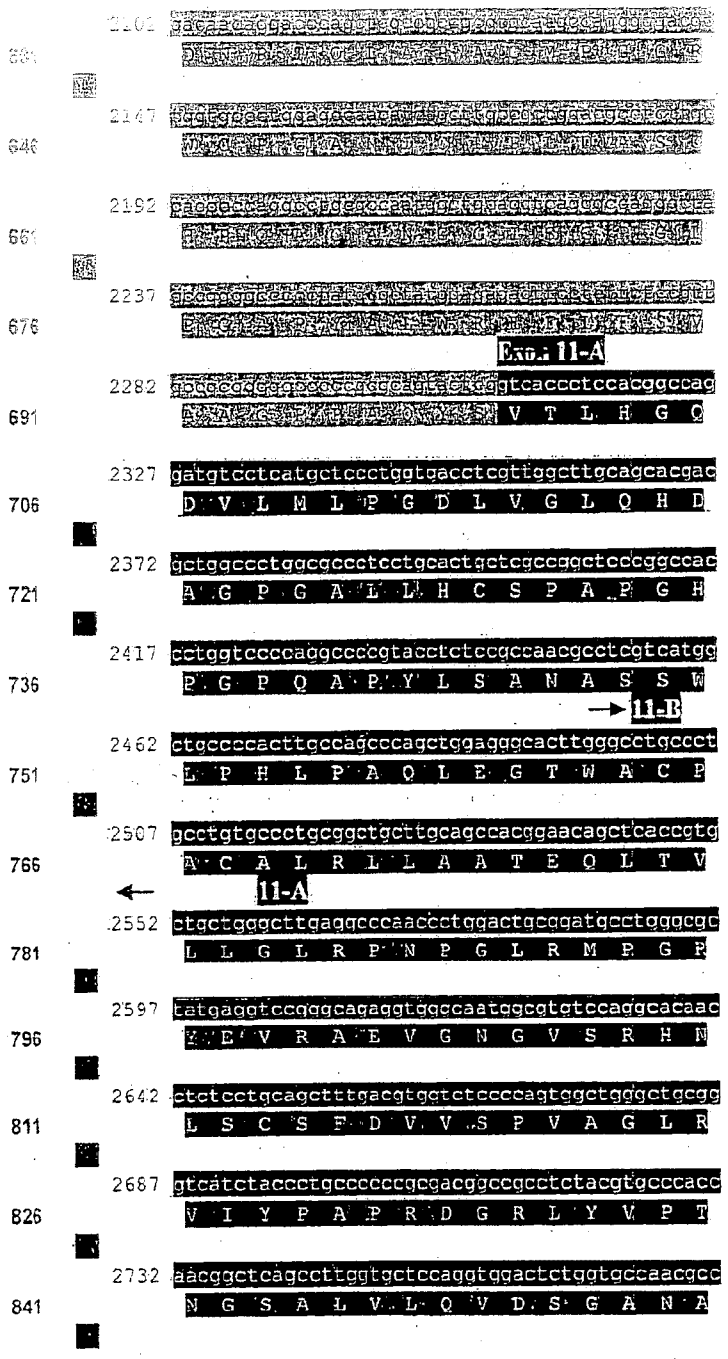


FIG. 3D

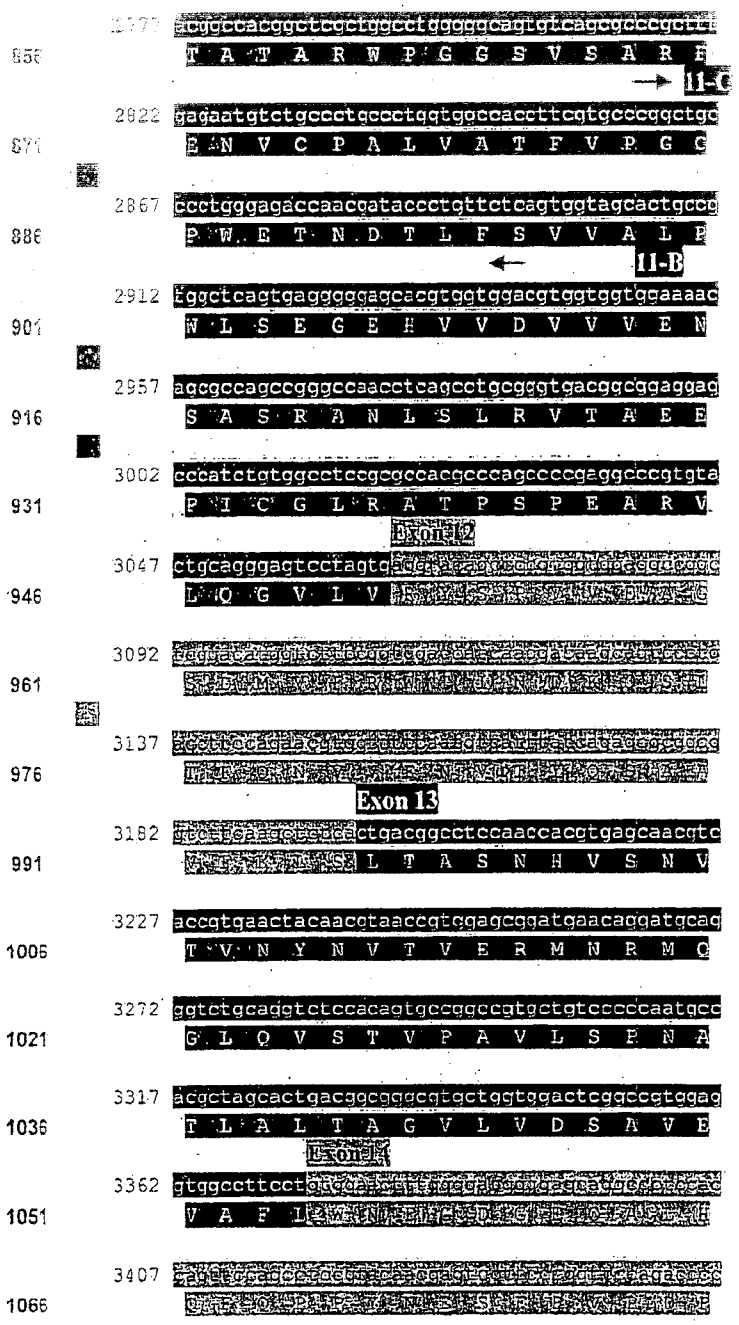


FIG. 3E

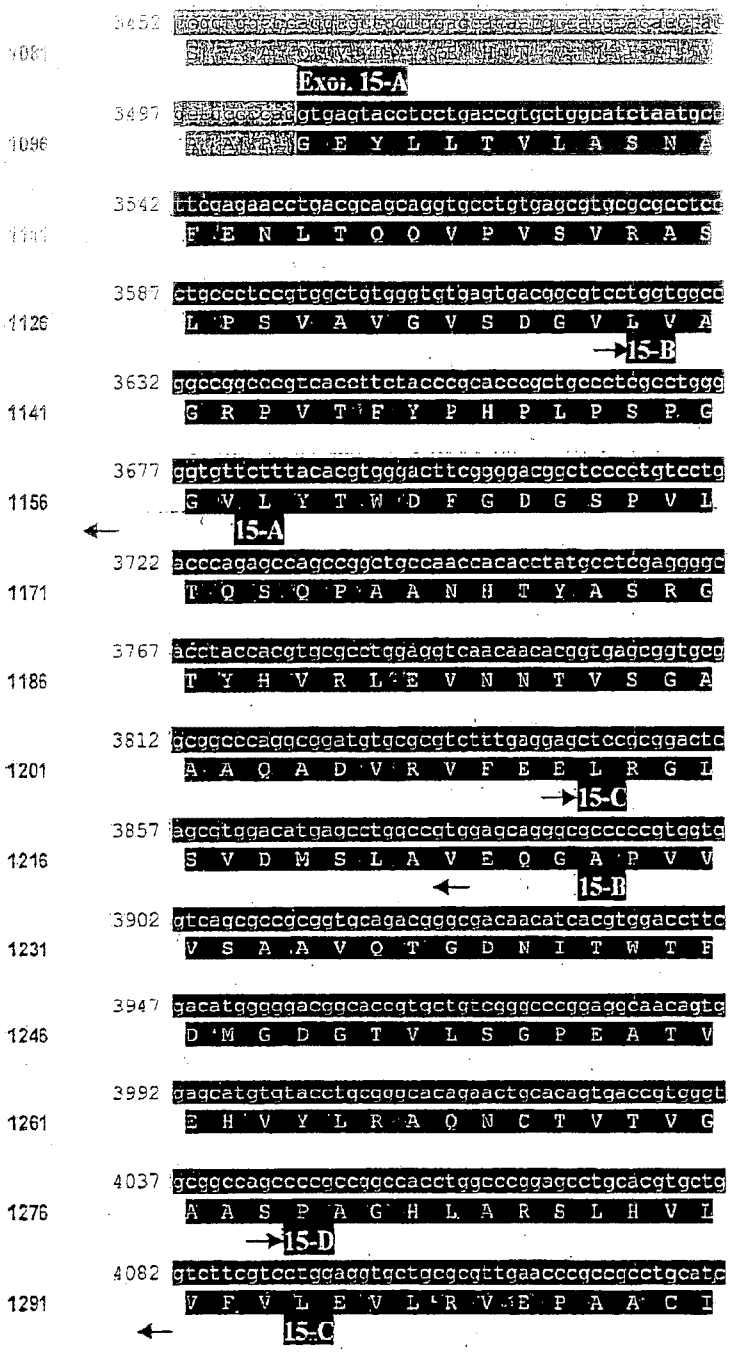


FIG. 3F

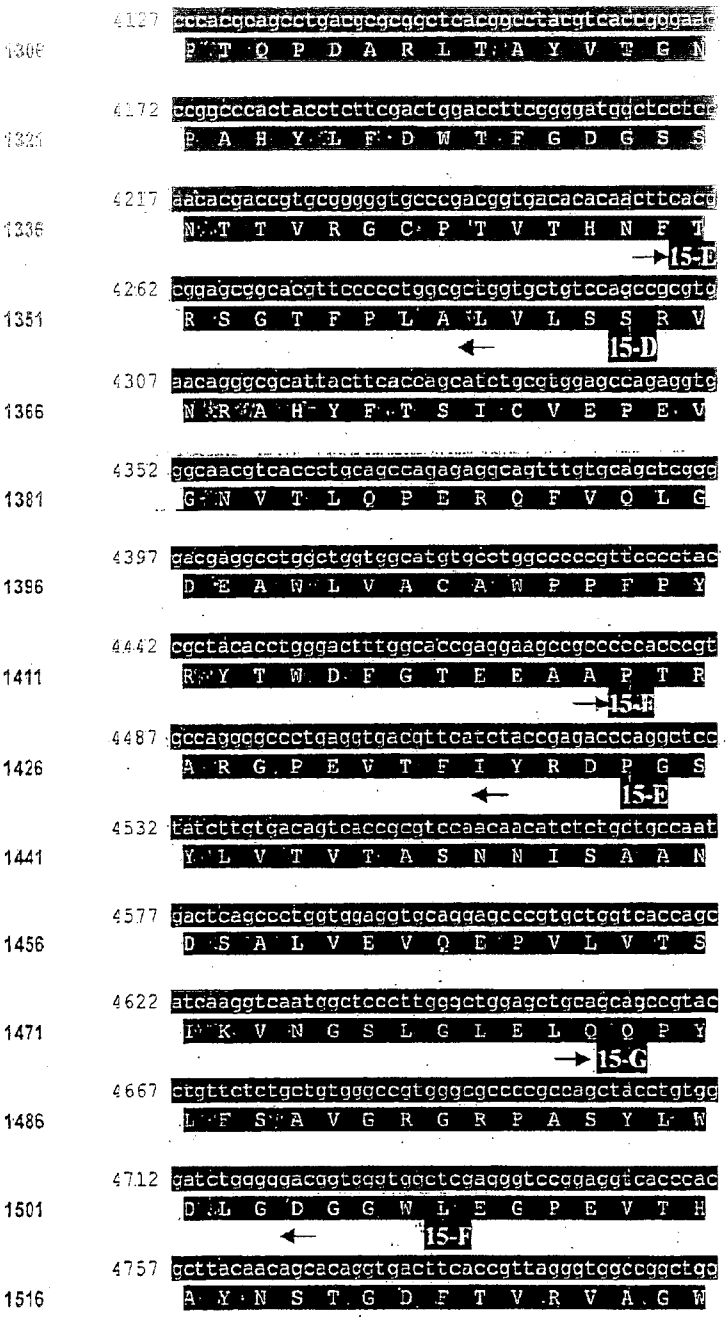


FIG. 3G

Date Reçue/Date Received 2021-09-14

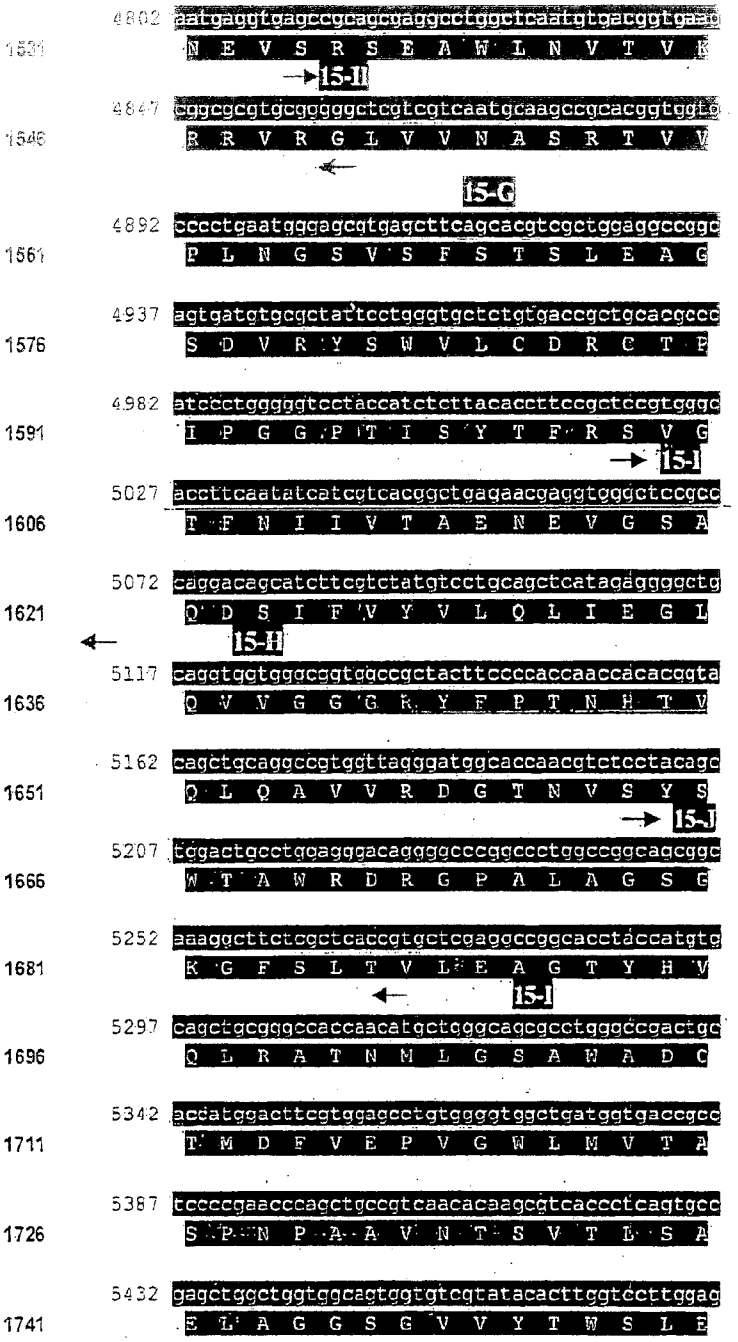

4802
1531
N E V S R S E A W L N V T V E
15-II
4847
1546
R R V R G L V V N A S R T V V
15-G
4892
1561
P L N G S V S F S T S L E A G
4937
1576
S D V R Y S W V L C D R C T P
4982
1591
I P G G P T I S Y T F R S V G
15-I
5027
1606
T F N I I V T A E N E V G S A
5072
1621
Q D S I F V Y V L Q L I E G L
15-H
5117
1636
Q V V G G G R Y F P T N H T V
5162
1651
Q L Q A V V R D G T N V S Y S
15-J
5207
1666
W T A W R D R G P A L A G S G
5252
1681
K G F S L T V L E A G T Y H V
15-I
5297
1696
Q L R A T N M L G S A W A D C
5342
1711
T M D F V E P V G W L M V T A
5387
1726
S P N P A A V N T S V T L S A
5432
1741
E L A G G S G V V Y T W S L E

FIG. 3H

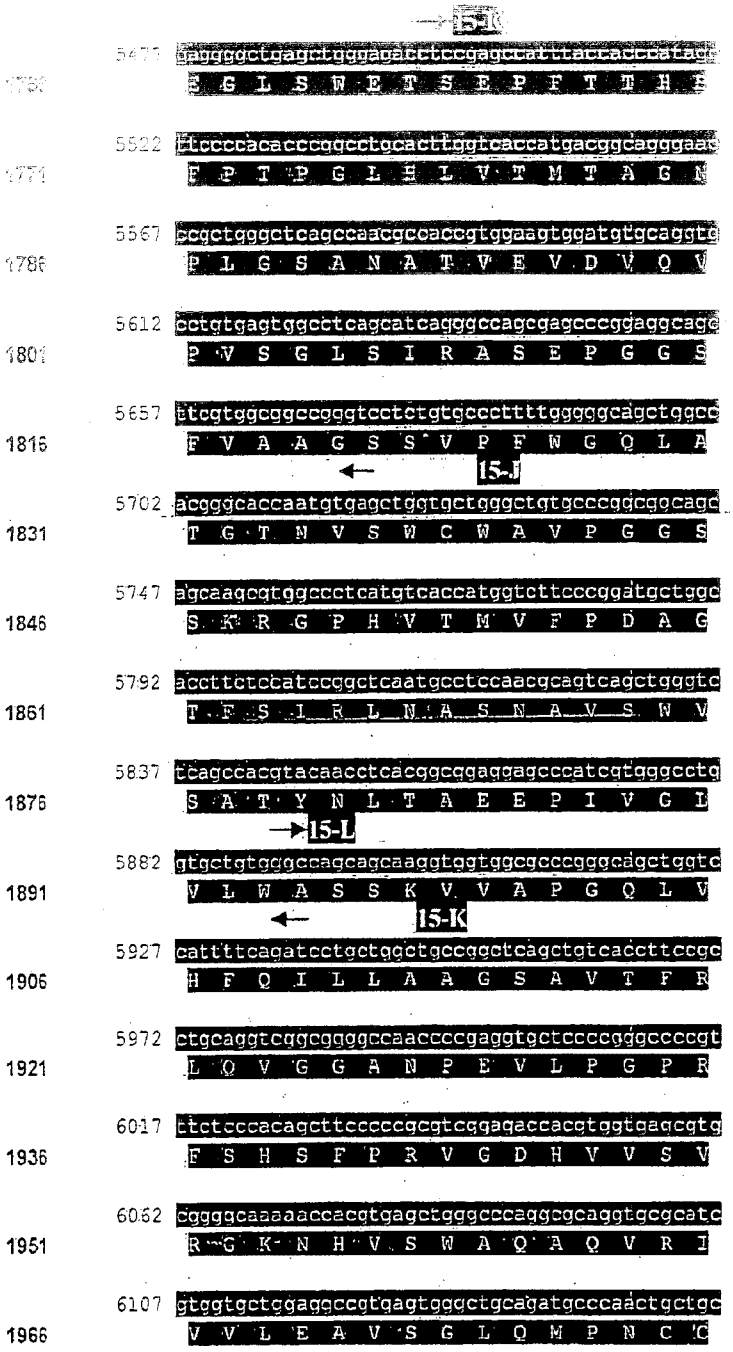


FIG. 3I

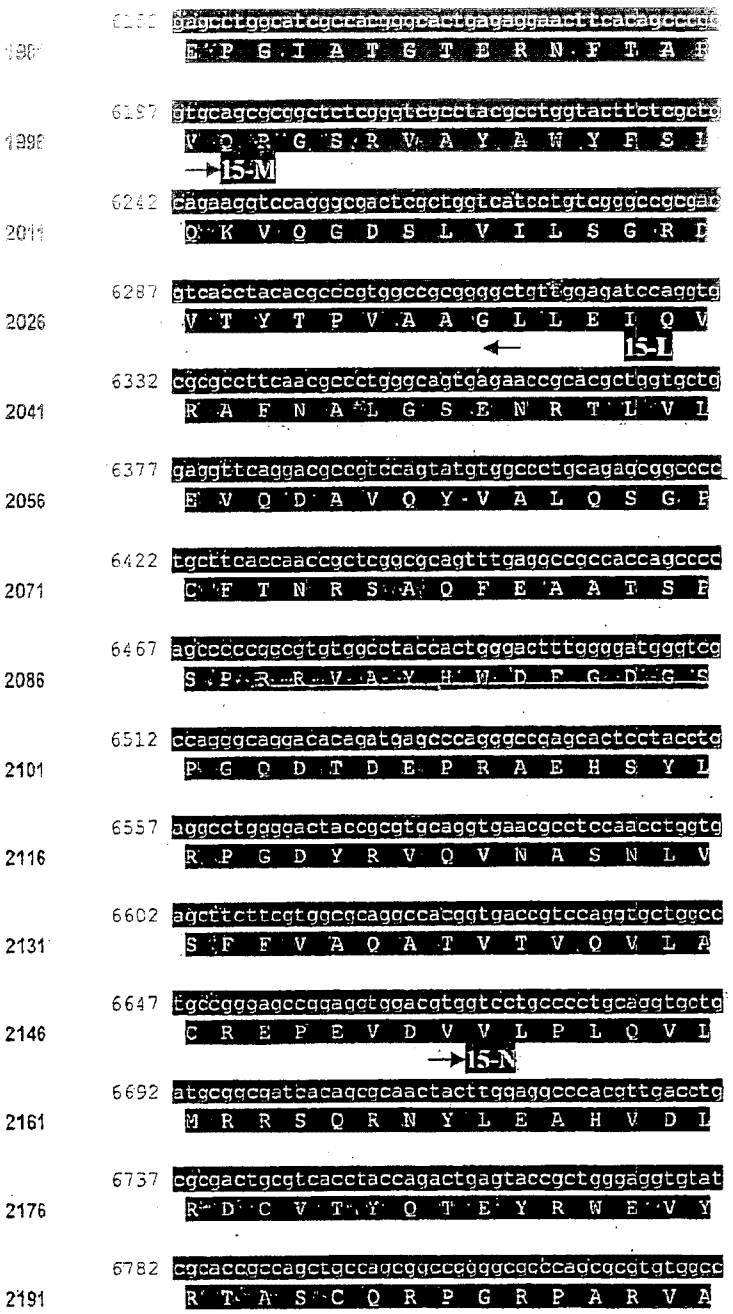


FIG. 3J

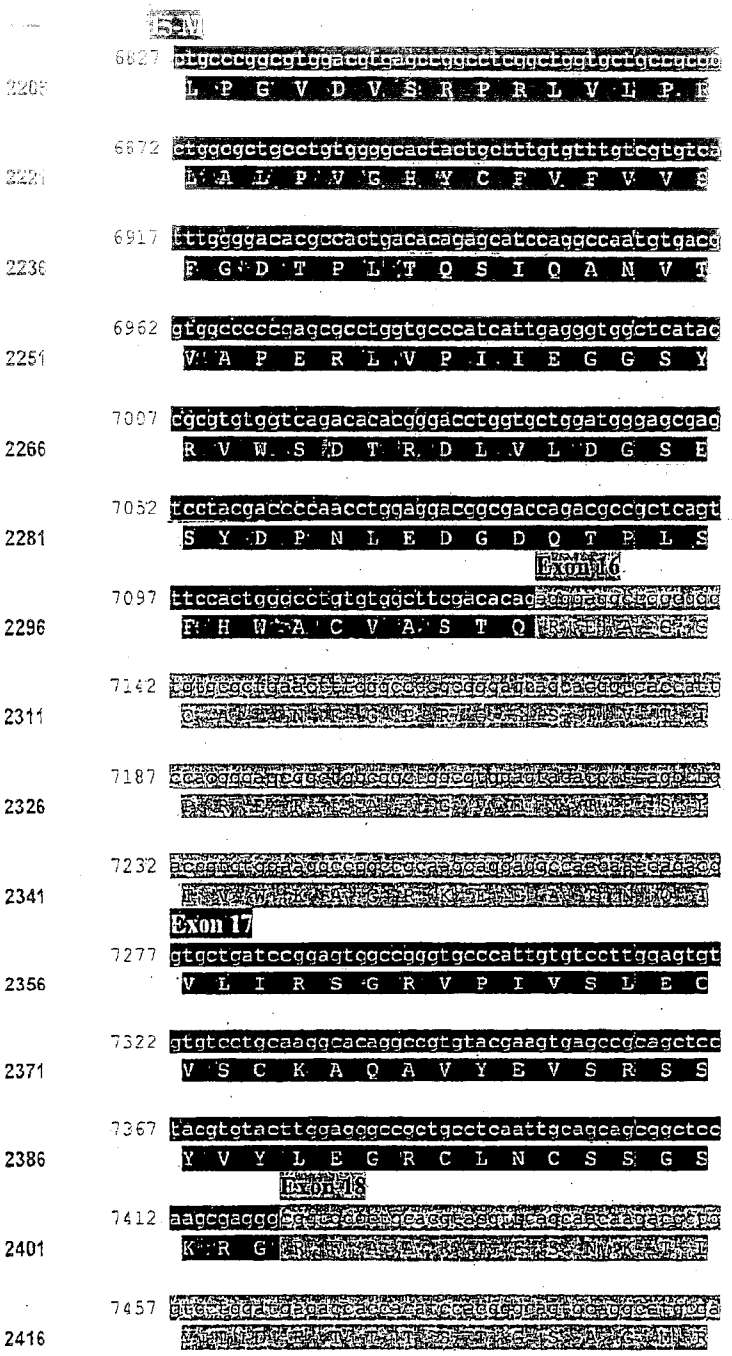


FIG. 3K

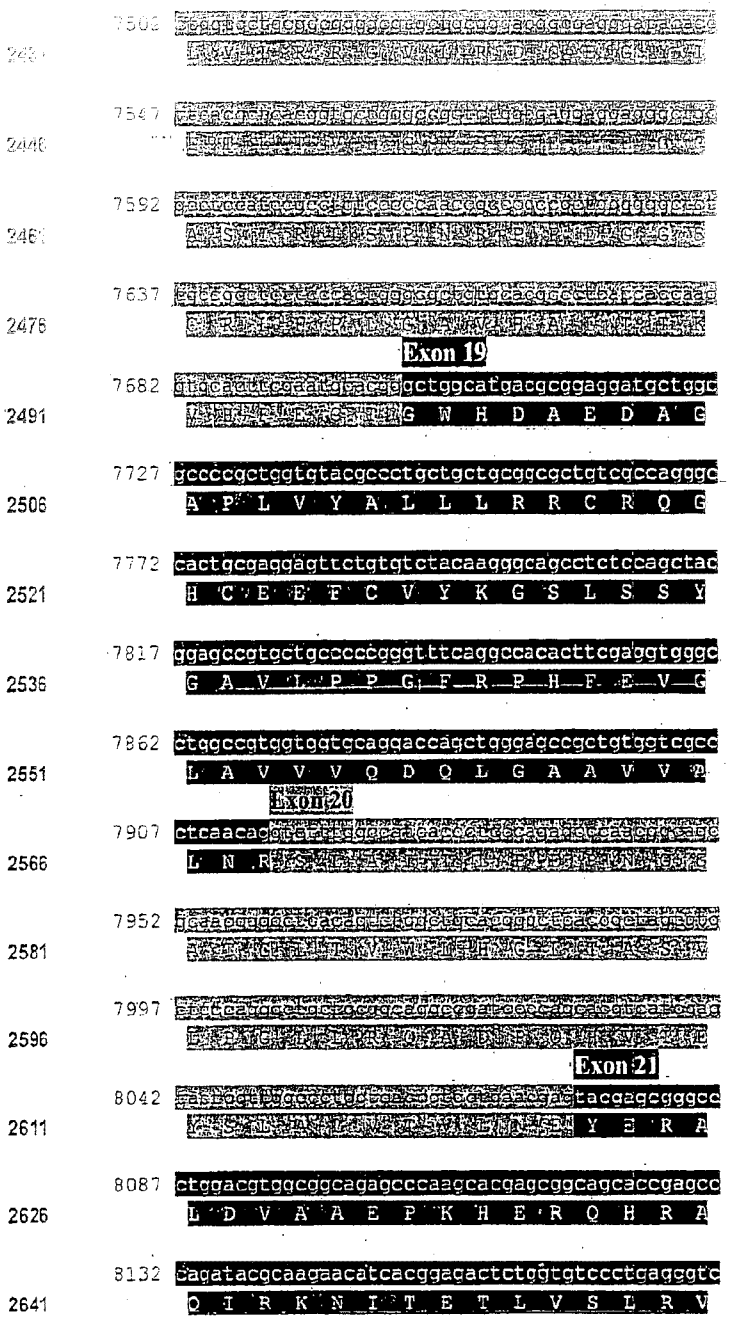


FIG. 3L

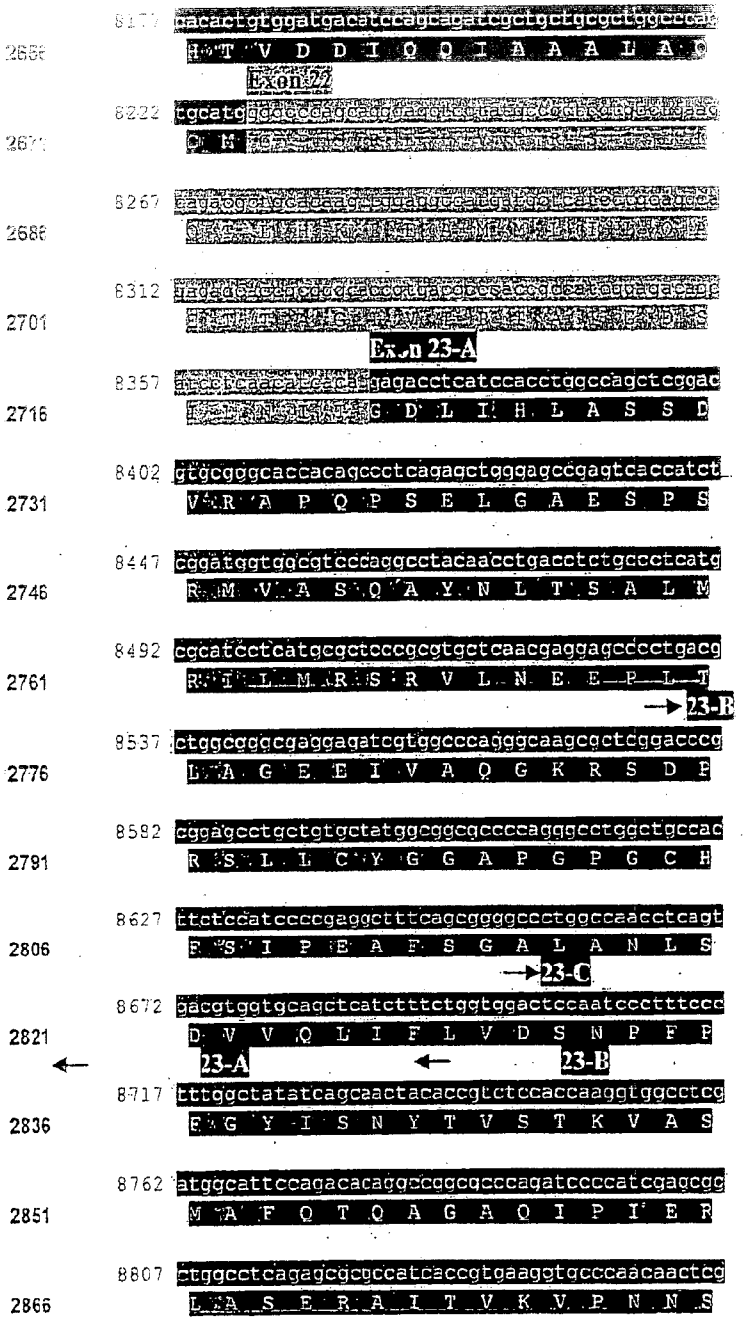


FIG. 3M

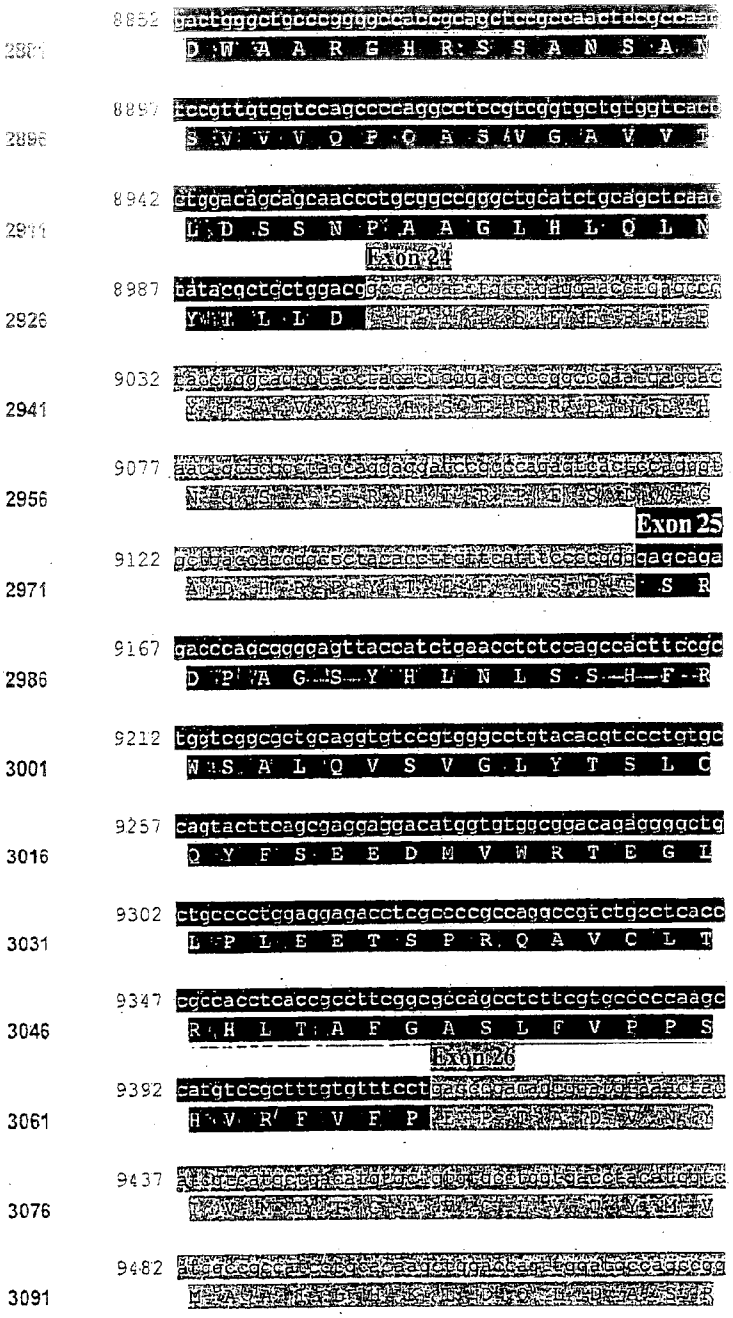


FIG. 3N

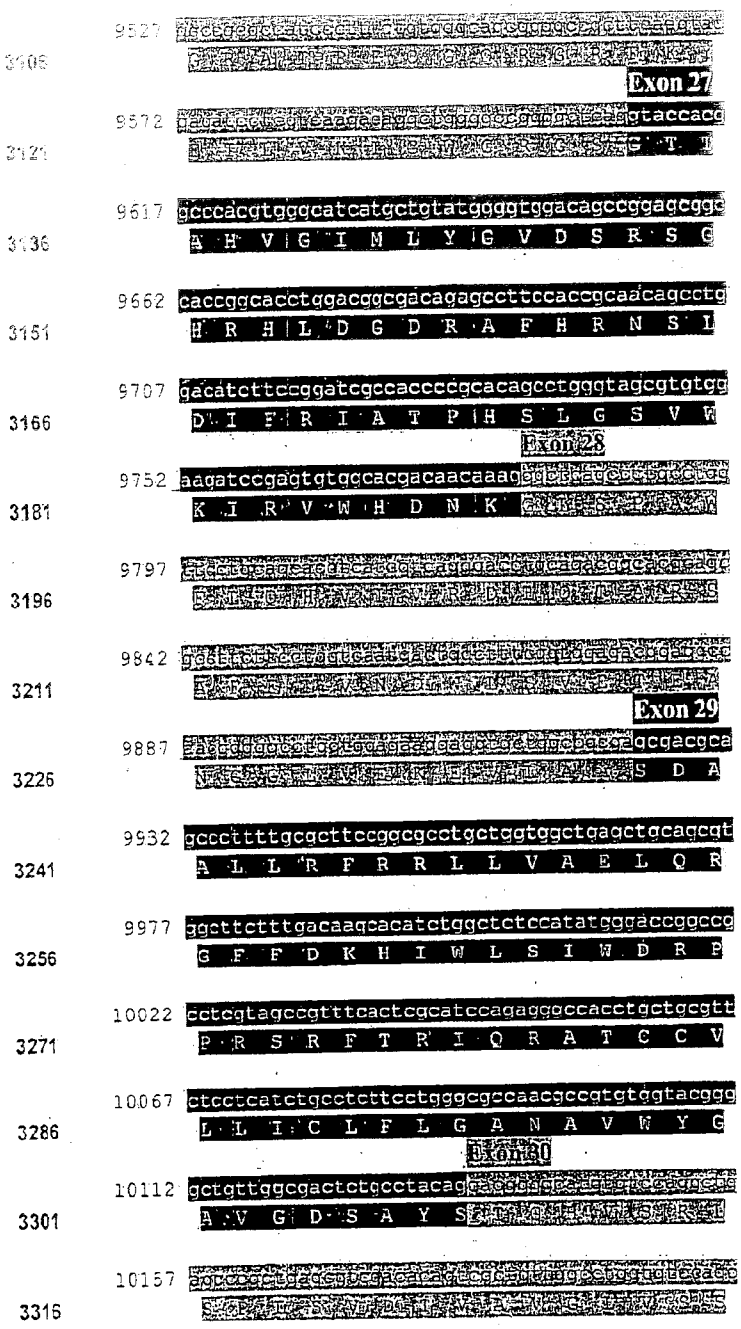


FIG. 3O

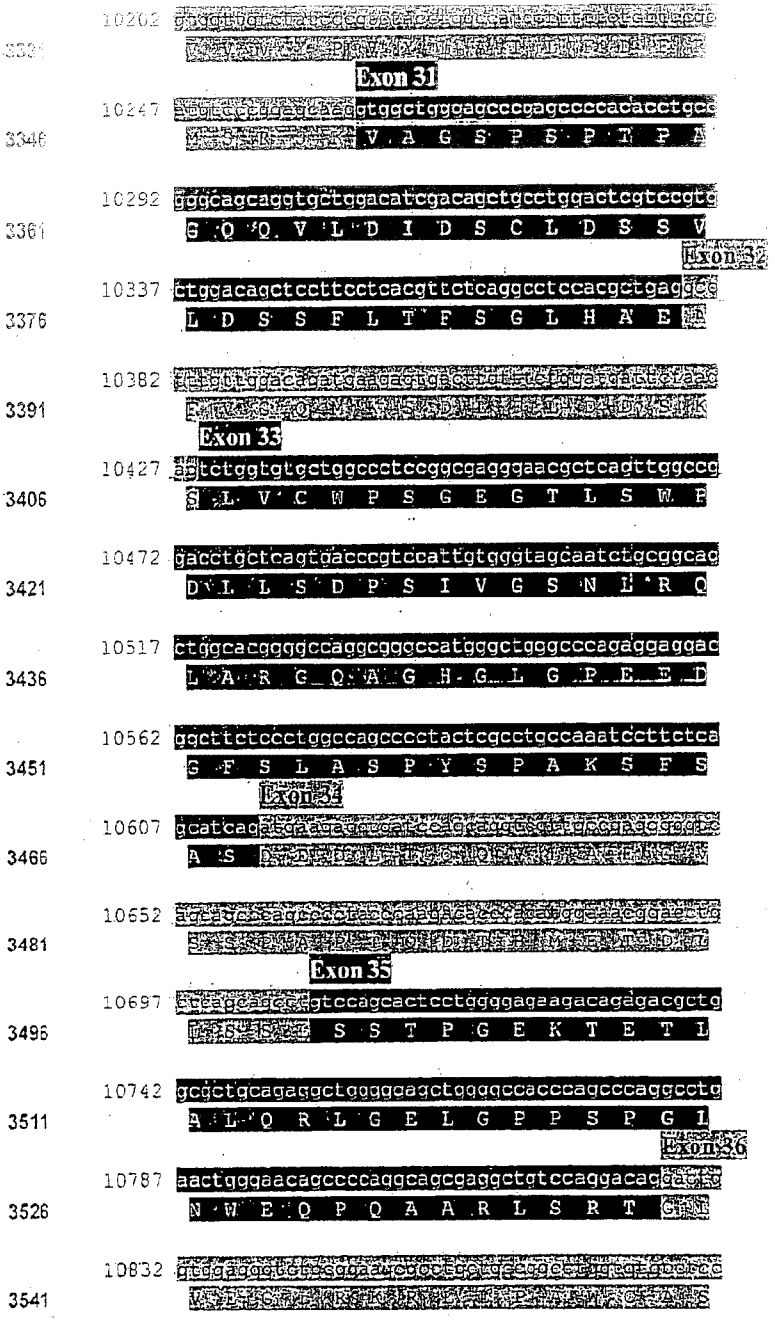


FIG. 3P

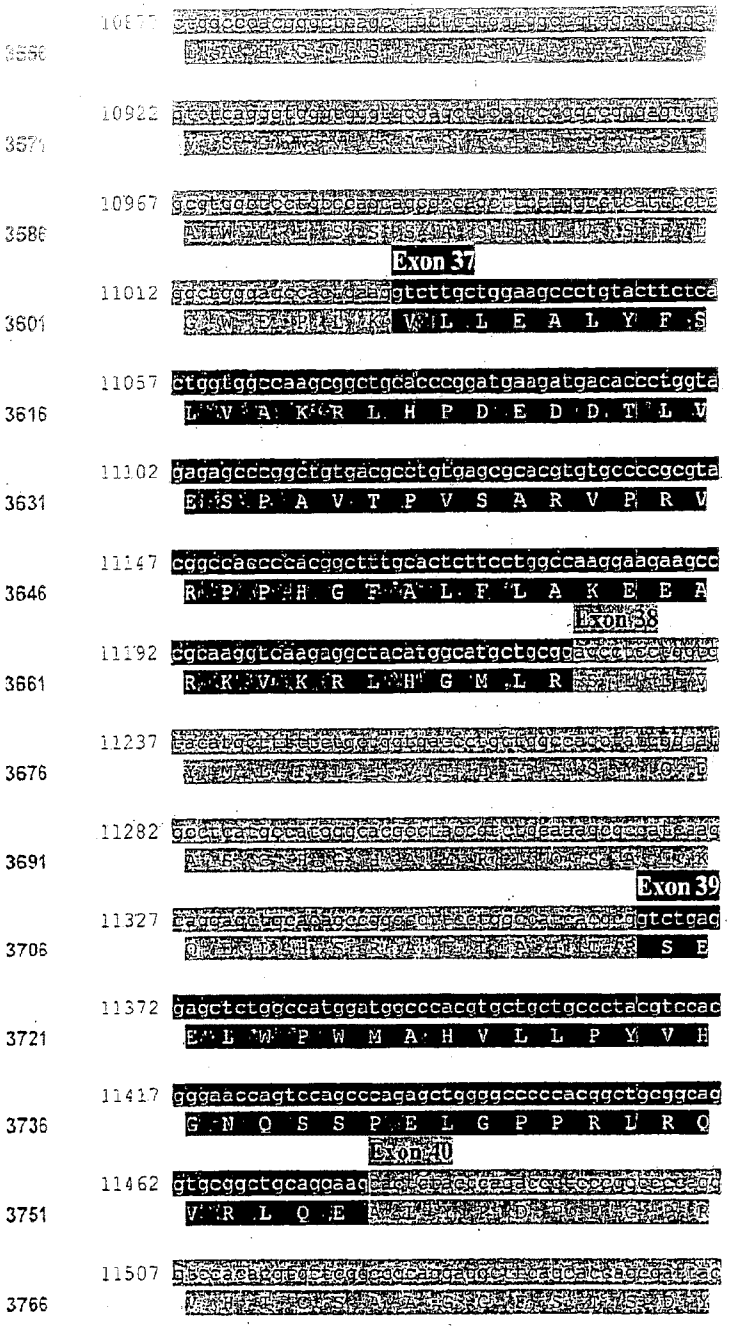


FIG. 3Q

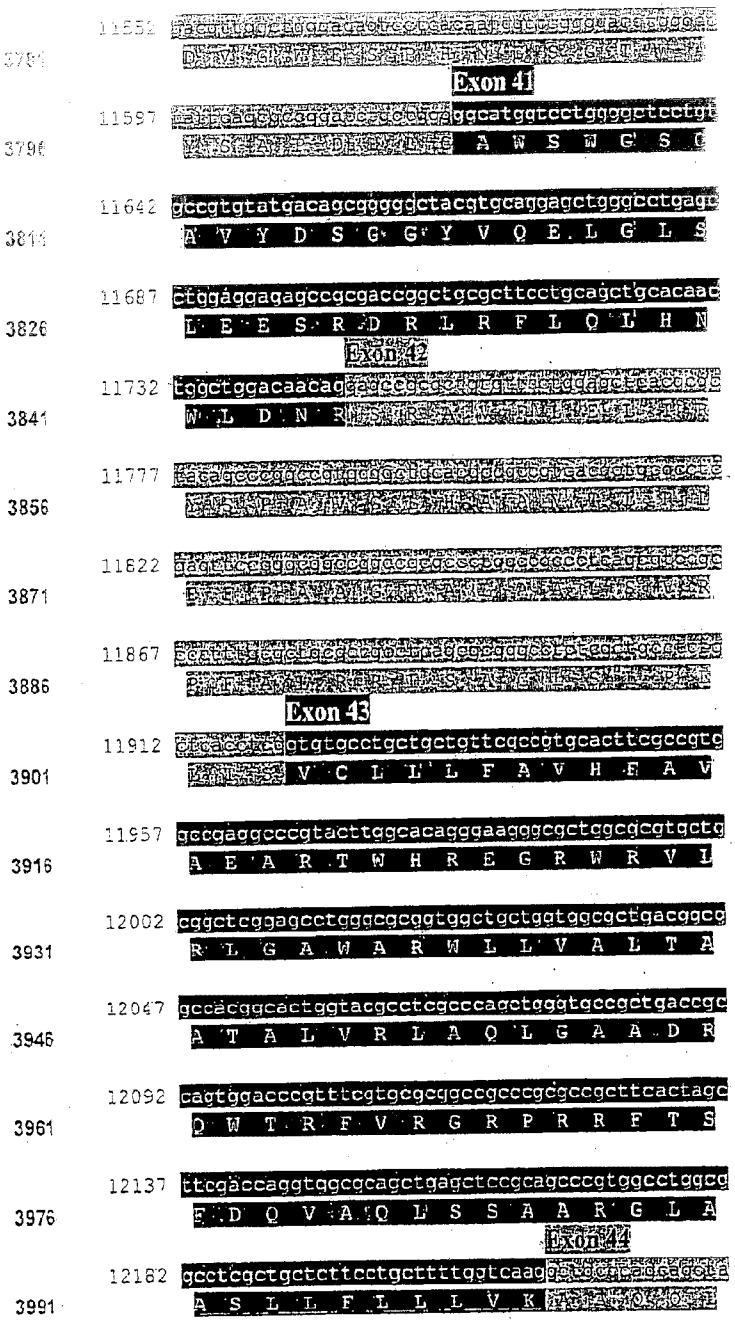


FIG. 3R

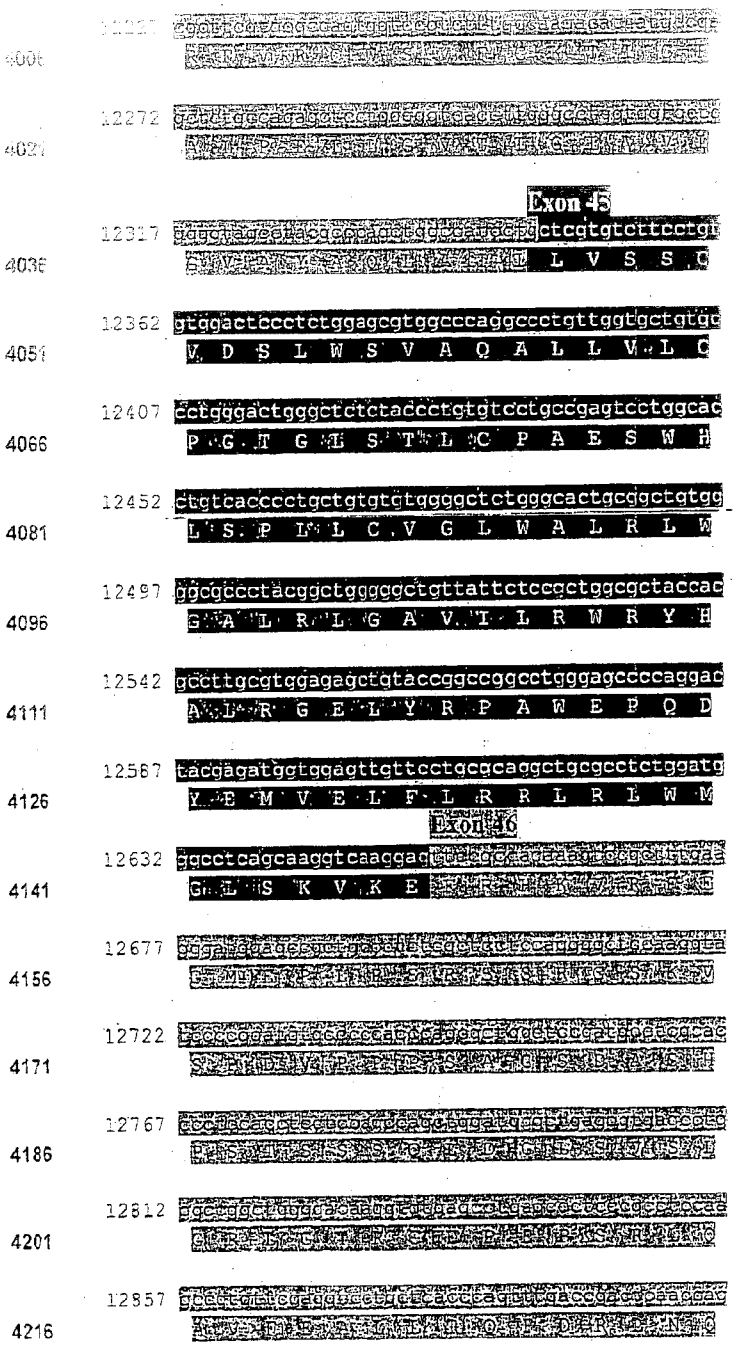


FIG. 3S

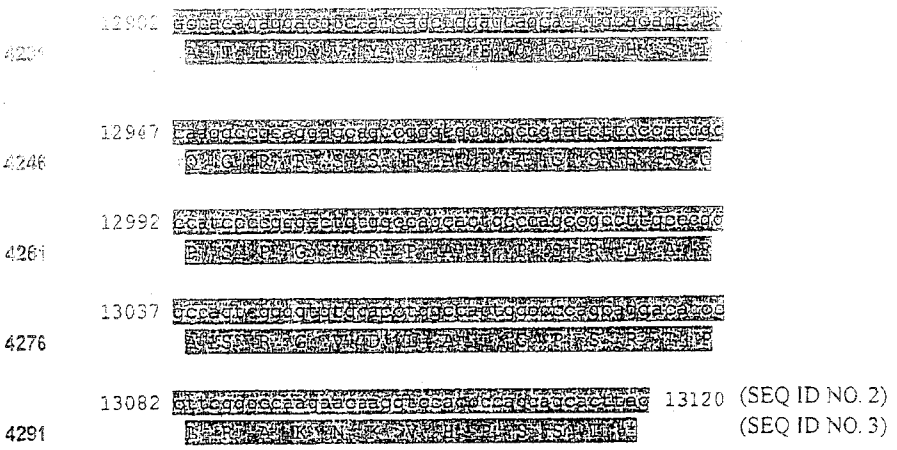

12902
4239
12947
4246
12992
4261
13037
4276
13082
4291
13120
(SEQ ID NO. 2)
(SEQ ID NO. 3)

FIG. 3T

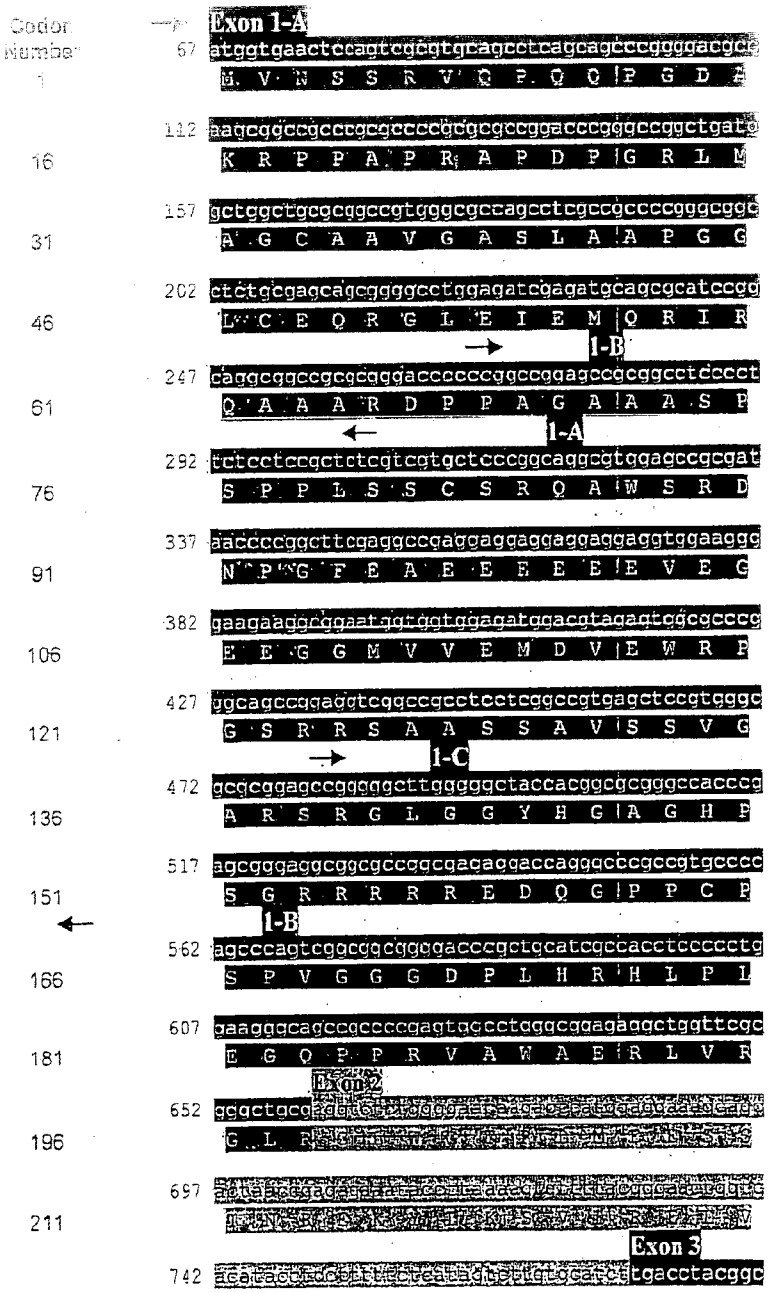


FIG. 4A

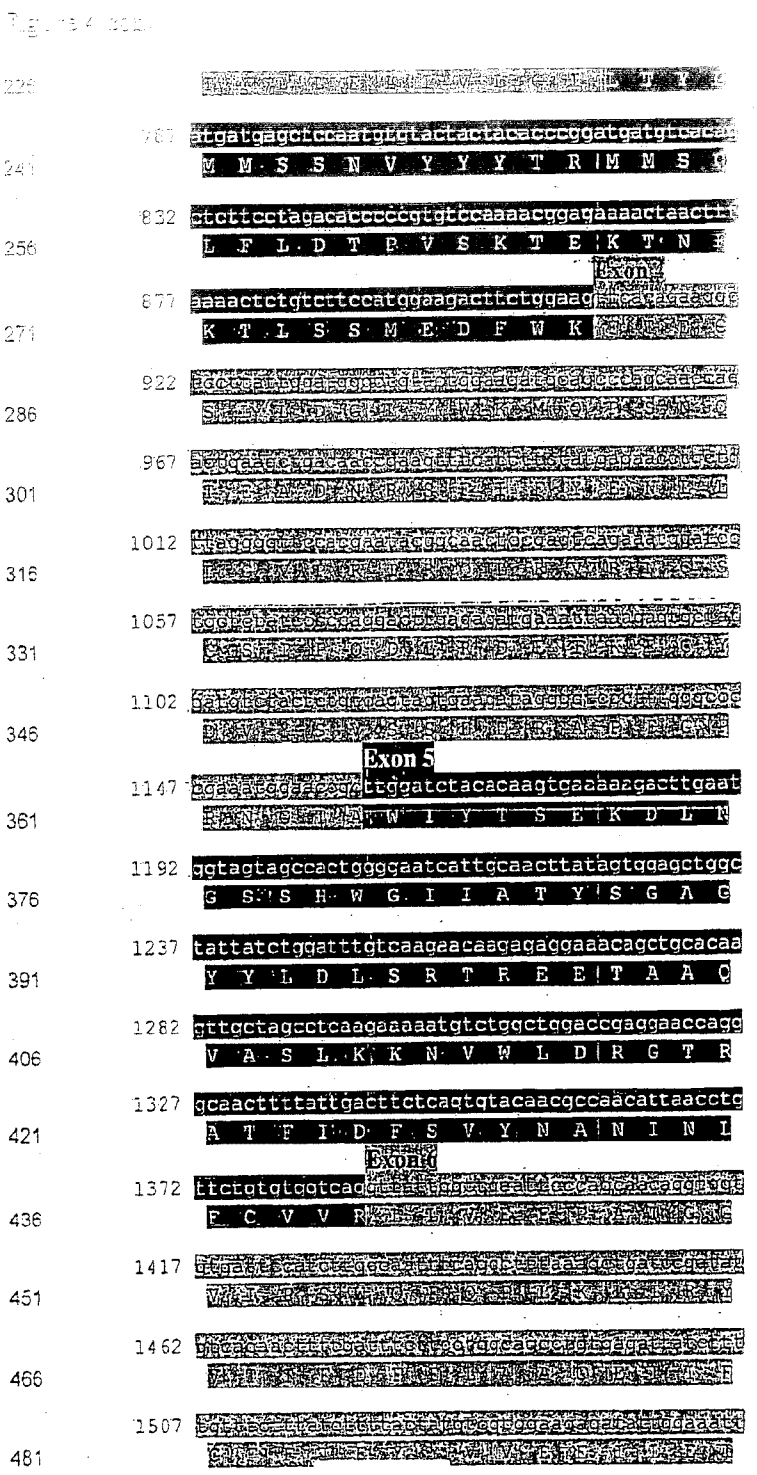


FIG. 4B

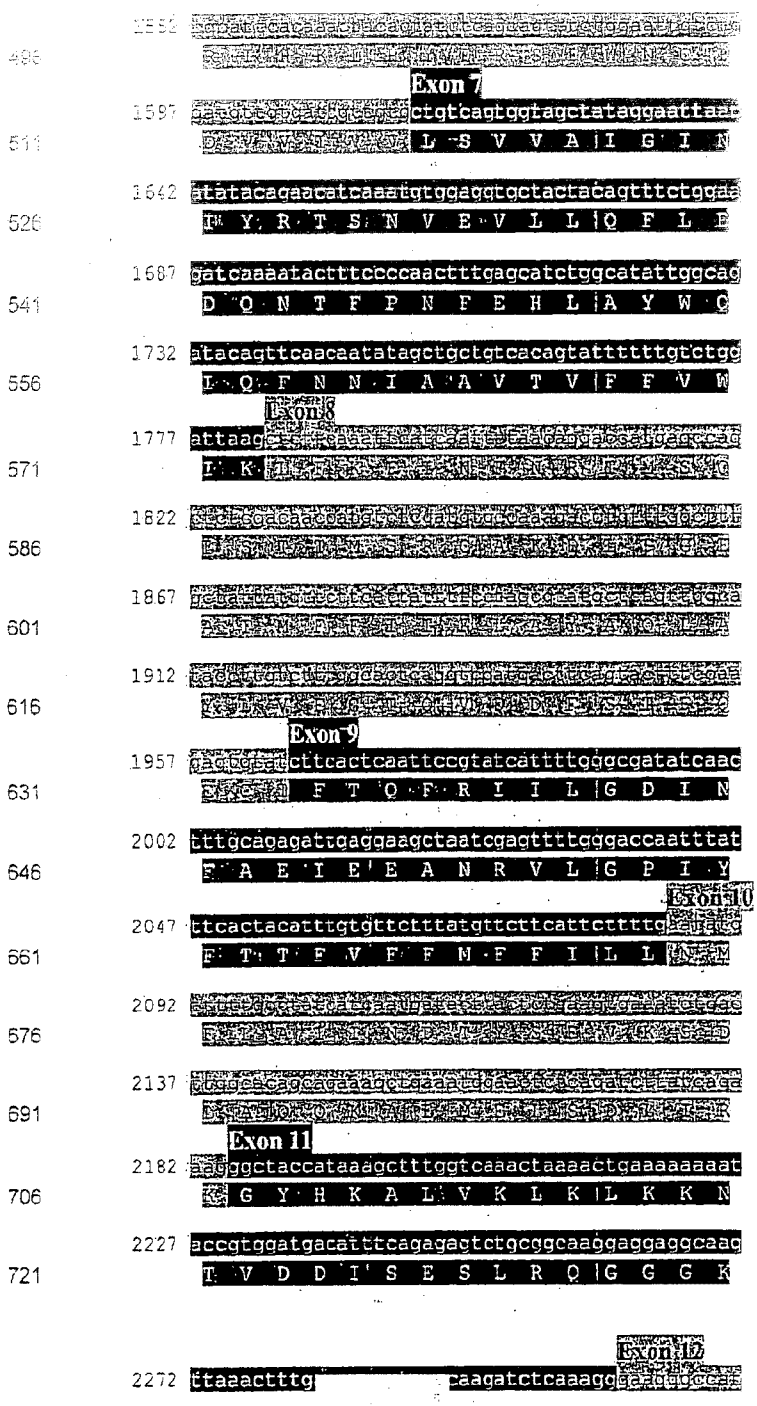


FIG. 4C

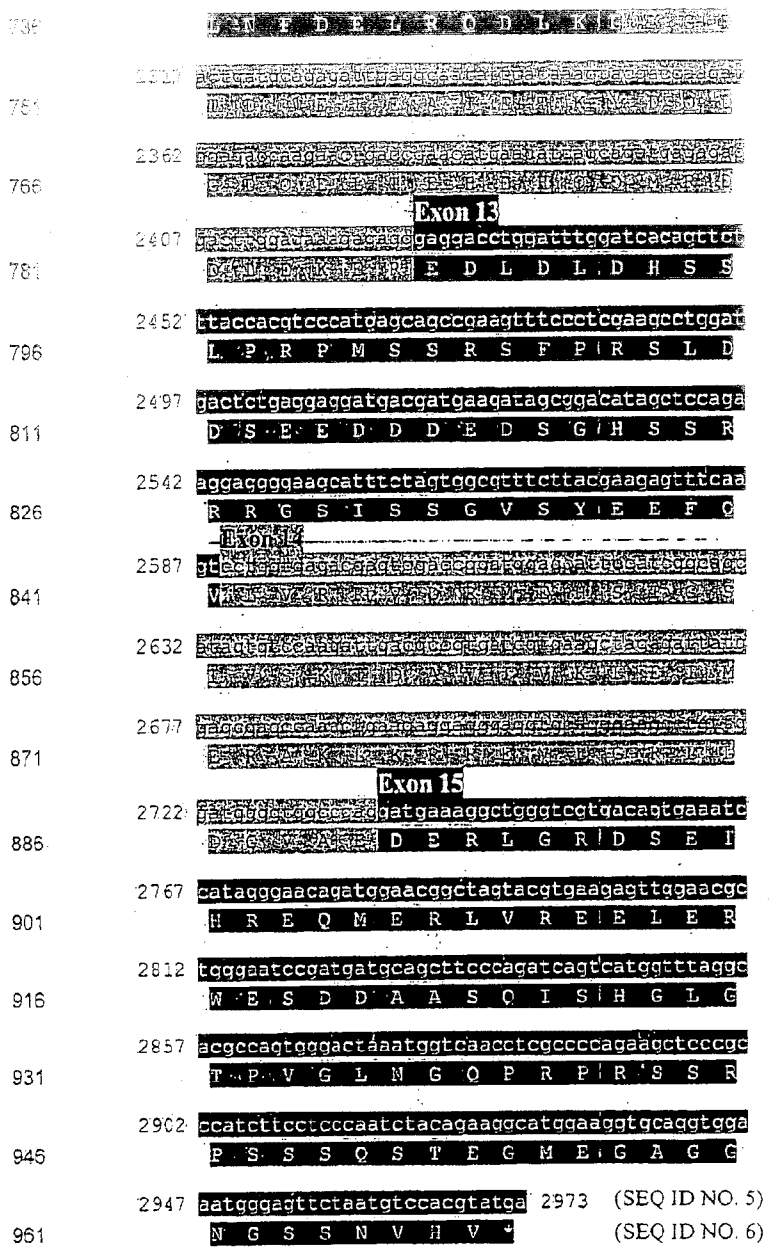
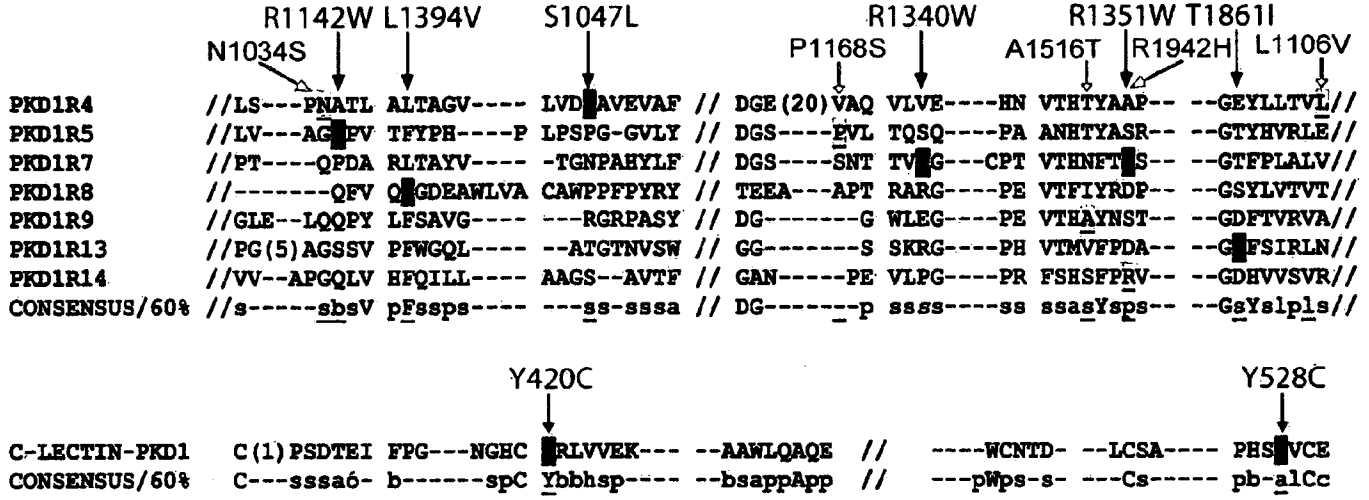


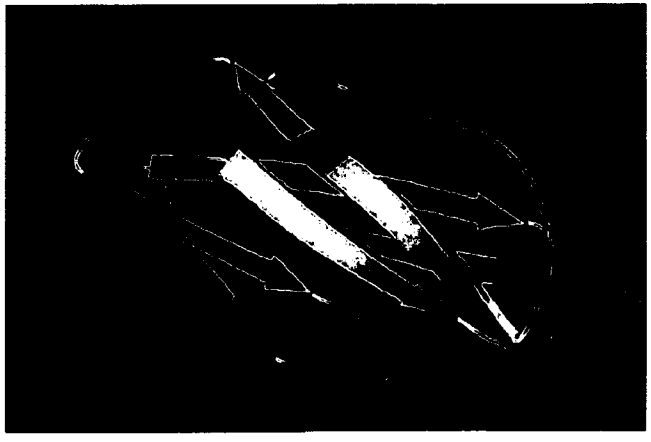
FIG. 4D

Figures 5A-5C

A)



B)



C)

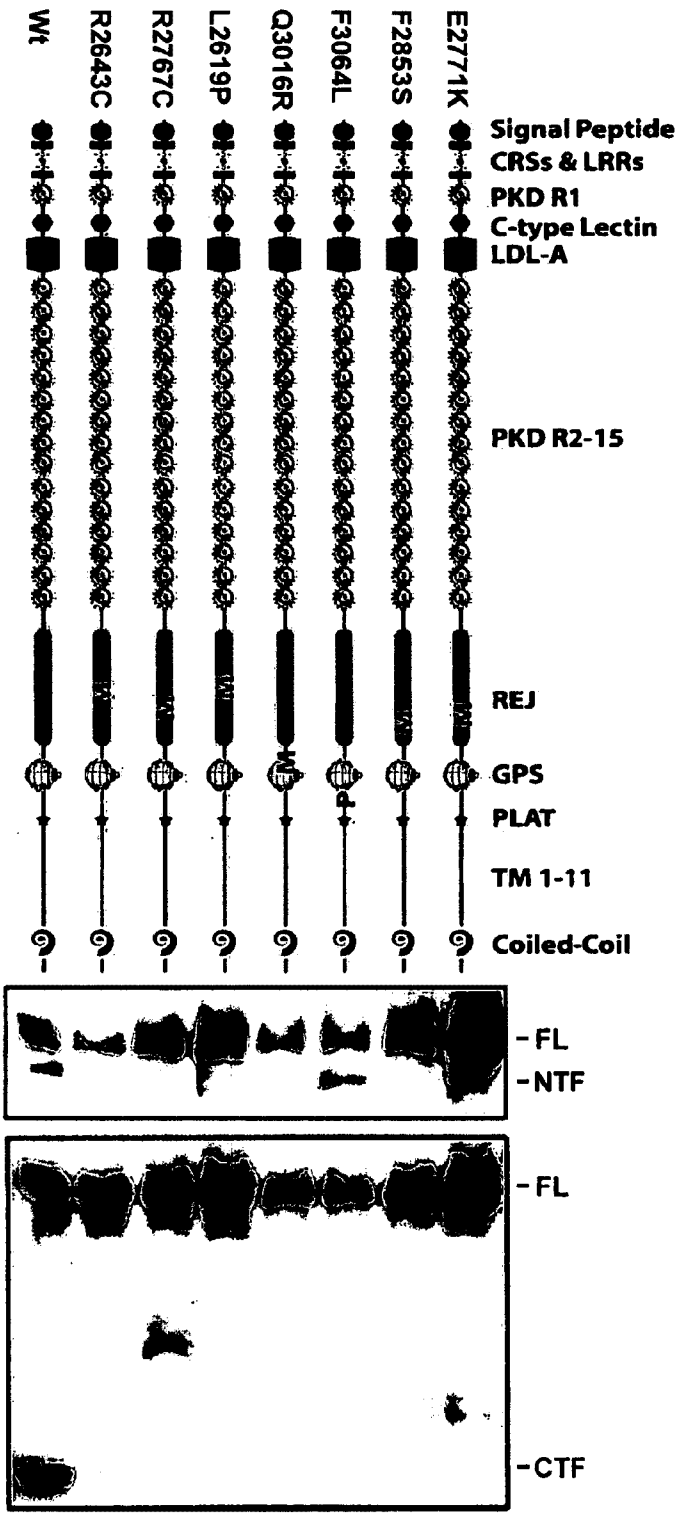


Figure 6

Date Reçue/Date Received 2021-09-14

Figure 7A

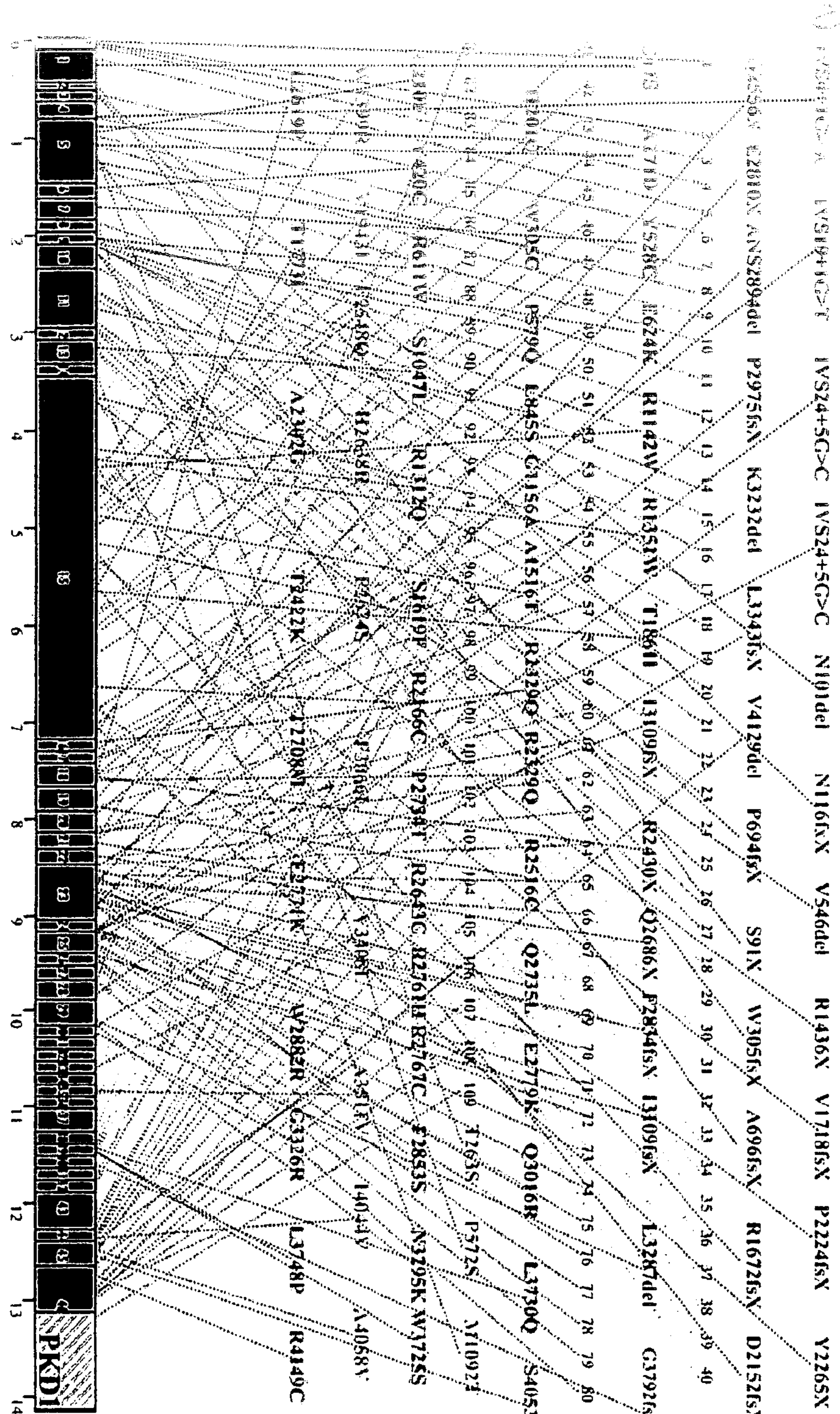
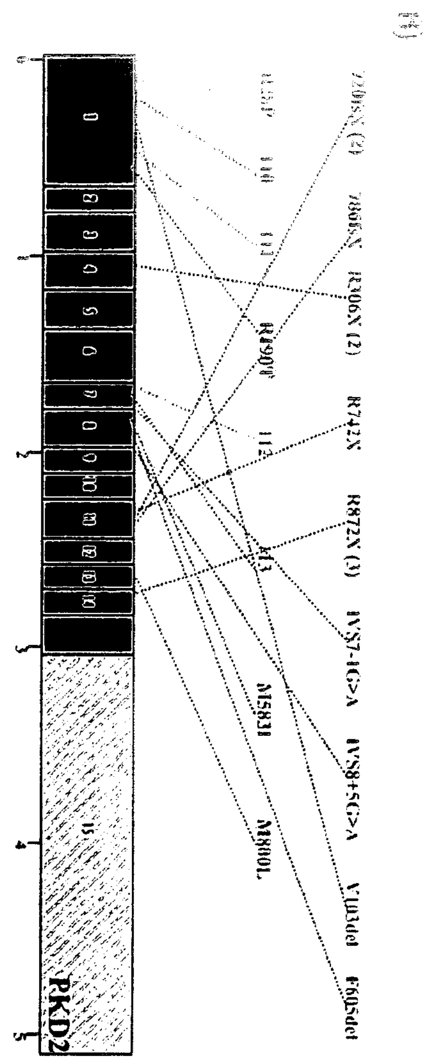

IVS24+5G>C
IVS24+5G>C
N101del
N116fsX
V546del
R1436X
V1718fsX
P2224fsX
E2810X
P2975fsX
K3232del
L3343fsX
V4129del
P694fsX
S91X
W305fsX
A696fsX
R1672fsX
D2152fsX
V526G
E624K
R1142W
R1351W
T1860I
I3109fsX
R2430X
Q2686X
F2834fsX
I3109fsX
L3287del
G3792fsX
W305C
P579Q
L845S
G1156A
A1516T
R2329Q
R2329Q
R2516C
Q2735L
E2779K
Q3016R
L3730Q
S4053F
R611W
S1047L
R1312Q
S1619P
R2166C
P2734T
R2649C
R2761H
R2767C
F2853S
N3295K
W3725S
T1870S
A2492T
T2422K
W2882R
C3326R
L3748P
R4149C
PKD1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14

Figure 7B


R306X (2)
R742X
R872X (3)
IVS7-1G>A
IVS8+5G>A
V1103del
F605del
M583I
M800L
PKD2

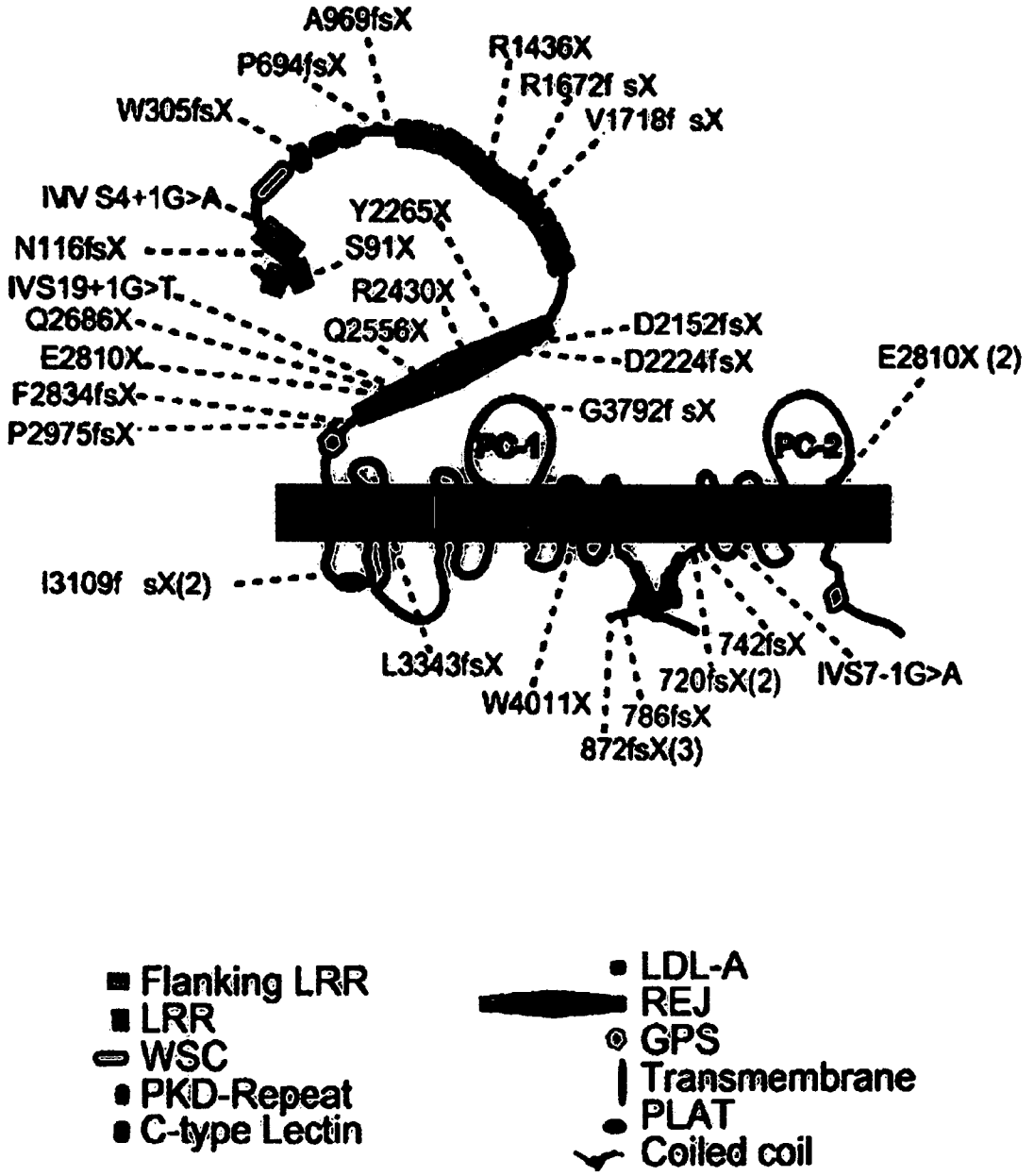

A969fsX
P694fsX
W305fsX
R1436X
R1672f sX
V1718f sX
IVS4+1G>A
N116fsX
Y2265X
S91X
IVS19+1G>T
R2430X
Q2686X
Q2556X
E2810X
F2834fsX
P2975fsX
D2152fsX
D2224fsX
E2810X (2)
G3792f sX
PC-1
PC-2
I3109f sX(2)
L3343fsX
W4011X
786fsX
872fsX(3)
720fsX(2)
742fsX
IVS7-1G>A
Flanking LRR
LRR
WSC
PKD-Repeat
C-type Lectin
LDL-A
REJ
GPS
Transmembrane
PLAT
Coiled coil

Figure 8A

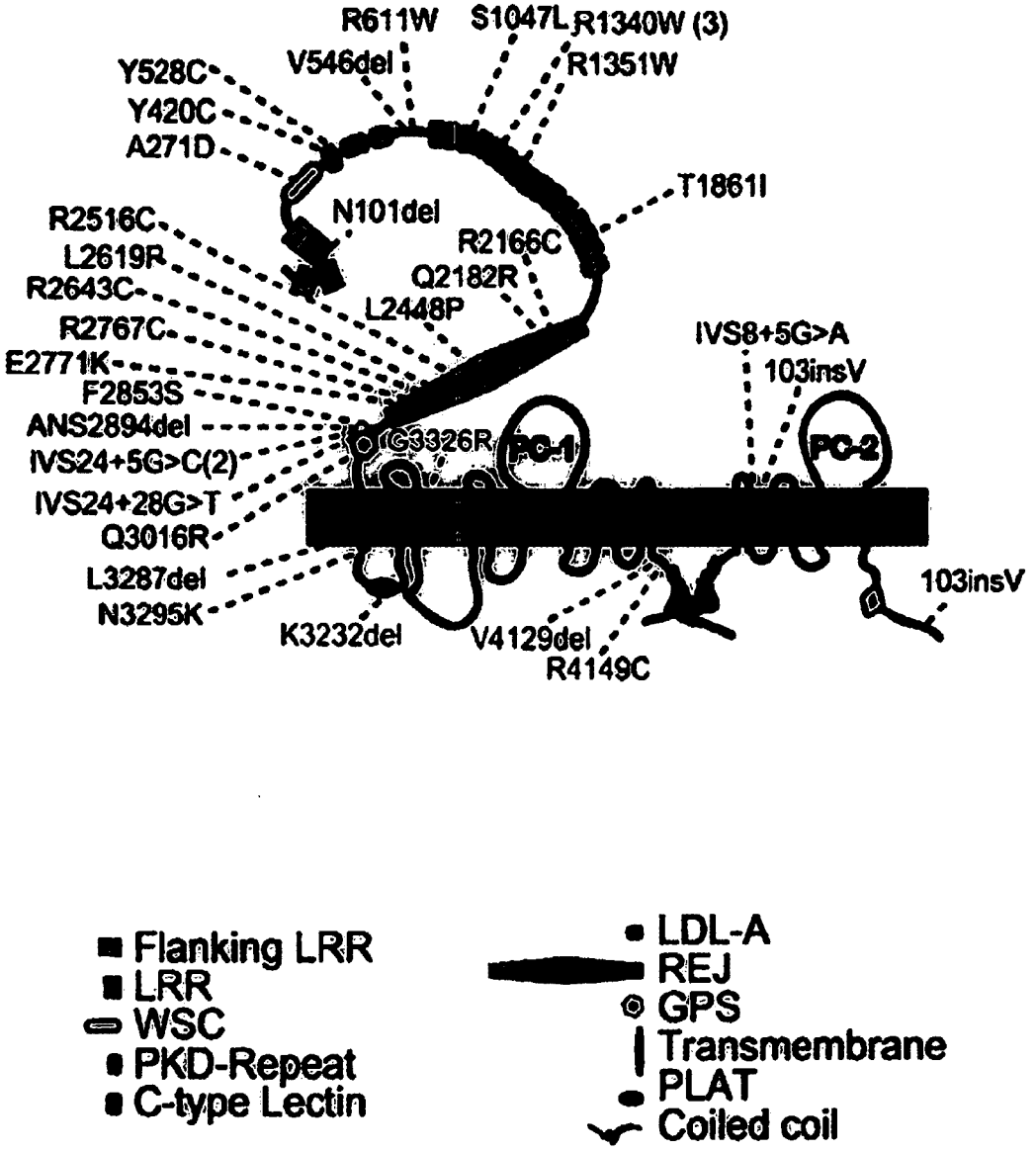


Figure 8B

Date Reçue/Date Received 2021-09-14