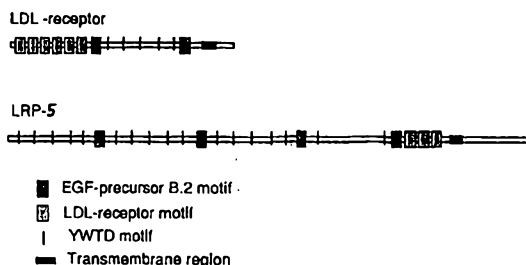| (12) | **PATENT** | (11) Application No. **AU 199870614 B2** |
| (19) | **AUSTRALIAN PATENT OFFICE** | (10) Patent No. **733722** |

(51) International Patent Classification 6 :

C12N 15/12, 15/11, C12Q 1/68, C07K 14/705, 16/28, A61K 38/17, 39/395, 48/00

A1

(11) International Publication Number: **WO 98/46743**

(43) International Publication Date: 22 October 1998 (22.10.98)

(21) International Application Number: PCT/GB98/01102

(22) International Filing Date: 15 April 1998 (15.04.98)

(30) Priority Data:
60/043,553	15 April 1997 (15.04.97)	US
60/048,740	5 June 1997 (05.06.97)	US

(71) Applicants (for all designated States except US): THE WELL-COME TRUST LIMITED AS TRUSTEE TO THE WELLCOME TRUST [GB/GB]; 183 Euston Road, London NW1 2BE (GB). MERCK & CO., INC. [US/US]; 126 Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): TODD, John, Andrew [GB/GB]; 64 Rousham Road, Tackley, Oxon OX5 3AJ (GB). HESS, John, Wilfred [US/US]; 566 Constitution Road, Lansdale, PA 19446 (US). CASKEY, Charles, Thomas [US/US]; 6402 Bellmont, Houston, TX 66005 (US). COX, Roger, David [GB/GB]; Peace Cottages, 7 Park Road, Banbury, Oxon OX16 0DW (GB). GERHOLD, David [US/US]; 730 Tranquility Lane, Lansdale, PA 19446 (US). HAMMOND, Holly [US/US]; 621 Melvins Road, Telford, PA 18969 (US). HEY, Patricia [US/US]; 1133 Bloomfield Circle, Lansdale, PA 19446 (US). KAWAGUCHI, Yoshihiko [JP/JP]; Yamadanishi 3-33-B-206, Suita-shi, Osaka 565 (JP). MERRIMAN, Tony, Raymond [NZ/GB]; 16A Cromwell Close, Marston, Oxford OX3 0RW (GB). METZKER, Michael, Lee [US/US]; 7400 Encampment Road, Lansdale, PA 19446 (US). NAKAGAWA, Yusuke [JP/GB]; 17 Latimer Grange, Latimer Road, Headington, Oxford OX3 7PQ (GB). PHILLIPS, Michael, Sean [CA/US]; 608 Poplar Court, Lansdale, PA 19446 (US). TWELLS, Rebecca, Christina, Joan [GB/GB]; 75 Park Street, Thame, Oxon OX9 3HU (GB).

(74) Agents: WALTON, Seán, M. et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NOVEL LDL-RECEPTOR

Motif Organization of the LDL- receptor and LRP-5

(57) Abstract

A novel receptor, "LDL-receptor related protein-3" ("LRP-3"), is provided, along with encoding nucleic acid. The gene is associated with type 1 diabetes (insulin dependent diabetes mellitus), and experimental evidence provides indication that it is the IDDM susceptibility gene *IDDM4*. In various aspects the invention provides nucleic acid, including coding sequences, oligonucleotide primers and probes, polypeptides, pharmaceutical compositions, methods of diagnosis or prognosis, and other methods relating to and based on the gene, including methods of treatment of diseases in which the gene may be implicated, including autoimmune diseases, such as glomerulonephritis, diseases and disorders involving disruption of endocytosis and/or antigen presentation, diseases and disorders involving cytokine clearance and/or inflammation, viral infection, elevation of free fatty acids or hypercholesterolemia, osteoporosis, Alzheimer's disease, and diabetes.

NOVEL LDL-RECEPTOR

FIELD OF THE INVENTION

The present invention relates to nucleic acids,
5 polypeptides, oligonucleotide probes and primers, methods of
diagnosis or prognosis, and other methods relating to and
based on the identification of a gene, which is characterised
as a member of the LDL-receptor family and for which there are
indications that some alleles are associated with
10 susceptibility to insulin-dependent diabetes mellitus
("IDDM"), also known as type 1 diabetes.

More particularly, the present invention is based on
cloning and characterisation of a gene which the present
inventors have termed "*LDL-receptor related protein-5 (LRP5)*"
15 (previously "*LRP-3*"), based on characteristics of the encoded
polypeptide which are revealed herein for the first time and
which identify it as a member of the LDL receptor family.
Furthermore, experimental evidence is included herein which
provides indication that *LRP5* is the IDDM susceptibility gene
20 *IDDM4*.

BACKGROUND OF THE INVENTION

Diabetes, the dysregulation of glucose homeostasis,
affects about 6% of the general population. The most serious
25 form, type 1 diabetes, which affects up to 0.4% of European-
derived population, is caused by autoimmune destruction of the
insulin producing β-cells of the pancreas, with a peak age of
onset of 12 years. The β-cell destruction is irreversible,
and despite insulin replacement by injection patients suffer
30 early mortality, kidney failure and blindness (Bach, 1994;
Tisch and McDevitt, 1996). The major aim, therefore, of
genetic research is to identify the genes predisposing to type
1 diabetes and to use this information to understand disease
mechanisms and to predict and prevent the total destruction
35 of β-cells and the disease.

The mode of inheritance of type 1 diabetes does not
follow a simple Mendelian pattern, and the concordance of

2

susceptibility genotype and the occurrence of disease is much
less than 100%, as evidenced by the 30-70% concordance of
identical twins (Matsuda and Kuzuya, 1994; Kyvik et al, 1995).
Diabetes is caused by a number of genes or polygenes acting
5 together in concert, which makes it particularly difficult to
identify and isolate individual genes.

The main IDDM locus is encoded by the major histo-
compatibility complex (MHC) on chromosome 6p21 (*IDDM1*). The
degree of familial clustering at this locus, $\lambda s = 2.5$, where
10 $\lambda s = P$ expected [sharing of zero alleles at the locus
identical-by-descent (IBD)]/P observed [sharing of zero
alleles IBD] (Risch 1987; Todd, 1994), with a second locus on
chromosome 11p15, *IDDM2*, the insulin minisatellite $\lambda s = 1.25$
(Bell et al, 1984; Thomson et al, 1989; Owerbach et al, 1990;
15 Julier et al, 1991; Bain et al, 1992; Spielman et al, 1993;
Davies et al, 1994; Bennett et al, 1995). These loci were
initially detected by small case control association studies,
based on their status as functional candidates, which were
later confirmed by further case-control, association and
20 linkage studies.

These two loci, however, cannot account for all the
observed clustering of disease in families ($\lambda s = 15$), which is
estimated from the ratio of the risk for siblings of patients
and the population prevalence (6%/0.4%) (Risch, 1990). We
25 initiated a positional cloning strategy in the hope of
identifying the other loci causing susceptibility to type 1
diabetes, utilising the fact that markers linked to a disease
gene will show excess of alleles shared identical-by-descent
in affected sibpairs (Penrose, 1953; Risch, 1990; Holmans,
30 1993).

The initial genome-wide scan for linkage utilising 289
microsatellite markers, in 96 UK sibpair families, revealed
evidence of linkage to an additional eighteen loci (Davies et
al, 1994). Confirmation of linkage to two of these loci was
35 achieved by analysis of two additional family sets (102 UK
families and 84 USA families), *IDDM4* on chromosome 11q13 (MLS
1.3 , P = 0.003 at *FGF3*) and *IDDM5* on chromosome 6q (MLS 1.8

at *ESR*). At *IDDM4* the most significant linkage was obtained in the subset of families sharing 1 or 0 alleles IBD at HLA (MLS = 2.8; P=0.001; λs = 1.2) (Davies et al, 1994). This linkage was also observed by Hashimoto *et al* (1994) using 251 affected sibpairs, obtaining P= 0.0008 in all sibpairs. Combining these results, with 596 families, provides substantial support for *IDDM4* (P = 1.5X10-6) (Todd and Farrall, 1996; Luo et al, 1996).

BRIEF DESCRIPTION OF THE INVENTION

The present inventors now disclose for the first time a gene encoding a novel member of the LDL-receptor family, which they term "*LRP5*" (previously "*LRP-3*"). Furthermore, evidence indicates that the gene represents the IDDM susceptibility locus *IDDM4*, the identification and isolation of which is a major scientific breakthrough.

Over the last 10 years many genes for single gene or monogenic diseases, which are relatively rare in the population, have been positioned by linkage analysis in families, and localised to a small enough region to allow identification of the gene. The latter sublocalisation and fine mapping can be carried out in single gene rare diseases because recombinations within families define the boundaries of the minimal interval beyond any doubt. In contrast, in common diseases such as diabetes or asthma the presence of the disease mutation does not always coincide with the development of the disease: disease susceptibility mutations in common disorders provide risk of developing of the disease, and this risk is usually much less than 100%. Hence, susceptibility genes in common diseases cannot be localised using recombination events within families, unless tens of thousands of families are available to fine map the locus. Because collections of this size are impractical, investigators are contemplating the use of association mapping, which relies on historical recombination events during the history of the population from which the families came from.

4

Association mapping has been used in over a dozen examples of rare single gene traits, and particularly in genetically isolated populations such as Finland to fine map disease mutations. Nevertheless, association mapping is
5 fundamentally different from straightforward linkage mapping because even though the degree of association between two markers or a marker and a disease mutation is proportional to the physical distance along the chromosome this relationship can be unpredictable because it is dependent on the allele
10 frequencies of the markers, the history of the population and the age and number of mutations at the disease locus. For rare, highly penetrant single gene diseases there is usually one major founder chromosome in the population under study, making it relatively feasible to locate an interval that is
15 smaller than one that can be defined by standard recombination events within living families. The resolution of this method in monogenic diseases in which there is one main founder chromosome is certainly less than 2cM, and in certain examples the resolution is down to 100 kb of DNA (Hastbacka et al.
20 (1994) Cell 78,1-20).

In common diseases like type 1 diabetes, which are caused by a number of genes or polygenes acting together in concert the population frequency of the disease allele may be very high, perhaps exceeding 50%, and there are likely to be
25 several founder chromosomes, all of which impart risk, and not a 100% certainty of disease development. Because association mapping is dependent on unpredictable parameters, and because founder chromosomes will be several and common in frequency in the general population, the task of fine mapping polygenes is
30 currently one of some controversy, and many doubt the feasibility at all of a systematic genetic approach using a combination of linkage and association mapping. Recently, Risch and Marakandis have provided some mathematical background to the feasibility of association mapping in
35 complex diseases (Science 273 1516-1517, 1996) but they did not take into account the effect of multiple founder chromosomes.

As a result of these uncertainties, extremely large
numbers of diabetic families are required for genotyping, with
a large number of markers across a specific region, giving a
linkage disequilibrium curve which may have several peaks.
5 The question is, which peak identifies the aetiological
mutation, and in what ways can we establish this? To our
knowledge, the linkage disequilibrium curves and haplotype
association maps shown in Figures 3, 4, 19 and 20 are the
first of their kind for any complex polygenic disease for any
10 locus. Curves of this nature have not been published yet in
the literature, even for the well-established *IDDM1/MHC* locus.
In this respect the work described here is entirely novel and
at the cutting edge of research into the genetics of
polygenes.
15

BRIEF DESCRIPTION OF THE FIGURES
      Figure 1 illustrates approximate localisation of *IDDM4* on
chromosome 11q13. Multipoint linkage map of maximum
likelihood IBD in a subgroup of HLA 1:0 sharers in 150
20 families. MLS of 2.3 at *FGF3* and *D11S1883* (λs = 1.19) were
obtained (Davies *et al* (1994) *Nature* 371: 130-136).
      Figure 2 shows a physical map of the region *D11S987* -
Galanin on chromosome 11q13. The interval was cloned in pacs,
bacs and cosmids, and restriction mapped using a range of
25 restriction enzymes to determine the physical distance between
each marker.
      Figure 3 shows a single-point linkage disequilibrium
curve at the *IDDM4* region. 1289 families were analysed by
TDT, with a peak at *H0570POLYA*,) P=0.001. x-axis: physical
30 distance in kb; y-axis: TDT $\chi2$ statistic (tdf).
      Figure 4 shows a three-point rolling linkage
disequilibrium curve at *IDDM4*, with 1289 families, from four
different populations (UK, USA, Sardinia and Norway). In
order to minimise the effects of variation in allele frequency
35 at each polymorphism, the TDT data was obtained at three
consecutive markers, and expressed as an average of the three.
x-axis: physical distance in kb; y-axis: TDT $\chi2$ statistic.
      Figure 5(a) shows DNA sequence of the *LRP5* isoform 1

cDNA.

Figure 5(b) shows the DNA sequence of the longest open
reading frame present in the LRP5 cDNA.

Figure 5(c) amino acid sequence translation (in standard
5 single letter code) of the open reading frame in Figure 5 (b).

Figure 5(d) motifs of LRP5 isoform 1, encoded by the open
reading frame contained in Figure 5(b). Symbols: Underlined
residues 1-24 contain a signal for protein export and
cleavage, ▼ indicates the position of an intron/exon boundary,
10 * indicates a putative N-linked glycosylation site in the
proposed extracellular portion of the receptor. The EGF-
binding motifs are shaded light gray, LDL-receptor ligand
motifs are shaded a darker gray. The spacer regions are
indicated by the underlined four amino acids with high
15 similarity to the YWTD motif. A putative transmembrane
spanning domain is underlined with a heavy line. Areas shaded
in the cytoplasmic domain (1409 to end) may be involved in
endocytosis.

Figure 5(e) amino acid sequence of the mature LRP5
20 protein.

Figure 5(f) shows the comparison of the nucleotide
sequence of the first 432 nucleotides of the 5' end of the
human isoform1 cDNA sequence (Figure 5(a)) on the upper line
with the first 493 nucleotides of the 5' end of the mouse Lrp5
25 cDNA sequence (Figure 16(a)) on the lower line. The
comparison was performed using the GCG algorithm GAP (Genetics
Computer Group, Madison, WI).

Figure 5(g) shows the comparison of the first 550 amino
acids of human LRP5 isoform 1 with the first 533 amino acids
30 of mouse Lrp5 using the GCG algorithm GAP (Genetics Computer
Group, Madison, WI).

Figure 6(a) shows the amino acid sequence of LRP5 motifs.
A comparison was made using the program crossmatch (obtained
from Dr. Phil Green, University of Washington) between the
35 motifs present in LRP1 and the LRP5 amino acid sequence. The
best match for each LRP5 motif is shown. For each motif, the
top line is the LRP5 isoform 1 amino acid sequence, the

middle line is amino acids that are identical in the two
motifs, the lower line is the amino acid sequence of the best
match LRP1 motif. Of particular note are the conserved
cysteine (C) residues that are the hallmark of both the EGF-
5 precursor and LDL-receptor ligand binding motifs.

Figure 6(b) illustrates the motif organization of the
LDL-receptor and LRP5. The LDL-receptor ligand binding motif
are represented by the light gray boxes, the EGFlike motifs
are represented by the dark gray boxes. The YWTD spacer
10 motifs are indicated by the vertical lines. The putative
transmembrane domains are represented by the black box.

Figure 7 shows *LRP5* gene structure. The DNA sequence of
contiguous pieces of genomic DNA is represented by the heavy
lines and are according to the indicated scale. The position
15 of the markers *D11S1917(UT5620)*, *H0570POLYA*, *L3001CA*,
D11S1337, and *D11S970* are indicated. The exons are indicated
by the small black boxes with their numerical or alphabetical
name below, the size of the exons is not to scale.

Figure 8 illustrates different *LRP5* gene isoforms.
20 Alternatively spliced 5' ends of the *LRP5* gene are indicated
with the isoform number for each alternatively spliced form.
The light gray arrow indicates the start of translation which
occurs in exon 6 in isoform 1, may occur upstream of exon 1 in
isoform 3 and occurs in exon B in isoforms 2, 4, 5. and 6.
25 The core 22 exons (A to V) are represented by the box.

Figure 9 is a SNP map of Contig 57. Polymorphisms were
identified by the comparison of the DNA sequence of BAC 14-1-
15 with cosmids EO 864 and BO 7185. Corresponding Table 6
indicates a PCR amplicon that includes the site of the
30 polymorphism, the nature of the single nucleotide polymorphis
(SNP), its location and the restriction site that is altered,
if any. The line represents the contiguous genomic DNA with
the relative location of the polymorphisms and the amplicons
used to detect them. The large thin triangles represent the
35 site of putative exons. The marker *H0570POLYA* is indicated.

Figure 10 is a SNP map of Contig 58. Polymorphisms were
identified by the comparison of the DNA sequence of BAC 14-1-

15 with cosmid BO 7185. Corresponding Table 6 indicates a PCR
amplicon that includes the site of the polymorphism, the
nature of the single nucleotide polymorphism (SNP), its
location and the restriction site that is altered, if any.
5 The line represents the contiguous genomic DNA with the
relative location of the polymorphisms and the amplicons used
to detect them. The large thin triangle at the very end of
the line represents exon A of *LRP5*.

Figure 11(a) shows the DNA sequence of the isoform 2
10 cDNA.

Figure 11(b) shows the longest open reading frame of
isoform 2 (also isoform 4,5,6).

Figure 11(c) shows the amino acid sequence of isoform 2
(also isoform 4,5,6), encoded by the open reading frame of
15 Figure 12(b).

Figure 12(a) shows the DNA sequence of isoform 3 cDNA.

Figure 12(b) shows sequence obtained by GRAIL and a
putative extension of isoform 3.

Figure 12(c) shows a putative open reading frame for
20 isoform 3.

Figure 12(d) shows the amino acid sequence of isoform 3.

Figure 12(e) shows the GRAIL predicted promoter sequence
for isoform 3.

Figure 13 shows the DNA sequence of the isoform 4 cDNA,
25 which contains an open reading frame encoding isoform 2
(Figure 11 (b)).

Figure 14 shows the DNA sequence of the present in cDNA
isoform 5, which contains an open reading frame encoding
isoform 2 (Figure 11 (b)).

30   Figure 15 shows the DNA sequence of isoform 6, which
contains an open reading frame encoding isoform 2(Figure 11
(b)).

Figure 15(b) shows the GRAIL predicted promoter sequence
associated with isoform6.

35   Figure 16(a) shows the DNA sequence of a portion of the
mouse Lrp5 cDNA.

Figure 16(b) shows the DNA sequence of the 5' extension

9

of the mouse clone.

Figure 16(c) shows the DNA sequence of a portion of the open reading frame of mouse *Lrp5*.

Figure 16(d) show the amino acid sequence of the open reading frame encoding a portion of mouse Lrp5.

Figure 17(a) shows DNA sequence of exons A to V.

Figure 17(b) shows the amino acid sequence encoded by an open reading frame contained in Figure 17(a).

Figure 18 (a) shows the nucleotide sequence of the full length mouse *Lrp5* cDNA.

Figure 18 (b) shows the nucleotide sequence for the longest open reading frame present in the mouse *Lrp5* cDNA.

Figure 18 (c) shows the amino acid sequence translation (in single letter code) of the open reading frame in Figure 18 (b).

Figure 18 (d) shows an alignment of the amino acid sequence of the human LRP5 protein and the mouse Lrp5 protein program using the GCG algorithm GAP (Genetics Computer Group, Madison, WI).

Figure 18 (e) shows an alignment of the amino acid sequence of the mature human LRP5 protein with the mature mouse LRP5 program using the GCG algorithm GAP (Genetics Computer Group, Madison, WI).

Figure 19 shows a schematic representation of haplotypes across the *IDDM4* region. Three distinct haplotypes are shown. Haplotype A is protective against IDDM whereas haplotypes B and C are susceptible/non-protective for IDDM.

Figure 20 shows a schematic representation of single nucleotide polymorphism (SNP) haplotypes across the *IDDM4* region. Haplotype A is protective whereas haplotypes B, C, D, and E are susceptible/non-protective. A minimal region of 25 kb which is Identical By Descent (IBD) for the four susceptible haplotypes is indicated. The SNP designations, e.g. 57-3, are as described in Table 6 and Figures 9 and 10.

LRP5 Gene Structure

The gene identified contains 22 exons, termed A-V, which

encode most of the mature LRP5 protein.   The 22 exons account
for 4961 nucleotides of the *LRP5* gene transcript (Figure 5(a))
and are located in an approximately 110 kb of genomic DNA.
The genomic DNA containing these exons begins downstream of
5 the genetic marker *L3001CA* and includes the genetic markers
D11S1337, 141ca5, and *D11S970* (Figure 7).   Several different
5' ends of the *LRP5* transcript have been identified.   Of
particular interest is isoform 1 with a 5' end encoding a
signal peptide sequence for protein export (secretory leader
10 peptide) across the plasma membrane.   As discussed below the
LRP5 protein is likely to contain a large extracellular
domain, therefore it would be anticipated that this protein
would have a signal sequence.   The exon encoding the signal
sequence, termed exon 6, lies near the genetic marker
15 *H0570POLYA*.   This exon is 35 kb upstream of exon A and thus
extends the genomic DNA comprising the *LRP5* gene to at least
160kb.

Several additional isoforms of the *LRP5* gene that arise
from alternative splicing of the 5' end have been identified
20 by PCR (Figure 8).   The functional relevance of these
additional isoforms is not clear.   Two of these *LRP5*
transcripts contain exon 1 which is located upstream of the
genetic marker *D11S1917(UT5620)* and expands the *LRP5* gene to
approximately 180 kb of genomic DNA.   The transcript termed
25 isoform 3 consists of exon 1 spliced directly to exon A.   The
reading frame is open at the 5' end and thus there is the
potential for additional coding information present in exons
upstream of exon 1.   Alternatively, centromeric extension of
exon 1 to include all of the open reading frame associated
30 with this region yields the open reading frame for isoform 3.
The second transcript that contains exon 1 also contains
exon 5, which is located near the genetic marker *H0570POLYA*.
The open reading frame for this isoform, isoform 2, begins in
exon B and thus encodes a truncated LRP5 protein which lacks
35 any predicted secretory leader peptide in the first 100 amino
acids.   There are three additional transcripts each with an
open reading frame beginning in exon B and with 5' ends near

the genetic marker L3001CA.

Expression Profile of LRP5

Northern blot analysis indicates that the major mRNA
5 transcript for the LRP5 gene is approximately 5 to 5.5 kb and
is most highly expressed in liver, pancreas, prostate, and
placenta. Expression is also detected in skeletal muscle,
kidney, spleen, thymus, ovary, lung, small intestine, and
colon. Minor bands both larger and smaller than 5 kb are
10 detected and may represent alternative splicing events or
related family members.

LRP5 is a Member of the LDL-receptor Family

The gene identified in the IDDM4 locus, lrp5, is a member
15 of the LDL-receptor family. This family of proteins has
several distinguishing characteristics, a large extracellular
domain containing cysteine rich motifs which are involved in
ligand binding, a single transmembrane spanning domain, and an
"NPXY" internalization motif (Krieger and Herz (1994) Ann.
20 Rev. Biochem. 63: 601-637). The functional role of the
members of this family is the clearance of their ligands by
the mechanism of receptor mediated endocytosis. This is
illustrated by the most highly characterized member of the
family, the LDL-receptor which is responsible for the
25 clearance of LDL cholesterol from plasma (Goldstein, *et. al* .
(1985) Ann. Rev. Cell Biol. 1: 1-39).

LRP5 is most closely related to the LDL-receptor related
protein (LRP) which is also know as the alpha2-macroglobulin
receptor. Translation of the open reading frame (ORF) of
30 isoform 1 yields the LRP5 protein. Comparison of the LRP5
protein to human LRP1 using the algorithm GAP (Genetics
Computer Group, Madison, WI) reveals an overall amino acid
similarity of 55% and 34% identity to the region of the human
LRP1 protein from amino acids 1236 to 2934. The DNA of this
35 ORF is 45% identical to LRP1 encoding DNA as indicated by GAP.
A slightly lower but significant level of similarity is seen
with the megalin receptor also termed LRP2 and gp330 (Saito,

12

et .al.. (1994) Proc. Natl. Acad. Sci. 91: 9725-9729), as well
as the *Drosophilla* vitellogenin receptor (Schonboum *et. al*.
(1995) Proc. Natl. Acad. Sci. 92: 1485-1489). Similarity is
also observed with other members of the LDL-receptor family
5 including the LDL-receptor (Suedhof *et. al*. (1985) Science
228: 815-822) and the VLDL receptor (Oka *et. al*. (1994)
Genomics 20: 298-300). Due to the presence of EGF-like motifs
in LRP5 similarity is also observed with the EGF precursor and
nidogen precursor which are not members of the LDL-receptor
10 family.

Properties and Motifs of LRP5
    The N-terminal portion of LRP5 likely has the potential
for a signal sequence cleavage site. Signal sequences are
15 frequently found in proteins that are exported across the
plasma membrane (von Heijne (1994) Ann. Rev. Biophys. Biomol.
Struc. 23: 167-192). In addition, other members of the LDL-
receptor family contain a signal sequence for protein export.
    The presence of a signal sequence cleavage site was
20 initially identified by a comparison of the human *LRP5* with a
mouse cDNA sequence that we obtained. The initial mouse
partial cDNA sequence that we obtained, 1711 nucleotides
(Figure 16(a)), is 87% identical over an approximately 1500
nucleotide portion to the human *LRP5* cDNA and thus is likely
25 to be the mouse ortholog (*Lrp5*) of the human *LRP5*. The cloned
portion of the mouse cDNA contains an open reading frame
(Figure 16(c)) encoding 533 amino acids. The initiating codon
has consensus nucleotides for efficient translation at both
the -3 (purine) and +4 (G nucleotide) positions (Kozak, M.
30 1996, Mamalian Genome 7:563-574). A 500 amino acid of the
portion of the mouse Lrp5 (Figure 5(g) and Figure 16(d)) is
96% identical to human LRP5, further supporting the proposal
that this is the mouse ortholog of *LRP5*.
    Significantly, the first 200 nucleotides of the mouse
35 cDNA have very little similarity to the 5' extensions present
in isoforms 2-6 discussed below. By contrast this sequence is
75% identical with the human sequence for exon 6 that

comprises the 5' end of isoform 1. Thus isoform 1 which
encodes a signal peptide for protein export likely represents
the most biologically relevant form of LRP5.

Importantly, both the human *LRP5* and mouse *Lrp5* open
5 reading frames encodes a peptide with the potential to act as
a eukaryotic signal sequence for protein export (von Heijne,
1994, Ann. Rev. Biophys. Biomol. Struc. 23:167-192). The
highest score for the signal sequence as determined by using
the SigCleave program in the GCG analysis package (Genetics
10 Computer Group, Madison WI) generates a mature peptide
beginning at residue 25 of human LRP5 and residue 29 of mouse
Lrp5 (Figure 5(d and g)). Additional sites that may be
utilized produce mature peptides in the human LRP5 beginning
at amino acid residues 22, 23, 23, 26, 27, 28, 30 or 32.
15 Additional cleavage sites in the mouse Lrp5 result in mature
peptides beginning at amino acid residue 31, 32, 33, or 38
(Figure 5(g)). The mature human LRP5 protein is show in Figure
5(e).

The other alternative isoforms of LRP5 lack a signal
20 sequence near the N-terminus of the encoded protein. The
functional relevance of these additional isoforms is not
known, however there are several exported proteins which lack
a signal sequence and are transported by a signal peptide
independent mechanism (Higgins, C.F. (1992) Ann. Rev. Cell
25 Biol. 8: 67-113). Thus it is possible that the putative
extracellular domain of these isoforms is translocated across
the plasma membrane.

The extracellular domain of members of the LDL receptor
family contains multiple motifs containing six cysteine
30 residues within an approximately 40 amino acid region.
(Krieger and Herz (1994) Ann. Rev. Biochem. 63: 601-637).
Several classes of these cysteine rich motifs have been
defined based on the spacing of the cysteine residues and the
nature of other conserved amino acids within the motif. The
35 LDL-receptor ligand binding (class A) motif is distinguished
by a cluster of acidic residues in the C-terminal portion of
the motif which includes a highly conserved SDE sequence. The

14

importance of this acidic region in ligand binding has been
demonstrated by mutagenesis studies (Russell et. al. (1989) J.
Biol. Chem. 264: 21682-21688). Three LDL-receptor ligand
binding motifs are found in the LRP5 protein (Figure 6(a)).
5 The EGF-like (class B) motif lacks the cluster of acidic
residues present in the LDL-receptor ligand binding motif. In
addition, the spacing of the cysteine residues differs in the
EGF-like motifs relative to the LDL-receptor ligand binding
motif. The LRP5 protein contains 4 EGF-precursor (B.2)
10 motifs,which have the property of an NGGCS motif between the
first and second cysteine residue (Figure 6(a)).

The size of the members of the LDL receptor family and
the number of the cysteine-rich repeats in the extracellular
domain varies greatly. LRP1 is a large protein of 4544 amino
15 acids and contains 31 LDL-receptor ligand binding motifs
(class A) and 22 EGF-like motifs (class B) (Herz et. al.,
(1988) EMBO 7: 4119-4127). Similarly the megalin receptor,
LRP2, is a protein of 4660 amino acids and consists of 36 LDL-
receptor ligand binding motifs and 17 EGF-like motifs (Saito
20 et. al. (1994) PNAS 91: 9725-9729). In contrast, the LDL
receptor is a relatively small protein of 879 amino acids
which contains 7 LDL-ligand binding motifs and 3 EGF-like
motifs. The predicted size of the mature LRP5 protein, 1591
amino acids, is intermediate between LRP1 and the LDL
25 receptor. As indicated above the LRP5 protein contains four
EGF-like motifs and three LDL-ligand binding motifs. It has
been postulated that the multiple motif units, particularly
evident in LRP1 and LRP2, account for the ability of these
proteins to bind multiple lipoprotein and protein ligands
30 (Krieger and Herz (1994) Ann. Rev. Biochem. 63: 601-637).

The arrangement of the LDL-receptor ligand binding and
EGF-like motifs relative to each other is similar in both the
LDL receptor, LRP1, and LRP2. In each of these proteins
multiple LDL-ligand binding motifs are grouped together and
35 followed by at least one EGF-like motif (Herz et. al., (1988)
EMBO 7: 4119-4127, 1988). By contrast, in the LRP5 protein an
EGF-like motif precedes the group of three LDL-ligand binding

motifs (Figure 6(b)). An additional property unique to LRP5 is that the LDL-ligand binding motifs in LRP5 are followed by the putative transmembrane domain. The different arrangement of the motifs may define LRP5 as a member of a new subfamily within the LDL-receptor related protein family.

LRP5 has a signal peptide for protein export at the N-terminus of the protein. Signal peptide cleavage yields a mature LRP5 protein which begins with an EGF precursor spacer domain from amino acids 31-297 (amino acid residue numbers are based upon the LRP5 precursor). The EGF precursor spacer domain is composed of five approximately 50 amino acid repeats that each contain the characteristic sequence motif Tyr-Trp-Thr-Asp (YWTD). There are three additional spacer domains from amino acids 339-602, 643-903, and 944-1214. Each spacer domain is followed by an EGF repeat from amino acids 297-338 (egf1), 603-642 (egf2), 904-943 (egf3), and 1215-1255 (egf4). The EGF repeats contain six conserved cysteine residues and are of the B.2 class which has an Asn-Gly-Gly-Cys (NGGC) motif as a feature (Herz et al. 1988, EMBO J 7:4119-27) (Figure 6(a)). A single unit defined as an EGF precursor spacer domain and an EGF repeat, is repeated four times in LRP5. The last EGF repeat is adjacent to three consecutive LDLR repeats from amino acids 1257-1295(ldlr1), 1296-1333 (ldlr2), and 1334-1372 (ldlr3). The LDLR repeats have the conserved cysteine residues, as well as, the motif Ser-Asp-Glu (SDE) as a characteristic feature (Figure 6(a)). There are thirteen amino acids separating the LDLR repeats from the putative transmembrane spanning domain of 23 amino acids from 1386-1408. The putative extracellular domain of LRP5 has six potential sites for N-linked glycosylation at amino acid residues 93, 138, 446, 499, 705, and 878 (Figure 5(d)).

The intracellular domain of LRP5 is comprised of 207 amino acids which is longer than most members of the family but similar in size to LRP2 (Saito et. al. (1994) PNAS 91:9725-9729). It does not exhibit similarity to the LDL-

receptor family, nor is it similar to any other known
proteins. The cytoplasmic domain of LRP5 is comprised of 16%
proline and 15% serine residues (Figure 5(d)). Most members
of the LDL-receptor family contain a conserved NPXY motif in
5 the cytoplasmic domain which has been implicated in
endocytosis by coated pits (Chen et. al. (1990) J. Biol. Chem.
265: 3116-3123). Mutagenesis studies have indicated that the
critical residue for recognition by components of the
endocytotic process is the tyrosine residue (Davis, et al.
10 (1987) Cell 45: 15-24). Replacement of the tyrosine residue
by phenylalanine or tryptophan is tolerated, thus the minimal
requirement for this residue appears to be that it is aromatic
amino acid (Davis, et al. (1987) Cell 45: 15-24). Structural
studies have indicated that the critical function of the NP
15 residues is to provide a beta-turn that presents the aromatic
residue (Bansal and Gierasch (1991) Cell 67: 1195-1201).

Although the cytoplasmic domain of LRP5 does not contain
an NPXY motif, there are several aromatic residues in the LRP5
cytoplasmic domain that lie in putative turn regions (Figure
20 5(d)) and thus may be involved in facilitating endocytosis.
In particular tyrosine 1473 which occurs in the sequence VPLY
motif has the proline and tyrosine in the correct position,
relative to the consensus motif. Although the NPXY motif has
been implicated in endocytosis in several proteins it is not
25 an absolute requirement as there are proteins that lack the
NPXY motif, e.g. the transferrin receptor, that undergo
endocytosis by coated pits (Chen, et. al. (1990) J. Biol.
Chem. 265: 3116-3123). In any event, we anticipate that the
primary function of this protein will be receptor mediated
30 endocytosis of its ligand.

Potential Roles of LRP5
The ability of members of the LDL-receptor family to bind
multiple ligands suggests that LRP5 may function to bind one
35 or more ligands. Moreover, in a fashion analogous to other
members of the family, once bound the LRP5 receptor ligand
complex would endocytose resulting in clearance of the ligand

from the extracellular milieu. The nature of the LRP5 ligand
may be a lipid, a protein, a protein complex, or a lipoprotein
and may possess a variety of functions. Although the
physiological function of the most closely related member of
5 the LDL-receptor family, LRP1, is uncertain, it does possess a
number of biochemical activities. LRP1 binds to alpha-2
macroglobulin. Alpha-2 macroglobulin is a plasma complex that
contains a "bait" ligand for a variety of proteinases e.g.
trypsin, chymotrypsin, pancreatic elastase and plasma
10 kallikrein (Jensen (1989) J. Biol. Chem. 20:11539-11542).
Once the proteinase binds and enzymatically cleaves the "bait"
alpha-2 macroglobulin undergoes a conformational change and
"traps" the proteinase. The proteinase:alpha-2 macroglobulin
complex is rapidly cleared by LRP. This mechanism scavenges
15 proteinases that have the potential to mediate a variety of
biological functions e.g. antigen processing and proteinase
secretion (Strickland *et. al.* (1990) J. Biol. Chem. 265:
17401-17404). The importance of this function is evidenced by
the prenatal death of *Lrp1* knockout mice (Zee *et. al.* (1994)
20 Genomics 23: 256-259).

Antigen presentation is a critical component in the
development of IDDM as is evidenced by the pivotal role of MHC
haplotypes in conferring disease susceptibility (Tisch and
McDivitt (1996) Cell 85: 291-297). By analogy with LRP1, LRP5
25 may play a role in antigen presentation in which case
polymorphisms within this gene could affect the development of
autoimmunity in the type 1 diabetic patient.

The alpha-2 macroglobulin complex also binds cytokines
and growth factors such as interleukin-1 beta, interleukin 2,
30 interleukin 6, transforming growth factor-beta, and fibroblast
growth factor (Moestrup and Gliemann (1991) J. Biol. Chem.
266: 14011-14017). Thus the alpha-2 macroglobulin receptor
has the potential to play a role in the clearance of cytokines
and growth factors. The role of cytokines in mediating immune
35 and inflammatory responses is well established. For example,
the interleukin-2 gene is a strong candidate gene for the *Idd3*
locus in the non-obese diabetic mouse, an animal model for

type 1 diabetes (Denny et. al. (1977) Diabetes 46:695-700).
If LRP5 binds alpha-2 macroglobulin or related complexes then
it may play a role in the immune response by mediating
cytokine clearance. For example, the LRP5 which is expressed
5 in pancreas, the target tissue of IDDM, may play a role in
clearing cytokines from the inflammatory infiltrate
(insulitis) that is ongoing in the disease. A polymorphism in
LRP5 that reduces the ability of LRP5 to clear cytokines may
increase an individuals susceptibility to developing IDDM.
10 Furthermore an individual with a polymorphism that increases
the ability of LRP5 to clear cytokines may be protected from
developing IDDM. Conversely, certain cytokines counteract
other cytokines and thus removal of certain beneficial
cytokines by LRP5 may confer disease susceptibility and thus a
15 polymorphism that reduces LRP5 activity may confer protection
from developing the disease.

       Increases of free fatty acids (FFA) have been shown to
reduce insulin secretion in animals (Boden et. al. (1997)
Diabetes 46: 3-10). In addition, ApoE which is a ligand for
20 the LDL-receptor, has been associated with an antioxidant
activity (Miyata and Smith (1996) Nature Genet. 14: 55-61) and
oxidative damage is a central pathogenic mechanism in
pancreatic β-cell destruction in type 1 diabetes (Bac (1994)
Endocrin. Rev. 15: 516-542). Thus alterations in the ability
25 of LRP5 to bind ApoE and related lipoproteins may influence
the susceptibility to oxidative damage in pancreatic β-cells.
Transfection of forms of LRP5 into β-cells may facilitate
resistance of β cells to damage by the immune system in
autoimmunity and in transplantation.

30   A pharmacological entity termed the lipolysis-stimulated
receptor (LSR) which binds and endocytoses chylomicron
remnants in the presence of FFA has been described (Mann et.
al.. (1995) Biochemistry 34: 10421-10431. One possible role
for the LRP5 gene product is that it is responsible for this
35 activity.

       Another member of the LRP family is LRP2, also known as
megalin and gp330, this protein has been implicated in

Heymann's nephritis, an autoimmune disease of the kidney in
rats (Saito et. al. (1994) PNAS 91: 9725-9729). Heymann's
nephritis is a model of glomerularnephritis and is
characterized by the development of autoantibodies to the
5 alpha-2 macroglobulin receptor associated protein, also known
as the Heymann nephritis antigen. The Heymann nephritis
antigen binds to LRP2 (Strickland et. al. (1991) J. Biol.
Chem. 266: 13364-13369). LRP2 may play a role in this disease
by clearance of this pathogenic protein. In an analogous
10 manner the function of LRP5 may be to bind and clear proteins
in the pancreas to which the IDDM patient has generated
autoantibodies. Alternatively LRP5 itself may be an
autoantigen in the IDDM patient.

LRP1 has been identified as the receptor for certain
15 bacterial toxins (Krieger and Herz (1994) Ann. Rev. Biochem.
63: 601-637) and the human rhinovirus (Hofer et. al. (1994)
Proc. Natl. Acad. Sci. 91: 1839-42). It is possible that a
viral infection alters an individuals susceptibility to IDDM
(Epstein (1994) N. Eng. J. Med. 331: 1428-1436). If certain
20 viruses utilize LRP5 as a mode of entry into the cell then
polymorphisms in LRP5 may alter the individuals susceptibility
to type 1 diabetes.

Alterations in LRP5 may participate in the pathogenesis
of other diseases. LRP1 binds lipoproteins such as apoE and
25 C-apolipoproteins. The clearance of lipoproteins such as apoE
and apoB by the LDL receptor is its primary role, mutations in
the LDL receptor lead to hypercholesterolemia (Chen et. al.
(1990) J. Biol. Chem. 265: 3116-3123). Therefore mutations in
LRP5 that decrease the ability of the protein to scavenge
30 lipoproteins may cause an elevation in cholesterol.
Variations in LRP5 could predispose to the development of
macrovascular complications in diabetics, the major cause of
death. In type 2 diabetics, pancreatic pathology is
characterised by the deposition of amyloid. Amyloid
35 deposition may decrease pancreatic β-cell function. LRP5
could function in the metabolism of islet amyloid and
influence susceptibility to type 2 diabetes as well as type 1

diabetes. The role of ApoE in Alzheimer's disease indicates
that proteins such as LRP1 and possibly LRP5 have the
potential to contribute to the pathogenesis of this disease.

Polymorphism in genes involved in the development of
5 osteoporosis-pseudoglioma syndrome have been mapped to a 3-cM
region of chromosome 11 which includes the gene encoding LRP5
(Gong et. al. (1996) Am. J. Hum. Genet. 59: 146-151). The
pathogenic mechanism of this disease is unknown but is
believed to involve a regulatory role, patients with have
10 aberrant vascular growth in the vitero-retina. The potential
role of LRP5 in the clearance of fibroblast growth factor, a
mediator of angiogenesis, and the chromosomal location of the
gene suggests that it may play a role in this disease. This
proposed function could also be connected with the development
15 of retinopathy in diabetes.

Polymorphisms in the LRP5 Gene

The exons of the *LRP5* gene are being scanned for
polymorphisms. There are several polymorphisms that change an
20 amino acid in LRP5 that have been identified in IDDM patients
(Table 5). Of particular interest is a C to T transition,
which changes an Ala codon to Val, in one of the three
conserved LDL receptor ligand binding motifs. In addition to
this polymorphism described above, a C to T transition was
25 identified in the codon for Asn^{709} (with no effect on the
encoded amino acid), and three polymorphisms were identified
in intronic sequences flanking the exons. An additional set
of polymorphisms has been identified by comparing
experimentally derived cDNA sequences with the genomic DNA
30 sequence (Table 5). Some of these polymorphism will be
analyzed in a large number of IDDM patients and control
individuals to determine their association with IDDM.

A number of (approximately 30) single nucleotide
polymorphisms (SNPs) were identified in the genomic DNA
35 sequences of overlapping BAC and cosmid clones surrounding the
genetic marker poly A. The contiguous genomic sequences
containing these polymorphism have been termed contig 57
(Figure 9), which contains exons 1 and 5 along with the

genetic markers poly A and *D11S1917(UT5620)*, and contig 58
(Figure 10) which contains the genetic marker *L3001ca* and part
of exon A.

Additional Experimental Evidence

A region of identity-by-descent associated with type 1
diabetes has been identified in the 5' portion of the *LRP5*
gene. By combining data from SNPs and microsatellite markers
we have identified a region identical-by-descent in
susceptible haplotypes, the minimal region consists of 25 kb
which contains the putative regulatory regions of *LRP5* and the
first exon. This strengthens the genetic evidence for *LRP5*
being a diabetes risk gene. Therefore therapies that affect
LRP5 may be useful in the prevention and treatment of type 1
diabetes.

Overexpression of *LRP5* in mice provides evidence for LRP5
affecting lipoprotein metabolism. Statistically significant
evidence for modulation of triglycerides by LRP5 has been
obtained. Thus therapies that affect *LRP5* may be useful in
the treatment of cardiovascular disease and conditions where
serum triglycerides are elevated.

Suggestive evidence was obtained for LRP5 reducing serum
cholesterol when it is above normal. There is also evidence
for the ability of LRP5 to interact with very low-density
lipoprotein particles and reduce their levels in serum.
Therefore therapies that affect *LRP-5* may be useful in the
treatment of cardiovascular disease and conditions where serum
cholesterol levels are elevated.

Biochemical studies indicate that LRP5 has the capacity
to function in the uptake of low-density lipoprotein (LDL)
particles. Thus therapies that affect *LRP5* may be useful in
the treatment of cardiovascular disease where LDL levels are
elevated.

Overexpression of *LRP5* in mice provided statistically
significant evidence for a reduction in serum alkaline
phosphatase. A reduction in serum alkaline phosphatase is
consistent with LRP5 playing a role in modulation of the
immune response. This provides evidence for *LRP5*
participating in the pathogenesis of type 1 diabetes.

Therefore therapies that affect LRP5 may be useful in the
treatment of autoimmune diseases.

     Cellular localization of LRP5 indicates that it is
expressed in a particular subtype, the phagocytic macrophages,
5 of mature tissue macrophages. Evidence from the literature
indicates that this class of macrophages is involved in
autoimmune disease, supporting a role for LRP5 in autoimmune
disease and type 1 diabetes. Therefore therapies that affect
LRP5 may be useful in the treatment of autoimmune diseases.
10   Full length cDNAs for both human and mouse LRP5 have been
obtained. Antibodies directed against LRP5 have been
developed. These reagents provide tools to further analyze
the biological function of LRP5.

15   Irrespective of LRP5's actual mode of action and
involvement in IDDM and other diseases, the experimental work
described herein establishes and supports the practical
applications which are disclosed as aspects and embodiments of
the present invention.
20

     According to one aspect of the present invention there is
provided a nucleic acid molecule which has a nucleotide
sequence encoding a polypeptide which includes the amino acid
sequence shown in Figure 5(c), Figure 5(d) or Figure 5(e).
25 The amino acid sequence of Figure 5(c) includes that of Figure
5(e) and a signal sequence.
     The coding sequence may be that shown included in Figure
5(a) or Figure 5(b) or it may be a mutant, variant, derivative
or allele of the sequence shown. The sequence may differ
30 from that shown by a change which is one or more of addition,
insertion, deletion and substitution of one or more
nucleotides of the sequence shown. Changes to a nucleotide
sequence may result in an amino acid change at the protein
level, or not, as determined by the genetic code.
35   Thus, nucleic acid according to the present invention may
include a sequence different from the sequence shown in Figure
5(a) or Figure 5(b) yet encode a polypeptide with the same
amino acid sequence. The amino acid sequence shown in Figure
5(c) consists of 1615 residues.

On the other hand the encoded polypeptide may comprise an
amino acid sequence which differs by one or more amino acid
residues from the amino acid sequence shown in Figure 5(c).
Nucleic acid encoding a polypeptide which is an amino acid
5 sequence mutant, variant, derivative or allele of the sequence
shown in Figure 5(c) is further provided by the present
invention. Such polypeptides are discussed below. Nucleic
acid encoding such a polypeptide may show at the nucleotide
sequence and/or encoded amino acid level greater than about
10 60% homology with the coding sequence shown in Figure 5(a)
and/or the amino acid sequence shown in Figure 5(c), greater
than about 70% homology, greater than about 80% homology,
greater than about 90% homology or greater than about 95%
homology. For amino acid "homology", this may be understood
15 to be similarity (according to the established principles of
amino acid similarity, e.g. as determined using the algorithm
GAP (Genetics Computer Group, Madison, WI) or identity. GAP
uses the Needleman and Wunsch algorithm to align two complete
sequences that maximizes the number of matches and minimizes
20 the number of gaps. Generally, the default parameters are
used, with a gap creation penalty = 12 and gap extension
penalty = 4. Use of either of the terms "homology" and
"homologous" herein does not imply any necessary evolutionary
relationship between compared sequences, in keeping for
25 example with standard use of terms such as "homologous
recombination" which merely requires that two nucleotide
sequences are sufficiently similar to recombine under the
appropriate conditions. Further discussion of polypeptides
according to the present invention, which may be encoded by
30 nucleic acid according to the present invention, is found
below.

The present invention extends to nucleic acid that
hybridizes with any one or more of the specific sequences
disclosed herein under stringent conditions. Suitable
35 conditions include, e.g. for detection of sequences that are
about 80-90% identical such as detection of mouse LRP5 with a
human probe or vice versa, hybridization overnight at 42°C in
0.25M Na_2HPO_4, pH 7.2, 6.5% SDS, 10% dextran sulfate and a
final wash at 55°C in 0.1X SSC, 0.1% SDS. For detection of

sequences that are greater than about 90% identical, suitable
conditions include hybridization overnight at 65°C in 0.25M
Na$_2$HPO$_4$, pH 7.2, 6.5% SDS, 10% dextran sulfate and a final wash
at 60°C in 0.1X SSC, 0.1% SDS.

5   The coding sequence may be included within a nucleic acid
molecule which has the sequence shown in Figure 5(a) (isoform
1) or Figure 5(b) and encode the full polypeptide of isoform 1
(Figure 5(c)). Mutants, variants, derivatives and alleles of
these sequences are included within the scope of the present
10 invention in terms analogous to those set out in the preceding
paragraph and in the following disclosure.

       Also provided by the present invention in various aspects
and embodiments is a nucleic acid molecule encoding a
polypeptide which includes the amino acid sequence shown in
15 Figure 17(b)This sequence forms a substantial part of the
amino acid sequence shown in Figure 5(e). Nucleic acid
encoding a polypeptide which includes the amino acid sequence
shown in Figure 17(b) may include the coding sequence shown in
Figure 17(b), or an allele, variant, mutant or derivative in
20 similar terms to those discussed above and below for other
aspects and embodiments of the present invention.

       According to various aspects of the present invention
there are also provided various isoforms of the LRP5
polypeptide and gene. The gene of Figure 5 is known as
25 isoform 1. Included within the present invention is a nucleic
acid molecule which has a nucleotide sequence encoding a
polypeptide which includes the amino acid sequence of a
polypeptide shown in Figure 11(c) (isoform 2). The coding
sequence may be as shown in Figure 11(b) (which may be
30 included within a molecule which has the sequence shown in
Figure 11(a) (isoform 2) or the sequence shown in Figure
12(a) (isoform 3)), Figure 13 (isoform 4), Figure 14 (isoform
5) and Figure 15 (isoform 6). Mutants, derivatives, variants
and alleles of these sequences are also provided by the
35 present invention, as disclosed.

       Further nucleic acid molecules according to the present
invention include the nucleotide sequence of any of Figure
5(a), Figure 12(b), Figure 12(e), Figure 15(b), Figure 16(a)
and Figure 16(b) and nucleic acid encoding the amino acid

sequences encoded by Figure 5(a), Figure 11(b), Figure 12(c) or Figure 16(c), along with mutants, alleles, variants and derivatives of these sequences. Further included are nucleic acid molecules encoding the amino acid sequence of Figure
5 18(c), particularly including the coding sequence shown in Figure 18(b).

Particular alleles according to the present invention have sequences have a variation indicated in Table 5 or Table 6. One or more of these may be associated with susceptibility
10 to IDDM or other disease. Alterations in a sequence according to the present invention which are associated with IDDM or other disease may be preferred in accordance with embodiments of the present invention. Implications for screening, e.g. for diagnostic or prognostic purposes, are discussed below.
15

Generally, nucleic acid according to the present invention is provided as an isolate, in isolated and/or purified form, or free or substantially free of material with which it is naturally associated, such as free or
20 substantially free of nucleic acid flanking the gene in the human genome, except possibly one or more regulatory sequence(s) for expression. Nucleic acid may be wholly or partially synthetic and may include genomic DNA, cDNA or RNA. The coding sequence shown herein is a DNA sequence. Where
25 nucleic acid according to the invention includes RNA, reference to the sequence shown should be construed as encompassing reference to the RNA equivalent, with U substituted for T.

Nucleic acid may be provided as part of a replicable
30 vector, and also provided by the present invention are a vector including nucleic acid as set out above, particularly any expression vector from which the encoded polypeptide can be expressed under appropriate conditions, and a host cell containing any such vector or nucleic acid. An expression
35 vector in this context is a nucleic acid molecule including nucleic acid encoding a polypeptide of interest and appropriate regulatory sequences for expression of the polypeptide, in an *in vitro* expression system, e.g. reticulocyte lysate, or *in vivo*, e.g. in eukaryotic cells such

as COS or CHO cells or in prokaryotic cells such as *E. coli*.
This is discussed further below.

The nucleic acid sequence provided in accordance with the
5 present invention is useful for identifying nucleic acid of
interest (and which may be according to the present invention)
in a test sample. The present invention provides a method of
obtaining nucleic acid of interest, the method including
hybridisation of a probe having the sequence shown in any of
10 Figures 5(a), 11(a), 11(b), 12(a), 12(b), 12(c), 12(e), 13,
14, 15, 15(b) 16(a), 16(b), and 16(c), or a complementary
sequence, to target nucleic acid. Hybridisation is generally
followed by identification of successful hybridisation and
isolation of nucleic acid which has hybridised to the probe,
15 which may involve one or more steps of PCR. It will not
usually be necessary to use a probe with the complete sequence
shown in any of these figures. Shorter fragments,
particularly fragments with a sequence encoding the conserved
motifs (Figure 5(c,d), and Figure 6(a)) may be used.
20   Nucleic acid according to the present invention is
obtainable using one or more oligonucleotide probes or primers
designed to hybridise with one or more fragments of the
nucleic acid sequence shown in any of the figures,
particularly fragments of relatively rare sequence, based on
25 codon usage or statistical analysis. A primer designed to
hybridise with a fragment of the nucleic acid sequence shown
in any of the figures may be used in conjunction with one or
more oligonucleotides designed to hybridise to a sequence in a
cloning vector within which target nucleic acid has been
30 cloned, or in so-called "RACE" (rapid amplification of cDNA
ends) in which cDNA's in a library are ligated to an
oligonucleotide linker and PCR is performed using a primer
which hybridises with a sequence shown and a primer which
hybridises to the oligonucleotide linker.
35   Such oligonucleotide probes or primers, as well as the
full-length sequence (and mutants, alleles, variants and
derivatives) are also useful in screening a test sample
containing nucleic acid for the presence of alleles, mutants
and variants, with diagnostic and/or prognostic implications

as discussed in more detail below.

Nucleic acid isolated and/or purified from one or more
cells (e.g. human) or a nucleic acid library derived from
nucleic acid isolated and/or purified from cells (e.g. a cDNA
5 library derived from mRNA isolated from the cells), may be
probed under conditions for selective hybridisation and/or
subjected to a specific nucleic acid amplification reaction
such as the polymerase chain reaction (PCR) (reviewed for
instance in "PCR protocols; A Guide to Methods and
10 Applications", Eds. Innis et al, 1990, Academic Press, New
York, Mullis et al, Cold Spring Harbor Symp. Quant. Biol.,
51:263, (1987), Ehrlich (ed), PCR technology, Stockton Press,
NY, 1989, and Ehrlich et al, Science, 252:1643-1650, (1991)).
PCR comprises steps of denaturation of template nucleic acid
15 (if double-stranded), annealing of primer to target, and
polymerisation. The nucleic acid probed or used as template
in the amplification reaction may be genomic DNA, cDNA or RNA.
Other specific nucleic acid amplification techniques include
strand displacement activation, the QB replicase system, the
20 repair chain reaction, the ligase chain reaction and ligation
activated transcription. For convenience, and because it is
generally preferred, the term PCR is used herein in contexts
where other nucleic acid amplification techniques may be
applied by those skilled in the art.   Unless the context
25 requires otherwise, reference to PCR should be taken to cover
use of any suitable nucleic amplification reaction available
in the art.

In the context of cloning, it may be necessary for one or
more gene fragments to be ligated to generate a full-length
30 coding sequence. Also, where a full-length encoding nucleic
acid molecule has not been obtained, a smaller molecule
representing part of the full molecule, may be used to obtain
full-length clones. Inserts may be prepared from partial cDNA
clones and used to screen cDNA libraries. The full-length
35 clones isolated may be subcloned into expression vectors and
activity assayed by transfection into suitable host cells,
e.g. with a reporter plasmid.

A method may include hybridisation of one or more (e.g.
two) probes or primers to target nucleic acid. Where the

nucleic acid is double-stranded DNA, hybridisation will
generally be preceded by denaturation to produce single-
stranded DNA. The hybridisation may be as part of a PCR
procedure, or as part of a probing procedure not involving
5 PCR. An example procedure would be a combination of PCR and
low stringency hybridisation. A screening procedure, chosen
from the many available to those skilled in the art, is used
to identify successful hybridisation events and isolated
hybridised nucleic acid.

10   Binding of a probe to target nucleic acid (e.g. DNA) may
be measured using any of a variety of techniques at the
disposal of those skilled in the art. For instance, probes
may be radioactively, fluorescently or enzymatically labelled.
Other methods not employing labelling of probe include
15 examination of restriction fragment length polymorphisms,
amplification using PCR, RN'ase cleavage and allele specific
oligonucleotide probing. Probing may employ the standard
Southern blotting technique. For instance DNA may be
extracted from cells and digested with different restriction
20 enzymes. Restriction fragments may then be separated by
electrophoresis on an agarose gel, before denaturation and
transfer to a nitrocellulose filter. Labelled probe may be
hybridised to the DNA fragments on the filter and binding
determined. DNA for probing may be prepared from RNA
25 preparations from cells.

      Preliminary experiments may be performed by hybridising
under low stringency conditions various probes to Southern
blots of DNA digested with restriction enzymes. Suitable
conditions would be achieved when a large number of
30 hybridising fragments were obtained while the background
hybridisation was low. Using these conditions nucleic acid
libraries, e.g. cDNA libraries representative of expressed
sequences, may be searched. Those skilled in the art are well
able to employ suitable conditions of the desired stringency
35 for selective hybridisation, taking into account factors such
as oligonucleotide length and base composition, temperature
and so on. On the basis of amino acid sequence information,
oligonucleotide probes or primers may be designed, taking into
account the degeneracy of the genetic code, and, where

appropriate, codon usage of the organism from the candidate
nucleic acid is derived. An oligonucleotide for use in
nucleic acid amplification may have about 10 or fewer codons
(e.g. 6, 7 or 8), i.e. be about 30 or fewer nucleotides in
5 length (e.g. 18, 21 or 24). Generally specific primers are
upwards of 14 nucleotides in length, but need not be than 18-
20. Those skilled in the art are well versed in the design of
primers for use processes such as PCR. Various techniques for
synthesizing oligonucleotide primers are well known in the
10 art, including phosphotriester and phosphodiester synthesis
methods.

Preferred amino acid sequences suitable for use in the
design of probes or PCR primers may include sequences
conserved (completely, substantially or partly) encoding the
15 motifs present in LRP5 (Figure 5(d).

A further aspect of the present invention provides an
oligonucleotide or polynucleotide fragment of the nucleotide
sequence shown in any of the figures herein providing nucleic
acid according to the present invention, or a complementary
20 sequence, in particular for use in a method of obtaining
and/or screening nucleic acid. Some preferred
oligonucleotides have a sequence shown in Table 2, Table 4,
Table 7, Table 8 or Table 9, or a sequence which differs from
any of the sequences shown by addition, substitution,
25 insertion or deletion of one or more nucleotides, but
preferably without abolition of ability to hybridise
selectively with nucleic acid in accordance with the present
invention, that is wherein the degree of similarity of the
oligonucleotide or polynucleotide with one of the sequences
30 given is sufficiently high.

In some preferred embodiments, oligonucleotides according
to the present invention that are fragments of any of the
sequences shown, or any allele associated with IDDM or other
disease susceptibility, are at least about 10 nucleotides in
35 length, more preferably at least about 15 nucleotides in
length, more preferably at least about 20 nucleotides in
length. Such fragments themselves individually represent
aspects of the present invention. Fragments and other
oligonucleotides may be used as primers or probes as discussed

but may also be generated (e.g. by PCR) in methods concerned
with determining the presence in a test sample of a sequence
indicative of IDDM or other disease susceptibility.

Methods involving use of nucleic acid in diagnostic
5 and/or prognostic contexts, for instance in determining
susceptibility to IDDM or other disease, and other methods
concerned with determining the presence of sequences
indicative of IDDM or other disease susceptibility are
discussed below.
10

Further embodiments of oligonucleotides according to the
present invention are anti-sense oligonucleotide sequences
based on the nucleic acid sequences described herein. Anti-
sense oligonucleotides may be designed to hybridise to the
15 complementary sequence of nucleic acid, pre-mRNA or mature
mRNA, interfering with the production of polypeptide encoded
by a given DNA sequence (e.g. either native polypeptide or a
mutant form thereof), so that its expression is reduce or
prevented altogether. Anti-sense techniques may be used to
20 target a coding sequence, a control sequence of a gene, e.g.
in the 5' flanking sequence, whereby the antisense
oligonucleotides can interfere with control sequences. Anti-
sense oligonucleotides may be DNA or RNA and may be of around
14-23 nucleotides, particularly around 15-18 nucleotides, in
25 length. The construction of antisense sequences and their use
is described in Peyman and Ulman, Chemical Reviews, 90:543-
584, (1990), and Crooke, Ann. Rev. Pharmacol. Toxicol.,
32:329-376, (1992).

Nucleic acid according to the present invention may be
30 used in methods of gene therapy, for instance in treatment of
individuals with the aim of preventing or curing (wholly or
partially) IDDM or other disease. This may ease one or more
symptoms of the disease. This is discussed below.

Nucleic acid according to the present invention, such as
35 a full-length coding sequence or oligonucleotide probe or
primer, may be provided as part of a kit, e.g. in a suitable
container such as a vial in which the contents are protected
from the external environment. The kit may include
instructions for use of the nucleic acid, e.g. in PCR and/or a

method for determining the presence of nucleic acid of
interest in a test sample. A kit wherein the nucleic acid is
intended for use in PCR may include one or more other reagents
required for the reaction, such as polymerase, nucleosides,
5 buffer solution etc. The nucleic acid may be labelled. A kit
for use in determining the presence or absence of nucleic acid
of interest may include one or more articles and/or reagents
for performance of the method, such as means for providing the
test sample itself, e.g. a swab for removing cells from the
10 buccal cavity or a syringe for removing a blood sample (such
components generally being sterile).

    According to a further aspect, the present invention
provides a nucleic acid molecule including a *LRP5* gene
promoter.

15   In another aspect, the present invention provides a
nucleic acid molecule including a promoter, the promoter
including the sequence of nucleotides shown in Figure 12(e) or
Figure 15(b). The promoter may comprise one or more fragments
of the sequence shown in Figure 12(e) or Figure 15(b),
20 sufficient to promote gene expression. The promoter may
comprise or consist essentially of a sequence of nucleotides
5' to the *LRP5* gene in the human chromosome, or an equivalent
sequence in another species, such as the mouse.

    Any of the sequences disclosed in the figures herein may
25 be used to construct a probe for use in identification and
isolation of a promoter from a genomic library containing a
genomic *LRP5* gene. Techniques and conditions for such probing
are well known in the art and are discussed elsewhere herein.
To find minimal elements or motifs responsible for tissue
30 and/or developmental regulation, restriction enzyme or
nucleases may be used to digest a nucleic acid molecule,
followed by an appropriate assay (for example using a reporter
gene such as luciferase) to determine the sequence required.
A preferred embodiment of the present invention provides a
35 nucleic acid isolate with the minimal nucleotide sequence
shown in Figure 12(e) or Figure 15(b) required for promoter
activity.

    As noted, the promoter may comprise one or more sequence
motifs or elements conferring developmental and/or tissue-

specific regulatory control of expression. Other regulatory
sequences may be included, for instance as identified by
mutation or digest assay in an appropriate expression system
or by sequence comparison with available information, e.g.
5 using a computer to search on-line databases.

By "promoter" is meant a sequence of nucleotides from
which transcription may be initiated of DNA operably linked
downstream (i.e. in the 3' direction on the sense strand of
double-stranded DNA).

10   "Operably linked" means joined as part of the same
nucleic acid molecule, suitably positioned and oriented for
transcription to be initiated from the promoter. DNA operably
linked to a promoter is "under transcriptional initiation
regulation" of the promoter.

15   The present invention extends to a promoter which has a
nucleotide sequence which is allele, mutant, variant or
derivative, by way of nucleotide addition, insertion,
substitution or deletion of a promoter sequence as provided
herein. Preferred levels of sequence homology with a provided
20 sequence may be analogous to those set out above for encoding
nucleic acid and polypeptides according to the present
invention. Systematic or random mutagenesis of nucleic acid
to make an alteration to the nucleotide sequence may be
performed using any technique known to those skilled in the
25 art. One or more alterations to a promoter sequence according
to the present invention may increase or decrease promoter
activity, or increase or decrease the magnitude of the effect
of a substance able to modulate the promoter activity.

"Promoter activity" is used to refer to ability to
30 initiate transcription. The level of promoter activity is
quantifiable for instance by assessment of the amount of mRNA
produced by transcription from the promoter or by assessment
of the amount of protein product produced by translation of
mRNA produced by transcription from the promoter. The amount
35 of a specific mRNA present in an expression system may be
determined for example using specific oligonucleotides which
are able to hybridise with the mRNA and which are labelled or
may be used in a specific amplification reaction such as the
polymerase chain reaction. Use of a reporter gene facilitates

determination of promoter activity by reference to protein
production.

    Further provided by the present invention is a nucleic
acid construct comprising a LRP5 promoter region or a
5 fragment, mutant, allele, derivative or variant thereof able
to promoter transcription, operably linked to a heterologous
gene, e.g. a coding sequence. A "heterologous" or "exogenous"
gene is generally not a modified form of LRP5. Generally, the
gene may be transcribed into mRNA which may be translated into
10 a peptide or polypeptide product which may be detected and
preferably quantitated following expression. A gene whose
encoded product may be assayed following expression is termed
a "reporter gene", i.e. a gene which "reports" on promoter
activity.
15   The reporter gene preferably encodes an enzyme which
catalyses a reaction which produces a detectable signal,
preferably a visually detectable signal, such as a coloured
product. Many examples are known, including β-galactosidase
and luciferase. β- galactosidase activity may be assayed by
20 production of blue colour on substrate, the assay being by eye
or by use of a spectro-photometer to measure absorbance.
Fluorescence, for example that produced as a result of
luciferase activity, may be quantitated using a
spectrophotometer. Radioactive assays may be used, for
25 instance using chloramphenicol acetyltransferase, which may
also be used in non-radioactive assays. The presence and/or
amount of gene product resulting from expression from the
reporter gene may be determined using a molecule able to bind
the product, such as an antibody or fragment thereof. The
30 binding molecule may be labelled directly or indirectly using
any standard technique.

    Those skilled in the art are well aware of a multitude of
possible reporter genes and assay techniques which may be used
to determine gene activity. Any suitable reporter/assay may
35 be used and it should be appreciated that no particular choice
is essential to or a limitation of the present invention.

    Nucleic acid constructs comprising a promoter (as
disclosed herein) and a heterologous gene (reporter) may be

employed in screening for a substance able to modulate
activity of the promoter. For therapeutic purposes, e.g. for
treatment of IDDM or other disease, a substance able to up-
regulate expression of the promoter may be sought. A method
5 of screening for ability of a substance to modulate activity
of a promoter may comprise contacting an expression system,
such as a host cell, containing a nucleic acid construct as
herein disclosed with a test or candidate substance and
determining expression of the heterologous gene.

10   The level of expression in the presence of the test
substance may be compared with the level of expression in the
absence of the test substance. A difference in expression in
the presence of the test substance indicates ability of the
substance to modulate gene expression. An increase in
15 expression of the heterologous gene compared with expression
of another gene not linked to a promoter as disclosed herein
indicates specificity of the substance for modulation of the
promoter.

        A promoter construct may be introduced into a cell line
20 using any technique previously described to produce a stable
cell line containing the reporter construct integrated into
the genome. The cells may be grown and incubated with test
compounds for varying times. The cells may be grown in 96
well plates to facilitate the analysis of large numbers of
25 compounds. The cells may then be washed and the reporter gene
expression analysed. For some reporters, such as luciferase
the cells will be lysed then analysed.

        Following identification of a substance which modulates
or affects promoter activity, the substance may be
30 investigated further. Furthermore, it may be manufactured
and/or used in preparation, i.e. manufacture or formulation,
of a composition such as a medicament, pharmaceutical
composition or drug. These may be administered to
individuals.

35   Thus, the present invention extends in various aspects
not only to a substance identified using a nucleic acid
molecule as a modulator of promoter activity, in accordance
with what is disclosed herein, but also a pharmaceutical
composition, medicament, drug or other composition comprising

such a substance, a method comprising administration of such a
composition to a patient, e.g. for increasing *LRP5* expression
for instance in treatment (which may include preventative
treatment) of IDDM or other disease, use of such a substance
5 in manufacture of a composition for administration, e.g. for
increasing *LRP5* expression for instance in treatment of IDDM
or other disease, and a method of making a pharmaceutical
composition comprising admixing such a substance with a
pharmaceutically acceptable excipient, vehicle or carrier, and
10 optionally other ingredients.

A further aspect of the present invention provides a
polypeptide which has the amino acid sequence shown in Figure
5(c), which may be in isolated and/or purified form, free or
15 substantially free of material with which it is naturally
associated, such as other polypeptides or such as human
polypeptides other than that for which the amino acid sequence
is shown in Figure 5(c), or (for example if produced by
expression in a prokaryotic cell) lacking in native
20 glycosylation, e.g. unglycosylated. Further polypeptides
according to the present invention have an amino acid
sequence selected from that shown in the polypeptide shown in
Figure 11(c), that shown in 12(d), and the partial polypeptide
shown in Figure 16(d).

25   Polypeptides which are amino acid sequence variants,
alleles, derivatives or mutants are also provided by the
present invention. A polypeptide which is a variant, allele,
derivative or mutant may have an amino acid sequence which
differs from that given in a figure herein by one or more of
30 addition, substitution, deletion and insertion of one or more
amino acids. Preferred such polypeptides have LRP5 function,
that is to say have one or more of the following properties:
immunological cross-reactivity with an antibody reactive the
polypeptide for which the sequence is given in a figure
35 herein; sharing an epitope with the polypeptide for which the
amino acid sequence is shown in a figure herein (as determined
for example by immunological cross-reactivity between the two
polypeptides; a biological activity which is inhibited by an
antibody raised against the polypeptide whose sequence is

shown in a figure herein; ability to reduce serum
triglyceride; ability to reduce serum cholesterol; ability to
interact with and/or reduce serum levels of very low-density
lipoprotein particles; ability to affect serum alkaline
5 phosphatase levels. Alteration of sequence may change the
nature and/or level of activity and/or stability of the LRP5
protein.

A polypeptide which is an amino acid sequence variant,
10 allele, derivative or mutant of the amino acid sequence shown
in a figure herein may comprise an amino acid sequence which
shares greater than about 35% sequence identity with the
sequence shown, greater than about 40%, greater than about
50%, greater than about 60%, greater than about 70%, greater
15 than about 80%, greater than about 90% or greater than about
95%. The sequence may share greater than about 60%
similarity, greater than about 70% similarity, greater than
about 80% similarity or greater than about 90% similarity with
the amino acid sequence shown in the relevant figure. Amino
20 acid similarity is generally defined with reference to the
algorithm GAP (Genetics Computer Group, Madison, WI) as noted
above, or the TBLASTN program, of Altschul et al. (1990) J.
Mol. Biol. 215: 403-10. Similarity allows for "conservative
variation", i.e. substitution of one hydrophobic residue such
25 as isoleucine, valine, leucine or methionine for another, or
the substitution of one polar residue for another, such as
arginine for lysine, glutamic for aspartic acid, or glutamine
for asparagine. Particular amino acid sequence variants may
differ from that shown in a figure herein by insertion,
30 addition, substitution or deletion of 1 amino acid, 2, 3, 4,
5-10, 10-20 20-30, 30-50, 50-100, 100-150, or more than 150
amino acids.
Sequence comparison may be made over the full-length of
the relevant sequence shown herein, or may more preferably be
35 over a contiguous sequence of about or greater than about 20,
25, 30, 33, 40, 50, 67, 133, 167, 200, 233, 267, 300, 333,
400, 450, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300,
1400, 1500, 1600, or more amino acids or nucleotide triplets,
compared with the relevant amino acid sequence or nucleotide

sequence as the case may be.

The present invention also includes active portions, fragments, derivatives and functional mimetics of the polypeptides of the invention. An "active portion" of a
5 polypeptide means a peptide which is less than said full length polypeptide, but which retains a biological activity, such as a biological activity selected from binding to ligand, involvement in endocytosis. Thus an active portion of the LRP5 polypeptide may, in one embodiment, include the
10 transmembrane domain and the portion of the cytoplasmic tail involved in endocytosis. Such an active fragment may be included as part of a fusion protein, e.g. including a binding portion for a different ligand. In different embodiments, combinations of LDL and EGF motifs may be included in a
15 molecule to confer on the molecule different binding specificities.

A "fragment" of a polypeptide generally means a stretch of amino acid residues of at least about five contiguous amino acids, often at least about seven contiguous amino acids,
20 typically at least about nine contiguous amino acids, more preferably at least about 13 contiguous amino acids, and, more preferably, at least about 20 to 30 or more contiguous amino acids. Fragments of the LRP5 polypeptide sequence may include antigenic determinants or epitopes useful for raising
25 antibodies to a portion of the amino acid sequence. Alanine scans are commonly used to find and refine peptide motifs within polypeptides, this involving the systematic replacement of each residue in turn with the amino acid alanine, followed by an assessment of biological activity.

30   Preferred fragments of LRP5 include those with any of the following amino acid sequences:

SYFHLFPPPPSPCTDSS

VDGRQNIKRAKDDGT

EVLFTTGLIRPVALVVDN

35   IQGHLDFVMDILVFHS,

which may be used for instance in raising or isolating antibodies. Variant and derivative peptides, peptides which have an amino acid sequence which differs from one of these sequences by way of addition, insertion, deletion or

substitution of one or more amino acids are also provided by
the present invention, generally with the proviso that the
variant or derivative peptide is bound by an antibody or other
specific binding member which binds one of the peptides whose
5 sequence is shown. A peptide which is a variant or
derivative of one of the shown peptides may compete with the
shown peptide for binding to a specific binding member, such
as an antibody or antigen-binding fragment thereof.

A "derivative" of a polypeptide or a fragment thereof may
10 include a polypeptide modified by varying the amino acid
sequence of the protein, e.g. by manipulation of the nucleic
acid encoding the protein or by altering the protein itself.
Such derivatives of the natural amino acid sequence may
involve one or more of insertion, addition, deletion or
15 substitution of one or more amino acids, which may be without
fundamentally altering the qualitative nature of biological
activity of the wild type polypeptide. Also encompassed
within the scope of the present invention are functional
mimetics of active fragments of the LRP5 polypeptides provided
20 (including alleles, mutants, derivatives and variants). The
term "functional mimetic" means a substance which may not
contain an active portion of the relevant amino acid sequence,
and probably is not a peptide at all, but which retains in
qualitative terms biological activity of natural LRP5
25 polypeptide. The design and screening of candidate mimetics
is described in detail below.

Sequences of amino acid sequence variants representative
of preferred embodiments of the present invention are shown in
Table 5 and Table 6. Screening for the presence of one or
30 more of these in a test sample has a diagnostic and/or
prognostic use, for instance in determining IDDM or other
disease susceptibility, as discussed below.

Other fragments of the polypeptides for which sequence
35 information is provided herein are provided as aspects of the
present invention, for instance corresponding to functional
domains. One such functional domain is the putative
extracellular domain, such that a polypeptide fragment
according to the present invention may include the

extracellular domain of the polypeptide of which the amino
acid sequence is shown in Figure 5(e) or Figure 5(c). This
runs to amino acid 1385 of the precursor sequence of Figure
5(c). Another useful LRP5 domain is the cytoplasmic domain,
5 207 amino acids shown in Figure 5(d). This may be used in
targeting proteins to move through the endocytotic pathway.

A polypeptide according to the present invention may be
isolated and/or purified (e.g. using an antibody) for instance
10 after production by expression from encoding nucleic acid (for
which see below). Thus, a polypeptide may be provided free or
substantially free from contaminants with which it is
naturally associated (if it is a naturally-occurring
polypeptide). A polypeptide may be provided free or
15 substantially free of other polypeptides. Polypeptides
according to the present invention may be generated wholly or
partly by chemical synthesis. The isolated and/or purified
polypeptide may be used in formulation of a composition, which
may include at least one additional component, for example a
20 pharmaceutical composition including a pharmaceutically
acceptable excipient, vehicle or carrier. A composition
including a polypeptide according to the invention may be used
in prophylactic and/or therapeutic treatment as discussed
below.
25   A polypeptide, peptide fragment, allele, mutant,
derivative or variant according to the present invention may
be used as an immunogen or otherwise in obtaining specific
antibodies. Antibodies are useful in purification and other
manipulation of polypeptides and peptides, diagnostic
30 screening and therapeutic contexts. This is discussed further
below.

A polypeptide according to the present invention may be
used in screening for molecules which affect or modulate its
35 activity or function, e.g. binding to ligand, involvement in
endocytosis, movement from an intracellular compartment to the
cell surface, movement from the cell surface to an
intracellular compartment. Such molecules may interact with
the ligand binding portion of LRP5, the cytoplasmic portion of

LRP5, or with one or more accessory molecules e.g. involved in
movement of vesicles containing LRP5 to and from the cell
surface, and may be useful in a therapeutic (possibly
including prophylactic) context.

5   It is well known that pharmaceutical research leading to
the identification of a new drug may involve the screening of
very large numbers of candidate substances, both before and
even after a lead compound has been found. This is one factor
which makes pharmaceutical research very expensive and time-
10 consuming. Means for assisting in the screening process can
have considerable commercial importance and utility. Such
means for screening for substances potentially useful in
treating or preventing IDDM or other disease is provided by
polypeptides according to the present invention. Substances
15 identified as modulators of the polypeptide represent an
advance in the fight against IDDM and other diseases since
they provide basis for design and investigation of
therapeutics for in vivo use. Furthermore, they may be useful
in any of a number of conditions, including autoimmune
20 diseases, such as glomerulonephritis, diseases and disorders
involving disruption of endocytosis and/or antigen
presentation, diseases and disorders involving cytokine
clearance and/or inflammation, viral infection, pathogenic
bacterial toxin contamination, elevation of free fatty acids
25 or hypercholesterolemia, type 2 diabetes, osteoporosis, and
Alzheimer's disease, given the functional indications for
LRP5, discussed elsewhere herein. As noted elsewhere, LRP5,
fragments thereof, and nucleic acid according to the invention
may also be useful in combatting any of these diseases and
30 disorders.

    A method of screening for a substance which modulates
activity of a polypeptide may include contacting one or more
test substances with the polypeptide in a suitable reaction
35 medium, testing the activity of the treated polypeptide and
comparing that activity with the activity of the polypeptide
in comparable reaction medium untreated with the test
substance or substances. A difference in activity between the
treated and untreated polypeptides is indicative of a

41

modulating effect of the relevant test substance or
substances.

Combinatorial library technology (Schultz, JS (1996)
Biotechnol. Prog. 12:729-743) provides an efficient way of
5 testing a potentially vast number of different substances for
ability to modulate activity of a polypeptide. Prior to or as
well as being screened for modulation of activity, test
substances may be screened for ability to interact with the
polypeptide, e.g. in a yeast two-hybrid system (which requires
10 that both the polypeptide and the test substance can be
expressed in yeast from encoding nucleic acid). This may be
used as a coarse screen prior to testing a substance for
actual ability to modulate activity of the polypeptide.

15   Following identification of a substance which modulates
or affects polypeptide activity, the substance may be
investigated further. Furthermore, it may be manufactured
and/or used in preparation, i.e. manufacture or formulation,
of a composition such as a medicament, pharmaceutical
20 composition or drug. These may be administered to
individuals.

Thus, the present invention extends in various aspects
not only to a substance identified as a modulator of
polypeptide activity, in accordance with what is disclosed
25 herein, but also a pharmaceutical composition, medicament,
drug or other composition comprising such a substance, a
method comprising administration of such a composition to a
patient, e.g. for treatment (which may include preventative
treatment) of IDDM or other disease, use of such a substance
30 in manufacture of a composition for administration, e.g. for
treatment of IDDM or other disease, and a method of making a
pharmaceutical composition comprising admixing such a
substance with a pharmaceutically acceptable excipient,
vehicle or carrier, and optionally other ingredients.

35

A substance identified using as a modulator of
polypeptide or promoter function may be peptide or non-peptide
in nature. Non-peptide "small molecules" are often preferred
for many in vivo pharmaceutical uses. Accordingly, a mimetic

or mimick of the substance (particularly if a peptide) may be
designed for pharmaceutical use. The designing of mimetics to
a known pharmaceutically active compound is a known approach
to the development of pharmaceuticals based on a "lead"
5 compound. This might be desirable where the active compound
is difficult or expensive to synthesise or where it is
unsuitable for a particular method of administration, e.g.
peptides are not well suited as active agents for oral
compositions as they tend to be quickly degraded by proteases
10 in the alimentary canal. Mimetic design, synthesis and
testing may be used to avoid randomly screening large number
of molecules for a target property.

There are several steps commonly taken in the design of
a mimetic from a compound having a given target property.
15 Firstly, the particular parts of the compound that are
critical and/or important in determining the target property
are determined. In the case of a peptide, this can be done by
systematically varying the amino acid residues in the peptide,
e.g. by substituting each residue in turn. These parts or
20 residues constituting the active region of the compound are
known as its "pharmacophore".

Once the pharmacophore has been found, its structure is
modelled to according its physical properties, e.g.
stereochemistry, bonding, size and/or charge, using data from
25 a range of sources, e.g. spectroscopic techniques, X-ray
diffraction data and NMR. Computational analysis, similarity
mapping (which models the charge and/or volume of a
pharmacophore, rather than the bonding between atoms) and
other techniques can be used in this modelling process.
30   In a variant of this approach, the three-dimensional
structure of the ligand and its binding partner are modelled.
This can be especially useful where the ligand and/or binding
partner change conformation on binding, allowing the model to
take account of this the design of the mimetic.
35   A template molecule is then selected onto which chemical
groups which mimic the pharmacophore can be grafted. The
template molecule and the chemical groups grafted on to it can
conveniently be selected so that the mimetic is easy to
synthesise, is likely to be pharmacologically acceptable, and

does not degrade in vivo, while retaining the biological
activity of the lead compound. The mimetic or mimetics found
by this approach can then be screened to see whether they have
the target property, or to what extent they exhibit it.
5 Further optimisation or modification can then be carried out
to arrive at one or more final mimetics for in vivo or
clinical testing.

Mimetics of substances identified as having ability to
modulate LRP5 polypeptide or promoter activity using a
10 screening method as disclosed herein are included within the
scope of the present invention. A polypeptide, peptide or
substance able to modulate activity of a polypeptide according
to the present invention may be provided in a kit, e.g. sealed
in a suitable container which protects its contents from the
15 external environment. Such a kit may include instructions for
use.

A convenient way of producing a polypeptide according to
the present invention is to express nucleic acid encoding it,
20 by use of the nucleic acid in an expression system.
Accordingly, the present invention also encompasses a method
of making a polypeptide (as disclosed), the method including
expression from nucleic acid encoding the polypeptide
(generally nucleic acid according to the invention). This
25 may conveniently be achieved by growing a host cell in
culture, containing such a vector, under appropriate
conditions which cause or allow expression of the polypeptide.
Polypeptides may also be expressed in in vitro systems, such
as reticulocyte lysate.
30   Systems for cloning and expression of a polypeptide in a
variety of different host cells are well known. Suitable host
cells include bacteria, eukaryotic cells such as mammalian and
yeast, and baculovirus systems. Mammalian cell lines
available in the art for expression of a heterologous
35 polypeptide include Chinese hamster ovary cells, HeLa cells,
baby hamster kidney cells, COS cells and many others. A
common, preferred bacterial host is E. coli. Suitable vectors
can be chosen or constructed, containing appropriate
regulatory sequences, including promoter sequences, terminator

fragments, polyadenylation sequences, enhancer sequences,
marker genes and other sequences as appropriate. Vectors may
be plasmids, viral e.g. 'phage, or phagemid, as appropriate.
For further details see, for example, Molecular Cloning: a
5 Laboratory Manual: 2nd edition, Sambrook et al., 1989, Cold
Spring Harbor Laboratory Press. Many known techniques and
protocols for manipulation of nucleic acid, for example in
preparation of nucleic acid constructs, mutagenesis,
sequencing, introduction of DNA into cells and gene
10 expression, and analysis of proteins, are described in detail
in Current Protocols in Molecular Biology, Ausubel et al.
eds., John Wiley & Sons, 1992.

        Thus, a further aspect of the present invention provides
a host cell containing nucleic acid as disclosed herein. The
15 nucleic acid of the invention may be integrated into the
genome (e.g. chromosome) of the host cell. Integration may be
promoted by inclusion of sequences which promote recombination
with the genome, in accordance with standard techniques. The
nucleic acid may be on an extra-chromosomal vector within the
20 cell.

        A still further aspect provides a method which includes
introducing the nucleic acid into a host cell. The
introduction, which may (particularly for in vitro
introduction) be generally referred to without limitation as
25 "transformation", may employ any available technique. For
eukaryotic cells, suitable techniques may include calcium
phosphate transfection, DEAE-Dextran, electroporation,
liposome-mediated transfection and transduction using
retrovirus or other virus, e.g. vaccinia or, for insect cells,
30 baculovirus. For bacterial cells, suitable techniques may
include calcium chloride transformation, electroporation and
transfection using bacteriophage.

        Marker genes such as antibiotic resistance or sensitivity
genes may be used in identifying clones containing nucleic
35 acid of interest, as is well known in the art.

        The introduction may be followed by causing or allowing
expression from the nucleic acid, e.g. by culturing host cells
(which may include cells actually transformed although more
likely the cells will be descendants of the transformed

cells) under conditions for expression of the gene, so that
the encoded polypeptide is produced. If the polypeptide is
expressed coupled to an appropriate signal leader peptide it
may be secreted from the cell into the culture medium.
5 Following production by expression, a polypeptide may be
isolated and/or purified from the host cell and/or culture
medium, as the case may be, and subsequently used as desired,
e.g. in the formulation of a composition which may include one
or more additional components, such as a pharmaceutical
10 composition which includes one or more pharmaceutically
acceptable excipients, vehicles or carriers (e.g. see below).

   Introduction of nucleic acid may take place *in vivo* by
way of gene therapy, as discussed below. A host cell
containing nucleic acid according to the present invention,
15 e.g. as a result of introduction of the nucleic acid into the
cell or into an ancestor of the cell and/or genetic alteration
of the sequence endogenous to the cell or ancestor (which
introduction or alteration may take place *in vivo* or *ex vivo*),
may be comprised (e.g. in the soma) within an organism which
20 is an animal, particularly a mammal, which may be human or
non-human, such as rabbit, guinea pig, rat, mouse or other
rodent, cat, dog, pig, sheep, goat, cattle or horse, or which
is a bird, such as a chicken. Genetically modified or
transgenic animals or birds comprising such a cell are also
25 provided as further aspects of the present invention.

   Thus, in various further aspects, the present invention
provides a non-human animal with a human *LRP5* transgene
within its genome. The transgene may have the sequence of any
of the isoforms identified herein or a mutant, derivative,
30 allele or variant thereof as disclosed. In one preferred
embodiment, the heterologous human *LRP5* sequence replaces the
endogenous animal sequence. In other preferred embodiments,
one or more copies of the human *LRP5* sequence are added to
the animal genome.

35   Preferably the animal is a rodent, and most preferably
mouse or rat.

   This may have a therapeutic aim. (Gene therapy is
discussed below.) The presence of a mutant, allele or variant
sequence within cells of an organism, particularly when in

46

place of a homologous endogenous sequence, may allow the
organism to be used as a model in testing and/or studying the
role of the LRP5 gene or substances which modulate activity
of the encoded polypeptide and/or promoter in vitro or are
5 otherwise indicated to be of therapeutic potential.

An animal model for LRP5 deficiency may be constructed
using standard techniques for introducing mutations into an
animal germ-line. In one example of this approach, using a
mouse, a vector carrying an insertional mutation within the
10 LRP5 gene may be transfected into embryonic stem cells. A
selectable marker, for example an antibiotic resistance gene
such as neoR, may be included to facilitate selection of
clones in which the mutant gene has replaced the endogenous
wild type homologue. Such clones may be also be identified or
15 further investigated by Southern blot hybridisation. The
clones may then be expanded and cells injected into mouse
blastocyst stage embryos. Mice in which the injected cells
have contributed to the development of the mouse may be
identified by Southern blotting. These chimeric mice may then
20 be bred to produce mice which carry one copy of the mutation
in the germ line. These heterozygous mutant animals may then
be bred to produce mice carrying mutations in the gene
homozygously. The mice having a heterozygous mutation in the
LRP5 gene may be a suitable model for human individuals having
25 one copy of the gene mutated in the germ line who are at risk
of developing IDDM or other disease.

Animal models may also be useful for any of the various
diseases discussed elsewhere herein.

30   Instead of or as well as being used for the production of
a polypeptide encoded by a transgene, host cells may be used
as a nucleic acid factory to replicate the nucleic acid of
interest in order to generate large amounts of it. Multiple
copies of nucleic acid of interest may be made within a cell
35 when coupled to an amplifiable gene such as dihyrofolate
reductase (DHFR), as is well known. Host cells transformed
with nucleic acid of interest, or which are descended from
host cells into which nucleic acid was introduced, may be
cultured under suitable conditions, e.g. in a fermentor, taken

from the culture and subjected to processing to purifiy the
nucleic acid. Following purification, the nucleic acid or one
or more fragments thereof may be used as desired, for instance
in a diagnostic or prognostic assay as discussed elsewhere
5 herein.

        The provision of the novel LRP-5 polypeptide isoforms and
mutants, alleles, variants and derivatives enables for the
first time the production of antibodies able to bind these
10 molecules specifically.
        Accordingly, a further aspect of the present invention
provides an antibody able to bind specifically to the
polypeptide whose sequence is given in a figure herein. Such
an antibody may be specific in the sense of being able to
15 distinguish between the polypeptide it is able to bind and
other human polypeptides for which it has no or substantially
no binding affinity (e.g. a binding affinity of about 1000x
less). Specific antibodies bind an epitope on the molecule
which is either not present or is not accessible on other
20 molecules. Antibodies according to the present invention may
be specific for the wild-type polypeptide. Antibodies
according to the invention may be specific for a particular
mutant, variant, allele or derivative polypeptide as between
that molecule and the wild-type polypeptide, so as to be
25 useful in diagnostic and prognostic methods as discussed
below. Antibodies are also useful in purifying the
polypeptide or polypeptides to which they bind, e.g. following
production by recombinant expression from encoding nucleic
acid.
30   Preferred antibodies according to the invention are
isolated, in the sense of being free from contaminants such as
antibodies able to bind other polypeptides and/or free of
serum components. Monoclonal antibodies are preferred for
some purposes, though polyclonal antibodies are within the
35 scope of the present invention.
        Antibodies may be obtained using techniques which are
standard in the art. Methods of producing antibodies include
immunising a mammal (e.g. mouse, rat, rabbit, horse, goat,
sheep or monkey) with the protein or a fragment thereof.

Antibodies may be obtained from immunised animals using any of
a variety of techniques known in the art, and screened,
preferably using binding of antibody to antigen of interest.
For instance, Western blotting techniques or
5 immunoprecipitation may be used (Armitage et al., 1992,
Nature 357: 80-82). Isolation of antibodies and/or antibody-
producing cells from an animal may be accompanied by a step of
sacrificing the animal.

As an alternative or supplement to immunising a mammal
10 with a peptide, an antibody specific for a protein may be
obtained from a recombinantly produced library of expressed
immunoglobulin variable domains, e.g. using lambda
bacteriophage or filamentous bacteriophage which display
functional immunoglobulin binding domains on their surfaces;
15 for instance see WO92/01047. The library may be naive, that
is constructed from sequences obtained from an organism which
has not been immunised with any of the proteins (or
fragments), or may be one constructed using sequences obtained
from an organism which has been exposed to the antigen of
20 interest.

Suitable peptides for use in immunising an animal and/or
isolating anti-LRP5 antibody include any of the following
amino acid sequences:

SYFHLFPPPPSPCTDSS
25   VDGRQNIKRAKDDGT
        EVLFTTGLIRPVALVVDN
        IQGHLDFVMDILVFHS.

Antibodies according to the present invention may be
30 modified in a number of ways. Indeed the term "antibody"
should be construed as covering any binding substance having a
binding domain with the required specificity. Thus the
invention covers antibody fragments, derivatives, functional
equivalents and homologues of antibodies, including synthetic
35 molecules and molecules whose shape mimicks that of an
antibody enabling it to bind an antigen or epitope.

Example antibody fragments, capable of binding an antigen
or other binding partner are the Fab fragment consisting of
the VL, VH, Cl and CH1 domains; the Fd fragment consisting of

the VH and CH1 domains; the Fv fragment consisting of the VL
and VH domains of a single arm of an antibody; the dAb
fragment which consists of a VH domain; isolated CDR regions
and F(ab')2 fragments, a bivalent fragment including two Fab
5 fragments linked by a disulphide bridge at the hinge region.
Single chain Fv fragments are also included.

A hybridoma producing a monoclonal antibody according to
the present invention may be subject to genetic mutation or
other changes. It will further be understood by those skilled
10 in the art that a monoclonal antibody can be subjected to the
techniques of recombinant DNA technology to produce other
antibodies or chimeric molecules which retain the specificity
of the original antibody. Such techniques may involve
introducing DNA encoding the immunoglobulin variable region,
15 or the complementarity determining regions (CDRs), of an
antibody to the constant regions, or constant regions plus
framework regions, of a different immunoglobulin. See, for
instance, EP184187A, GB 2188638A or EP-A-0239400. Cloning and
expression of chimeric antibodies are described in EP-A-
20 0120694 and EP-A-0125023.

Hybridomas capable of producing antibody with desired
binding characteristics are within the scope of the present
invention, as are host cells, eukaryotic or prokaryotic,
containing nucleic acid encoding antibodies (including
25 antibody fragments) and capable of their expression. The
invention also provides methods of production of the
antibodies including growing a cell capable of producing the
antibody under conditions in which the antibody is produced,
and preferably secreted.

30   The reactivities of antibodies on a sample may be
determined by any appropriate means. Tagging with individual
reporter molecules is one possibility. The reporter molecules
may directly or indirectly generate detectable, and preferably
measurable, signals. The linkage of reporter molecules may be
35 directly or indirectly, covalently, e.g. via a peptide bond or
non-covalently. Linkage via a peptide bond may be as a result
of recombinant expression of a gene fusion encoding antibody
and reporter molecule.

One favoured mode is by covalent linkage of each antibody

with an individual fluorochrome, phosphor or laser dye with
spectrally isolated absorption or emission characteristics.
Suitable fluorochromes include fluorescein, rhodamine,
phycoerythrin and Texas Red. Suitable chromogenic dyes
5 include diaminobenzidine.

Other reporters include macromolecular colloidal
particles or particulate material such as latex beads that are
coloured, magnetic or paramagnetic, and biologically or
chemically active agents that can directly or indirectly cause
10 detectable signals to be visually observed, electronically
detected or otherwise recorded. These molecules may be
enzymes which catalyse reactions that develop or change
colours or cause changes in electrical properties, for
example. They may be molecularly excitable, such that
15 electronic transitions between energy states result in
characteristic spectral absorptions or emissions. They may
include chemical entities used in conjunction with
biosensors. Biotin/avidin or biotin/streptavidin and alkaline
phosphatase detection systems may be employed.

20   The mode of determining binding is not a feature of the
present invention and those skilled in the art are able to
choose a suitable mode according to their preference and
general knowledge. Particular embodiments of antibodies
according to the present invention include antibodies able to
25 bind and/or which bind specifically, e.g. with an affinity of
at least 10^{-7} M, to one of the following peptides:

SYFHLFPPPPSPCTDSS

VDGRQNIKRAKDDGT

EVLFTTGLIRPVALVVDN

30   IQGHLDFVMDILVFHS.

Antibodies according to the present invention may be used
in screening for the presence of a polypeptide, for example in
a test sample containing cells or cell lysate as discussed,
and may be used in purifying and/or isolating a polypeptide
35 according to the present invention, for instance following
production of the polypeptide by expression from encoding
nucleic acid therefor. Antibodies may modulate the activity
of the polypeptide to which they bind and so, if that

polypeptide has a deleterious effect in an individual, may be
useful in a therapeutic context (which may include
prophylaxis).

5   An antibody may be provided in a kit, which may include
instructions for use of the antibody, e.g. in determining the
presence of a particular substance in a test sample. One or
more other reagents may be included, such as labelling
molecules, buffer solutions, elutants and so on. Reagents may
be provided within containers which protect them from the
10 external environment, such as a sealed vial.

        The identification of the *LRP5* gene and indications of
its association with IDDM and other diseases paves the way for
aspects of the present invention to provide the use of
materials and methods, such as are disclosed and discussed
15 above, for establishing the presence or absence in a test
sample of an variant form of the gene, in particular an allele
or variant specifically associated with IDDM or other disease.
This may be for diagnosing a predisposition of an individual
to IDDM or other disease. It may be for diagnosing IDDM of a
20 patient with the disease as being associated with the IDDM4
gene.

        This allows for planning of appropriate therapeutic
and/or prophylactic treatment, permitting stream-lining of
treatment by targeting those most likely to benefit.

25   A variant form of the gene may contain one or more
insertions, deletions, substitutions and/or additions of one
or more nucleotides compared with the wild-type sequence (such
as shown in Table 5 or Table 6) which may or may not disrupt
the gene function. Differences at the nucleic acid level are
30 not necessarily reflected by a difference in the amino acid
sequence of the encoded polypeptide. However, a mutation or
other difference in a gene may result in a frame-shift or stop
codon, which could seriously affect the nature of the
polypeptide produced (if any), or a point mutation or gross
35 mutational change to the encoded polypeptide, including
insertion, deletion, substitution and/or addition of one or
more amino acids or regions in the polypeptide. A mutation in
a promoter sequence or other regulatory region may prevent or
reduce expression from the gene or affect the processing or

stability of the mRNA transcript. For instance, a sequence
alteration may affect alternative splicing of mRNA. As
discussed, various *LRP5* isoforms resulting from alternative
splicing are provided by the present invention.

5   There are various methods for determining the presence or
absence in a test sample of a particular nucleic acid
sequence, such as the sequence shown in any figure herein, or
a mutant, variant or allele thereof, e.g. including an
alteration shown in Table 5 or Table 6.

10   Tests may be carried out on preparations containing
genomic DNA, cDNA and/or mRNA. Testing cDNA or mRNA has the
advantage of the complexity of the nucleic acid being reduced
by the absence of intron sequences, but the possible
disadvantage of extra time and effort being required in making
15 the preparations. RNA is more difficult to manipulate than
DNA because of the wide-spread occurrence of RN'ases. Nucleic
acid in a test sample may be sequenced and the sequence
compared with the sequence shown in any of the figures herein,
to determine whether or not a difference is present. If so,
20 the difference can be compared with known susceptibility
alleles (e.g. as shown in Table 5 or Table 6) to determine
whether the test nucleic acid contains one or more of the
variations indicated, or the difference can be investigated
for association with IDDM or other disease.

25   Since it will not generally be time- or labour-efficient
to sequence all nucleic acid in a test sample or even the
whole *LRP5* gene, a specific amplification reaction such as PCR
using one or more pairs of primers may be employed to amplify
the region of interest in the nucleic acid, for instance the
30 *LRP5* gene or a particular region in which polymorphisms
associated with IDDM or other disease susceptibility occur.
The amplified nucleic acid may then be sequenced as above,
and/or tested in any other way to determine the presence or
absence of a particular feature. Nucleic acid for testing may
35 be prepared from nucleic acid removed from cells or in a
library using a variety of other techniques such as
restriction enzyme digest and electrophoresis.

        Nucleic acid may be screened using a variant- or allele-
specific probe. Such a probe corresponds in sequence to a

region of the *LRP5* gene, or its complement, containing a
sequence alteration known to be associated with IDDM or other
disease susceptibility. Under suitably stringent conditions,
specific hybridisation of such a probe to test nucleic acid is
5 indicative of the presence of the sequence alteration in the
test nucleic acid. For efficient screening purposes, more
than one probe may be used on the same test sample.

Allele- or variant-specific oligonucleotides may
similarly be used in PCR to specifically amplify particular
10 sequences if present in a test sample. Assessment of whether
a PCR band contains a gene variant may be carried out in a
number of ways familiar to those skilled in the art. The PCR
product may for instance be treated in a way that enables one
to display the polymorphism on a denaturing polyacrylamide DNA
15 sequencing gel, with specific bands that are linked to the
gene variants being selected.

SSCP heteroduplex analysis may be used for screening DNA
fragments for sequence variants/mutations. It generally
involves amplifying radiolabelled 100-300 bp fragments of the
20 gene, diluting these products and denaturing at 95°C. The
fragments are quick-cooled on ice so that the DNA remains in
single stranded form. These single stranded fragments are run
through acrylamide based gels. Differences in the sequence
composition will cause the single stranded molecules to adopt
25 difference conformations in this gel matrix making their
mobility different from wild type fragments, thus allowing
detecting of mutations in the fragments being analysed
relative to a control fragment upon exposure of the gel to X-
ray film. Fragments with altered mobility/conformations may be
30 directly excised from the gel and directly sequenced for
mutation.

Sequencing of a PCR product may involve precipitation
with isopropanol, resuspension and sequencing using a TaqFS+
Dye terminator sequencing kit. Extension products may be
35 electrophoresed on an ABI 377 DNA sequencer and data analysed
using Sequence Navigator software.

A further possible screening approach employs a PTT assay
in which fragments are amplified with primers that contain
the consensus Kozak initiation sequences and a T7 RNA

polymerase promoter. These extra sequences are incorporated
into the 5' primer such that they are in frame with the native
coding sequence of the fragment being analysed. These PCR
products are introduced into a coupled
5 transcription/translation system. This reaction allows the
production of RNA from the fragment and translation of this
RNA into a protein fragment. PCR products from controls make
a protein product of a wild type size relative to the size of
the fragment being analysed. If the PCR product analysed has
10 a frame-shift or nonsense mutation, the assay will yield a
truncated protein product relative to controls. The size of
the truncated product is related to the position of the
mutation, and the relative region of the gene from this
patient may be sequenced to identify the truncating mutation.
15   An alternative or supplement to looking for the presence
of variant sequences in a test sample is to look for the
presence of the normal sequence, e.g. using a suitably
specific oligonucleotide probe or primer. Use of
oligonucleotide probes and primers has been discussed in more
20 detail above.

        Allele- or variant-specific oligonucleotide probes or
primers according to embodiments of the present invention may
be selected from those shown in Table 4, Table 7 or Table 8.

        Approaches which rely on hybridisation between a probe
25 and test nucleic acid and subsequent detection of a mismatch
may be employed. Under appropriate conditions (temperature,
pH etc.), an oligonucleotide probe will hybridise with a
sequence which is not entirely complementary. The degree of
base-pairing between the two molecules will be sufficient for
30 them to anneal despite a mis-match. Various approaches are
well known in the art for detecting the presence of a mis-
match between two annealing nucleic acid molecules.

        For instance, RN'ase A cleaves at the site of a mis-
match. Cleavage can be detected by electrophoresing test
35 nucleic acid to which the relevant probe or probe has annealed
and looking for smaller molecules (i.e. molecules with higher
electrophoretic mobility) than the full length probe/test
hybrid.

        Thus, an oligonucleotide probe that has the sequence of a

region of the normal *LRP5* gene (either sense or anti-sense
strand) in which mutations associated with IDDM or other
disease susceptibility are known to occur (e.g. see Table 5
and Table 6) may be annealed to test nucleic acid and the
5 presence or absence of a mis-match determined. Detection of
the presence of a mis-match may indicate the presence in the
test nucleic acid of a mutation associated with IDDM or other
disease susceptibility. On the other hand, an oligonucleotide
probe that has the sequence of a region of the gene including
10 a mutation associated with IDDM or other disease
susceptibility may be annealed to test nucleic acid and the
presence or absence of a mis-match determined. The presence
of a mis-match may indicate that the nucleic acid in the test
sample has the normal sequence (the absence of a mis-match
15 indicating that the test nucleic acid has the mutation). In
either case, a battery of probes to different regions of the
gene may be employed.

       The presence of differences in sequence of nucleic acid
molecules may be detected by means of restriction enzyme
20 digestion, such as in a method of DNA fingerprinting where the
restriction pattern produced when one or more restriction
enzymes are used to cut a sample of nucleic acid is compared
with the pattern obtained when a sample containing the normal
gene shown in a figure herein or a variant or allele, e.g. as
25 containing an alteration shown in Table 5 or Table 6 is
digested with the same enzyme or enzymes.

       The presence or absence of a lesion in a promoter or
other regulatory sequence may also be assessed by determining
the level of mRNA production by transcription or the level of
30 polypeptide production by translation from the mRNA.
Determination of promoter activity has been discussed above.

       A test sample of nucleic acid may be provided for example
by extracting nucleic acid from cells or biological tissues or
35 fluids, urine, saliva, faeces, a buccal swab, biopsy or
preferably blood, or for pre-natal testing from the amnion,
placenta or foetus itself.

       There are various methods for determining the presence or

absence in a test sample of a particular polypeptide, such as
the polypeptide with the amino acid sequence shown in any
figure herein or an amino acid sequence mutant, variant or
allele thereof.

5   A sample may be tested for the presence of a binding
partner for a specific binding member such as an antibody (or
mixture of antibodies), specific for one or more particular
variants of the polypeptide shown in a figure herein. A
sample may be tested for the presence of a binding partner for
10 a specific binding member such as an antibody (or mixture of
antibodies), specific for the polypeptide shown in a figure
herein. In such cases, the sample may be tested by being
contacted with a specific binding member such as an antibody
under appropriate conditions for specific binding, before
15 binding is determined, for instance using a reporter system as
discussed. Where a panel of antibodies is used, different
reporting labels may be employed for each antibody so that
binding of each can be determined.

        A specific binding member such as an antibody may be used
20 to isolate and/or purify its binding partner polypeptide from
a test sample, to allow for sequence and/or biochemical
analysis of the polypeptide to determine whether it has the
sequence and/or properties of the polypeptide whose sequence
is disclosed herein, or if it is a mutant or variant form.
25 Amino acid sequence is routine in the art using automated
sequencing machines.

        A test sample containing one or more polypeptides may be
provided for example as a crude or partially purified cell or
30 cell lysate preparation, e.g. using tissues or cells, such as
from saliva, faeces, or preferably blood, or for pre-natal
testing from the amnion, placenta or foetus itself.

        Whether it is a polypeptide, antibody, peptide, nucleic
35 acid molecule, small molecule or other pharmaceutically useful
compound according to the present invention that is to be
given to an individual, administration is preferably in a
"prophylactically effective amount" or a "therapeutically
effective amount" (as the case may be, although prophylaxis

57

may be considered therapy), this being sufficient to show
benefit to the individual. The actual amount administered,
and rate and time-course of administration, will depend on the
nature and severity of what is being treated. Prescription of
5 treatment, e.g. decisions on dosage etc, is within the
responsibility of general practioners and other medical
doctors.

A composition may be administered alone or in combination
with other treatments, either simultaneously or sequentially
10 dependent upon the condition to be treated.

Pharmaceutical compositions according to the present
invention, and for use in accordance with the present
invention, may include, in addition to active ingredient, a
pharmaceutically acceptable excipient, carrier, buffer,
15 stabiliser or other materials well known to those skilled in
the art. Such materials should be non-toxic and should not
interfere with the efficacy of the active ingredient. The
precise nature of the carrier or other material will depend on
the route of administration, which may be oral, or by
20 injection, e.g. cutaneous, subcutaneous or intravenous.

Pharmaceutical compositions for oral administration may
be in tablet, capsule, powder or liquid form. A tablet may
include a solid carrier such as gelatin or an adjuvant.
Liquid pharmaceutical compositions generally include a liquid
25 carrier such as water, petroleum, animal or vegetable oils,
mineral oil or synthetic oil. Physiological saline solution,
dextrose or other saccharide solution or glycols such as
ethylene glycol, propylene glycol or polyethylene glycol may
be included.

30   For intravenous, cutaneous or subcutaneous injection, or
injection at the site of affliction, the active ingredient
will be in the form of a parenterally acceptable aqueous
solution which is pyrogen-free and has suitable pH,
isotonicity and stability. Those of relevant skill in the art
35 are well able to prepare suitable solutions using, for
example, isotonic vehicles such as Sodium Chloride Injection,
Ringer's Injection, or Lactated Ringer's Injection.
Preservatives, stabilisers, buffers, antioxidants and/or other
additives may be included, as required.

Targeting therapies may be used to deliver the active
agent more specifically to certain types of cell, by the use
of targeting systems such as antibody or cell specific
ligands. Targeting may be desirable for a variety of reasons;
5 for example if the agent is unacceptably toxic, or if it would
otherwise require too high a dosage, or if it would not
otherwise be able to enter the target cells.

Instead of administering an agent directly, it may be be
produced in target cells by expression from an encoding gene
10 introduced into the cells, e.g. in a viral vector (see below).
The vector may be targeted to the specific cells to be
treated, or it may contain regulatory elements which are
switched on more or less selectively by the target cells.
Viral vectors may be targeted using specific binding
15 molecules, such as a sugar, glycolipid or protein such as an
antibody or binding fragment thereof. Nucleic acid may be
targeted by means of linkage to a protein ligand (such as an
antibody or binding fragment thereof) via polylysine, with the
ligand being specific for a receptor present on the surface of
20 the target cells.

An agent may be administered in a precursor form, for
conversion to an active form by an activating agent produced
in, or targeted to, the cells to be treated. This type of
approach is sometimes known as ADEPT or VDEPT; the former
25 involving targeting the activating agent to the cells by
conjugation to a cell-specific antibody, while the latter
involves producing the activating agent, e.g. an enzyme, in a
vector by expression from encoding DNA in a viral vector (see
for example, EP-A-415731 and WO 90/07936).

30

Nucleic acid according to the present invention, e.g.
encoding the authentic biologically active LRP-5 polypeptide
or a functional fragment thereof, may be used in a method of
gene therapy, to treat a patient who is unable to synthesize
35 the active polypeptide or unable to synthesize it at the
normal level, thereby providing the effect provided by the
wild-type with the aim of treating and/or preventing one or
more symptoms of IDDM and/or one or more other diseases.
Vectors such as viral vectors have been used to introduce

genes into a wide variety of different target cells.
Typically the vectors are exposed to the target cells so that
transfection can take place in a sufficient proportion of the
cells to provide a useful therapeutic or prophylactic effect
5 from the expression of the desired polypeptide. The
transfected nucleic acid may be permanently incorporated into
the genome of each of the targeted cells, providing long
lasting effect, or alternatively the treatment may have to be
repeated periodically.

10   A variety of vectors, both viral vectors and plasmid
vectors, are known in the art, see e.g. US Patent No.
5,252,479 and WO 93/07282. In particular, a number of viruses
have been used as gene transfer vectors, including adenovirus,
papovaviruses, such as SV40, vaccinia virus, herpesviruses,
15 including HSV and EBV, and retroviruses, including gibbon ape
leukaemia virus, Rous Sarcoma Virus, Venezualian equine
enchephalitis virus, Moloney murine leukaemia virus and murine
mammary tumourvirus. Many gene therapy protocols in the prior
art have used disabled murine retroviruses.

20   Disabled virus vectors are produced in helper cell lines
in which genes required for production of infectious viral
particles are expressed. Helper cell lines are generally
missing a sequence which is recognised by the mechanism which
packages the viral genome and produce virions which contain no
25 nucleic acid. A viral vector which contains an intact
packaging signal along with the gene or other sequence to be
delivered (e.g. encoding the LRP5 polypeptide or a fragment
thereof) is packaged in the helper cells into infectious
virion particles, which may then be used for the gene
30 delivery.
Other known methods of introducing nucleic acid into cells
include electroporation, calcium phosphate co-precipitation,
mechanical techniques such as microinjection, transfer
mediated by liposomes and direct DNA uptake and receptor-
35 mediated DNA transfer. Liposomes can encapsulate RNA, DNA and
virions for delivery to cells. Depending on factors such as
pH, ionic strength and divalent cations being present, the
composition of liposomes may be tailored for targeting of
particular cells or tissues. Liposomes include phospholipids

and may include lipids and steroids and the composition of
each such component may be altered. Targeting of liposomes
may also be achieved using a specific binding pair member such
as an antibody or binding fragment thereof, a sugar or a
5 glycolipid.

The aim of gene therapy using nucleic acid encoding the
polypeptide, or an active portion thereof, is to increase the
amount of the expression product of the nucleic acid in cells
in which the level of the wild-type polypeptide is absent or
10 present only at reduced levels. Such treatment may be
therapeutic or prophylactic, particularly in the treatment of
individuals known through screening or testing to have an
IDDM4 susceptibility allele and hence a predisposition to the
disease.

15   Similar techiques may be used for anti-sense regulation
of gene expression, e.g. targeting an antisense nucleic acid
molecule to cells in which a mutant form of the gene is
expressed, the aim being to reduce production of the mutant
gene product. Other approaches to specific down-regulation of
20 genes are well known, including the use of ribozymes designed
to cleave specific nucleic acid sequences. Ribozymes are
nuceic acid molecules, actually RNA, which specifically cleave
single-stranded RNA, such as mRNA, at defined sequences, and
their specificity can be engineered. Hammerhead ribozymes may
25 be preferred because they recognise base sequences of about
11-18 bases in length, and so have greater specificity than
ribozymes of the Tetrahymena type which recognise sequences of
about 4 bases in length, though the latter type of ribozymes
are useful in certain circumstances. References on the use of
30 ribozymes include Marschall, et al. Cellular and Molecular
Neurobiology, 1994. 14(5) : 523; Hasselhoff, Nature 334: 585
(1988) and Cech, J. Amer. Med. Assn., 260: 3030 (1988).

Aspects of the present invention will now be illustrated
35 with reference to the accompanying figures described already
above and experimental exemplification, by way of example and
not limitation. Further aspects and embodiments will be
apparent to those of ordinary skill in the art. All documents
mentioned in this specification are hereby incorporated herein

by reference.

EXAMPLE 1

CLONING OF *LRP5*

5   As noted above, confirmation of linkage to two of the 18
potential loci for IDDM predisposition was achieved by
analysis of two family sets (102 UK families and 84 USA
families), *IDDM4* on chromosome 11q13 (MLS 1.3 , P = 0.01 at
FGF3) and *IDDM5* on chromosome 6q (MLS 1.8 P= 0.003 at *ESR*).
10 At *IDDM4* the most significant linkage was obtained in the
subset of families sharing 1 or 0 alleles IBD at HLA (MLS =
2.8; P=0.0002; ls = 1.2) (Davies et al, 1994). This linkage
was also observed by Hashimoto et al (1994) using 251 affected
sibpairs, obtaining P= 0.0008 in all sibpairs. Combining
15 these results, with 596 families, provides substantial support
for *IDDM4* (P = 1.5 X 10^{-6}) (Todd and Farrall, 1996; Luo et al,
1996).

       Multipoint analysis with other markers in the *FGF3*
region produced an MLS of 2.3 at *FGF3* and *D11S1883* (ls =
20 1.19), and delineated the interval to a 27cM region, flanked
by the markers *D11S903* and *D11S527* (Figure 1).

       Multipoint linkage analysis cannot localise the gene to a
small region unless several thousand multiplex families are
available. Instead, association mapping has been used for
25 rare single gene diseases which can narrow the interval
containing the disease gene to less than 2cM or 2M bases.
Nevertheless, this method is highly unpredictable and has not
previously been used to locate a polygene for a common
disease. Association mapping has been used to locate the
30 *IDDM2/INS* polygene but this relied on the selection of a
functional candidate polymorphism/gene and was restricted to a
very small (<30kb) region. Linkage disequilibrium (LD) or
association studies were carried out in order to delineate the
IDDM4 region to less than 2cM. In theory, association of a
35 particular allele very close to the founder mutation will be
detected in populations descended from that founder. The
transmission disequilibrium test (TDT, Spielman et al, 1993)
measures association by assessing the deviation from 50% of

the transmission of alleles from a marker locus from parents
to affected children. The detection of association is
dependent on the ancestry of each population studied to be as
homogeneous as possible, in order to reduce the possiblity
5 that the presence of several founder-chromosomes, decreasing
the power to detect the association. These parameters are
highly unpredictable.

Analysis of markers spanning the *IDDM4* linkage interval,
LD was detected at *D11S1917(UT5620)* in 554 families, P=0.01.
10 A physical map of this region, comprising approximately 500kb,
was achieved by constructing a pac, bac and cosmid contig
(Figure 2). The region was physically mapped by hybridisation
of markers onto restriction-enzyme digested clones resolved
through agarose, and Southern blotted.

15   Further microsatellites (both published, and those
isolated from the clones by microsatellite rescue) were
analysed within 1289 families, from four different populations
(UK, USA, Sardinia and Norway). A LD graph was constructed,
with a peak at *H0570POLYA*, P = 0.001, flanked by the markers
20 *D11S987* and *18018AC* (Figure 3). The LD detected at a
polymorphic marker is influenced by allele frequency, and
whether the mutation causing susceptibility to type 1 diabetes
arose on a chromosome where the allele in LD is the same
allele as that on protective or neutral chromosomes. In the
25 case where the marker being analysed has the same allele in LD
with both susceptible and protective genotypes, these will
remain undetected by single point analysis, in effect
cancelling each other out, and showing little or no evidence
for LD with the disease locus. Unpredictability of the method
30 arising from this has been noted already above.

In order to maximise the information obtained with each
marker, a three point rolling LD curve was produced with the
IDDM4 markers (Figure 4). In this case the percentage
transmission (%T) was calculated from a marker, and its two
35 immediate flanking markers, and averaged between them to
minimise the effects of fluctuating allele frequency. This
also produced a peak at *H0570POLYA*, with P=0.04, and indicates
that the *IDDM4* mutation is more likely to be in the interval
E0864CA - *D11S1337* (75kb).

63

By the identification of this 75kb interval which shows association with type 1 diabetes, disease associated haplotypes were identified. These are derived from the original founder chromosomes on which the diabetes mutation or
5 mutations *IDDM4* arose. In order to identify the mutation causing susceptibility to type 1 diabetes, a refined linkage disequilibrium curve, based on single nucleotide polymorphisms (SNPs) and haplotypes, is constructed. SNPs are identified by sequencing individuals with specific haplotypes which have
10 been identified from the microsatellite analysis: homozygous susceptible to type 1 diabetes, homozygous protective for type 1 diabetes, and controls. One of these SNPs may be the etiological mutation *IDDM4*, or may be in very strong linkage disequilibrium with the primary disease locus, and hence be at
15 a peak of the refined curve. Cross-match analysis further reduces the number of candidate SNPs, as shown by the localisation of the *IDDM2* mutation by this method (Bennett et al, 1995; Bennett and Todd, 1996). This requires identification of distinct haplotypes or founder chromosomes,
20 which have a different arrangement of alleles from the main susceptible or protective haplotypes, so that association or transmission of candidate SNP alleles can be tested in different haplotype backgrounds. The candidate mutations can be assessed for effects on gene function or regulation.
25   In different populations different *IDDM4* mutations may have arisen in the same gene. We are sequencing several putative founder chromosome or disease associated haplotypes from several unrelated individuals from different populations to identify candidate mutations for *IDDM4*, and which cluster
30 in the same gene.

To carry out an extensive search for DNA mutations or polymorphisms, the entire region and flanking regions of the associated region was sequenced (the 75kb core region and 125 kb of flanking DNA). The DNA sequence also aids in gene
35 identification and is complementary to other methods of gene identification such as cDNA selection or gene identification by DNA sequencing and comparative analysis of homologous mouse genomic DNA.

Various strategies were used in the hope of identifying

potential coding sequences within this region: sequencing,
computer prediction of putative exons and promoters, and cDNA
selection, to try to increase the likelihood of identifying
all the genes within this interval.
5

Construction of Libraries for Shotgun Sequencing
        DNA was prepared from either cosmids, BACs (Bacterial
Artificial Chromosomes), or PACs (P1 Artificial Chromosomes).
Cells containing the vector were streaked on Luria-Bertani
10 (LB) agar plates supplemented with the appropriate antibiotic.
A single colony was used to inoculate 200 ml of LB media
supplemented with the appropriate antibiotic and grown
overnight at 37°C. The cells were pelleted by centrifugation
and plasmid DNA was prepared by following the QIAGEN
15 (Chatsworth, CA) Tip500 Maxi plasmid/cosmid purification
protocol with the following modifications; the cells from 100
ml of culture were used for each Tip500 column, the NaCl
concentration of the elution buffer was increased from 1.25M
to 1.7M, and the elution buffer was heated to 65°C.
20   Purified BAC and PAC DNA was digested with Not I
restriction endonuclease and then subjected to pulse field gel
electrophoresis using a BioRad CHEF Mapper system. (Richmond,
CA). The digested DNA was electrophoresed overnight in a 1%
low melting temperature agarose (BioRad, Richmond CA) gel that
25 was prepared with 0.5X Tris Borate EDTA (10X stock solution,
Fisher, Pittsburg, PA). The CHEF Mapper autoalgorithm default
settings were used for switching times and voltages.
Following electrophoresis the gel was stained with ethidium
bromide (Sigma, St. Louis, MO) and visualized with a
30 ultraviolet transilluminator. The insert band(s) was excised
from the gel. The DNA was eluted from the gel slice by beta-
Agarase (New England Biolabs, Beverly MA) digestion according
to the manufacturer's instructions. The solution containing
the DNA and digested agarose was brought to 50 mM Tris pH
35 8.0, 15 mM MgCl2, and 25% glycerol in a volume of 2 ml and
placed in a AERO-MIST nebulizer (CIS-US, Bedford MA). The
nebulizer was attatched to a nitrogen gas source and the DNA
was randomly sheared at 10 psi for 30 sec. The sheared DNA
was ethanol precipitated and resuspended in TE (10 mM Tris, 1

mM EDTA). The ends were made blunt by treatment with Mung
Bean Nuclease (Promega, Madison, WI) at 30°C for 30 min,
followed by phenol/chloroform extraction, and treatment with
T4 DNA polymerase (GIBCO/BRL, Gaithersburg, MD) in multicore
5 buffer (Promega, Madison, WI) in the presence of 40 uM dNTPs
at 16°C. To facilitate subcloning of the DNA fragments, BstX I
adapters (Invitrogen, Carlsbad, CA) were ligated to the
fragments at 14°C overnight with T4 DNA ligase (Promega,
Madison WI). Adapters and DNA fragments less than 500 bp were
10 removed by column chromatography using a cDNA sizing column
(GIBCO/BRL, Gaithersburg, MD) according to the instructions
provided by the manufacturer. Fractions containing DNA
greater than 1 kb were pooled and concentrated by ethanol
precipitation. The DNA fragments containing BstX I adapters
15 were ligated into the BstX I sites of pSHOT II which was
constructed by subcloning the BstX I sites from pcDNA II
(Invitrogen, Carlsbad, CA) into the BssH II sites of
pBlueScript (Stratagene, La Jolla, CA). pSHOT II was prepared
by digestion with BstX I restriction endonuclease and purified
20 by agarose gel electrophoresis. The gel purified vector DNA
was extracted from the agarose by following the Prep-A-Gene
(BioRad, Richmond, CA) protocol. To reduce ligation of the
vector to itself, the digested vector was treated with calf
intestinal phosphatase (GIBCO/BRL, Gaithersburg, MD).
25 Ligation reactions of the DNA fragments with the cloning
vector were transformed into ultra-competent XL-2 Blue cells
(Stratagene, La Jolla, CA), and plated on LB agar plates
supplemented with 100 ug/ml ampicillin. Individual colonies
were picked into a 96 well plate containing 100 ul/well of LB
30 broth supplemented with ampicillin and grown overnight at
37°C. Approximately 25 ul of 80% sterile glycerol was added
to each well and the cultures stored at -80°C.

Preparation of plasmid DNA
35   Glycerol stocks were used to inoculate 5 ml of LB broth
supplemented with 100 ug/ml ampicillin either manually or by
using a Tecan Genesis RSP 150 robot (Tecan AG, Hombrechtikon,
Switzerland) programmed to inoculate 96 tubes containing 5 ml
broth from the 96 wells. The cultures were grown overnight at

37°C with shaking to provide aeration. Bacterial cells were
pelleted by centrifugation, the supernatant decanted, and the
cell pellet stored at -20°C. Plasmid DNA was prepared with a
QIAGEN Bio Robot 9600 (QIAGEN, Chatsworth CA) according to the
5 Qiawell Ultra protocol. To test the frequency and size of
inserts plasmid DNA was digested with the restriction
endonuclease Pvu II. The size of the restriction endonuclease
products was examined by agarose gel electrophoresis with the
average insert size being 1 to 2 kb.
10

DNA Sequence Analysis of Shotgun clones
        DNA sequence analysis was performed using the ABI PRISM™
dye terminator cycle sequencing ready reaction kit with
AmpliTaq DNA polymerase, FS (Perkin Elmer, Norwalk, CT). DNA
15 sequence analysis was performed with M13 forward and reverse
primers. Following amplification in a Perkin-Elmer 9600 the
extension products were purified and analyzed on an ABI PRISM
377 automated sequencer (Perkin Elmer, Norwalk, CT).
Approximately 12 to 15 sequencing reactions were performed per
20 kb of DNA to be examined e.g. 1500 reactions would be
performed for a PAC insert of 100 kb.

Assembly of DNA sequences
        Phred/Phrap was used for DNA sequences assembly. This
25 program was developed by Dr. Phil Green and licensed from the
University of Washington (Seattle, WA). Phred/Phrap consists
of the following programs: Phred for base-calling, Phrap for
sequence assembly, Crossmatch for sequence comparisons, Consed
and Phrapview for visualization of data, and Repeatmasker for
30 screening repetitive sequences. Vector and *E. coli* DNA
sequences were identified by Crossmatch and removed from the
DNA sequence assembly process. DNA sequence assembly was on a
SUN Enterprise 4000 server running Solaris 2.51 operating
system (Sun Microsystems Inc., Mountain View, CA) using
35 default Phrap parameters. The sequence assemblies were
further analyzed using Consed and Phrapview.

BioInformatic Analysis of Assembled DNA Sequences
        When the assembled DNA sequences approached five to six

fold coverage of the region of interest the exon and promoter
prediction abilities of the program GRAIL (ApoCom, Oak Ridge)
were utilized to aid in gene identification. ApoCom GRAIL is
a commercial version of the Department of Energy developed
5 GRAIL Gene Characterization Software licensed to ApoCom Inc.
by Lockheed Martin Energy Research Corporation and ApoCom
Client Tool for Genomics (ACTG) TM.

The DNA sequences at various stages of assembly were
queried against the DNA sequences in the GenBank database
10 (subject) using the BLAST algorithm (S. F. Altschul, et al.
(1990) J. Mol. Biol. 215, 403-410), with default parameters.
When examining large contiguous sequences of DNA repetitive
elements were masked following identification by crossmatch
with a database of mammalian repetitive elements. Following
15 BLAST analysis the results were compiled by a parser program
written by Dr. Guochun Xie (Merck Research Lab). The parser
provided the following information from the database for each
DNA sequence having a similarity with a P value greater than
10^{-6}; the annotated name of the sequence, the database from
20 which it was derived, the length and percent identity of the
region of similarity, and the location of the similarity in
both the query and the subject.

The BLAST analysis identified a high degree of
similarities (90-100% identical) over a length of greater than
25 100 bp between DNA sequences we obtained and a number of human
EST sequences present in the database. These human EST
sequences clustered into groups that are represented by
accession numbers; R73322, R50627, F07016. In general, each
EST cluster is presumed to represent a single gene. The DNA
30 sequences in R73322 cluster of 424 nucleotides had a lower but
significant degree of DNA sequence similarity to the gene
encoding the LDL receptor related protein (GenBank accession
number X13916) and several other members of the LDL receptor
family. Therefore it was concluded that the sequences that
35 were highly similar to EST R73322 encoded a member of the LDL
receptor family.

Members of each EST cluster were assembled using the
program Sequencher (Perkin Elmer, Norwalk CT). To increase

the accuracy of the EST sequence data extracted from the database relevent chromatogram trace files from the genomic DNA sequences obtained from shotgun sequencing were included in the assembly. The corrected EST sequences were reanalyzed

5 by BLAST and BLASTX. For EST cluster 3, represented by accession number R50627 analysis of the edited EST assembly revealed that this cluster was similar to members of the LDL receptor family. This result suggested the possibility that these two EST clusters were components of the same gene.

10   Experimentally derived cDNA sequences were assembled using the program Sequencher (Perkin Elmer, Norwalk CT). Genomic DNA sequences and cDNA sequences were compared by using the program Crossmatch which allowed for a rapid and sensitive detection of the location of exons. The

15 identification of intron/exon boundaries was then accomplished by manually comparing the genomic and cDNA sequences by using the program GeneWorks (Intelligenetics Inc., Campbell CA).

Northern Blot Analysis

20   Primers 256F and 622R (Table 2) were used to amplify a PCR product of 366 bp from a fetal brain cDNA library. This product was purified on an agarose gel, the DNA extracted, and subcloned into pCR2.1 (Invitrogen, Carlsbad, CA) . The 366 bp probe was labeled by random priming with the Amersham

25 Rediprime kit (Arlington Heights, IL) in the presence of 50-100 uCi of 3000 Ci/mmole [alpha ^{32}P]dCTP (Dupont/NEN, Boston, MA). Unincorporated nucleotides were removed with a ProbeQuant G-50 spin column (Pharmacia/ Biotech, Piscataway, NJ). The radiolabeled probe at a concentration of greater

30 than 1×10^6 cpm/ml in rapid hybridization buffer (Clontech, Palo Alto, CA) was incubated overnight at 65°C with human multiple tissue Northern's I and II (Clontech, Palo Alto, CA). The blots were washed by two 15 min incubations in 2X SSC, 0.1% SDS (prepared from 20X SSC and 20 % SDS stock solutions,

35 Fisher, Pittsburg, PA) at room temperature, followed by two 15 min incubations in 1X SSC, 0.1% SDS at room temperature, and two 30 min incubations in 0.1X SSC, 0.1% SDS at 60°C. Autoradiography of the blots was done to visualize the bands

69

that specifically hybridized to the radiolabeled probe.

The probe hybridized to an approximately 5-5.5 kb mRNA transcript that is most highly expressed in placenta, liver, pancreas, and prostate. It is expressed at an intermediate
5 level in lung, skeletal muscle, kidney, spleen, thymus, ovary, small intestine, and colon. The message is expressed at a low level in brain, testis, and leukocytes. In tissues where the transcript is highly expressed, e.g. liver and pancreas, additional bands of 7kb and 1.3 kb are observed.
10

Isolation of full length cDNAs

PCR based techniques were used to extend regions that were highly similar to ESTs and regions identified by exon prediction software (GRAIL). The one technique utilized is a
15 variation on Rapid Amplification of cDNA Ends (RACE) termed Reduced Complexity cDNA Analysis (RCCA) similar procedures are reported by Munroe *et .al.* (1995) PNAS 92: 2209-2213 and Wilfinger *et. al.* (1997) BioTechniques 22: 481-486. This technique relies upon a PCR template that is a pool of
20 approximately 20,000 cDNA clones, this reduces the complexity of the template and increases the probability of obtaining longer PCR extensions. A second technique that was used to extend cDNAs was PCR between regions that were identified in the genomic sequence of having the potential to be portions of
25 a gene e.g. sequences that were very similar to ESTs or sequences that were identified by GRAIL. These PCR reactions were done on cDNA prepared from approximately 5 ug of mRNA (Clontech, Palo Alto, CA) with the SuperScript™ choice system (Gibco/BRL, Gaithersburg, MD). The first strand cDNA
30 synthesis was primed using 1 ug of oligo(dT)$_{12-18}$ primer and 25 ng of random hexamers per reaction. Second strand cDNA synthesis was performed according to the manufacturer's instructions.

35 *Identification of additional exons related to EST cluster 1*

We scanned 96 wells of a human fetal brain plasmid library, 20,000 clones per well, by amplifying a 366 bp PCR product using primers 256F and 622R. The reaction mix

consisted of 4 ul of plasmid DNA (0.2 ng/ml), 10 mM Tris-HCl
pH 8.3, 50 mM KCl, 10% sucrose, 2.5 mM $MgCl_2$, 0.1% Tetrazine,
200 mM dNTP's, 100 ng of each primer and 0.1 ul of Taq Gold
(Perkin-Elmer, Norwalk, CT). A total reaction volume of 11 ul

5 was incubated at 95°C for 12 min followed by 32 cycles of 95°C
for 30 sec, 60°C, for 30 sec and 72°C for 30 sec.
Approximately 20 wells were found to contain the correct 366
bp fragment by PCR analysis. 5' and 3' RACE was subsequently
performed on several of the positive wells containing the

10 plasmid cDNA library using a vector specific primer and a gene
specific primer. The vector specific primers, PBS 543R and
PBS 873F were both used in combination with gene specific
primers 117F and 518R because the orientation of the insert
was not known. PCR amplification conditions consisted of 1X

15 TaKaRa Buffer LA, 2.5 mM $MgCl_2$, 500 mM dNTP's, 0.2 ul of
TaKaRa LA Taq Polymerase (PanVera, Madison WI), 100 ng of each
primer and 5 ul of the plasmid library at 0.2 ng/ml. In a
total reaction volume of 20 ml, the thermal cycling conditions
were as follows: 92°C for 30 sec, followed by 32 cycles of

20 92°C for 30 sec, 1 min at 60°C and 10 min at 68°C. After the
initial PCR amplification, a nested or semi-nested PCR
reaction was performed using nested vector primers PBS 578R
and PBS 838F and various gene specific primers (256F, 343F,
623R and 657R). The PCR products were separated from the

25 unincorporated dNTP's and primers using QIAGEN, QIAquick PCR
purification spin columns using standard protocols and
resuspended in 30 ul of water. The amplification conditions
for the nested and semi-nested PCR were the same as the
initial PCR amplification except that 3 ul of the purified PCR

30 fragment was used as template and that the cycling conditions
were for only 20 cycles. Products obtained from this PCR
amplification were analyzed on 1% agarose gels, excised
fragments were purified using QIAGEN QIAquick spin columns and
sequenced using ABI dye-terminator sequencing kits. The

35 products were analyzed on ABI 377 sequencers according to
standard protocols.

Connection of EST clusters 1-3

As discussed above it is possible that each EST cluster represents a single gene, alternatively the EST clusters may be portions of the same gene. To distinguish between these two possiblities, primers were designed to the two other EST
5 clusters in the region represented by EST accession numbers F07016 (cluster 2, containing 272 nucleotides) and R50627 (cluster 3, containing 1177 nucleotides). Primers from cluster 1 (117F and 499F) were paired with a primer from EST cluster 3 (4034R) in a PCR reaction. A 50 ul reaction was
10 performed using the Takara LA Taq polymerase (Panvera, Madison, WI) in the reaction buffer supplied by the manufacturer with the addition of 0.32 mM dNTPs, primers, and approximately 30ng of lymph node cDNA. PCR products were amplified for 35 cycles of 94°C for 30 sec, 60°C for 30 sec,
15 and 72°C for 4 minutes. Products were electrophoresed on a 1% agarose gel and bands of 2.5 to 3 kb were excised, subcloned into pCR 2.1 (Invitrogen, Carlsbad, CA), and plasmid DNA was prepared for DNA sequence analysis.

The primary reaction described above generated by a
20 primer in EST cluster 1 (638F) and EST cluster 3 (4173R) was utilized as the template for a reaction with a primer from EST cluster 1 (638F) and from EST cluster 2 (3556R). This semi-nested PCR reaction was performed with Takara LA Taq polymerase as described in the previous paragraph. An
25 approximately 2 kb product was generated and subcloned for DNA sequence analysis. The assembly of the DNA sequence results of these PCR products indicated that EST clusters 1 to 3 were part of the same gene and established their orientation relative to each other in the mRNA transcript produced by this
30 gene.

PCR reactions were also performed between EST clusters 2 and 3. Amplification from liver cDNA using Takara LA Taq polymerase (Panvera, Madison, WI) with the primers 2519F, 3011F, or 3154F (EST cluster 2) in combination with 5061R (EST
35 cluster 3) was done for 35 cyles of 95°C for 30 sec, 60°C for 60 sec, and 72°C for 3 minutes. The PCR products were gel purified, subcloned, and the DNA sequence was determined. The DNA sequence analysis of the ends of all these PCR products resulted in most of the cDNA sequence however to provide for

complete DNA sequence of both strands oligonucleotide primers
were designed and used for DNA sequencing (Figure 5(a)).

Extension of the 5' end

5   RCCA analysis was utilized to obtain a number of clones
extended 5' by using the internal gene specific primers as
described previously. Several clonal extensions were isolated
however most of the clones analyzed stopped within exon A.
One clone extended past the 5' end of exon A but the sequence
10 was contiguous with genomic DNA, since a body of evidence
indicates an intron/exon boundary at the 5' end of exon A it
appeared likely that this extension is a result of unprocessed
intronic sequence. A second clone h10 extended past this
point but diverged from the genomic DNA sequence. It was
15 concluded that this represented a chimeric clone that was
present in the original fetal brain cDNA library.

Identification of 5' end of isoform 1

         As described above results from RCCA experiments yielded
20 a number of independent clones that terminated at the 5' end
of exon A. This suggested that the human *LRP5* gene contains a
region that the reverse transcriptase has difficulty
transcribing. To circumvent this problem we decided to
isolate the mouse ortholog of *LRP5*, since subtle differences
25 in DNA sequence content can alter the ability of an enzyme to
transcribe a region. To increase the probability of isolating
the 5' portion of the mouse gene a human probe of 366
nucleotides, described above and derived from exons A and B
was used.
30   A cDNA library was constructed from mouse liver mRNA
purchased from Clontech (Palo Alto, CA). cDNA was prepared
using the SuperScript Choice system (Gibco/BRL Gaithersburg,
MD) according to the manufacturer's instructions.
Phosphorylated *Bst* XI adapters (Invitrogen, San Diego, CA)
35 were ligated to approximately 2 ug of mouse liver cDNA. The
ligation mix was diluted and size-fractionated on a cDNA
sizing column (Gibco/BRL Gaithersburg, MD). Drops from the
column were collected and the eluted volume from the column
determined as described for the construction of shotgun

libraries. The size-fractionated cDNA with the *Bst* XI
linkers was ligated into the vector pSHOT II, described above,
cut with the restriction endonuclease *Bst* XI, gel purified,
and dephosphorylated with calf intestinal phosphatase
5 (Gibco/BRL, Gaithersburg, MD). The ligation containing
approximately 10-20 ng of cDNA and approximately 100 ng of
vector was incubated overnight at 14°C. The ligation was
transformed into XL-2 Blue Ultracompetent cells (Stratagene,
La Jolla, CA.). The transformed cells were spread on twenty
10 133 mm Colony/Plaque Screen filters (Dupont/NEN, Boston, MA.)
at a density of approximately 30,000 colonies per plate on
Luria Broth agar plates supplemented with 100 ug/ml ampicillin
(Sigma, St. Louis, MO.). The colonies were grown overnight
and then replica plated onto two duplicate filters. The
15 replica filters were grown for several hours at 37°C until the
colonies were visible and processed for in situ hybridization
of colonies according to established procedures (Maniatis,
Fritsch and Sambrook, 1982). A Stratalinker (Stratagene, La
Jolla, CA.) was used to crosslink the DNA to the filter. The
20 filters were hybridized overnight with greater than 1,000,000
cpm/ml probe in 1X hybridization buffer (Gibco/BRL,
Gaithersburg, MD) containing 50% formamide at 42°C. The probe
was generated from a PCR product derived from the human *LRP5*
cDNA using primers 512F and 878R. This probe was random prime
25 labeled with the Amersham Rediprime kit (Arlington Heights,
IL) in the presence of 50-100 uCi of 3000 Ci/mmole [alpha
32P]dCTP (Dupont/NEN, Boston, MA) and purified using a
ProbeQuant G-50 spin column (Pharmacia/Biotech, Piscataway,
NJ). The filters were washed with 0.1X SSC, 0.1% SDS at 42°C.
30 Following autoradiography individual regions containing
hybridization positive colonies were excised from the master
filter and placed into 0.5 ml Luria Broth plus 20% glycerol.
Each positive was replated at a density of approximate 50-200
colonies per 100 mm plate and screened by hybridization as
35 described above. Single colonies were isolated and plasmid
DNA was prepared for DNA sequence analysis.

Three clones were isolated from the mouse cDNA library
the assembled sequence of the clones (Figure 16(a)) that had a
high degree of similarity (87% identical over an approximately

1700 nucleotide portion) with the human *LRP5* gene and thus
likely represent the mouse ortholog of *LRP5*. The 500 amino
acid of the portion of the mouse LRP5 (Figure 16(d)) that we
initially obtained is 96% identical to human LRP5.

5 Significantly two of these clones had sequence that was 5' of
the region corresponding to exon A, clone 19a contained an
additional 200 bp and clone 9a contained an additional 180 bp
(Figure 16(b)). The additional 200 bp contains an open
reading frame that begins at bp 112 (Figure 16(c)). The
10 initiating codon has consensus nucleotides for efficient
initiation of translation at both the -3 (purine) and +4 (G
nucleotide) positions (Kozak, M. 1996, Mamalian Genome 7:563-
574). This open reading frame encodes a peptide with the
potential to act as a eukaryotic signal sequence for protein
15 export (von Heijne, 1994, Ann. Rev. Biophys. Biomol. Struc.
23:167-192). The highest score for the signal sequence as
determined by using the SigCleave program in the GCG analysis
package (Genetics Computer Group, Madison WI) generates a
mature peptide beginning at residue 29 of isoform 1.
20 Additional sites that may be utilized produce mature peptides
beginning at amino acid residue 31 (the first amino acid
encoded by exon A) or amino acid residues 32, 33, or 38.

Molecular cloning of the full length mouse Lrp3 *cDNA*
25   The mouse cDNA clones isolated by nucleic acid
hybridization contain 1.7 Kb of the 5' end of the *Lrp3* cDNA
(Figure 16(a)). This accounts for approximately one-third of
the full length cDNA when compared to the human cDNA sequence.
The remainder of the mouse *Lrp3* cDNA was isolated using PCR to
30 amplify products from mouse liver cDNA. PCR primers, Table 9,
were designed based upon DNA sequences identified by the
sequence skimming of mouse genomic clones, BACs 53-d-8 and
131-p-15, which contain the mouse *Lrp3* gene. BAC 53-d-8 was
mapped by FISH analysis to mouse chromosome 19 which is
35 syntenic with 11q13. Sequence skimming of these clones
identified DNA sequences that corresponded to the coding
region of human *LRP5* as well as the 3' untranslated region.
This strategy resulted in the determination of a mouse cDNA
sequence of 5059 nucleotides (Figure 18(a)) which contains an

open reading frame of 4842 nucleotides (Figure 18(b)) that
encodes a protein of 1614 amino acids (Figure 18(c)). The
putative ATG is in a sequence context favorable for initiation
of translation (Kozak, M. 1996, Mamalian Genome 7:563-574).

5

Comparison of human and mouse LRP5

        The cDNA sequences of human and mouse *LRP5* display 87%
identity. The open reading frame of the human *LRP5* cDNA
encodes a protein of 1615 amino acids that is 94% identical to
10 the 1614 amino acid protein encoded by mouse *Lrp3* (Figure
18(d)). The difference in length is due to a single amino
acid deletion in the mouse Lrp3 signal peptide sequence. The
signal peptide sequence is not highly conserved being less
than 50% identical between human and mouse. The location of
15 the putative signal sequence cleavage site is at amino acid
residue 25 in the human and amino acid 29 in the mouse.
Cleavage at these sites would result in mature human and mouse
proteins of 1591 and 1586 amino acids, respectively, which are
95% identical (Figure 18(e)). The high degree of overall
20 sequence similarity argues strongly that the identified
sequences are orthologs of the *LRP5* gene. This hypothesis is
further supported by the results of genomic Southern
experiments (data not shown).

25 *Identification of human signal peptide exon for isoform 1*
        The human exon encoding a signal peptide was isolated
from liver cDNA by PCR. The forward primer 1F (Table 9) was
used in combination with one of the following reverse primers:
218R, 265R, 318R, and 361R in a PCR reaction using Taq Gold
30 polymerase (Perkin-Elmer, Norwalk, CT) and supplemented with
either 3, 5, or 7% DMSO. Products were amplified for 40
cycles of 30 sec 95°C, 30 sec 58°C, and 1 min 72°C. The
products were analyzed on an agarose gel and some of the
reactions containing bands of the predicted size were selected
35 for DNA sequence analysis and subcloning into pCR2.1
(Invitrogen, San Diego, CA).
        The derived DNA sequence of 139 nucleotides upstream of
exon 2 (also known as exon A) contains an ATG that is in a
context for efficient initiation of translation: an adenine

(A) residue at the -3 position and a guanine (G) residue at
the +4 position (Kozak, M. 1996, Mamalian Genome 7:563-574).
The open reading frame for this ATG continues for 4854
nucleotides (Figure 5(b)) which encodes a polypeptide of 1615
5 amino acids (Figure 5(c)

       The sequence following the initiator ATG codon encodes a
peptide with the potential to act as a signal for protein
export. The highest score for the signal sequence (15.3)
indicated by the SigCleave program in the GCG analysis package
10 (Genetics Computer Group, Madison WI) generates a mature
polypeptide beginning at amino acid residue 25 (Figure 5(d,e).
Additional putative cleavage sites that may be utilized to
produce a mature LRP5 protein are predicted for residues 23,
24, 26, 27, 28, 30 and 32 (the first amino acid encoded by
15 exon A).

*Determination of the genomic DNA sequence containing and
flanking the signal peptide exon*
       The region that contained genomic DNA sequence identical
20 to the cDNA sequence encoding a signal peptide was in a gap
between two stretches of contiguous genomic DNA sequence known
as contigs 57 and 58. To close this gap four clones were
chosen from the shotgun library that were determined to span
this gap according to analysis by the program Phrapview
25 licensed from Dr. Phil Green of the University of Washington
(Seattle, WA). Direct DNA sequencing of these clones was
unsuccessful, i.e. high GC content significantly reduced the
efficiency of the cycle sequencing. To circumvent this
problem PCR products were generated by incorporating 7-deaza-
30 dGTP (Pharmacia, Pharmacia Biotech, Piscataway, NJ). The
conditions for these reactions consisted of a modification of
the Klentaq Advantage-GC polymerase kit (Clontech, Palo Alto,
CA). The standard reaction protocol was modified by
supplementing the reaction mix with 200 uM 7-deaza-dGTP.
35 Inserts were amplified with M13 forward and reverse primers
for 32 cycles of 30 sec at 92°C, 1 min at 60°C, and 5 min at
68°C. Products were gel purified using Qiaquick gel
extraction kit (Qiagen Inc., Santa Clarita, CA) and sequenced
as described previously. Assembly of the resulting sequences

closed the gap and generated a contiguous sequence of
approximately 78,000 bp of genomic DNA.

Extension of Isoforms 2 and 3

5   The software package GRAIL (supra) predicts exons and
promoter sequences from genomic DNA sequence. One region
identified by GRAIL is an exon originally designated G1 and
subsequently termed exon 1 that is approximately 55 kb
upstream of the beginning of exon A (Figure 12(c)). Three
10 primers designated G1 1f to 3f were designed based on this
sequence. This exon was of particular interest because GRAIL
also predicted a promoter immediately upstream of the exonic
sequence (Figure 12 (e)). Furthermore one of the open reading
frames in G1 encoded a peptide that had the characteristics of
15 a eukaryotic signal sequence.
       To determine whether the G1 predicted exon was part of
the *LRP5* gene, reverse transcriptase (RT) PCR was performed
using the Taqara RNA PCR kit (Panvera, Madison WI). Human
liver mRNA (50 ng) was used as the template for a 10 ul
20 reverse transcriptase reaction. The reverse transcriptase
reaction using one of the *LRP5* specific primers (622R, 361R,
or 318R) was incubated at 60°C for 30 min, followed by 99°C
for 5 min, and then the sample was placed on ice. One of the
forward primers, Table 2, (G1 1f, 2f, or 3f) was added along
25 with the reagents for PCR amplification and the reaction was
amplified for 30 cycles of 30 sec at 94°C, 30 sec at 60°C, and
2 min at 72°C. This primary PCR reaction was then diluted 1:2
in water and 1 ul of the reaction was used in a second 20 ul
reaction using nested primers. The reaction conditions for
30 the second round of amplification were 30 cycles of 94°C for
30 sec, 60°C for 30 sec and 72°C for 2 min. The products were
separated on an agarose gel and excised. The purified
fragments were subcloned into pCR 2.1 (Invitrogen, Carlsbad,
CA), plasmid DNA was prepared, and the DNA sequence was
35 determined.
       The DNA sequence of these products indicated that G1
(exon 1) was present on at least a portion of the *LRP5*
transcripts. Two different isoforms were identified. The
first, isoform 2 (Figure 11(a)), identified in this experiment

consists of exon 1 followed by an exon that we have given the
designation exon 5. This splice variant has an open reading
frame that initiates in exon B nucleotide 402 (Figure 11 (a)),
the initiator methionine at this location does not conform to
5 the consensus sequences for translation initiation (Kozak, M.
(1996) Mamalian Genome 7:563-574). A second potential
initiator methionine is present at nucleotide 453, this codon
is in a context for efficient initiation of translation
initiation (Kozak, M. (1996) Mamalian Genome 7:563-574). The
10 longest potential open reading frame for isoform 2 (Figure 11
(c)) encodes a splice variant contains a eukaryotic signal
sequence at amino acid 153. The mature peptide generated by
this splice variant would be lacking the first five spacer
domains and a portion of the first EGF-like motif.

15   The second isoform (isoform 3) consists of exon 1
followed by exon A (Figure 12 (a)). It is not known whether
exon 1 is the first exon of isoform 2. However the location
of a GRAIL predicted promoter upstream of G1 suggests the
possibility that exon 1 is the first exon. Futhermore there
20 is an open reading frame that extends past the 5' intron/exon
boundary postulated by GRAIL (Figure 12 (b)). Therefore we
have examined the possiblity of incorporating this extended
open reading frame into the *LRP5* transcript. The resulting
open reading frame (Figure 12 (c)) encodes a 1639 amino acid
25 protein (Figure 12 (d). The initiator methionine codon does
not contain either of the consensus nucleotides that are
thought to be important for efficient translation (Kozak, M.
1996, Mamalian Genome 7:563-574). Nor does the predicted
protein contain a predicted eukaryotic signal sequence within
30 the first 100 amino acids. Alternatively there may be
additional exons upstream of exon 1 which provide the
initiator methionine codon and/or a potential signal sequence.

RACE extension of the 5' end of lrp5: Isoforms 4 and 5
35   RACE is an established protocol for the analysis of cDNA
ends. This procedure was performed using the Marathon RACE
template purchased from Clontech (Palo Alto, CA). This was
performed according to instructions using Clontech "Marathon"
cDNA from fetal brain and mammary tissue. Two "nested" PCR

amplifications were performed using the ELONGASE™ long-PCR
enzyme mix & buffer from Gibco-BRL (Gaithersburg, MD).

Marathon primers
5 AP1:   CCATCCTAATACGACTCACTATAGGGC
   AP2:   ACTCACTATAGGGCTCGAGCGGC
        First round PCR used 2 microliters Marathon placenta cDNA
   template and 10 pmoles each of primers L217 and AP1. Thermal
   cycling was: 94°C 30 sec, 68°C 6 min, 5 cycles; 94°C 30 sec,
10 64°C 30 sec, 68°C 4 min, 5 cycles; 94°C 30 sec, 62°C 30 sec,
   68°C 4 min, 30 cycles. One microliter from a 1/20 dilution of
   this reaction was added to a second PCR reaction as DNA
   template. This PCR reaction also differed from the first PCR
   reaction in that nested primers L120 and AP2 were used. Two
15 products of approximately 1600 bp and 300 bp were observed and
   cloned into pCR2.1 (Invitrogen, Carlsbad CA). The DNA
   sequence of these clones indicated that they were generated by
   splicing of sequences to exon A. The larger 1.6 kb fragment
   (Figure 13) identified a region approximately 4365 nucleotides
20 upstream of exon A and appeared to be contiguous with genomic
   DNA for 1555 base pairs. The sequence identified by the 300
   bp fragment was approximately 5648 nucleotides upstream of
   exon A (Figure 14). This sequence had similarity to Alu
   repeats. The region identified by the 300 bp fragment was
25 internal to the region identified by the 1.6 kb fragment. The
   open reading frame for these isoforms designated 4 and 5 is
   the same as described for isoform 2 (Figure 11 (b)).

   Extension of Isoform 6
30   GRAIL (supra) analysis was used to predict potential
   promoter regions for the gene. Primers were designed to the
   isoform 6 promoter sequence (Figure 15 (b)) which was defined
   by GRAIL and is approximately 4 kb centromeric of exon A.
   This region was designated GRAIL promoter-1 (Gp-1).
35   The PCR primer Gp 1f (Table 2) was used in a PCR reaction
   with primer 574r and 599r using the polymerase Taq Gold in the
   reaction buffer supplied by the manufacturer (Perkin Elmer,
   Norwalk, CT). The reaction conditions were 12 min at 95°C
   followed by 35 cycles of 95°C for 30 sec, 60°C for 30 sec, and

72°C for 1 min 30 sec with approximately 10ng of liver cDNA
per 20 ul reaction. The primary reactions were diluted 20
fold in water and a second round of PCR using primer Gp 1f in
combination with either 474r or 521r was done. Products were
5 analyzed on a 2% agarose gel and bands of approximately 220 to
400 bp were subcloned into pCR 2.1 (Invitrogen, Carlsbad, CA)
and analyzed by DNA sequence analysis. The open reading frame
present in isoform 4 is the same as described for isoform 2
above (Figure 11 (b)).

10

Microsatellite Rescue

        A vectorette library was made from each clone by
restricting each clone and ligating on a specific bubble
linker (Munroe, D.J. et al. (1994) Genomics 19, 506). PCR was
15 carried out beween a primer (Not 1-A) specific for the linker,
and a repeat motif (AC)11N, (where N is not A), at an
annealing temperature of 65°C. The PCR products were gel
purified and sequenced using the ABI PRISM dye terminator
cycle sequencing kit as previously described. From this
20 sequence, a primer was designed, which was used in PCR with
the Not 1-A primer. This was also sequenced, and a second PCR
primer designed, (Table 8) so that both primers flanked the
repeat motif, and were used for genotyping.

25 *Mutation Scanning*

        Single nucleotide polymorphisms (SNP's) were identified
in type 1 diabetic patients using a sequencing scanning
approach (Table 5).

        Primers were designed to specifically amplify genomic
30 fragments, approximately 500 to 800 bp in length, containing
specific regions of interest (i.e. regions that contained *LRP5*
exons, previously identified SNP's or GRAIL predicted exons).
To facilitate fluorescent dye primer sequencing, forward and
reverse primer pairs were tailed with sequences that
35 correspond to the M13 Universal primer (5'-TGTAAAACGACGGCCAGT-
3') and a modified M13 reverse primer (5'-
GCTATGACCATGATTACGCC-3'), respectively. PCR products produced
using the primer sets, mentioned above, were amplified in 50
ul reactions consisting of Perkin-Elmer 10 x PCR Buffer, 200

mM dNTP's, 0.5 ul of Taq Gold (Perkin-Elmer Corp., Foster
City, CA), 50 ng of patient DNA and 20 pmol/ml of forward and
reverse primers. Cycling conditions were 95°C for 12 min; 35
cycles of 95°C for 30 sec, 57°C for 30 sec and 68°C for 2 min,
5 followed by an extension of 72°C for 6 min and a 4°C hold.

Conditions were optimized so that only single DNA
fragments were produced by these reaction. The PCR products
were then purified for sequencing using QiaQuick strips or
QiaQuick 96 well plates on the Qiagen robot (Qiagen Inc.,
10 Santa Clarita, CA). This purification step removes the
unincorporated primers and nucleotides.

Direct BODIPY dye primer cycle sequencing was the method
used to analyze the PCR products (Metzker et. al. (1996)
15 Science 271, 1420-1422). A Tecan robot (Tecan, Research
Triangle Park, NC) carried out the sequencing reactions using
standard dye primer sequencing protocols (ABI Dye Primer Cycle
Sequencing with AmpliTaq DNA Polymerase FS, Perkin-Elmer
Corp., Foster City, CA). The reactions were generated using
20 the following cycling conditions on a DNA Engine thermal
cycler (M.J. Research Inc., Watertown, MA), 15 cycles of 95°C
for 4 sec, 55°C for 10 sec, and 70°C for 60 sec; followed by
15 cycles of 95°C for 4 sec, and 70°C for 60 sec. After
cycling, samples were pooled, precipitated and dried down.
25 The samples were resuspended in 3 ul of loading buffer and 2
ml were run on an ABI 377 Automated DNA sequencer.

Once SNP's have been identified, scanning technologies
are employed to evaluate their informativeness as markers to
assist in the determination of association of the gene with
30 disease in the type 1 diabetic families. We are using
restriction fragment length polymorphisms (RFLP's) to assess
SNP's that change a restriction endonuclease site.
Furthermore, we are using forced RFLP PCR (Li and Hood (1995)
Genomics 26, 199-206; Haliassos et.al. (1989) Nuc. Acids Res.
35 17, 3608) and ARMS (Gibbs et.al. (1989) Nuc. Acids Res. 17,
2437-2448; Wu et. al. (1989) Proc. Natl. Acad. Sci. USA 86,
2757-2760) to evaluate SNP's that do not change a restriction
endonuclease site. We are also trying to scan larger regions
of the locus by developing fluorescent based Cleavase (CFLP)

(Life Technologies, Gaithersburg, MD) and Resolvase, (Avitech
Diagnostics, Malvern, PA) assays.

Haplotype analysis at IDDM4

5   Haplotype mapping (or identity-by-descent mapping) has
been used in conjunction with association mapping to identify
regions of identity-by-descent (IBD) in founder populations,
where (some) of the affected individuals in a founder
population share not only the mutation, but also a quite large
10 genomic haplotype (hence identical piece of DNA) surrounding
the disease locus. Recombinant haplotypes can be utilised to
delineate the region containing the mutation. These methods
have been used to map the genes of the recessive disorders:
Wilson's disease, Batten's disease, Hirschsprung's disease and
15 hereditary haemochromatosis (Tanzi, R., *et al.* (1993) *Nature
Genet* 5, 344-350; The International Batten Disease Consortium.
(1995) *Cell* 82, 949-957; Puffenberger, E., *et al.* (1994) *Hum
Mol Genet* 3, 1217-1225; and Feder, J., *et al.* (1996) *Nature
Genet* 13, 399-408). Similarly, in type 1 diabetes, for *IDDM1*,
20 comparative MHC haplotype mapping between specific Caucasian
and haplotypes of African origin identified both *HLA-DQA1* and
HLA-DQB1 as susceptibility loci for this disorder (Todd, J. *et
al* (1989) *Nature* 338, 587-589; and Todd, J. *et al* (1987)
Nature 329, 599-604).

25   On chromosome 11q13 haplotype analysis was undertaken
in conjunction with association analysis in order to identify
regions of IBD between haplotypes which are transmitted more
often than expected, hence contain a susceptible allele at the
aetiological locus; in contrast protective haplotypes will be
30 transmitted less often than expected and contain a different
(protective) allele at the aetiological locus. Evidence for a
deviation in the expected transmission of alleles was shown
with the two polymorphic markers *D11S1917* and *H0570POLYA* . In
2042 type 1 diabetic families from the UK, USA, Norway,
35 Sardinia, Romania, Finland, Italy and Denmark, transmission of
D11S1917-H0570POLYA haplotype 3-2 to affected offspring was
negative (46%), with a 2X2 test of heterogeneity between
affected and unaffected transmissions produced $\chi^2=23$, df=1,
p<1.5 x 10^{-6}, providing good evidence that this is a

protective haplotype. In contrast, the 2-3 haplotype was more
transmitted to affected than non-affected offspring (%T=51.3;
2X2 contingency test; χ^2=5.5, df=1, p<0.02), indicating that
this was a susceptible (or possibly neutral) chromosome. A
5 further haplotype, which is rare, has been identified which
appears to be susceptible to type 1 diabetes (*D11S1917-
H0570POLYA*, 3-3, %T affecteds = 62.4, 2X2 contingency test,
affecteds *vs* non-affecteds;chi^2=6.7, df=1, p<0.009).
Therefore, analysis of association in this region has produced
10 evidence for a haplotype which contains an allele protective
against type 1 diabetes, as it is significantly less
transmitted to the affected offspring in comparison to the
unaffected offspring, and evidence for two non-protective
haplotypes, which have a neutral or susceptible effect on type
15 1 diabetes.

Extending this haplotype analysis to include the 14
flanking microsatellite markers *255ca5*, *D11S987*, *255ca6*,
255ca3, *D11S1296*, *E0864CA*, *TAA*, *L3001CA*, *D11S1337*, *14LCA5*,
D11S4178, *D11S970*, *14LCA1*, *18O18*, as well as the single
20 nucleotide polymorphisms (SNPs) *58-1*, *Exon E* (intronic, 8bp 3'
of exon 6) and *Exon R* (Ala ^{1330}, exon 18) (Figure 19), revealed
highly conserved haplotypes within this interval in the
diabetic individuals. A distinct protective haplotype (A) has
been identified (encompassing the 3-2 haplotype at *D11S1917-
25 H0570POLYA*), as well as a distinct susceptible haplotype (B)
(encompassing the 2-3 haplotype at *D11S1917-H0570POLYA*). The
susceptible haplotype is IBD with the protective haplotype, 3'
of marker *D11S1337*, indicating that the aetiological variant
playing a role in type 1 diabetes does not lie within the
30 identical region, localising it 5' of *Exon E* of the *LRP-5*
gene.   This region that is IBD between the protective, and
susceptible haplotypes prevents association analysis being
undertaken, as no deviation in transmission to affected
offspring would be detected.   The rare susceptible haplotype
35 (C), 3-3 at *D11S1917-H0570POLYA*, can also be identified.
Haplotype analysis with the additional markers in the region
reveals that this rare susceptible haplotype is identical to
the susceptible haplotype between *UT5620* and *14L15CA*,
potentially localising the aetiological variant between *UT5620*

84

and *Exon E*, which is approximately 100kb. Therefore, the
susceptible and rare susceptible haplotypes may carry an
allele (or separate alleles) which confers a susceptible
effect on type 1 diabetes, whereas the protective haplotype
5 contains an allele protective against *IDDM*. The 5' region of
the *LRP5* gene lies within this interval, encompassing the 5'
regulatory regions of the *LRP5* gene and exons 1 to 6.

Analysis of the Italian and Sardinian haplotypes revealed
an additional two susceptible haplotypes. At *D11S1917-*
10 *H0570POLYA* in the Italian families haplotype 1-3, 63%T, 2X2
affected verses non-affecteds p=0.03 (haplotype D). At
H0570POLYA -L3001 in the Sardinian families haplotype 1-2
58%T, 2X2 affected verses non-affecteds, p=0.05 (haplotype E).

15   Samples containing the above five haplotypes were
genotyped with SNPs from the *IDDM4* region in order to
investigate regions of IBD (Figure B). These SNPs confirmed
the region of IBD between the susceptible haplotypes B and C
between *UT5620* and *14L15CA*. It also confirmed the region of
20 IBD between the protective and susceptible haplotypes A and B
3' of marker *D11S1337*, excluding this region from containing
the aetiological variant. The SNP analysis also revealed a
potential region of IBD between *UT5620* and *TAA*, between the
susceptible haplotypes B, C, D and E, which is distinct from
25 the protective haplotype A (a 25kb region). The marker
H0570POLYA lies within this interval, and is not identical in
haplotype E compared to the other susceptible haplotypes;
possibly this is due to mutation at this polymorphism, or it
delineates a boundary within this region and the aetiological
30 variant is either 5' or 3' of this marker. Further analysis of
additional SNPs within this interval will be necessary.

Therefore haplotype mapping within the *IDDM4* region has
identified a region of IBD between the susceptible haplotypes
B and C of 100kb, in the 5' region of the LRP5 gene. SNP
35 haplotype mapping has possibly further delineated this to a
25kb interval encompassing the 5' region of LRP5 which
includes possible regulatory sequences for this gene; a
putative promoter, and regions of homology with the mouse
syntenic region (Table 12), as well as exon 1 of LRP5.

Construction of Adenovirus vectors containing LRP5

    The full-length human *LRP5* gene was cloned into the
adenovirus transfer vector pdelE1sp1A-CMV-bGHPA containing
5 the human Cytomegalovirus immediate early promoter and the
bovine growth hormone polyadenylation signal to create
pdehlrp3. This vector was used to construct an adenovirus
containing the *LRP5* gene inserted into the E1 region of the
virus directed towards the 5' ITR. In order to accommodate a
10 cDNA of this length, the E3 region has been completely deleted
from the virus as it has been described for pBHG10 (Bett at
al.1994 Proc Natl Acad Sci 91: 8802-8806) An identical
strategy was used to construct an adenoviral vector containing
the full-length mouse *Lrp5* gene.
15   A soluble version of mouse *Lrp5* was constructed in which
a His tag and a translational stop signal replaced the
putative transmembrane spanning domain (primers listed in
Table 9). This should result in the secretion of the
extracellular domain of Lrp5 and facilitate the biochemical
20 characterization of the putative ligand binding domain of
Lrp5. Similarly a soluble version of human LRP5 can be
constructed using primers shown in Table 9. The extracellular
domain runs to amino acid 1385 of the precursor (immature)
protein sequence.
25

Identification of LRP5 ligands

    LRP5 demonstrates the ablility to bind and take up LDL
(see below), but this activity is not a high level.
Therefore, it is likely that LRP5 has the capacity to bind
30 additional ligand(s). To identify LRP5 ligands the
extracellular domain consisting of the first 1399 amino acids
of human LRP5, or the corresponding region of mouse Lrp5 will
be purified. A number of expression systems can be used these
include plasmid based systems in *Drosophila* S2 cells, yeast
35 and *E. coli* and viral based systems in mammalian cells and
SF9 insect cells. A histidine tag will be used to purify LRP5
on a nickel column (Novagen, Madison WI). A variety of resins
may be used in column chromatography to further enrich soluble
LRP5. LRP5 will be attached to a solid support e.g. a nickel

column. Solutions containing ligands from serum fractions,
urine fractions, or fractions from tissue extracts will be
fractionated over the LRP5 column. LRP5 complexed with bound
ligand will be eluted from the nickel column with imidizole.
5 The nature of the ligand(s) bound to LRP5 will be
characterized by gel electrophoresis, amino acid sequence,
amino acid composition, gas chromatography, and mass
spectrophotometer.

Attachment of purified LRP5 to a BiaCore 2000 (BiaCore,
10 Uppsula Sweden) chip will be used to determine whether ligands
that bind to LRP5 are present in test solutions. Once ligands
for LRP5 are identified the LRP5 chip will be used to
characterize the kinetics of the LRP5 ligand interaction.

Adenoviral vectors containing soluble versions of LRP5
15 will be used to infect animals, isolation of ligand/LRP5
complexes from serum or liver extracts will be facilitated by
the use of a histidine tag and antibodies directed against
this portion of LRP5.

20 *Treatment of animals with LRP5 virus*

A wide range of species may be treated with adenovirus
vectors carrying a transgene. Mice are the preferred species
for performing experiments due to the availability of a number
of genetically altered strains of mice, i.e. knockout,
25 transgenic and inbred mice. However larger animals e.g. rats
or rabbits may be used when appropriate. A preferred animal
model to test the ability of *LRP5* to modify the development of
type 1 diabetes is the non-obese diabetic (NOD) mouse.
Preferred animal models for examination of a potential role
30 for *LRP5* in lipoprotein metabolism are mice in which members
of the LDL-receptor family have been disrupted, e.g. the LDL-
receptor (*LDLR*), or in which genes involved in lipoprotein
metabolism, e.g. *Apo-E*, have been disrupted.

Adenoviruses are administered by injecting approximately
35 1 x 10^9 plaque forming units into the tail vein of a mouse.
Based on previous studies this form of treatment results in
the infection of hepatocytes at a relatively high frequency.
Three different adenovirus treatments were prepared, 1.)
adenovirus containing no insert (negative control), 2.)

adenovirus containing human *LDLR* (positive control) or 3.)
adenovirus containing human *LRP5*. Each of these viruses were
used to infect five C57 wild type and five C57 LDLR knockout
mice. A pretreatment bleed, 8 days prior to injection of the
5 virus was used to examine serum chemistry values prior to
treatment. The animals were injected with virus. On day
five following administration of the virus a second (treatment
) bleed was taken and the animals were euthanized for
collection of serum for lipoprotein fractionation. In
10 addition tissues were harvested for in situ analysis, immuno-
histochemistry, and histopathology.

     Throughout the experiment, animals were maintained in a
standard light/dark cycle and given a regular chow diet. The
animals were fasted prior to serum collection. In certain
15 experimental conditions it may be desirable to give animals a
high fat diet.

     Standard clinical serum chemistry assays were performed
to determine; serum triglycerides, total cholesterol, alkaline
phosphatase, aspartate aminotransferase, alanine
20 aminotransferase, urea nitrogen, and creatinine. Hematology
was performed to examine the levels of circulating leukocytes,
neutrophils, the percent lymphocytes, monocytes, and
eosinophils, erythrocytes, platelets, hemoglobin, and percent
hematocrit.

25   Serum lipoproteins were fractionated into size classes
using a Superose 6 FPLC sizing column and minor modifications
of the procedure described by Gerdes et al. (Clin. Chim. Acta
205:1-9 (1992)), the most significant difference from the
Gerdes procedure being that only one column was used. Column
30 fractions were collected and analyzed for cholesterol and
triglyceride. The "area under the curve" was calculated for
each lipoprotein class. The approximate peak fractions that
correspond to each of the classes defined by density are:
fraction 24 for VLDL, fraction 36 for LDL and fraction 51 for
35 HDL.

*LRP5 overexpression affects serum triglycerides and
lipoproteins*
     Statistical analysis of serum chemistry data indicated

that relative to control virus there was a 30% decrease, p
value = 0.025, in triglyceride levels in animals treated
with *LRP5* containing virus (Table 10). This decrease in
triglycerides occurred at a similar level in both wild type
5 and KO mice. By comparison, the *LDLR* virus reduced serum
triglycerides approximately 55%, relative to the contol
virus. This result indicates that *LRP5* has the potential to
modulate serum triglyceride levels.

The serum lipoprotein profile indicated that the VLDL
10 particle class was decreased in wild type mice treated with
LRP5 virus. Although the number of samples analyzed was not
sufficient for statistical analyses, this result is consistent
with the observed decrease in serum triglycerides. These
results suggest that LRP5 has the potential to bind and
15 internalize lipid rich particles, causing the decrease in
serum triglycerides and VLDL particles. Therefore treatment
with *LRP5* or with therapeutic agents that increase the
expression of *LRP5* or the biological activity of LRP5 may be
useful in reducing lipid rich particles and triglycerides in
20 patients with diseases that increase triglyceride levels, e.g.
type 2 diabetes and obesity.

Although not statistically significant there was an
observed trend towards a reduction in serum cholesterol levels
as a consequence of LRP5 treatment (28 %, p = 0.073) in mice
25 that have a high level of serum cholesterol (approximately 220
mg/dL), due to a disruption (knockout) of the LDL-receptor
(Table 10). An opposite trend, in that LRP5 treatment
elevated serum cholesterol (30%, p = 0.08) was not observed in
wild type mice which have a relatively low level of serum
30 cholesterol (approximately 70 mg/dL). The small treatment
groups, n = 4 , in these data sets limits the interpretation
of these results and indicates that further experimentation is
necessary. Nevertheless, these results suggest that in a
state of elevated cholesterol an increase in the activity of
35 LRP5 might reduce serum cholesterol levels. Therefore
treatment with *LRP5* or with therapeutic agents that increase
either the expression of *LRP5* or the biological activity of
LRP5 may be useful in reducing cholesterol in patients with
hypercholesterolemia.

*LRP5 overexpression may affect serum alkaline phosphatase
levels*

Serum alkaline phosphatase levels can be dramatically
elevated, e.g. 20 fold increase, as a consequence of an
5 obstruction of the bile duct (Jaffe, M. S. and McVan, B.,
1997, <u>Davis's laboratory and diagnostic test handbook.</u> pub.
F.A. Davis Philadelphia PA). However, lower levels, up to a
three fold increase of alkaline phosphatase can result from
the inflammatory response that take place in response to an
10 infectious agent in the liver, e.g. adenovirus.   In animals
treated with a control virus there was an approximately 2-fold
increase in alkaline phosphatase levels.   In contrast, there
was only a slight increase in alkaline phophatase levels in
animals treated with the *LRP5* virus.   Relative to the control
15 the alkaline phosphatase level was reduced 49% in the *LRP5*
treated animals, p value = 0.001 (Table 10).

The increase in alkaline phosphatase levels may be a
consequence of the level of infection with the adenovirus,
therefore, a possible explanation for the decrease in the
20 animals treated with the *LRP5* virus may simply be due to less
virus in this treatment group.   An indicator of the level of
the viral infection is the appearance in the serum of the
liver enzymes aspartate aminotransferase and alanine
aminotransferase.   These enzymes are normally found in the
25 cytoplasm of cells and elevated in the serum when cellular
damage occurs (Jaffe, M. S. and McVan, B., 1997, <u>Davis's
laboratory and diagnostic test handbook.</u> pub. F.A. Davis
Philadelphia PA). Therefore these enzymes serve as markers for
the level of toxicity that is a consequence of the adenoviral
30 infection.   These enzymes are present at a normally low level
prior to the infection and in animals that did not receive
virus.   Importantly, the levels of aspartate aminotransferase
and alanine aminotransferase are higher in the animals given
the *LRP5* virus indicating that these animals have more
35 cellular damage and thus a more extensive infection than the
animals given the control virus (Table 11).   Therefore, it is
unlikely that the reduced level of alkaline phosphatase is
simply owing to less *LRP5* virus being administered.   A second
possible explanation is that LRP5 modifies the nature of the

inflammatory response resulting from the adenovirus infection.
A possible role for *LRP5* in modulating the inflammatory
response is' consistent with the genetic data indicating that
this gene is associated with risk for developing type 1
5 diabetes. Chronic insulitis or inflammation is a precursor to
clinical onset of type 1 diabetes therefore *LRP5* treatment or
treatment with therapeutic agents that either increase the
transcription of *LRP5* may be of utility in preventing type 1
diabetes.   Type 1 diabetes is an autoimmune disease,
10 therefore treatment with *LRP5* or with therapeutics agents that
either increase the expression of *LRP5* or the biological
activity of LRP5 may be useful in treating other autoimmune
diseases.

15 *Expression of* LRP5 *in cell lines*
         Overexpression of *LRP5* under the control of a
heterologous promoter can be accomplished either by infection
with an adenovirus containing *LRP5* or by transfection with a
plasmid vector containing *LRP5*. Transfection with a plasmid
20 vector can lead to either transient or a stable expression of
the transgene.
         Endogenous LDL-receptors reduce the ability to detect the
uptake of LDL by other members of the LDL-receptor family. To
study lipoprotein uptake in the absence of the LDL-receptor,
25 primary cell lines from human patients with familial
hypercholesterolemia (FH) were used. These FH cell lines lack
any endogenous LDL-receptor.   FH fibroblasts were infected at
an MOI of. 500 plaque forming units per cell for 24 hours at
37°C. Following infection, cells were incubated with 40 μg/ml
30 ^{125}I-LDL at 37°C. After 4 hours, cells were washed and uptake
of LDL measured. A modest (approximately 60%) increase in the
level of LDL uptake was observed. By comparison, the
infection of FH cells with an adenovirus containing the LDL-
receptor resulted in a 20-fold increase in LDL uptake (p <
35 0.0001, n = 3).   To determine whether this modest level of
activity mediated by LRP5 was statistically significant, 24
individual wells were infected with *LRP5* virus and analyzed.
Statistical analysis of this experiment indicated that the
increase in LDL uptake was highly signficant, p < 0.0001.

Therefore LRP5 can mediate LDL uptake. However, based on the
modest level of activity, relative to the LDL-receptor, it
does not appear that the primary activity of LRP5 is to
mediate the uptake of LDL.

5   Additional cell lines exist that lack either the LDL-
receptor or other members of the LDL-receptor family. The
PEA-13 cell line (ATCC 2216-CRL) lacks the LRP1 receptor.
Mutant CHO cells lacking the LDL receptor have been described
by Kingsley and Krieger (Proceedings National Academy
10 Sciences USA (1984) 81:5454). This cell line, known as *ldlA7*,
is particularly useful for the creation of stable transfectant
cell lines expressing recombinant LRP5.

Anti-LRP5 Antibodies
15 Western Blot Analysis
         Antisera prepared in rabbits immunized with the human
    LRP5 MAP peptides
         SYFHLFPPPPSPCTDSS
         VDGRQNIKRAKDDGT
20   EVLFTTGLIRPVALVVDN
         IQGHLDFVMDILVFHS
    were evaluated by Western blot analysis.
         COS cells were infected with an adenovirus containing
    human *LRP5* cDNA. Three days after the infection the cells
25 were harvested by scraping into phosphate buffered saline
    (Gibco/BRL Gaithersburg, MD) containing the protease
    inhibitors PMSF (100ug/ml), aprotinin (2 ug /ml), and
    pepstatin A (1 ug/ml). The cells were pelleted by a low speed
    spin, resuspended in phosphate buffered saline containing
30 protease inhibitors and lysed by Dounce homogenization.
    Nuclei were removed with a low speed spin, 1000 rpm for 5 min
    in a Beckman J-9 rotor. The supernatant was collected and
    centrifuged at high speed, 100,000 x g for 3 hours, to pellet
    the membranes. Membranes were resuspended in SDS-sample
35 buffer (Novex, San Diego CA).
         Membrane proteins were fractionated by electrophoresis on
    a 10% Tris-glycine acrylamide gel (Novex, San Diego CA). The
    fractionated proteins were transferred to PVDF paper (Novex,
    San Diego CA) according to the manufacturer's instructions.

Standard Western blot analysis was performed on the membrane
with the primary antibody being a 1:200 dilution of crude
antisera and the secondary antibody a 1:3000 dilution of
antirabbit IgG HRP conjugate (Amersham, Arlington Heights,
5 IL).   ECL reagents (Amersham, Arlington Heights, IL) were used
to visualize proteins recognized by the antibodies present in
the sera.

A band of approximately 170-180 kD was detected by sera
from a rabbit immunized with the peptide SYFHLFPPPPSPCTDSS.
10 This band was only detected in the cells that were infected
with the adenovirus containing human LRP5 and was not present
in cells that were infected with a control virus.
Furthermore, the detection of this 170 kD band was blocked by
preadsorbing a 1:500 dilution of the sera with 0.1 ug/ml of
15 the peptide SYFHLFPPPPSPCTDSS but not with 0.1 ug/ml of the
peptide VDGRQNIKRAKDDGT.   Therefore this protein band of
approximately 170 kD detected by the antibody directed against
the peptide SYFHLFPPPPSPCTDSS is human LRP5.   The predicted
size of the mature human LRP5 protein is 176 kD.

20   The antisera from a rabbit immunized with the peptide
SYFHLFPPPPSPCTDSS was affinity purified with an Affigel 10
column (BioRad, Hercules CA) to which the MAP peptide
SYFHLFPPPPSPCTDSS was covalently attatched.   This results in
antisera with greater specificity for LRP5.

25   The antisera from a rabbit immunized with the peptide
IQGHLDFVMDILVFHS is able to detect a band of approximately 170
kD that is present in cells infected with an LRP5 containing
virus but not cells infected with a control virus.   This
antibody recognizes a peptide that is present in the putative
30 extracellular domain of LRP5 and thus will be useful in
detecting the soluble version of LRP5.   However, there is
greater background observed when using this antisera relative
to that from the rabbit immunized with the peptide
SYFHLFPPPPSPCTDSS.

35

LRP5 is expressed in tissue macrophages
The crude and affinity purified antisera to the LRP5
peptide SYFHLFPPPPSPCTDSS was used for immunocytochemistry
studies in human liver.   The antibody recognized tissue

macrophages, termed Kupfer cells in the liver, that stained
positive for LRP5 and positive for the marker RFD7 (Harlan
Bioproducts, Indianapolis IN) which recognizes mature tissue
phagocytes and negative for an MHC class II marker, RFD1
5 (Harlan Bioproducts, Indianapolis IN). This pattern of
staining (RFD1 - RFD7+) identifies a subpopulation of
macrophages, the effector phagocytes. This class of
macrophages has been implicated in the progression of disease
in a model for autoimmune disease, experimental autoimmune
10 neuritis (Jung. S. et al., 1993, J Neurol Sci 119: 195-202).
The expression in phagocytic tissue macrophages supports a
role for LRP3 in modulating the inflammatory component of the
immune response. This result is consistent with the proposed
role based on the differences observed in alkaline phoshatase
15 levels in animals treated with LRP5 virus and the genetic data
indicating that *LRP5* is a diabetes risk gene.

Determination of additional conserved regions of the LRP5 *gene*
        High throughput DNA sequencing of shotgun libraries
20 prepared from mouse BAC clones 131-p-15 and 53-d-8 was used to
identify regions of the *LRP5* gene that are conserved between
mouse and man. To identify these regions the mouse genomic
DNA, either unassembled sequences or assembled contigs, was
compared against an assembly of human genomic DNA. The
25 comparison was done by using the BLAST algorithm with a
cutoff of 80%. This analysis resulted in the identification
of a majority of the exons of the *LRP5* gene and identified a
number of patches of conserved sequences at other locations in
the gene (Table 12).
30   There are sequences conserved between human and mouse
located 4.3 kb and 168 bp upstream of the putative ATG. These
sequences may represent 5' untranslated sequences of the mRNA
transcript or promoter elements.
        Within the putative first intron of 36 kb there are
35 twelve patches that exhibit a degree of DNA sequence
conservation. Some of these regions, e.g. 41707-41903, are
quite extensive and have a high degreee of sequence
conservation, similar to that observed for the exons of the
LRP5 gene. Since these regions do not appear to be

transcribed it is likely that these conserved regions play a role in regulating either the transcription of the *LRP5* gene or the processing of the *LRP5* mRNA transcript. Regardless of exact nature of their role these newly identified regions

5 represent areas where sequence polymorphism may affect the biological activity of LRP5.

The BAC clone 131-p-15 which contains the first two exons of *LRP5* was sequenced extensively, i.e. approximately 6X coverage. BAC clone 53-d-8 contains sequences from exon D to

10 exon V, however the level of sequence coverage of this clone was only approximately 1X (skim sequencing). The skim sequencing of mouse BAC 53-d-8 resulted in 76% of the exons being detected, however in some instances only a portion of an exon was present in the mouse sequence data. In addition

15 to the exons, there were three patches in the BAC 53-d-8 sequences that exhibited a degree of sequence conservation with the human sequences (Table 12). All of these were located in the large 20 kb intron between exons D and E. These sequences may represent regions that are important for

20 the processing of this large intron and thus polymorphisms in these sequences may affect the expression level of *LRP5*.

Determination of relative abundance of alternatively spliced
LRP5 mRNA transcripts

25   Several techniques may be used to determine the relative abundance of the different alternatively spliced isoforms of LRP5.

Northern blot analysis of probes derived from specific transcripts is used to survey tissues for the abundance of a

30 particular transcript. More sensitive techniques such as RNase protection assays will be performed. Reagents from commercially available kits (Ambion, Inc. Austin TX) are used to prepare probes. The relative abundance of transcript that hybridizes to a probe radiolabeled with [alpha]32P-UTP is

35 analyzed by native and denaturing acrylamide gels (Novex Inc., San Diego, CA). Primer extension assays are performed according to established procedures (Sambrook *et. al.* (1989) Molecular Cloning, Cold Spring Harbour Press, NY) using reverse primers derived from the 5' portion of the transcript.

95

Isolation of other species homologs of <u>LRP5</u> gene

The *LRP5* gene from different species, e.g. rat, dog, are isolated by screening of a cDNA library with portions of the gene that have been obtained from cDNA of the species of
5 interest using PCR primers designed from the human *LRP5* sequence. Degenerate PCR is performed by designing primers of 17-20 nucleotides with 32-128 fold degeneracy by selecting regions that code for amino acids that have low codon degeneracy e.g. Met and Trp. When selecting these primers
10 preference is given to regions that are conserved in the protein e.g. the motifs shown in Figure 6b. PCR products are analyzed by DNA sequence analysis to confirm their similarity to human *LRP5*. The correct product is used to screen cDNA libraries by colony or plaque hybridization at high
15 stringency. Alternatively probes derived directly from the human *LRP5* gene are utilized to isolate the cDNA sequence of *LRP5* from different species by hybridization at reduced stringency. A cDNA library is generated as described above.

20 *REFERENCES*

1.   Bach, J.-F (1994). Endocrine. Rev. 15: 516-542.

2.   Bain, S., et al. (1992). Diabetes 41: 91A.

3.   Bell, G.I., et al. (1984). Diabetes 33: 176-83.

4.   Bennett, S.T., et al. (1995). Nature Genet. 9: 284-292.

25 5.   Bennett, S.T. and Todd, J.A (1996). Annu. Rev. Genet.30: 343-370.

6.   Buckler, A. et al. (1991). P.N.A.S USA 88: 4005-4009.

7.   Davies, J.L., et al. (1994). Nature 371: 130-136.

8.   Doria, A., et al (1996). Diabetologia 39: 594-599.

30 9.   Hashimoto, L., et al. (1994). Nature 371: 161-164.

10.   Holmans, P. (1993). Am. J. Hum. Genet. 52: 362-374.

11.   Julier, C., et al. (1991a). Nature 354: 155-159.

12.   Kennedy, G.C., et al. (1995). Nature Genet. 9: 293-298.

13.   Kyvik, K.O., et al. (1995). Brit. Med. J. 311: 913-917.

35 14.   Lucassen, A., et al. (1993). Nature Genet. 4: 305-310.

15.   Lucassen, A., et al. (1995). Hum. Mol. Genet. 4: 501-506.

16.   Luo, D.-F., et al. (1996). Hum. Mol. Genet. 5: 693-698.

17.   Matsuda, A. and Kuzuya, T. (1994). Diab. Res. Clin. Pract. 24: Suppl., S63-S67.

18. Risch (1987). Am. J. Hum. Genet. 40: 1-14.

19. Owerbach, D., et al. (1990). Diabetes 39: 1504-1509.

20. Parimoo, S., et al. (1991). P.N.A.S. USA 88: 9623-9627.

21. Penrose, L.S. (1953). Acta. Genet. Stat. Med. 4: 257-265.

5 22. Risch, S.S. (1990). Diabetes 39: 1315-19.

23. Spielman, R., et al. (1993). Am. J. Hum. Genet. 52: 506-516.

24. Thomson, G., et al. (1989). Genet. Epidemiol. 6: 155-160.

25. Tisch, R. and McDevitt, H.O. (1996). Cell 85: 291-297.

10 26. Todd, J.A. (1994). Diabetic Med. 11: 6-16.

27. Todd, J.A., et al. (1987). Nature 329: 599-604.

28. Todd, J.A. and Farrall, M. (1996). Hum. Mol. Genet. 5: 1443-1448.

29. Todd, J.A., et al. (1989). Nature 338: 587-589.

15 30. Vafiadis, P., et al. (1996). J. Autoimmunity 9: 397-403.

TABLE 1

Haplotype analysis at D11S1917(UT5620) - H0570POLYA, within 2582 families from UK, USA, Norway and Sardinia. 20 Susceptible, protective and neutral alleles were identified at each polymorphism, and transmission of recombinant haplotypes to diabetic offspring was calculated (t=transmission, nt=non transmission). Significant transmission of the haplotype 332-104 was detected (P=0.005), as well as significant non- 25 transmission of the haplotype 328-103 (P=0.03).

	D11S1917(UT5620)	*H0570POLYA*	t	nt	P
	328	104	539	474	
Protective	332	103	427	521	0.002
30 Susceptible	332	104	60	33	0.005
Protective	328	103	16	31	0.03

TABLE 2   PCR Primers for obtaining LRP5 cDNA

Primers located within *LRP5* cDNA:
The primers are numbered beginning at nucleotide 1 in Figure
5 5(a).

1F (muex 1f) : ATGGAGCCCGAGTGAGC

218R (27R) : ATGGTGGACTCCAGCTTGAC
10
256F (1F) : TTCCAGTTTTCCAAGGGAG

265R (26R) : AAAACTGGAAGTCCACTGCG

15 318R (4R) : GGTCTGCTTGATGGCCTC

343F (2F) : GTGCAGAACGTGGTCATCT

Vector Primers for RCCA
20
361R (21R) : GTGCAGAACGTGGTCATCT

622R (2R) : AGTCCACAATGATCTTCCGG

25 638F (4F) : CCAATGGACTGACCATCGAC

657R (1R) : GTCGATGGTCAGTCCATTGG

956R (22R) : TTGTCCTCCTCACAGCGAG
30
1713F (21F) : GGACTTCATCTACTGGACTG

1481R (23R) : CAGTCTGTCCAGTACATGAG

35 1981F (22F) : GCCTTCTTGGTCTTCACCAG

2261F (23F) :GGACCAACAGAATCGAAGTG

2484R (5R) : GTCAATGGTGAGGTCGT

2519F (5F) : ACACCAACATGATCGAGTCG

3011F (24F) : ACAAGTTCATCTACTGGGTG

5 3154F (25F) : CGGACACTGTTCTGGACGTG

3173R (25R) : CACGTCCAGAACAGTGTCCG

3556R (3R) : TCCAGTAGAGATGCTTGCCA
10
Vector Primers for RCCA

3577F (3F) : ATCGAGCGTGTGGAGAAGAC

15 4094F (30F) : TCCTCATCAAACAGCAGTGC

4173R (6R) : CGGCTTGGTGATTTCACAC

4687F (6F) : GTGTGTGACAGCGACTACAGC
20
4707R (30R) : GCTGTAGTCGCTGTCACACAC

5061R (7R) : GTACAAAGTTCTCCCAGCCC

25 PCR primers in Sequences identified by GRAIL

G1 1F: TCTTCTCCAGAGGATGCAGC

G1 2F: TTCGTCTTGAACTTCCCAGC
30
G1 3F: TCTTCTTCTCCAGAGGATGCA

Gp1 1F: AGGCTGGTCTCAAACTCCTG

35 PBS.543R: GGGGATGTGCTGCAAGGCGA

PBS.578R: CCAGGGTTTTCCCAGTCACGAC

PBS.838F: TTGTGTGGAATTGTGAGCGGATAAC

PBS.873F: CCCAGGCTTTACACTTTATGCTTCC

100

Table 3   Intron-Exon Organization of Human LRP5

3' Acceptor Sequence Intron / Exon	Exon Number	Exon Size (bp)	5' Donor Sequence Exon / Intron	Intron Number & Size (bp)
ccgggtcaac/ATGGAG	Ex 1 (6)	(91)	CCGCGG/gtaggtgggc	1 (35051)
tgccccacag/CCTCGC	Ex 2 (A)	(391)	TCACGG/gtaaaccctg	2 (9408)
cccgtcacag/GTACAT	Ex 3 (B)	(198)	GTTCCG/gtaggtaccc	3 (6980)
ctgactgcag/GCAGAA	Ex 4 (C)	(197)	CTTTCT/gtgagtgccg	4 (1640)
gttttcccag/TCCACA	Ex 5 (D)	(132)	AGGCAG/gtgaggcggt	5 (20823)
gtctccacag/GAGCCG	Ex 6 (E)	(397)	GATGGG/gtaagacggg	6 (3213)
tcttctccag/CCTCAT	Ex 7 (F)	(172)	ATCGAG/gtgaggctcc	7 (13445)
cgtcctgcag/GTGATC	Ex 8 (G)	(217)	TCGTCG/gtgagtccgg	8 (2826)
tcgcttccag/GAACCA	Ex 9 (H)	(290)	CTGAAG/gtagcgtggg	9 (5000+)
ctgctgccag/ACCATC	Ex 10 (I)	(227)	CAAGGG/gtaagtgttt	10 (1295)
tgccttccag/CTACAT	Ex 11 (J)	(185)	TGCTGG/gtgagggccg	11 (2068)
gttcatgcag/GTCAGG	Ex 12 (K)	(324)	GCAGCC/gtaagtgcct	12 (2005)
cctcctctag/CGCCCA	Ex 13 (L)	(200)	ACCCAG/gcaggtgccc	13 (6963)
tgtcttacag/CCCTTT	Ex 14 (M)	(209)	GCGAGG/gtaggaggcc	14 (1405)
cctcccgcag/GTACCT	Ex 15 (N)	(191)	TGTCAG/gtaaggggcc	15 (686)
ctgcttgcag/GGGCCA	Ex 16 (O)	(210)	AGTTCT/gtacgtgggg	16 (3894)
gtctttgcag/CAGCCC	Ex 17 (P)	(126)	GTGGAG/gtaggtgtga	17 (3903)
cctcccccag/AGCCGC	Ex 18 (Q)	(237)	GTGACG/gtgaggccct	18 (3042)
tcccttgcag/CCATCT	Ex 19 (R)	(111)	TGTGTG/gtgagccagc	19 (1448)
tctctggcag/AAATCA	Ex 20 (S)	(237)	TCACAG/gtaaggagcc	20 (1095)
tccctgccag/GCATCG	Ex 21 (T)	(140)	CCGCCG/gtgaggggcg	21 (6514)
ctctcctcag/ATCCTG	Ex 22 (U)	(98)	GTACAG/gtaggacatc	22 (2275)
tcccttcag/GCCCTA	Ex 23 (V)	(>262)		23 (19985)

101

Table 4   page 1 of 7

LRP-5 Exon primers

E1x1 1f   CAGGGTTTCATCCTTTGTGG
E1x1 1fU   TGTAAAACGACGGCCAGTCAGGGTTTCATCCTTTGTGG
E1x1 1fR   GCTATGACCATGATTACGCCCAGGGTTTCATCCTTTGTGG
E1x1 1r   TGACGGGAAGAGTTCCTCAG
E1x1 1rR   GCTATGACCATGATTACGCCTGACGGGAAGAGTTCCTCAG

E1x5 1f   TCTGCTCTTCCTGAACTGCC
E1x5 1fU   TGTAAAACGACGGCCAGTTCTGCTCTTCCTGAACTGCC
E1x5 1r   TTGAGTCCTTCAACAAGCCC
E1x5 1rR   GCTATGACCATGATTACGCCTTGAGTCCTTCAACAAGCCC

E1x6 1fU   TGTAAAACGACGGCCAGTTTCCCCACTCATAGAGGCTC
E1x6 1rR   GCTATGACCATGATTACGCCGCTCCCAACTCGCCAAGT

E1x6a 1fU   TGTAAAACGACGGCCAGTGGTCAACATGGAGGCAGC ·
E1x6a 1rR   GCTATGACCATGATTACGCCCAGGTGTCAGTCCGCTTG

E1x6b 1fU   TGTAAAACGACGGCCAGTGCAGAGAAGTTCTGAGC
E1x6b 1rR   GCTATGACCATGATTACGCCCACTTGGCCAGCCATACTC

E1x6c 1fU   TGTAAAACGACGGCCAGTCAAGCAAGCCTCTTGCTACC
E1x6c 1rR   GCTATGACCATGATTACGCCACTGCAATGAGGTGAAAGGC

E1x6d 1fU   TGTAAAACGACGGCCAGTCAGGTGAGAACAAGTGTCCG
E1x6d 1rR   GCTATGACCATGATTACGCCGCTGCCTCCATGTTGACC

E1x6e 1fU   TGTAAAACGACGGCCAGTTGTGCCTGGGTGAGATTCT
E1x6e 1rR   GCTATGACCATGATTACGCCTGTGGAGCCTCTATGAGTGG

E1x6f 1fU   TGTAAAACGACGGCCAGTGGGTGACAGGTGGCAGTAG
E1x6f 1rR   GCTATGACCATGATTACGCCGGAAGGAAGGACACTTGAGC

E1x6g 1fU   TGTAAAACGACGGCCAGTCCTGGTGTGTTTGAGAACCC
E1x6g 1rR   GCTATGACCATGATTACGCCCAATGGGAAGCCAGGCTAG

E1xA 1f   ATCTTGCTGGCTTAGCCAGT
E1xA 1fU   TGTAAAACGACGGCCAGTATCTTGCTGGCTTAGCCAGT
E1xA 1fR   GCTATGACCATGATTACGCCATCTTGCTGGCTTAGCCAGT
E1xA 1r   GCTCATGCAAATTCGAGAGAG
E1xA 1rR   GCTATGACCATGATTACGCCGCTCATGCAAATTCGAGAGAG

E1xB 1f   CCTGTTGGTTATTTCCGATGG
E1xB 1fU   TGTAAAACGACGGCCAGTCCTGTTGGTTATTTCCGATGG
E1xB 1fR   GCTATGACCATGATTACGCCCCTGTTGGTTATTTCCGATGG
E1xB 1r   CCTGAGTTAAGAAGGAACGCC
E1xB 1rR   GCTATGACCATGATTACGCCCCTGAGTTAAGAAGGAACGCC

E1xC 1f   AATTGGGTCAGCAGCAATG

Table 4   page 2 of 7   con't.

E1xC	1fR	GCTATGACCATGATTACGCC AATTGGGTCAGCAGCAATG
E1xC	2f	AATTGGGTCAGCAGCAATG
E1xC	2fU	TGTAAAACGACGGCCAGTAATTGGGTCAGCAGCAATG
E1xC	2fR	GCTATGACCATGATTACGCC AATTGGGTCAGCAGCAATG
E1xC	1r	TTGGATCGCTAGAGATTGGG
E1xC	1rR	GCTATGACCATGATTACGCCTTGGATCGCTAGAGATTGGG
E1xC	2r	GCACCCTAATTGGCACTCA
E1xC	2rR	GCTATGACCATGATTACGCCGCACCCTAATTGGCACTCA
E1xD	1f	TGACGGTCCTCTTCTGGAAC
E1xD	1fR	GCTATGACCATGATTACGCCTGACGGTCCTCTTCTGGAAC
E1xD	2f	CGAGGCAGGATGTGACTCAT
E1xD	2fU	TGTAAAACGACGGCCAGTCGAGGCAGGATGTGACTCAT
E1xD	2fR	GCTATGACCATGATTACGCC CGAGGCAGGATGTGACTCAT
E1xD	1r	AGTGGATCATTTCGAACGG
E1xD	1rR	GCTATGACCATGATTACGCCAGTGGATCATTTCGAACGG
E1xD	2r	CCAACTCAGCTTCCCGAGTA
E1xD	2rR	GCTATGACCATGATTACGCCCCAACTCAGCTTCCCGAGTA
E1xE	1f	TGGCTGAGTATTTCCCTTGC
E1xE	1fU	TGTAAAACGACGGCCAGTTGGCTGAGTATTTCCCTTGC
E1xE	1fR	GCTATGACCATGATTACGCCTGGCTGAGTATTTCCCTTGC
E1xE	1r	TTTAACAAGCCCTCCTCCG
E1xE	1rR	GCTATGACCATGATTACGCCTTTAACAAGCCCTCCTCCG
E1xF	1f	CAACGCCAGCATCTACTGA
E1xF	1fU	TGTAAAACGACGGCCAGTCAACGCCAGCATCTACTGA
E1xF	1fR	GCTATGACCATGATTACGCCCAACGCCAGCATCTACTGA
E1xF	1r	CAAATAGCAGAGCACAGGCA
E1xF	1rR	GCTATGACCATGATTACGCCCAAATAGCAGAGCACAGGCA
E1xG	1f	TGAAGTTGCTGCTCTTGGG
E1xG	1fU	TGTAAAACGACGGCCAGTTGAAGTTGCTGCTCTTGGG
E1xG	1fR	GCTATGACCATGATTACGCCTGAAGTTGCTGCTCTTGGG
E1xG	1r	CACTTCCTCCTCATGCAAGTC
E1xG	1rR	GCTATGACCATGATTACGCCCACTTCCTCCTCATGCAAGTC
E1xH	1f	AGACTGGAGCCTCTGTGTTCG
E1xH	1fU	TGTAAAACGACGGCCAGTAGACTGGAGCCTCTGTGTTCG
E1xH	1fR	GCTATGACCATGATTACGCCAGACTGGAGCCTCTGTGTTCG
E1xH	1r	TGTGTGTCTACCGGACTTGC
E1xH	1rR	GCTATGACCATGATTACGCCTGTGTGTCTACCGGACTTGC
E1xH	2r	GAACAGAGGCAAGGTTTTCCC
E1xH	2rR	GCTATGACCATGATTACGCCGAACAGAGGCAAGGTTTTCCC
E1xI	1f	AGAATCGCTTGAACCCAGG
E1xI	1fR	GCTATGACCATGATTACGCCAGAATCGCTTGAACCCAGG
E1xI	2f	GCTGGTTCCTAAAATGTGGC
E1xI	2fU	TGTAAAACGACGGCCAGTGCTGGTTCCTAAAATGTGGC
E1xI	2fR	GCTATGACCATGATTACGCCGCTGGTTCCTAAAATGTGGC

Table 4 page 3 of 7 con't.

E1xI 1r   CATACGAGGTGAACACAAGGAC
E1xI 1rR   GCTATGACCATGATTACGCCCATACGAGGTGAACACAAGGAC

E1xJ 1f   TGAAGAGGTGGGGACAGTTG
E1xJ 1fR   GCTATGACCATGATTACGCCTGAAGAGGTGGGGACAGTTG
E1xJ 2f   CTTGTGCCTTCCAGCTACATC
E1xJ 2fU   TGTAAAACGACGGCCAGTCTTGTGCCTTCCAGCTACATC
E1xJ 2fR   GCTATGACCATGATTACGCCCTTGTGCCTTCCAGCTACATC
E1xJ 1r   AGTCCTGGCACAGGGATTAG
E1xJ 1rR   GCTATGACCATGATTACGCCAGTCCTGGCACAGGGATTAG
E1xJ 2r   ATAACTGCAGCAAAGGCACC
E1xJ 2rR   GCTATGACCATGATTACGCCATAACTGCAGCAAAGGCACC

E1xK 1f   GCTTCAGTGGATCTTGCTGG
E1xK 1fU   TGTAAAACGACGGCCAGTGCTTCAGTGGATCTTGCTGG
E1xK 1fR   GCTATGACCATGATTACGCCGCTTCAGTGGATCTTGCTGG
E1xK 1r   TGTGCAGTGCACAACCTACC
E1xK 1rR   GCTATGACCATGATTACGCCTGTGCAGTGCACAACCTACC

E1xL 1f   GTTGTCGAGTGGCGTGCTAT
E1xL 1fU   TGTAAAACGACGGCCAGTGTTGTCGAGTGGCGTGCTAT
E1xL 1fR   GCTATGACCATGATTACGCCGTTGTCGAGTGGCGTGCTAT
E1xL 1r   AAAAGTCCTGTGGGGTCTGA
E1xL 1rR   GCTATGACCATGATTACGCCAAAAGTCCTGTGGGGTCTGA

E1xM 1f   AGAAGTGTGGCCTCTGCTGT
E1xM 1fU   TGTAAAACGACGGCCAGTAGAAGTGTGGCCTCTGCTGT
E1xM 1fR   GCTATGACCATGATTACGCCAGAAGTGTGGCCTCTGCTGT
E1xM 1r   GTGAAAGAGCCTGTGTTTGCT
E1xM 1rR   GCTATGACCATGATTACGCCGTGAAAGAGCCTGTGTTTGCT

E1xN 1f   AGACCCTGCTTCCAAATAAGC
E1xN 1fU   TGTAAAACGACGGCCAGTAGACCCTGCTTCCAAATAAGC
E1xN 1fR   GCTATGACCATGATTACGCCAGACCCTGCTTCCAAATAAGC
E1xN 1r   ACTCATTTTCTGCCTCTGCC
E1xN 1rR   GCTATGACCATGATTACGCCACTCATTTTCTGCCTCTGCC

E1xO 1f   TGGCAGTCCTGTCAACCTCT
E1xO 1fU   TGTAAAACGACGGCCAGTTGGCAGTCCTGTCAACCTCT
E1xO 1fR   GCTATGACCATGATTACGCCTGGCAGTCCTGTCAACCTCT
E1xO 1r   CACACAGGATCTTGCACTGG
E1xO 1rR   GCTATGACCATGATTACGCCCACACAGGATCTTGCACTGG

E1xP 1f   AGGGCCAGTTCTCATGAGTT
E1xP 1fU   TGTAAAACGACGGCCAGTAGGGCCAGTTCTCATGAGTT
E1xP 1fR   GCTATGACCATGATTACGCCAGGGCCAGTTCTCATGAGTT
E1xP 1r   GGGCAAAGGAAGACACAATC
E1xP 1rR   GCTATGACCATGATTACGCCGGGCAAAGGAAGACACAATC

E1xQ 1f   CAACTTCTGCTTTGAAGCCC

104

Table 4   page 4 of 7 con't.

E1xQ 1fU   TGTAAAACGACGGCCAGTCAACTTCTGCTTTGAAGCCC
E1xQ 1fR   GCTATGACCATGATTACGCCCAACTTCTGCTTTGAAGCCC
E1xQ 1r   GACAGACTTGGCAATCTCCC
E1xQ 1rR   GCTATGACCATGATTACGCCGACAGACTTGGCAATCTCCC

E1xR 1f   TCTGCTCTCTGTTTGGAGTCC
E1xR 1fU   TGTAAAACGACGGCCAGTTCTGCTCTCTGTTTGGAGTCC
E1xR 1fR   GCTATGACCATGATTACGCCTCTGCTCTCTGTTTGGAGTCC
E1xR 1r   CCCTAAACTCCACGTTCCTG
E1xR 1rR   GCTATGACCATGATTACGCCCCCTAAACTCCACGTTCCTG

E1xS 1f   GGGTTAATGTTGGCCACATC
E1xS 1fR   GCTATGACCATGATTACGCCGGGTTAATGTTGGCCACATC
E1xS 2f   TTGGCAGGGATGTGTTGAG
E1xS 2fU   TGTAAAACGACGGCCAGTTTGGCAGGGATGTGTTGAG
E1xS 2fR   GCTATGACCATGATTACGCCTTGGCAGGGATGTGTTGAG
E1xS 1r   GTCTGCCACATGTGCAAGAG
E1xS 1rR   GCTATGACCATGATTACGCCGTCTGCCACATGTGCAAGAG

E1xT 1f   TGGTCTGAGTCTCGTGGGTA
E1xT 1fU   TGTAAAACGACGGCCAGTTGGTCTGAGTCTCGTGGGTA
E1xT 1fR   GCTATGACCATGATTACGCCTGGTCTGAGTCTCGTGGGTA
E1xT 1r   GAGGTGGATTTGGGTGAGATT
E1xT 1rR   GCTATGACCATGATTACGCCGAGGTGGATTTGGGTGAGATT

E1xU 1f   AGCCCTCTCTGCAAGGAAAG
E1xU 1fU   TGTAAAACGACGGCCAGTAGCCCTCTCTGCAAGGAAAG
E1xU 1fR   GCTATGACCATGATTACGCCAGCCCTCTCTGCAAGGAAAG
E1xU 1r   CAGAACGTGGAGTTCTGCTG
E1xU 1rR   GCTATGACCATGATTACGCCCAGAACGTGGAGTTCTGCTG

E1xV 1f   TACCGAATCCCACTCCTCTG
E1xV 1fU   TGTAAAACGACGGCCAGTTACCGAATCCCACTCCTCTG
E1xV 1fR   GCTATGACCATGATTACGCCTACCGAATCCCACTCCTCTG
E1xV 2f   CATGGTAGAGGTGGGACCAT
E1xV 2fU   TGTAAAACGACGGCCAGTCATGGTAGAGGTGGGACCAT
E1xV 2fR   GCTATGACCATGATTACGCCCATGGTAGAGGTGGGACCAT
E1xV 1r   GATATCCACCTCTGCCCAAG
E1xV 1rR   GCTATGACCATGATTACGCCGATATCCACCTCTGCCCAAG
E1xV 2r   TTACAGGGGCACAGAGAAGC
E1xV 2rR   GCTATGACCATGATTACGCCTTACAGGGGCACAGAGAAGC

Table 4 page 5 of 7 con't.

SNP primers

```
57-1 1f    GCAACAGAGCAAGACCCTGT
57-1 1fR   GCTATGACCATGATTACGCCGCAACAGAGCAAGACCCTGT
57-1 1r    AAATTAGCCAGGCATGGTG
57-1 1rR   GCTATGACCATGATTACGCCAAATTAGCCAGGCATGGTG
57-1 1fU   TGTAAAACGACGGCCAGTGCAACAGAGCAAGACCCTGT


57-2 1f    CCTGCAGAAGGAAACCTGAC
57-2 1fR   GCTATGACCATGATTACGCCCCTGCAGAAGGAAACCTGAC
57-2 1r    CTGCATCTTTGCCACCATG
57-2 1rR   GCTATGACCATGATTACGCCCTGCATCTTTGCCACCATG
57-2 1fU   TGTAAAACGACGGCCAGTCCTGCAGAAGGAAACCTGAC


57-3 1f    TTCCCAGGAGGCAAGTTATG
57-3 1fR   GCTATGACCATGATTACGCCTTCCCAGGAGGCAAGTTATG
57-3 1r    TGGGCTTAGGTGATCCTCAC
57-3 1rR   GCTATGACCATGATTACGCCTGGGCTTAGGTGATCCTCAC
57-3 1fU   TGTAAAACGACGGCCAGTTTCCCAGGAGGCAAGTTATG


57-4 1f    ACCAAGCCCAACTAATCAGC
57-4 1fR   GCTATGACCATGATTACGCCACCAAGCCCAACTAATCAGC
57-4 1r    ATGCCTGTAATCCCAGCACT
57-4 1rR   GCTATGACCATGATTACGCCATGCCTGTAATCCCAGCACT
57-4 1fU   TGTAAAACGACGGCCAGTACCAAGCCCAACTAATCAGC


57-5 1f    ACTGCAAGCCCTCTCTGAAC
57-5 1r    CGAAGACTGCGAAACAGACA


58-1 1f    CTAGTGCCGTGCAGAATGAG
58-1 1r    GGCCACTGCAATGAGATACA


58-2 1f    GAGAAACAGTTCCAGGGTGG
58-2 1fR   GCTATGACCATGATTACGCCGAGAAACAGTTCCAGGGTGG
58-2 1r    AAACTGAGGCTGGGAGAGGT
58-2 1rR   GCTATGACCATGATTACGCCAAACTGAGGCTGGGAGAGGT


58-3 1f    TGTTCTTCCTCACAGGGAGG
58-3 1fR   GCTATGACCATGATTACGCCTGTTCTTCCTCACAGGGAGG
58-3 1r    TCCCCAAATCTGTCCAGTTC
58-3 1rR   GCTATGACCATGATTACGCCTCCCCAAATCTGTCCAGTTC


58-4 1f    CATACCTGGAGGGATGCTTG
58-4 1fR   GCTATGACCATGATTACGCCCATACCTGGAGGGATGCTTG
58-4 1r    TAGGTTGCTGTGTGGCTTCA
58-4 1rR   GCTATGACCATGATTACGCCTAGGTTGCTGTGTGGCTTCA


58-5 1f    CTTCTGACAAAGCAGAGGCC
58-5 1fR   GCTATGACCATGATTACGCCCTTCTGACAAAGCAGAGGCC
```

106

Table 4   page 6 of 7   con't.


```
58-5  1r     GCTGTTAGGGTTACCATCGC
58-5  1rR    GCTATGACCATGATTACGCCGCTGTTAGGGTTACCATCGC


58-6  1f     CCACAGGGTGATATGCTGTC
58-6  1fR    GCTATGACCATGATTACGCCCCACAGGGTGATATGCTGTC
58-6  1r     CGCCTGGCTACTTTGGTACT
58-6  1rR    GCTATGACCATGATTACGCCCGCCTGGCTACTTTGGTACT


58-7  1f     CCAAATGAACCTGGGCAAC
58-7  1fR    GCTATGACCATGATTACGCCCCAAATGAACCTGGGCAAC
58-7  1r     GTCTTGGCTCACTGCAACCT
58-7  1rR    GCTATGACCATGATTACGCCGTCTTGGCTCACTGCAACCT


58-8  1f     GCCAAGACTGTGCTACTGCA
58-8  1r     CAGGGAGCAGATCTTACCCA


58-9  1f     TGGGATTAACTAGGGAGGGG
58-9  1fR    GCTATGACCATGATTACGCCTGGGATTAACTAGGGAGGGG
58-9  1r     TGCTGCTGTCTCCATCTCTG
58-9  1rR    GCTATGACCATGATTACGCCTGCTGCTGTCTCCATCTCTG


58-10 1f     ACAGACCAGCAGTGAAACCTG
58-10 1fR    GCTATGACCATGATTACGCCACAGACCAGCAGTGAAACCTG
58-10 1r     GTTCACTGCAACCTCTGCCT
58-10 1rR    GCTATGACCATGATTACGCCGTTCACTGCAACCTCTGCCT


58-11 1f     GTTCTCGTAGATGCTTGCAGG
58-11 1fR    GCTATGACCATGATTACGCCGTTCTCGTAGATGCTTGCAGG
58-11 1r     GAGGCAGGAGGATCACTTGA
58-11 1rR    GCTATGACCATGATTACGCCGAGGCAGGAGGATCACTTGA


58-12 1f     TGAGCTGAGATCACACCGCT
58-12 1fR    GCTATGACCATGATTACGCCTGAGCTGAGATCACACCGCT
58-12 1r     AGTTGACACTTTGCTGGCCT
58-12 1rR    GCTATGACCATGATTACGCCAGTTGACACTTTGCTGGCCT


58-13 1f     CTCTGCATGGCTTAGGGACA
58-13 1fR    GCTATGACCATGATTACGCCCTCTGCATGGCTTAGGGACA
58-13 1r     GGCTGCTCTCTGCATTCTCT
58-13 1rR    GCTATGACCATGATTACGCCGGCTGCTCTCTGCATTCTCT


58-14 1f     CTGGCTTTAGCTTGCATTTCC
58-14 1fR    GCTATGACCATGATTACGCCCTGGCTTTAGCTTGCATTTCC
58-14 1r     TGCCTCAGTTTTCTCACCTGT
58-14 1rR    GCTATGACCATGATTACGCCTGCCTCAGTTTTCTCACCTGT


58-15 1f     CAAACAGCCACTGAGCATGT
58-15 1fR    GCTATGACCATGATTACGCCCAAACAGCCACTGAGCATGT
58-15 1r     TCCTCCTGTAGATGCCCAAG
```

Table 4   page 7 of 7

58-15 1rR GCTATGACCATGATTACGCCTCCTCCTGTAGATGCCCAAG

Table 5

LRP-5 exon SNPs

Exon	Polymorphism	Amino Acid Change	Location
exon E	G to A	Intronic	10 bp 3' of exon E
exon E	C to T	none	Phe^{331}, exon E
exon F	G to A	Intronic	50 bp 5' of exon F
exon G	C to T	none	Phe^{518}, exon G
exon I	C to T	none	Asn^{709}, exon I
exon P	C to T	Intronic	82 bp 5' of exon P
exon N	C to T	none	Asp^{1068}, exon N
exon N	A to G	none	Val^{1088}, exon N
exon Q	C to T	Ala^{1299} to Val	Ala^{1299}, exon Q
exon U	T to C	Val^{1494} to Ala	Val^{1494}, exon U

Table 6
SNP's Identified in the IDDM 4 Locus

List of PCR Fragments and available RFLP Sites for Analysis:

PCR Product	SNP	Location	Enzyme
Contig 57			
57-1	a/t	13363	none
57-1	a/g	13484	Bst XI
57-2	a/g	14490	none
57-2	a/g	14885	none
57-3	c/g	18776	Mae II
57-3	t/c	18901	Msp I
57-3	a/g	19313	Afl II
57-4	22T/25T	20800	none
57-5	g/a	23713	Msp I
Contig 58			
58-15	c/t	3015	none
58-14	g/c	3897	Pfl MI
58-13	c/g	5574	Eco NI
58-12	t/g	6051	none
58-11	a/g	8168	none
58-10	a/g	8797	none
58-9	g/t	9445	none
58-9	c/t	9718	none
58-8	insert T	10926	Pst I
58-7	t/a	11449	Bst XI
58-7	t/c	11468	none
58-6	t/c	11878	none
58-6	g/a	12057	none
58-6	a/g	12180	Hga I
58-5	c/t	14073	none
58-4	a/g	15044	Mae II
58-4	t/c	15354	none
58-3	insert G	16325	none
58-2	g/a	17662	none
58-1	g/t	18439	Bgl II

Table 7   page 1 of 3

SNP primers

```
57-1  1f     GCAACAGAGCAAGACCCTGT
57-1  1fR    GCTATGACCATGATTACGCCGCAACAGAGCAAGACCCTGT
57-1  1r     AAATTAGCCAGGCATGGTG
57-1  1rR    GCTATGACCATGATTACGCCAAATTAGCCAGGCATGGTG
57-1  1fU    TGTAAAACGACGGCCAGTGCAACAGAGCAAGACCCTGT


57-2  1f     CCTGCAGAAGGAAACCTGAC
57-2  1fR    GCTATGACCATGATTACGCCCCTGCAGAAGGAAACCTGAC
57-2  1r     CTGCATCTTTGCCACCATG
57-2  1rR    GCTATGACCATGATTACGCCCTGCATCTTTGCCACCATG
57-2  1fU    TGTAAAACGACGGCCAGTCCTGCAGAAGGAAACCTGAC


57-3  1f     TTCCCAGGAGGCAAGTTATG
57-3  1fR    GCTATGACCATGATTACGCCTTCCCAGGAGGCAAGTTATG
57-3  1r     TGGGCTTAGGTGATCCTCAC
57-3  1rR    GCTATGACCATGATTACGCCTGGGCTTAGGTGATCCTCAC
57-3  1fU    TGTAAAACGACGGCCAGTTTCCCAGGAGGCAAGTTATG


57-4  1f     ACCAAGCCCAACTAATCAGC
57-4  1fR    GCTATGACCATGATTACGCCACCAAGCCCAACTAATCAGC
57-4  1r     ATGCCTGTAATCCCAGCACT
57-4  1rR    GCTATGACCATGATTACGCCATGCCTGTAATCCCAGCACT
57-4  1fU    TGTAAAACGACGGCCAGTACCAAGCCCAACTAATCAGC


57-5  1f     ACTGCAAGCCCTCTCTGAAC
57-5  1r     CGAAGACTGCGAAACAGACA


58-1  1f     CTAGTGCCGTGCAGAATGAG
58-1  1r     GGCCACTGCAATGAGATACA


58-2  1f     GAGAAACAGTTCCAGGGTGG
58-2  1fR    GCTATGACCATGATTACGCCGAGAAACAGTTCCAGGGTGG
58-2  1r     AAACTGAGGCTGGGAGAGGT
58-2  1rR    GCTATGACCATGATTACGCCAAACTGAGGCTGGGAGAGGT


58-3  1f     TGTTCTTCCTCACAGGGAGG
58-3  1fR    GCTATGACCATGATTACGCCTGTTCTTCCTCACAGGGAGG
58-3  1r     TCCCCAAATCTGTCCAGTTC
58-3  1rR    GCTATGACCATGATTACGCCTCCCCAAATCTGTCCAGTTC


58-4  1f     CATACCTGGAGGGATGCTTG
58-4  1fR    GCTATGACCATGATTACGCCCATACCTGGAGGGATGCTTG
58-4  1r.    TAGGTTGCTGTGTGGCTTCA
58-4  1rR    GCTATGACCATGATTACGCCTAGGTTGCTGTGTGGCTTCA


58-5  1f     CTTCTGACAAAGCAGAGGCC
58-5  1fR    GCTATGACCATGATTACGCCCTTCTGACAAAGCAGAGGCC
```

SUBSTITUTE SHEET (RULE 26)

Table 7   page 2 of 3   con't.

```
58-5 1r     GCTGTTAGGGTTACCATCGC
58-5 1rR    GCTATGACCATGATTACGCCGCTGTTAGGGTTACCATCGC


58-6 1f     CCACAGGGTGATATGCTGTC
58-6 1fR    GCTATGACCATGATTACGCCCCACAGGGTGATATGCTGTC
58-6 1r     CGCCTGGCTACTTTGGTACT
58-6 1rR    GCTATGACCATGATTACGCCCGCCTGGCTACTTTGGTACT


58-7 1f     CCAAATGAACCTGGGCAAC
58-7 1fR    GCTATGACCATGATTACGCCCCAAATGAACCTGGGCAAC
58-7 1r     GTCTTGGCTCACTGCAACCT
58-7 1rR    GCTATGACCATGATTACGCCGTCTTGGCTCACTGCAACCT


58-8 1f     GCCAAGACTGTGCTACTGCA
58-8 1r     CAGGGAGCAGATCTTACCCA


58-9 1f     TGGGATTAACTAGGGAGGGG
58-9 1fR    GCTATGACCATGATTACGCCTGGGATTAACTAGGGAGGGG
58-9 1r     TGCTGCTGTCTCCATCTCTG
58-9 1rR    GCTATGACCATGATTACGCCTGCTGCTGTCTCCATCTCTG


58-10 1f    ACAGACCAGCAGTGAAACCTG
58-10 1fR   GCTATGACCATGATTACGCCACAGACCAGCAGTGAAACCTG
58-10 1r    GTTCACTGCAACCTCTGCCT
58-10 1rR   GCTATGACCATGATTACGCCGTTCACTGCAACCTCTGCCT


58-11 1f    GTTCTCGTAGATGCTTGCAGG
58-11 1fR   GCTATGACCATGATTACGCCGTTCTCGTAGATGCTTGCAGG
58-11 1r    GAGGCAGGAGGATCACTTGA
58-11 1rR   GCTATGACCATGATTACGCCGAGGCAGGAGGATCACTTGA


58-12 1f    TGAGCTGAGATCACACCGCT
58-12 1fR   GCTATGACCATGATTACGCCTGAGCTGAGATCACACCGCT
58-12 1r    AGTTGACACTTTGCTGGCCT
58-12 1rR   GCTATGACCATGATTACGCCAGTTGACACTTTGCTGGCCT


58-13 1f    CTCTGCATGGCTTAGGGACA
58-13 1fR   GCTATGACCATGATTACGCCCTCTGCATGGCTTAGGGACA
58-13 1r    GGCTGCTCTCTGCATTCTCT
58-13 1rR   GCTATGACCATGATTACGCCGGCTGCTCTCTGCATTCTCT


58-14 1f    CTGGCTTTAGCTTGCATTTCC
58-14 1fR   GCTATGACCATGATTACGCCCTGGCTTTAGCTTGCATTTCC
58-14 1r    TGCCTCAGTTTTCTCACCTGT
58-14 1rR   GCTATGACCATGATTACGCCTGCCTCAGTTTTCTCACCTGT


58-15 1f    CAAACAGCCACTGAGCATGT
58-15 1fR   GCTATGACCATGATTACGCCCAAACAGCCACTGAGCATGT
58-15 1r    TCCTCCTGTAGATGCCCAAG
```

Table 7   page 3 of 3   con't.

58-15 1rR GCTATGACCATGATTACGCCTCCTCCTGTAGATGCCCAAG

TABLE 8

Primers designed by microsatellite rescue for genotyping and restriction mapping of
the IDDM4 region on chromosome 11q13. The other primers used are published,
and are also in the Genome Database.

255CA3F   GCCGAGAATTGTCATCTTAACT
255CA3R   GGATTGAAAGCTGCAAACTACA

255CA5F   GGAGCCACCACATCCAGTTA
255CA5R   TGGAGGGATTGCTTGAGG

255CA6F   AGGTGTACACCACCATGCCT
255CA6R   TGGTGCCAATTATTGCTGC

14LCA5F   AGATCTTATACACATGTGCGCG
14LCA5R   AGGTGACATCACTTACAGCGG

L15CA1F   ATTACCCAGGCATGGTGC
L15CA1R   CAGGCACTTCTTCCAGGTCT

18018ACF   AGGGTTACACTGGAGTTTGC
18018ACR   AAACCTTCAATGTGTTCATTAAAAC

E0864CAF   TCAACTTTATTGGGGGTTTA
E0864CAR   AAGGTAAAAGTCCAAAATGG

H0570POLYAF   GGACAGTCAGTTATTGAAATG
H0560POLYAR   TTTCCTCTCTGGGAGTCTCT

E0864CA was obtained from the cosmid E0864
H0570POLYA was obtained from the cosmid H0570
255CA5, 255CA3 and 255CA6 were obtained from the PAC255_m_19
14LCA5 and L15CA1 were obtained from the BAC 14_1_15
18018AC was obtained from the PAC 18_o_18

TABLE 9   PCR Primers for obtaining *LRP-3* cDNA

A.) Primers located within human*LRP-3* cDNA:
The primers are numbered beginning at nucleotide 1 in Fig. 17 (a)

1F (muex 1f)
ATGGAGCCCGAGTGAGC

200f
TCAAGCTGGAGTCCACCATC

218R (27R)
ATGGTGGACTCCAGCTTGAC

256F (1F)
TTCCAGTTTTCCAAGGGAG

265R (26R)
AAAACTGGAAGTCCACTGCG

318R (4R)
GGTCTGCTTGATGGCCTC

343F (2F)
GTGCAGAACGTGGTCATCT

361R (21R)
GTGCAGAACGTGGTCATCT

622R (2R)
AGTCCACAATGATCTTCCGG

638F (4F)
CCAATGGACTGACCATCGAC

657R (1R)
GTCGATGGTCAGTCCATTGG

936f
CACTCGCTGTGAGGAGGAC

956R (22R)
TTGTCCTCCTCACAGCGAG

TABLE 9 (Continued - Page 2 of 7)

1040f (51f)
  ACAACGGCAGGACGTGTAAG

1174f (40f)
ATTGCCATCGACTACGACC

1277f (52f)
  TGGTCAACACCGAGATCAAC

1333f
  AACCTCTACTGGACCGACAC

1462f (41f)
  CTCATGTACTGGACAGACT

1481R (23R)
CAGTCTGTCCAGTACATGAG

1607f (50f)
GAGACGCCAAGACAGACAAG

1713F (21F)
GGACTTCATCTACTGGACTG

1732r (40r)
CAGTCCAGTAGATGAAGTCC

1904r (k275r)
GTGAAGAAGCACAGGTGGCT

1960r
  TCATGTCACTCAGCAGCTCC

1981F (22F)
GCCTTCTTGGTCTTCACCAG

2261F (23F)
GGACCAACAGAATCGAAGTG

2484R (5R)
GTCAATGGTGAGGTCGT

2519F (5F)
ACACCAACATGATCGAGTCG

TABLE 9 (Continued - Page 3 of 7)

2780r
CCGTTGTTGTGCATACAGTC

3011F (24F)
ACAAGTTCATCTACTGGGTG

3154F (25F)
CGGACACTGTTCTGGACGTG

3173R (25R)
CACGTCCAGAACAGTGTCCG

3556R (3R)
TCCAGTAGAGATGCTTGCCA

3577F (3F)
ATCGAGCGTGTGGAGAAGAC

3851r
GTGGCACATGCAAACTGGTC

4094F (30F)
TCCTCATCAAACAGCAGTGC

4173R (6R)
CGGCTTGGTGATTTCACAC

4687F (6F)
GTGTGTGACAGCGACTACAGC

4707R (30R)
GCTGTAGTCGCTGTCACACAC

5061R (7R)
GTACAAAGTTCTCCCAGCCC

3' end with XbaI site
5069r
GCTCTAGAGTACAAAGTTCTCCCAGCCC

Soluble/HSV/His primers
HLRP3_His_primer1 (4203r)
ATCCTCGGGGTCTTCCGGGGCGAGTTCTGGCTGGCTACTGCTGTGGGCCGGGCT

TABLE 9 (Continued - Page 4 of 7)

HLRP3_His_primer2
TGGATATCTCAGTGGTGGTGGTGGTGGTGCTCGACATCCTCGGGGTCTTCCGG
G

HLRP3_5'_primer (49f)
TAGAATTCGCCGCCACCATGGAGGCAGCGCCGCCC

B.) Mouse *Lrp-3* cDNA primers.
The primers are numbered beginning at nucleotide 1 in Figure 18(a).

13f (mulrp3 5f)
GAGGCGGGAGCAAGAGG

68f (MucD 1f)
GC Hind 3 CATGGAGCCCGAGTGAGC

69f (muex 1f)
ATGGAGCCCGAGTGAGC

83r (muex 1r)
TCACTCGGGCTCCATGG

171f (MucD 2f)
TGCTGTACTGCAGCTTGGTC

300f (MucD 10F)
ATGCAGCTGCTGTAGACTTCC

378r (mulrp3 3r)
GTCTGTTTGATGGCCTCCTC

414r (MucD 7R)
ATGTTCTGTGCAGCACCTCC

445r (mulrp3 4r)
GCCATCAGGTGACACGAG

536f (MucD 11F)
AAGGTTCTCTTCTGGCAGGAC

619r (MucD 12R)
CCAGTCAGTCCAGTACATG

118

TABLE 9 (Continued - Page 5 of 7)

714f (museq 1f)
TCGACCTGGAGGAACAGAAG

752f (mulrpAb 1f)
AAGCTCAGCTTCATCCACCG

765r (MucD 8R)
ATGAAGCTGAGCTTGGCATC

915f (MucD 12F)
AGCAGAGGAAGGAGATCCTTAG

957r (MucD 9R)
TCCATGGGTGAGTACAGAGC

1105r (museq 1r)
ATTGTCCTGCAACTGCACAC

1232f (MucD 13F)
GCCATTGCCATTGACTACG

1254r (MucD 10R)
GGATCGTAGTCAATGGCAATG

1425f (MucD 14F)
GAATTGAGGTGACTCGCCTC

1433r (MucD 18R)
CCTCAATTCTGTAGTGCCTG

1501f (muxt 4f)
TGTGTTGCACCCTGTGATG

1579r (MucD 11R)
ATCTAGGTTGGCGCATTCG

1610r (MucD 13R)
AGGTGTTCACCAGGACATG

1710r (mulrpAb 1r)
GCGAGCTCCCGTCTATGTTGATCACCTCG

1868f (MucD 3f)
GACCTGATGGGACTCAAAGC

TABLE 9 (Continued - Page 6 of 7)

2062r (MucD 2r)
GCTGGTGAATACCAGGAAGG

2103f (MucD 4f)
ACGATGTGGCTATCCCACTC

2422r (MucD 14R)
AGTAGGATCCAGAGCCAGAG

2619f (MucD 5f)
AGCGCATGGTGATAGCTGAC

2718r (MucD 3r)
CGTTCAATGCTATGCAGGTTC

2892f (MucD 15F)
GTGCTTCACACTACACGCTG

2959f (MucD 6f)
CAGCCAGAAATTTGCCATC

3218r (MucD 4r)
TCCGGCTGTAGATGTCAATG

3237f (MucD 7f)
AGGCCACCAACACTATCAATG

3348r (MucD 52R)
TACCCTCGCTCAGCATTGAC

3554f (MucD 8f)
CTGGAAGATGCCAACATCG

3684r (MucD 5r)
TGAACCCTAGTCCGCTTGTC

3848f (MucD 18F)
CTGCAGAACCTGCTGACTTG

3973f (MucD 19F)
CCAGAGTGATGAAGAAGGCTG

3981r (MucD 15R)
TCACTCTGGTCAGCACACTC

TABLE 9 (Continued - Page 7 of 7)

4079f (MucD 16F)
CAGGATCGCTCTGATGAAGC

4105r (MucD 53R)
GCAGTTAGCTTCATCAGAGCG

4234f (MucD 9f)
ACCCTCTGATGACATCCCAG

4270r (MucD 16R)
AATGGCACTGCTGTGGGC

4497r (MucD 6r)
AGGCTCATGGAGCTCATCAC

4589r (MucD 54R)
ATAGTGTGGCCTTTGTGCTG

4703f (MucD 17F)
GTCATTCGAGGTATGGCACC

4799r (MucD 17R)
GGTAGTATTTGCTGCTCTTCC

5114r (MucD 1r)
GC Xba I AAAGTTTCCCAGCCCTGCC

Soluble/adeno primers

3554f (MsolF)
CTGGAAGATGCCAACATCG

4264r (MHisR)
GCTCTAGACTAGTGATGGTGATGGTGATGACTGCTGTGGGCTGGGATGTCATC
AGAGGGTGG

Table 10 Summary of Serum Chemistry Comparison of LRP3 treatment vs
control

Variable	Mouse Type	Treatment (%diff± SE)	p-value (Treatment)
triglycerides	WT+KO	-30± 14	0.025
alkaline phosphatase#	WT+KO	-49± 15	0.001
total cholesterol	KO only	-28± 15	0.073
total cholesterol	WT only	30± 13	0.080
AST#	WT+KO	8± 66	0.912
ALT#	WT+KO	-34± 51	0.431
BUN	WT+KO	-19± 15	0.195

statistically significantly higher baseline values for controls.

Table 11: Summary for Blood Chemistry Variables Pooled over Knockout and Wild-Type Mice

Variable	Treat Group	Animal Type	n	baseline (mean±%CV)	post-treat (mean±%CV)t	change	%change (95%CI)	p-value (%chg)
trigly(mg/dL)	Control	POOLED	10	86± 13%	186± 35%	100	115% (61, 189)	<0.001
trigly(mg/dL)	LDL	POOLED	9	92± 31%	81± 55%	-12	-13% (-35, 17)	0.321
trigly(mg/dL)	LRP3	POOLED	8	99± 24%	128± 36%	29	30% (-10, 86)	0.133
alkphos(U/L)	Control	POOLED	10	190± 19%	374± 30%	184	97% (68, 130)	<0.001
alkphos(U/L)	LDL	POOLED	9	162± 12%	193± 29%	31	19% (-1, 43)	0.061
alkphos(U/L)	LRP3	POOLED	8	154± 13%	146± 35%	-8	-5% (-24, 19)	0.604
totchol(mg/dL)	Control	POOLED	10	116± 69%	176± 86%	60	51% (21, 89)	0.002
totchol(mg/dL)	LDL	POOLED	9	124± 58%	87± 68%	-37	-30% (-41, -17)	0.001
totchol(mg/dL)	LRP3	POOLED	8	127± 62%	166± 57%	39	30% (9, 56)	0.009
AST(U/L)	Control	POOLED	9	41± 22%	821± 69%	780	1894% (1142, 3101)	<0.001
AST(U/L)	LDL	POOLED	8	41± 25%	362± 61%	320	772% (369, 1520)	<0.001
AST(U/L)	LRP3	POOLED	8	33± 21%	989± 129%	955	2888% (953, 8380)	<0.001
ALT(U/L)	Control	POOLED	10	33± 15%	624± 59%	591	1798% (1203, 2665)	<0.001
ALT(U/L)	LDL	POOLED	8	32± 36%	331± 42%	299	938% (447, 1872)	<0.001
ALT(U/L)	LRP3	POOLED	8	25± 35%	1020± 157%	994	3944% (861, 16921)	<0.001
BUN(U/L)	Control	POOLED	8	29± 12%	23± 11%	-5	-19% (-29, -7)	0.008
BUN(U/L)	LDL	POOLED	9	28± 19%	25± 14%	-3	-12% (-22, 1)	0.062
BUN(U/L)	LRP3	POOLED	8	28± 12%	19± 41%	-9	-31% (-53, 2)	0.058

Note means given are geometric means.
p-value is from a 2-sided paired t-test.

123

Table 12   Regions of
Sequence Similarity
Between Human and
Mouse LRP-3

Location in Human Sequence	Nucleotide Length	Percent Identity	BLAST Score	Exon Name
Contig 31				
20235-20271	37	86	140	
24410-24432	23	86	88	
24464-24667	204	82	168, 223	6
24904-24995	52	82	179	
25489-25596	108	81	360	
26027-26078	52	80	170	
26192-26261	70	84	251	
26385-26486	102	87	393	
28952-28993	42	85	156	
41707-41903	197	90	823	
42827-42898	66	81	222	
43468-43585	117	85	316	
50188-50333	146	86	550	
54455-54494	40	80	128	
54718-54750	33	87	129	
59713-60123	411	87	1587	A
78536-78680	145	80	473	D
87496-87548	53	88	211	
87598-87717	120	84	429	
90772-90819	48	85	177	
99457-99795	339	83	1182	E
103094-103281	188	83	661	F
116659 -116954	296	81	985	G
119754-120089	336	83	1167	H
Contig 30				
8920-9256	337	89	1026	K
11238-11353	116	84	*418	L
18394-18648	255	80	825	M
20020-20224	205	84	746	N
20926-21153	228	83	807	O
24955-25155	201	82	672	P
29126-19288	163	74	*437	Q
33874-34033	160	85	*593	S
35205-35340	136	86	509	T
41911-41911	55	80	*176	U
44629-44681	53	73	*249	V

- 123 A-

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgment or any form of suggestion that that prior art forms part of the common general knowledge in Australia.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

- 124 -

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. An isolated nucleic acid molecule encoding a polypeptide which includes the amino acid sequence shown in Figure 5(e).

2. A nucleic acid molecule according to claim 1 wherein the polypeptide includes the amino acid sequence shown in Figure 5 (c).

3. A nucleic acid molecule according to claim 2 including the coding sequence shown in Figure 5(a).

4. An isolated nucleic acid molecule which hybridizes under stringent conditions to a nucleic acid according to claim 3 and which encodes a polypeptide which has LRP5 activity.

5. An isolated nucleic acid molecule encoding a polypeptide which has LRP5 activity and which is a mutant, allele, variant or derivative of the amino acid sequence of Figure 5(e) or Figure 5(c), by way of addition, deletion, insertion and/or substitute of one or more amino acids.

6. A nucleic acid molecule according to claim 5 wherein said polypeptide includes the amino acid sequence of a polypeptide selected from that shown in Figure 11(c), Figure 12(d) and Figure 18(c).

7. A nucleic acid molecule according to claim 6 including a coding sequence selected from the nucleotide sequences shown in Figure 11(b), Figure 12(d), Figure 13, Figure 14, Figure 15(a) and Figure 18(b).

8. A nucleic acid molecule according to claim 7 wherein the coding sequence is that shown in Figure 11(b), included within a molecule which has the sequence shown in Figure 11(a).

9.   An oligonucleotide fragment of a nucleic acid molecule according to any preceding claim of at least about 20 nucleotides.

10.   An oligonucleotide with a sequence shown in any of Tables 2, 4, 7, 8 and 9.

11.   An isolated nucleic acid molecule according to any one of claims 1 to 8 including a LRP-5 gene promoter.

12.   A nucleic acid molecule according to claim 11 including a promoter, the promoter including the sequence or nucleotides shown in Figure 12(e) or Figure 15(b).

13.   An isolated polypeptide to including the amino acid sequence shown in Figure 5(c).

14.   A polypeptide according to claim 13 including the amino acid sequence shown in Figure 5(c).

15.   An isolated polypeptide which has LRP5 activity which is an amino acid sequence mutant, variant, allele or derivative of the amino acid sequence of Figure 5(e) or Figure 5(c), by way of addition, deletion, insertion and/or substitution of one or more amino acids.

16.   A polypeptide according to claim 15 wherein said polypeptide includes the amino acid sequence of a polypeptide selected from that shown in Figure 11(c), Figure 12(d) and Figure 18(c).

17.   A fragment of a polypeptide according to any of claims 13 to 16 including at least 5 contiguous amino acids of an amino acid sequence selected from the amino acid sequences of Figure 5(c), Figure 11(c), Figure 12(d) and Figure 18(c).

18.   A fragment according to claim 17 which has an amino acid

sequence selected from:

SYFHLFPPPPSPCTDSS,

VDGRQNIKRAKDDGT,

EVLFTTGLIRPVALVVDN, and

IQGHLDFVMDILVFHS.

19. A fragment according to claim 17 which includes the LRP5 extracellular domain.

20. A fragment according to claim 17 which includes the LRP5 cytoplasmic domain.

21. A method of production of a polypeptide according to any of claims 13 to 16 which includes expression of the polypeptide from encoding nucleic acid.

22. A method according to claim 21 further including isolating and/or purifying the polypeptide.

23. A method according to claim 21 or claim 22 further including formulating the polypeptide into a composition which includes at least one additional component.

24. A composition including a polypeptide according to any of claims 13 to 16 and a pharmaceutically acceptable excipient.

25. A method of production of a fragment according to any of claims 17 to 20 which includes expression of the fragment from encoding nucleic acid.

26. A method according to claim 25 further including isolating and/or purifying the polypeptide.

27. A method according to claim 25 or claim 26 further including formulating the polypeptide into a composition which includes at least one additional component.

28.   A composition including a fragment according to any of claims 17 to 20, or a functional mimetic thereof, and a pharmaceutically acceptable excipient.

29.   A composition including a nucleic acid molecule according to any of claims 1 to 8 and a pharmaceutically acceptable excipient.

30.   An isolated antibody specific for a polypeptide according to any of claims 13 to 16.

31.   An isolated antibody according to claim 30 which binds an amino acid sequence selected from:

SYFHLFPPPPSPCTDSS,
VDGRQNIKRAKDDGT,
EVLFTTGLIRPVALVVDN, and
IQGHLDFVMDILVFHS.

32.   A composition including an antibody according to claim 30 or claim 31 and a pharmaceutically acceptable excipient.

33.   A method of reducing triglyceride levels in serum of an individual, the method including administering to the individual a polypeptide according to any of claims 13 to 16, or nucleic acid encoding a said polypeptide.

34.   A nucleic acid molecule according to any one of claims 1 to 12, a polypeptide according to any one of claims 13 to 20, a method according to any one of claims 21 to 23 or 25 to 27 or 33, a composition according to any one of claims 24 or 28 to 29 or 32 or an antibody according to claims 30 to 31 substantially as hereinbefore described with reference to the Figures and/or Examples.

DATED this 27th day of MARCH 2001

MERCK & CO., INC. AND THE WELLCOME TRUST LIMITED AS TRUSTEE TO THE WELLCOME TRUST
by DAVIES COLLISON CAVE
Patent Attorneys for the Applicant(s)

Fig.1.

Fig.2.

B= BamHI
H= HindIII
E= EcoRI
Ea= EagI
Hp= HpaI
N= NruI
S= SaII
p= pac
b= bac
c=cosmid

Fig.3.

Fig.4.

Figure 5(a)

ATGGAGCCCGAGTGAGCGCGGCGCGGGCCCGTCCGGCCGCCGGACAACATGGAGG
CAGCGCCGCCCGGGCCGCCGTGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGGCG
CTGTGCGGCTGCCCGGCCCCCGCCGCGGCCTCGCCGCTCCTGCTATTTGCCAACCG
CCGGGACGTACGGCTGGTGGACGCCGGCGGAGTCAAGCTGGAGTCCACCATCGTG
GTCAGCGGCCTGGAGGATGCGGCCGCAGTGGACTTCCAGTTTTCCAAGGGAGCCGT
GTACTGGACAGACGTGAGCGAGGAGGCCATCAAGCAGACCTACCTGAACCAGACG
GGGGCCGCCGTGCAGAACGTGGTCATCTCCGGCCTGGTCTCTCCCGACGGCCTCGC
CTGCGACTGGGTGGGCAAGAAGCTGTACTGGACGGACTCAGAGACCAACCGCATC
GAGGTGGCCAACCTCAATGGCACATCCCGGAAGGTGCTCTTCTGGCAGGACCTTGA
CCAGCCGAGGGCCATCGCCTTGGACCCCGCTCACGGGTACATGTACTGGACAGACT
GGGGTGAGACGCCCCGGATTGAGCGGGCAGGGATGGATGGCAGCACCCGGAAGAT
CATTGTGGACTCGGACATTTACTGGCCCAATGGACTGACCATCGACCTGGAGGAGC
AGAAGCTCTACTGGGCTGACGCCAAGCTCAGCTTCATCCACCGTGCCAACCTGGAC
GGCTCGTTCCGGCAGAAGGTGGTGGAGGGCAGCCTGACGCACCCCTTCGCCCTGAC
GCTCTCCGGGGACACTCTGTACTGGACAGACTGGCAGACCCGCTCCATCCATGCCT
GCAACAAGCGCACTGGGGGGAAGAGGAAGGAGATCCTGAGTGCCCTCTACTCACC
CATGGACATCCAGGTGCTGAGCCAGGAGCGGCAGCCTTTCTTCCACACTCGCTGTG
AGGAGGACAATGGCGGCTGCTCCCACCTGTGCCTGCTGTCCCCAAGCGAGCCTTTC
TACACATGCGCCTGCCCCACGGGTGTGCAGCTGCAGGACAACGGCAGGACGTGTA
AGGCAGGAGCCGAGGAGGTGCTGCTGCTGGCCCGGCGGACGGACCTACGGAGGAT
CTCGCTGGACACGCCGGACTTTACCGACATCGTGCTGCAGGTGGACGACATCCGGC
ACGCCATTGCCATCGACTACGACCCGCTAGAGGGCTATGTCTACTGGACAGATGAC
GAGGTGCGGGCCATCCGCAGGGCGTACCTGGACGGGTCTGGGGCGCAGACGCTGG
TCAACACCGAGATCAACGACCCCGATGGCATCGCGGTCGACTGGGTGGCCCGAAA
CCTCTACTGGACCGACACGGGCACGGACCGCATCGAGGTGACGCGCCTCAACGGC
ACCTCCCGCAAGATCCTGGTGTCGGAGGACCTGGACGAGCCCCGAGCCATCGCACT
GCACCCCGTGATGGGCCTCATGTACTGGACAGACTGGGGAGAGAACCCTAAAATCG
AGTGTGCCAACTTGGATGGGCAGGAGCGGCGTGTGCTGGTCAATGCCTCCCTCGGG
TGGCCCAACGGCCTGGCCCTGGACCTGCAGGAGGGGAAGCTCTACTGGGGAGACG
CCAAGACAGACAAGATCGAGGTGATCAATGTTGATGGGACGAAGAGGCGGACCCT
CCTGGAGGACAAGCTCCCGCACATTTTCGGGTTCACGCTGCTGGGGGACTTCATCT
ACTGGACTGACTGGCAGCGCCGCAGCATCGAGCGGGTGCACAAGGTCAAGGCCAG
CCGGGACGTCATCATTGACCAGCTGCCCGACCTGATGGGGCTCAAAGCTGTGAATG
TGGCCAAGGTCGTCGGAACCAACCCGTGTGCGGACAGGAACGGGGGGTGCAGCCA
CCTGTGCTTCTTCACACCCCACGCAACCCGGTGTGGCTGCCCCATCGGCCTGGAGC
TGCTGAGTGACATGAAGACCTGCATCGTGCCTGAGGCCTTCTTGGTCTTCACCAGC
AGAGCCGCCATCCACAGGATCTCCCTCGAGACCAATAACAACGACGTGCCATCCCG
CTCACGGGCGTCAAGGAGGCCTCAGCCCTGGACTTTGATGTGTCCAACAACCACAT
CTACTGGACAGACGTCAGCCTGAAGACCATCAGCCGCGCCTTCATGAACGGGAGCT
CGGTGGAGCACGTGGTGGAGTTTGGCCTTGACTACCCCGAGGGCATGGCCGTTGAC
TGGATGGGCAAGAACCTCTACTGGGCCGACACTGGGACCAACAGAATCGAAGTGG
CGCGGCTGGACGGGCAGTTCCGGCAAGTCCTCGTGTGGAGGGACTTGGACAACCCG
AGGTCGCTGGCCCTGGATCCCACCAAGGGCTACATCTACTGGACCGAGTGGGGCGG

Figure 5(a) (Continued)

CAAGCCGAGGATCGTGCGGGCCTTCATGGACGGGACCAACTGCATGACGCTGGTGG
ACAAGGTGGGCCGGGCCAACGACCTCACCATTGACTACGCTGACCAGCGCCTCTAC
TGGACCGACCTGGACACCAACATGATCGAGTCGTCCAACATGCTGGGTCAGGAGCG
GGTCGTGATTGCCGACGATCTCCCGCACCCGTTCGGTCTGACGCAGTACAGCGATT
ATATCTACTGGACAGACTGGAATCTGCACAGCATTGAGCGGGCCGACAAGACTAGC
GGCCGGAACCGCACCCTCATCCAGGGCCACCTGGACTTCGTGATGGACATCCTGGT
GTTCCACTCCTCCCGCCAGGATGGCCTCAATGACTGTATGCACAACAACGGGCAGT
GTGGGCAGCTGTGCCTTGCCATCCCCGGCGGCCACCGCTGCGGCTGCGCCTCACAC
TACACCCTGGACCCCAGCAGCCGCAACTGCAGCCCGCCCACCACCTTCTTGCTGTT
CAGCCAGAAATCTGCCATCAGTCGGATGATCCCGGACGACCAGCACAGCCCGGATC
TCATCCTGCCCCTGCATGGACTGAGGAACGTCAAAGCCATCGACTATGACCCACTG
GACAAGTTCATCTACTGGGTGGATGGGCGCCAGAACATCAAGCGAGCCAAGGACG
ACGGGACCCAGCCCTTTGTTTTGACCTCTCTGAGCCAAGGCCAAAACCCAGACAGG
CAGCCCCACGACCTCAGCATCGACATCTACAGCCGGACACTGTTCTGGACGTGCGA
GGCCACCAATACCATCAACGTCCACAGGCTGAGCGGGGAAGCCATGGGGGTGGTG
CTGCGTGGGGACCGCGACAAGCCCAGGGCCATCGTCGTCAACGCGGAGCGAGGGT
ACCTGTACTTCACCAACATGCAGGACCGGGCAGCCAAGATCGAACGCGCAGCCCTG
GACGGCACCGAGCGCGAGGTCCTCTTCACCACCGGCCTCATCCGCCCTGTGGCCCT
GGTGGTAGACAACACACTGGGCAAGCTGTTCTGGGTGGACGCGGACCTGAAGCGC
ATTGAGAGCTGTGACCTGTCAGGGGCCAACCGCCTGACCCTGGAGGACGCCAACAT
CGTGCAGCCTCTGGGCCTGACCATCCTTGGCAAGCATCTCTACTGGATCGACCGCC
AGCAGCAGATGATCGAGCGTGTGGAGAAGACCACCGGGGACAAGCGGACTCGCAT
CCAGGGCCGTGTCGCCCACCTCACTGGCATCCATGCAGTGGAGGAAGTCAGCCTGG
AGGAGTTCTCAGCCCACCCATGTGCCCGTGACAATGGTGGCTGCTCCCACATCTGT
ATTGCCAAGGGTGATGGGACACCACGGTGCTCATGCCCAGTCCACCTCGTGCTCCT
GCAGAACCTGCTGACCTGTGGAGAGCCGCCCACCTGCTCCCCGGACCAGTTTGCAT
GTGCCACAGGGGAGATCGACTGTATCCCCGGGGCCTGGCGCTGTGACGGCTTTCCC
GAGTGCGATGACCAGAGCGACGAGGAGGGCTGCCCCGTGTGCTCCGCCGCCCAGTT
CCCCTGCGCGCGGGGTCAGTGTGTGGACCTGCGCCTGCGCTGCGACGGCGAGGCAG
ACTGTCAGGACCGCTCAGACGAGGCGGACTGTGAGGCCATCTGCCTGCCCAACCAG
TTCCGGTGTGCGAGCGGCCAGTGTGTCCTATCAAACAGCAGTGCGACTCCTTCCCC
GACTGTATCGACGGCTCCGACGAGCTCATGTGTGAAATCACCAAGCCGCCCTCAGA
CGACAGCCCGGCCCACAGCAGTGCCATCGGGCCCGTCATTGGCATCATCCTCTCTC
TCTTCGTCATGGGTGGTGTCTATTTTGTGTGCCAGCGCGTGGTGTGCCAGCGCTATG
CGGGGGGCCAACGGGCCCTTCCCGCACGAGTATGTCAGCGGGACCCCGCACGTGCCC
CTCAATTTCATAGCCCCGGGCGGTTCCCAGCATGGCCCCTTCACAGGCATCGCATG
CGGAAAGTCCATGATGAGCTCCGTGAGCCTGATGGGGGGCCGGGGCGGGGTGCCC
CTCTACGACCGGAACCACGTCACAGGGGCCTCGTCCAGCAGCTCGTCCAGCACGAA
GGCCACGCTGTACCCGCCGATCCTGAACCCGCCGCCCTCCCCGGCCACGGACCCCT
CCCTGTACAACATGGACATGTTCTACTCTTCAAACATTCCGGCCACTGTGAGACCG
TACAGGCCCTACATCATTCGAGGAATGGCGCCCCCGACGACGCCCTGCAGCACCGA
CGTGTGTGACAGCGACTACAGCGCCAGCCGCTGGAAGGCCAGCAAGTACTACCTG
GATTTGAACTCGGACTCAGACCCCTATCCACCCCCACCCACGCCCCACAGCCAGTA

Figure 5(a) (Continued)

CCTGTCGGCGGAGGACAGCTGCCCGCCCTCGCCCGCCACCGAGAGGAGCTACTTCC
ATCTCTTCCCGCCCCCTCCGTCCCCCTGCACGGACTCATCCTGACCTCGGCCGGGCC
ACTCTGGCTTCTCTGTGCCCCTGTAAATAGTTTTAAATATGAACAAAGAAAAAAAT
ATATTTTATGATTTAAAAAATAAATATAATTGGGATTTTAAAAACATGAGAAATGT
GAACTGTGATGGGGTGGGCAGGGCTGGGAGAACTTTGTACAGTGGAACAAATATTT
ATAAACTTAATTTTGTAAAACAG

Figure 5(b)

ATGGAGGCAGCGCCGCCCGGGCCGCCGTGGCCGCTGCTGCTGCTGCTGCTGCTG
CTGCTGGCGCTGTGCGGCTGCCCGGCCCCCGCCGCGGCCTCGCCGCTCCTGCTA
TTTGCCAACCGCCGGGACGTACGGCTGGTGGACGCCGGCGGAGTCAAGCTGGA
GTCCACCATCGTGGTCAGCGGCCTGGAGGATGCGGCCGCAGTGGACTTCCAGTT
TTCCAAGGGAGCCGTGTACTGGACAGACGTGAGCGAGGAGGCCATCAAGCAGA
CCTACCTGAACCAGACGGGGGCCGCCGTGCAGAACGTGGTCATCTCCGGCCTGG
TCTCTCCCGACGGCCTCGCCTGCGACTGGGTGGGCAAGAAGCTGTACTGGACGG
ACTCAGAGACCAACCGCATCGAGGTGGCCAACCTCAATGGCACATCCCGGAAG
GTGCTCTTCTGGCAGGACCTTGACCAGCCGAGGGCCATCGCCTTGGACCCCGCT
CACGGGTACATGTACTGGACAGACTGGGGTGAGACGCCCCGGATTGAGCGGGC
AGGGATGGATGGCAGCACCCGGAAGATCATTGTGGACTCGGACATTTACTGGCC
CAATGGACTGACCATCGACCTGGAGGAGCAGAAGCTCTACTGGGCTGACGCCA
AGCTCAGCTTCATCCACCGTGCCAACCTGGACGGCTCGTTCCGGCAGAAGGTGG
TGGAGGGCAGCCTGACGCACCCCTTCGCCCTGACGCTCTCCGGGGACACTCTGT
ACTGGACAGACTGGCAGACCCGCTCCATCCATGCCTGCAACAAGCGCACTGGGG
GGAAGAGGAAGGAGATCCTGAGTGCCCTCTACTCACCCATGGACATCCAGGTGC
TGAGCCAGGAGCGGCAGCCTTTCTTCCACACTCGCTGTGAGGAGGACAATGGCG
GCTGCTCCCACCTGTGCCTGCTGTCCCCAAGCGAGCCTTTCTACACATGCGCCT
GCCCCACGGGTGTGCAGCTGCAGGACAACGGCAGGACGTGTAAGGCAGGAGCC
GAGGAGGTGCTGCTGCTGGCCCGGCGGACGGACCTACGGAGGATCTCGCTGGA
CACGCCGGACTTTACCGACATCGTGCTGCAGGTGGACGACATCCGGCACGCCAT
TGCCATCGACTACGACCCGCTAGAGGGCTATGTCTACTGGACAGATGACGAGGT
GCGGGCCATCCGCAGGGCGTACCTGGACGGGTCTGGGGCGCAGACGCTGGTCA
ACACCGAGATCAACGACCCCGATGGCATCGCGGTCGACTGGGTGGCCCGAAAC
CTCTACTGGACCGACACGGGCACGGACCGCATCGAGGTGACGCGCCTCAACGG
CACCTCCCGCAAGATCCTGGTGTCGGAGGACCTGGACGAGCCCCGAGCCATCGC
ACTGCACCCCGTGATGGGCCTCATGTACTGGACAGACTGGGGAGAGAACCCTAA
AATCGAGTGTGCCAACTTGGATGGGCAGGAGCGGCGTGTGCTGGTCAATGCCTC
CCTCGGGTGGCCCAACGGCCTGGCCCTGGACCTGCAGGAGGGGAAGCTCTACTG
GGGAGACGCCAAGACAGACAAGATCGAGGTGATCAATGTTGATGGGACGAAGA
GGCGGACCCTCCTGGAGGACAAGCTCCCGCACATTTTCGGGTTCACGCTGCTGG
GGGACTTCATCTACTGGACTGACTGGCAGCGCCGCAGCATCGAGCGGGTGCACA
AGGTCAAGGCCAGCCGGGACGTCATCATTGACCAGCTGCCCGACCTGATGGGGC
TCAAAGCTGTGAATGTGGCCAAGGTCGTCGGAACCAACCCGTGTGCGGACAGG
AACGGGGGGTGCAGCCACCTGTGCTTCTTCACACCCCACGCAACCCGGTGTGGC
TGCCCCATCGGCCTGGAGCTGCTGAGTGACATGAAGACCTGCATCGTGCCTGAG
GCCTTCTTGGTCTTCACCAGCAGAGCCGCCATCCACAGGATCTCCCTCGAGACC
AATAACAACGACGTGGCCATCCCGCTCACGGGCGTCAAGGAGGCCTCAGCCCTG
GACTTTGAGTGTCCAACAACCACATCTACTGGACAGACGTCAGCCTGAAGACCA
TCAGCCGCGCCTTCATGAACGGGAGCTCGGTGGAGCACGTGGTGGAGTTTGGCC
TTGACTACCCCGAGGGCATGGCCGTTGACTGGATGGGCAAGAACCTCTACTGGG
CCGACACTGGGACCAACAGAATCGAAGTGGCGCGGCTGGACGGGCAGTTCCGG
CAAGTCCTCGTGTGGAGGGACTTGGACAACCCGAGGTCGCTGGCCCTGGATCCC
ACCAAGGGCTACATCTACTGGACCGAGTGGGGCGGCAAGCCGAGGATCGTGCG
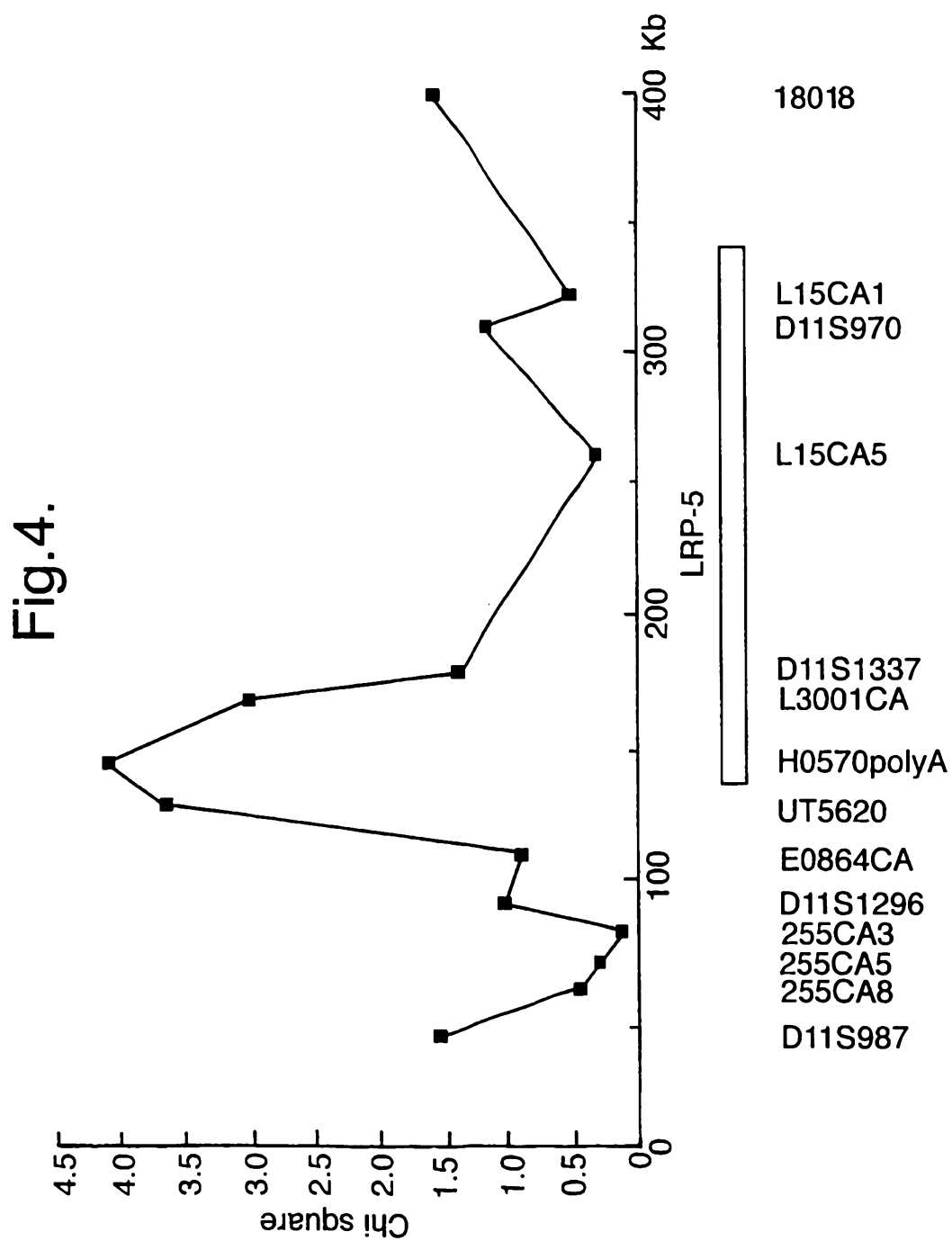GGCCTTCATGGACGGGACCAACTGCATGACGCTGGTGGACAAGGTGGGCCGGG
CCAACGACCTCACCATTGACTACGCTGACCAGCGCCTCTACTGGACCGACCTGG

Figure 5(b) (Continued)

ACACCAACATGATCGAGTCGTCCAACATGCTGGGTCAGGAGCGGGTCGTGATTG
CCGACGATCTCCCGCACCCGTTCGGTCTGACGCAGTACAGCGATTATATCTACT
GGACAGACTGGAATCTGCACAGCATTGAGCGGGCCGACAAGACTAGCGGCCGG
AACCGCACCCTCATCCAGGGCCACCTGGACTTCGTGATGGACATCCTGGTGTTC
CACTCCTCCCGCCAGGATGGCCTCAATGACTGTATGCACAACAACGGGCAGTGT
GGGCAGCTGTGCCTTGCCATCCCCGGCGGCCACCGCTGCGGCTGCGCCTCACAC
TACACCCTGGACCCCAGCAGCCGCAACTGCAGCCCGCCCACCACCTTCTTGCTG
TTCAGCCAGAAATCTGCCATCAGTCGGATGATCCCGGACGACCAGCACAGCCCG
GATCTCATCCTGCCCCTGCATGGACTGAGGAACGTCAAAGCCATCGACTATGAC
CCACTGGACAAGTTCATCTACTGGGTGGATGGGCGCCAGAACATCAAGCGAGCC
AAGGACGACGGGACCCAGCCCTTTGTTTTGACCTCTCTGAGCCAAGGCCAAAAC
CCAGACAGGCAGCCCCACGACCTCAGCATCGACATCTACAGCCGGACACTGTTC
TGGACGTGCGAGGCCACCAATACCATCAACGTCCACAGGCTGAGCGGGGAAGC
CATGGGGGTGGTGCTGCGTGGGGACCGCGACAAGCCCAGGGCCATCGTCGTCA
ACGCGGAGCGAGGGTACCTGTACTTCACCAACATGCAGGACCGGGCAGCCAAG
ATCGAACGCGCAGCCCTGGACGGCACCGAGCGCGAGGTCCTCTTCACCACCGGC
CTCATCCGCCCTGTGGCCCTGGTGGTAGACAACACACTGGGCAAGCTGTTCTGG
GTGGACGCGGACCTGAAGCGCATTGAGAGCTGTGACCTGTCAGGGGCCAACCG
CCTGACCCTGGAGGACGCCAACATCGTGCAGCCTCTGGGCCTGACCATCCTTGG
CAAGCATCTCTACTGGATCGACCGCCAGCAGCAGATGATCGAGCGTGTGGAGAA
GACCACCGGGGACAAGCGGACTCGCATCCAGGGCCGTGTCGCCCACCTCACTGG
CATCCATGCAGTGGAGGAAGTCAGCCTGGAGGAGTTCTCAGCCCACCCATGTGC
CCGTGACAATGGTGGCTGCTCCCACATCTGTATTGCCAAGGGTGATGGGACACC
ACGGTGCTCATGCCCAGTCCACCTCGTGCTCCTGCAGAACCTGCTGACCTGTGG
AGAGCCGCCCACCTGCTCCCCGGACCAGTTTGCATGTGCCACAGGGGAGATCGA
CTGTATCCCCGGGGCCTGGCGCTGTGACGGCTTTCCCGAGTGCGATGACCAGAG
CGACGAGGAGGGCTGCCCCGTGTGCTCCGCCGCCCAGTTCCCCTGCGCGCGGGG
TCAGTGTGTGGACCTGCGCCTGCGCTGCGACGGCGAGGCAGACTGTCAGGACCG
CTCAGACGAGGCGGACTGTGACGCCATCTGCCTGCCCAACCAGTTCCGGTGTGC
GAGCGGCCAGTGTGTCCTCATCAAACAGCAGTGCGACTCCTTCCCCGACTGTAT
CGACGGCTCCGAGAGCTCATGTGTGAAATCACCAAGCCGCCCTCAGACGACAGC
CCGGCCCACAGCAGTGCCATCGGGCCCGTCATTGGCATCATCCTCTCTCTCTTC
GTCATGGGTGGTGTCTATTTTGTGTGCCAGCGCGTGGTGTGCCAGCGCTATGCG
GGGGCCAACGGGCCCTTCCCGCACGAGTATGTCAGCGGGACCCCGCACGTGCCC
CTCAATTTCATAGCCCCGGGCGGTTCCCAGCATGGCCCCTTCACAGGCATCGCA
TGCGGAAAGTCCATGATGAGCTCCGTGAGCCTGATGGGGGGGCCGGGGCGGGGT
GCCCCTCTACGACCGGAACCACGTCACAGGGGCCTCGTCCAGCAGCTCGTCCAG
CACGAAGGCCACGCTGTACCCGCCGATCCTGAACCCGCCGCCCTCCCCGGCCAC
GGACCCCTCCCTGTACAACATGGACATGTTCTACTCTTCAAACATTCCGGCCAC
TGTGAGACCGTACAGGCCCTACATCATTCGAGGAATGGCGCCCCCGACGACGCC
CTGCAGCACCGACGTGTGTGACAGCGACTACAGCGCCAGCCGCTGGAAGGCCA
GCAAGTACTACCTGGATTTGAACTCGGACTCAGACCCCTATCCACCCCCACCCA
CGCCCCACAGCCAGTACCTGTCGGCGGAGGACAGCTGCCCGCCCTCGCCCGCCA
CCGAGAGGAGCTACTTCCATCTCTTCCCGCCCCCCTCCGTCCCCCTGCACGGA
CTCATCC

Figure 5(c)

MEAAPPGPPWPLLLLLLLLLALCGCPAPAAASPLLLFANRRDVRLVDAGGVKLESTIV
VSGLEDAAAVDFQFSKGAVYWTDVSEEAIKQTYLNQTGAAVQNVVISGLVSPDGLAC
DWVGKKLYWTDSETNRIEVANLNGTSRKVLFWQDLDQPRAIALDPAHGYMYWTDW
GETPRIERAGMDGSTRKIIVDSDIYWPNGLTIDLEEQKLYWADAKLSFIHRANLDGSFR
QKVVEGSLTHPFALTLSGDTLYWTDWQTRSIHACNKRTGGKRKEILSALYSPMDIQVLS
QERQPFFHTRCEEDNGGCSHLCLLSPSEPFYTCACPTGVQLQDNGRTCKAGAEEVLLL
ARRTDLRRISLDTPDFTDIVLQVDDIRHAIAIDYDPLEGYVYWTDDEVRAIRRAYLDGS
GAQTLVNTEINDPDGIAVDWVARNLYWTDTGTDRIEVTRLNGTSRKILVSEDLDEPRAI
ALHPVMGLMYWTDWGENPKIECANLDGQERRVLVNASLGWPNGLALDLQEGKLYW
GDAKTDKIEVINVDGTKRRTLLEDKLPHIFGFTLLGDFIYWTDWQRRSIERVHKVKASR
DVIIDQLPDLMGLKAVNVAKVVGTNPCADRNGGCSHLCFFTPHATRCGCPIGLELLSD
MKTCIVPEAFLVFTSRAAIHRISLETNNNDVAIPLTGVKEASALDFDVSNNHIYWTDVSL
KTISRAFMNGSSVEHVVEFGLDYPEGMAVDWMGKNLYWADTGTNRIEVARLDGQFR
QVLVWRDLDNPRSLALDPTKGYIYWTEWGGKPRIVRAFMDGTNCMTLVDKVGRAND
LTIDYADQRLYWTDLDTNMIESSNMLGQERVVIADDLPHPFGLTQYSDYIYWTDWNL
HSIERADKTSGRNRTLIQGHLDFVMDILVFHSSRQDGLNDCMHNNGQCGQLCLAIPGG
HRCGCASHYTLDPSSRNCSPPTTFLLFSQKSAISRMIPDDQHSPDLILPLHGLRNVKAIDY
DPLDKFIYWVDGRQNIKRAKDDGTQPFVLTSLSQGQNPDRQPHDLSIDIYSRTLFWTCE
ATNTINVHRLSGEAMGVVLRGDRDKPRAIVVNAERGYLYFTNMQDRAAKIERAALDG
TEREVLFTTGLIRPVALVVDNTLGKLFWVDADLKRIESCDLSGANRLTLEDANIVQPLG
LTILGKHLYWIDRQQQMIERVEKTTGDKRTRIQGRVAHLTGIHAVEEVSLEEFSAHPCA
RDNGGCSHICIAKGDGTPRCSCPVHLVLLQNLLTCGEPPTCSPDQFACATGEIDCIPGA
WRCDGFPECDDQSDEEGCPVCSAAQFPCARGQCVDLRLRCDGEADCQDRSDEADCD
AICLPNQFRCASGQCVLIKQQCDSFPDCIDGSDELMCEITKPPSDDSPAHSSAIGPVIGIIL
SLFVMGGVYFVCQRVVCQRYAGANGPFPHEYVSGTPHVPLNFIAPGGSQHGPFTGIAC
GKSMMSSVSLMGGRGGVPLYDRNHVTGASSSSSSSTKATLYPPILNPPPSPATDPSLYN
MDMFYSSNIPATVRPYRPYIIRGMAPPTTPCSTDVCDSDYSASRWKASKYYLDLNSDSD
PYPPPPTPHSQYLSAEDSCPPSPATERSYFHLFPPPPSPCTDSS

Figure 5(d)

```
MEAAPPGPPW PLLLLLLLLL ALCGCPAPAA ASPLLLFANR RDVRLVDAGG      50
VKLESTIVVS GLEDAAAVDF QFSKGAVYWT DVSEEAIKQT YLNQTGAAVQ     100
NVVISGLVSP DGLACDWVGK KLYWIDSETN RIEVANLNGT SRKVLFWQDL     150
DQPRAIALDP AHGYMYWIDW GETPRIERAG MDGSTRKIIV DSDIYWPNGL     200
TIDLEEQKLY WADAKLSFIH RANLDGSFRQ KVVEGSLTHP FALTLSGDTL     250
YWIDWQTRSI HACNKRTGGK RKEILSALYS PMDIQVLSQE RQPFFHTRCE     300
EDNGGCSHLC LSPSEPSA CACPIGVQLQ DNGRTCKAGA EEVLLLARRT      350
DLRRISLDTP DFTDIVLQVD DIRHAIAIDY DPLEGYVYWT DDEVRAIRRA     400
YLDGSGAQTL VNTEINDPDG IAVDWVARNL YWIDTGTDRI EVTRLNGTSR     450
KILVSEDLDE PRAIALHPVM GLMYWIDWGE NPKIECANLD GQERRVLVNA     500
SLGWPNGLAL DLQEGKLYWG DAKTDKIEVI NVDGTKRRTL LEDKLPHIFG     550
FTLLGDFTYW TDWQRRSIER VHKVKASRDV IIDQLPDLMG LKAVNVAKVV     600
GINREADRNS CGSHLCFEEP HATRGCCPIG DDLGSRMKIC IVPEAFLVFT     650
SRAAIHRISL EINNNDVAIP LTGVKEASAL DFDVSNNHIY WIDVSLKTIS     700
RAFMNGSSVE HVVEFGLDYP EGMAVDWMGK NLYWADIGTN RIEVARLDGQ     750
FRQVLVWRDL DNPRSIALDP TKGYIYWTEW GGKPRIVRAF MDGTNCMTLV     800
DKVGRANDLT IDYADQRLYW TDLDINMIES SNMLGQERVV IADDLPHPFG     850
LTQYSDYIYW TDWNLHSIER ADKTSGRNRT LIQGHLDFVM DILVFHSSRQ     900
DGTNRCMINN GGCQQLCLAS PGGHRSGCAS HMLDPSSRN CSPPTTFLLF      950
SQKSAISRMI PDDQHSPDLI LPLHGLRNVK AIDYDPLDKF IYWVDGRQNI    1000
KRAKDDGTQP FVLTSLSQGQ NPDRQPHDLS IDIYSRTLFW TCEATNTINV    1050
HRLSGEAMGV VLRGDRDKPR AIVVNAERGY LYFINMQDRA AKIERAALDG    1100
TEREVLFTTG LIRPVALVVD NTLGKLFWVD ADLKRIESCD LSGANRLTLE    1150
DANIVQPLGL TILGKHLYWI DRQQQMIERV EKTTGDKRTR IQGRVAHLTG    1200
IHAVEEVSLE EFSAHPCARD NGGCSHICIA KGDNTPKCSC PZHIMKLGN     1250
CHGGEHTAGS PDQPAGATPH IDSLPCARTC IDCPPGDDCS DHHCPVGSA     1300
AGRHGARGCC VILPPRDCELA ADGGCRSHEA CCDALCLPNG KRQSGGCVV    1350
TKCGCKSFEE CEPCSHDLV ETTKPPSDDS PAHSSAIGPV IGIILSLFVM    1400
GGVYFVCQRV VCQRYAGANG PFPHEYVSGT PHVPLNFIAP GGSQHGPFIG    1450
IACGKSMMSS VSLMGGRGGV PLYDRNHVTG ASSSSSSSTK ATLYPHILNP    1500
PPSPATDPSL YNMDMFYSSN IPATVRPRESSI IRGMAPPT TPCSIDVCDS    1550
DYSASRWKAS KYYLDLNSDS DPYPPPPTPH SQYLSAEDSC PPSPATERSY    1600
FHLFPPPPSP CTDSS                                         1615
```

Figure 5(e)

```
CPAPAAASPLLLFANRRDVRLVDAGGVKLESTIVVSGLEDAAAVDFQFSKGAVYWTD
VSEEAIKQTYLNQTGAAVQNVVISGLVSPDGLACDWVGKKLYWTDSETNRIEVANLN
GTSRKVLFWQDLDQPRAIALDPAHGYMYWTDWGETPRIERAGMDGSTRKIIVDSDIY
WPNGLTIDLEEQKLYWADAKLSFIHRANLDGSFRQKVVEGSLTHPFALTLSGDTLYWT
DWQTRSIHACNKRTGGKRKEILSALYSPMDIQVLSQERQPFFHTRCEEDNGGCSHLCLL
SPSEPFYTCACPTGVQLQDNGRTCKAGAEEVLLLARRTDLRRISLDTPDFTDIVLQVDDI
RHAIAIDYDPLEGYVYWTDDEVRAIRRAYLDGSGAQTLVNTEINDPDGIAVDWVARNL
YWTDTGTDRIEVTRLNGTSRKILVSEDLDEPRAIALHPVMGLMYWTDWGENPKIECAN
LDGQERRVLVNASLGWPNGLALDLQEGKLYWGDAKTDKIEVINVDGTKRRTLLEDKL
PHIFGFTLLGDFIYWTDWQRRSIERVHKVKASRDVIIDQLPDLMGLKAVNVAKVVGTN
PCADRNGGCSHLCFFTPHATRCGCPIGLELLSDMKTCIVPEAFLVFTSRAAIHRISLETN
NNDVAIPLTGVKEASALDFDVSNNHIYWTDVSLKTISRAFMNGSSVEHVVEFGLDYPE
GMAVDWMGKNLYWADTGTNRIEVARLDGQFRQVLVWRDLDNPRSLALDPTKGYIY
WTEWGGKPRIVRAFMDGTNCMTLVDKVGRANDLTIDYADQRLYWTDLDTNMIESSN
MLGQERVVIADDLPHPFGLTQYSDYIYWTDWNLHSIERADKTSGRNRTLIQGHLDFVM
DILVFHSSRQDGLNDCMHNNGQCGQLCLAIPGGHRCGCASHYTLDPSSRNCSPPTTFLL
FSQKSAISRMIPDDQHSPDLILPLHGLRNVKAIDYDPLDKFIYWVDGRQNIKRAKDDGT
QPFVLTSLSQGQNPDRQPHDLSIDIYSRTLFWTCEATNTINVHRLSGEAMGVVLRGDRD
KPRAIVVNAERGYLYFTNMQDRAAKIERAALDGTEREVLFTTGLIRPVALVVDNTLGK
LFWVDADLKRIESCDLSGANRLTLEDANIVQPLGLTILGKHLYWIDRQQQMIERVEKTT
GDKRTRIQGRVAHLTGIHAVEEVSLEEFSAHPCARDNGGCSHICIAKGDGTPRCSCPVH
LVLLQNLLTCGEPPTCSPDQFACATGEIDCIPGAWRCDGFPECDDQSDEEGCPVCSAAQ
FPCARGQCVDLRLRCDGEADCQDRSDEADCDAICLPNQFRCASGQCVLIKQQCDSFPD
CIDGSDELMCEITKPPSDDSPAHSSAIGPVIGIILSLFVMGGVYFVCQRVVCQRYAGANG
PFPHEYVSGTPHVPLNFIAPGGSQHGPFTGIACGKSMMSSVSLMGGRGGVPLYDRNHV
TGASSSSSSSTKATLYPPILNPPPSPATDPSLYNMDMFYSSNIPATVRPYRPYIIRGMAPPT
TPCSTDVCDSDYSASRWKASKYYLDLNSDSDPYPPPPTPHSQYLSAEDSCPPSPATERSY
FHLFPPPPSPCTDSS
```

Figure 5f

```
Human   1 ..................ATGGAGCCCGAGTGAGCGCGGCGCGGGCCCCGT  32
            |||||||||||||||||||||||||       ||| |
Mouse  51 ACCGCCGCCGCGCGCGCCATGGAGCCCGAGTGAGCGCG....CGGCGCTC  96

       33 CCGGCCGCCGGACAACATGGAGGCAGCGCCGCCCGGGCCGCCGTGGCCGC  82
            ||||||||||||| |||||||| | |||||| || || | ||    ||||
       97 CCGGCCGCCGGACGACATGGAAACGGCGCCGACCCGGGCCCC...TCCGC 143

       83 TGCTGCTGCTGCTGCTGCTGCTGCTGGCGCTGTGCGGCTGCCCGGCCCCC 132
            || || || || |||||||||||||| ||||| | || ||  || ||||
      144 CGCCGCCGCCGCCGCTGCTGCTGCTGGTGCTGTACTGCAGCTTGGTCCCC 193

      133 GCCGCGGCCTCGCCGCTCCTGCTATTTGCCAACCGCCGGACGTACGGCT 182
            |||||||||| ||||||||| | ||||||||||||||||| || |||||
      194 GCCGCGGCCTCACCGCTCCTGTTGTTTGCCAACCGCCGGGATGTGCGGCT 243

      183 GGTGGACGCCGGCGGGAGTCAAGCTGGAGTCCACCATCGTGGTCAGCGGCC 232
            ||||| |||||||||| |||||||||||||||||| |||| ||| ||||
      244 AGTGGATGCCGGCGGGAGTGAAGCTGGAGTCCACCATTGTGGCCAGTGGCC 293

      233 TGGAGGATGCGGCCGCAGTGGACTTCCAGTTTTCCAAGGGAGCCGTGTAC 282
            |||||||||| || || || ||||||||||| ||||||||| || ||||||
      294 TGGAGGATGCAGCTGCTGTAGACTTCCAGTTCTCCAAGGGTGCTGTGTAC 343

      283 TGGACAGACGTGAGCGAGGAGGCCATCAAGCAGACCTACCTGAACCAGAC 332
            |||||||| |||||||||||||||||||| |||||||||||||||||||||
      344 TGGACAGATGTGAGCGAGGAGGCCATCAAACAGACCTACCTGAACCAGAC 393

      333 GGGGGCCGCCGTGCAGAACGTGGTCATCTCCGGCCTGGTCTCTCCCGACG 382
            || | || | |||||| | ||||||||| ||||| || || || ||| |
      394 TGGAGGTGCTGCACAGAACATTGTCATCTCGGCCTCGTGTCACCTGATG 443

      383 GCCTCGCCTGCGACTGGGTGGGCAAGAAGCTGTACTGGACGGACTCAGAG 432
            |||| |||||| |||||||| ||||||||||||||||||||||||||| |||
      444 GCCTGGCCTGTGACTGGGTTGGCAAGAAGCTGTACTGGACGGACTCCGAG 493
```

Figure 5g

```
Human   1 MEAAPFGPPWPLLLLLLLLLLAICGCPAPAAASPLLLFANRRDVRLVDAGG 50
            ||  ||     | |       ||||.  |       |||||||||||||||||||||||||||||
Mouse   1 METAPTRAPPPPPPPLLLLVLYCSL.VPAAASPLLLFANRRDVRLVDAGG 49

       51 VKLESTIVVSGLEDAAAVDFQFSKGAVYWIDVSEEAIKQTYLNQTGAAVQ 100
            ||||||||| |||||||||||||||||||||||||||||||||||||| | |
       50 VKLESTIVASGLEDAAAVDFQFSKGAVYWIDVSEEAIKQTYLNQTGGAAQ 99

      101 NVVISGLVSPDGLACDWVGKKLYWIDSEINRIEVANLNGTSRKVLFWQDL 150
            |:||||||||||||||||||||||||||||||||||||||||||||||||||
      100 NIVISGLVSPDGLACDWVGKKLYWIDSEINRIEVANLNGTSRKVLFWQDL 149

      151 DQPRAIALDPAHGYMYWIDWGETPRIERAGMDGSIRKIIVDSDIYWPNGL 200
            ||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
      150 DQPRAIALDPAHGYMYWIDWGEAPRIERAGMDGSIRKIIVDSDIYWPNGL 199

      201 TIDLEEQKLYWADAKLSFIHRANLDGSFRQKVVEGSLTHPFALTLSGDTL 250
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
      200 TIDLEEQKLYWADAKLSFIHRANLDGSFRQKVVEGSLTHPFALTLSGDTL 249

      251 YWIDWQTRSIHACNKRTGGKRKEILSALYSPMDIQVLSQERQPFFHTRCE 300
            ||||||||||||||| ||  .||||||||||||||||||||||| ||| ||
      250 YWIDWQTRSIHACNKWTGEQRKEILSALYSPMDIQVLSQERQPPFHTPCE 299

      301 EDNGGCSHLCLLSPSEPFYTCACPTGVQLQDNGRTCKAGAEEVLLLARRT 350
            ||||||||||||| ||||.||||||||||||||:||| ||||||||||||
      300 EDNGGCSHLCLLSPREPFYSCACPTGVQLQDNGKTCKTGAEEVLLLARRT 349

      351 DLRRISLDTPDFTDIVLQVDDIRHAIAIDYDPLEGYVYWIDDEVRAIRRA 400
            |||||||||||||||||  |||||||||||||||||||||||||||||||||
      350 DLRRISLDTPDFTDIVLQVGDIRHAIAIDYDPLEGYVYWIDDEVRAIRRA 399

      401 YLDGSGAQTLVNTEINDPDGIAVDWARNLYWIDTGTDRIEVTRLNGTSR 450
            ||||||||||||||||||||||||||||||||||||||||||||||||||
      400 YLDGSGAQTLVNTEINDPDGIAVDWARNLYWIDTGTDRIEVTRLNGTSR 449

      451 KILVSEDLDEPRAIALHPVMGLMYWIDWGENPKIECANLDGQERRVLVNA 500
            ||||||||||||  |||||||||||||||||||||||||||L|||.:| ||||
      450 KILVSEDLDEPRAIVLHPVMGLMYWIDWGENPKIECANLDGRDRHVLVNT 499

      501 SLGWPNGLALDLQEGKLYWGDAKTDKIEVINVDGTKRRTLLEDKLPHIFG 550
            ||||||||||||||||||||||||||||||||:||
      500 SLGWPNGLALDLQEGKLYWGDAKTDKIEVINIDG............... 533
```

Figure 6(a)

EGF precursor motifs in LRP-5 isoform 1

Isoform 1   268   CEEDNGGCSHLCLLSPSEPFYTCACPTGVQLQDNGRTC 345
                    C NGGCS LCLLSP   - CACPT   L   GRTC
LRP-EGF repeat   CKVNNGGCSNLCLLSPGGG-HKCACPINFYLGSDGRTC

Isoform 1   570 GINPCADRNGGCSHLCFFTPHATRCGCPIGLELLSDMKTCI 650
                    GIN C NGGCS LC TP   C C   L D TC
LRP-EGF repeat   GINKCRVNNGGCSSLCLATPGSRQCACAEDQVLDADGVTCL

Isoform 1   871 GLNDCMHNNGQCGQLCLAIPGGHRCGCASHYTLDPSSRNC 950
                    G N C NNG C LCLA PG   C CA   LD   C
LRP-EGF repeat   GINKCRVNNGGCSSLCLATPGSRQCACAEDQVLDADGVTC

Isoform 1   1184   HPCARDNGGCSHICIAKGDGTPRCSCPVHLVLLQNLLTC 1262
                    HPC   NGGCS C   G   C CP   L   TC
LRP-EGF repeat   HPCKVNNGGCSNLCLLSPGGGHKCACPINFYLGSDGRTC

LDL-receptor motifs in LRP-5 isoform 1

Isoform 1   1226   PTCSPDQFACATGEIDCIPGAWRCDGFPECDDQSDEEGC 1304
                    P C DQF C   G --CIP WRCD   C D SDEE C
LRP-LDL repeat   PRCDMDQFQCKSGH--CIPLRWRCDADADCMDGSDEEAC
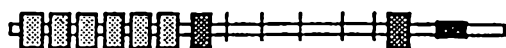

Isoform 1   1267   CSAAQFPCARGQCVDLRLRCDGEADCQDRSDEADCD 1342
                    C QF C G C   CDG DCQD SDEA CD
LRP-LDL repeat   CRPGQFQCSTGICTNPAFICDGDNDCQDNSDEANCD

Isoform 1   1305   CLPNQFRCASGQCVLIKQQCDSFPDCIDGSDELMC 1379
                    C QF C SG C   CD   DC DGSDE C
LRP-LDL repeat   CDMDQFQCKSGHCIPLRWRCDADADCMDGSDEEAC 37

Figure 6(b)

Motif Organization of the LDL- receptor and LRP-5

LDL -receptor

LRP-5
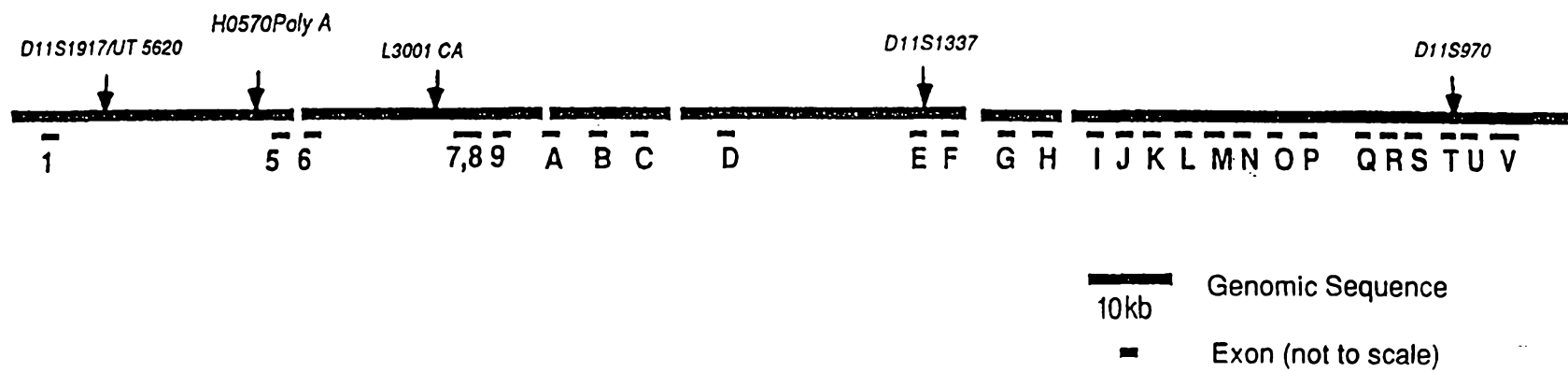

🔲 EGF-precursor B.2 motif

🔲 LDL-receptor motif

| YWTD motif

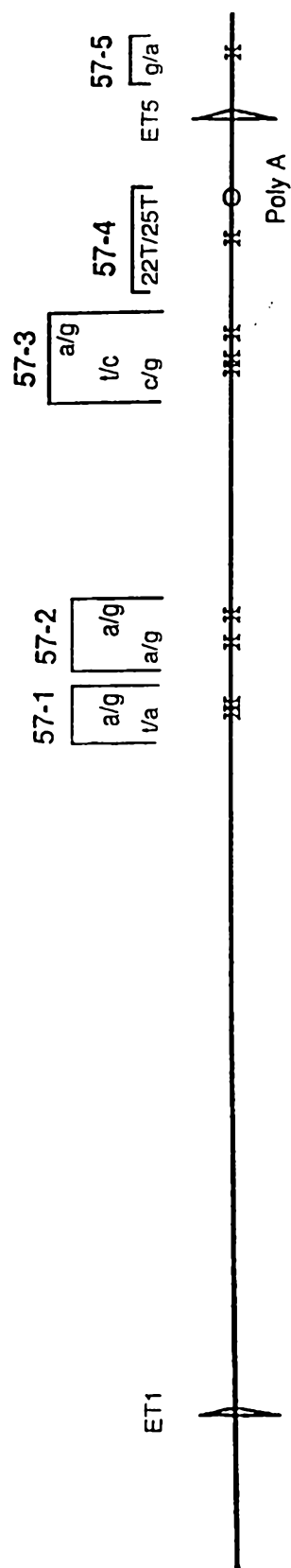▬▬ Transmembrane region
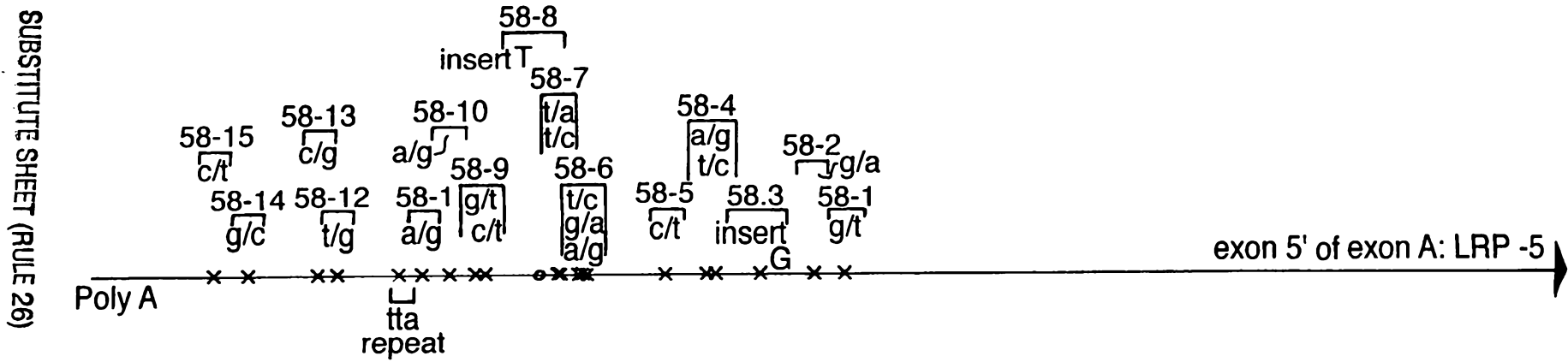
Figure 7

Genomic Sequence

10kb

Exon (not to scale)
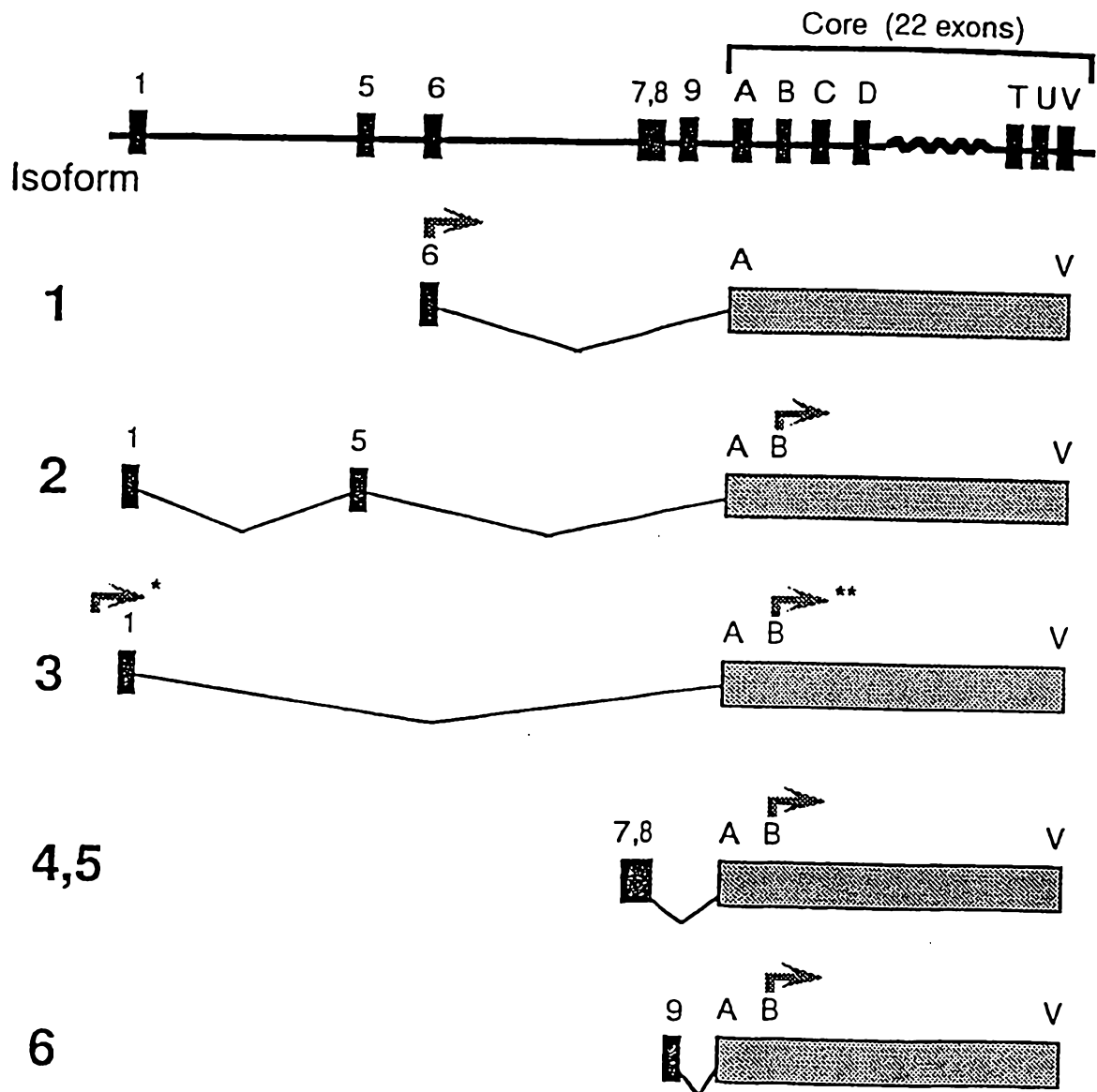
Figure 8

Figure 9

Fig.10.

Figure 11(a)

GAGAGGACACCGCATTCTTCTTCTCCAGAGGATGCAGCAGCAAGGCGCCATC
TTGAAACCAGAGACCAAACCAACCAGCAWTTTTGTCTTGAACTTCCCAGCC
TCCACAACTAATATAAACCCCATGAGGGCAGAGGCGTTCAGCCTGACTCCAG
CCTGGCAAAGCTGTCACAAATCTGGAGGAACACACACGTTCACGGGCACTCA
GTTCTGTGAGCCTCGCCGCTCCTGCTATTTGCCAACCGCCGGGACGTACGGCTG
GTGGACGCCGGCGGAGTCAAGCTGGAGTCCACCATCGTGGTCAGCGGCCTGGA
GGATGCGGCCGCAGTGGACTTCCAGTTTTCCAAGGGAGCCGTGTACTGGACAG
ACGTGAGCGAGGAGGCCATCAAGCAGACCTACCTGAACCAGACGGGGGCCGC
CGTGCAGAACGTGGTCATCTCCGGCCTGGTCTCTCCCGACGGCCTCGCCTGCGAC
TGGGTGGGCAAGAAGCTGTACTGGACGGACTCAGAGACCAACCGCATCGAGG
TGGCCAACCTCAATGGCACATCCCGGAAGGTGCTCTTCTGGCAGGACCTTGAC
CAGCCGAGGGCCATCGCCTTGGACCCCGCTCACGGGTACATGTACTGGACAGA
CTGGGGTGAGACGCCCCGGATTGAGCGGGCAGGGATGGATGGCAGCACCCGGA
AGATCATTGTGGACTCGGACATTTACTGGCCCAATGGACTGACCATCGACCT
GGAGGAGCAGAAGCTCTACTGGGCTGACGCCAAGCTCAGCTTCATCCACCGTG
CCAACCTGGACGGCTCGTTCCGGCAGAAGGTGGTGGAGGGCAGCCTGACGCAC
CCCTTCGCCCTGACGCTCTCCGGGGACACTCTGTACTGGACAGACTGGCAGACC
CGCTCCATCCATGCCTGCAACAAGCGCACTGGGGGGGAAGAGGAAGGAGATCC
TGAGTGCCCTCTACTCACCCATGGACATCCAGGTGCTGAGCCAGGAGCGGCAG
CCTTTCTTCCACACTCGCTGTGAGGAGGACAATGGCGGCTGCTCCCACCTGTGC
CTGCTGTCCCCAAGCGAGCCTTTCTACACATGCGCCTGCCCCACGGGTGTGCAG
CTGCAGGACAACGGCAGGACGTGTAAGGCAGGAGCCGAGGAGGTGCTGCTGC
TGGCCCGGCGGACGGACCTACGGAGGATCTCGCTGGACACGCCGGACTTTACCG
ACATCGTGCTGCAGGTGGACGACATCCGGCACGCCATTGCCATCGACTACGAC
CCGCTAGAGGGCTATGTCTACTGGACAGATGACGAGGTGCGGGCCATCCGCAG
GGCGTACCTGGACGGGTCTGGGGCGCAGACGCTGGTCAACACCGAGATCAACG
ACCCCGATGGCATCGCGGTCGACTGGGTGGCCCGAAACCTCTACTGGACCGAC
ACGGGCACGGACCGCATCGAGGTGACGCGCCTCAACGGCACCTCCCGCAAGAT
CCTGGTGTCGGAGGACCTGGACGAGCCCCGAGCCATCGCACTGCACCCCGTGAT
GGGCCTCATGTACTGGACAGACTGGGGAGAGAACCCTAAAATCGAGTGTGCC
AACTTGGATGGGCAGGAGCGGCGTGTGCTGGTCAATGCCTCCCTCGGGTGGCCC
AACGGCCTGGCCCTGGACCTGCAGGAGGGGAAGCTCTACTGGGGAGACGCCA
AGACAGACAAGATCGAGGTGATCAATGTTGATGGGACGAAGAGGCGGACC
CTCCTGGAGGACAAGCTCCCGCACATTTTCGGGTTCACGCTGCTGGGGGACTTC
ATCTACTGGACTGACTGGCAGCGCCGCAGCATCGAGCGGGTGCACAAGGTCA
AGGCCAGCCGGGACGTCATCATTGACCAGCTGCCCGACCTGATGGGGCTCAAA
GCTGTGAATGTGGCCAAGGTCGTCGGAACCAACCCGTGTGCGGACAGGAACG
GGGGGTGCAGCCACCTGTGCTTCTTCACACCCCACGCAACCCGGTGTGGCTGCCC
CATCGGCCTGGAGCTGCTGAGTGACATGAAGACCTGCATCGTGCCTGAGGCCT

Figure 11(a), Ctd.

TCTTGGTCTTCACCAGCAGAGCCGCCATCCACAGGATCTCCCTCGAGACCAAT
AACAACGACGTGGCCATCCCGCTCACGGGCGTCAAGGAGGCCTCAGCCCTGGA
CTTTGATGTGTCCAACAACCACATCTACTGGACAGACGTCAGCCTGAAGACC
ATCAGCCGCGCCTTCATGAACGGGAGCTCGGTGGAGCACGTGGTGGAGTTTGG
CCTTGACTACCCCGAGGGCATGGCCGTTGACTGGATGGGCAAGAACCTCTACT
GGGCCGACACTGGGACCAACAGAATCGAAGTGGCGCGGCTGGACGGGCAGTT
CCGGCAAGTCCTCGTGTGGAGGGACTTGGACAACCCGAGGTCGCTGGCCCTGG
ATCCCACCAAGGGCTACATCTACTGGACCGAGTGGGGCGGCAAGCCGAGGAT
CGTGCGGGCCTTCATGGACGGGACCAACTGCATGACGCTGGTGGACAAGGTGG
GCCGGGCCAACGACCTCACCATTGACTACGCTGACCAGCGCCTCTACTGGACCG
ACCTGGACACCAACATGATCGAGTCGTCCAACATGCTGGGTCAGGAGCGGGT
CGTGATTGCCGACGATCTCCCGCACCCGTTCGGTCTGACGCAGTACAGCGATT
ATATCTACTGGACAGACTGGAATCTGCACAGCATTGAGCGGGCCGACAAGA
CTAGCGGCCGGAACCGCACCCTCATCCAGGGCCACCTGGACTTCGTGATGGAC
ATCCTGGTGTTCCACTCCTCCCGCCAGGATGGCCTCAATGACTGTATGCACAA
CAACGGGCAGTGTGGGCAGCTGTGCCTTGCCATCCCCGGCGGCCACCGCTGCGGC
TGCGCCTCACACTACACCCTGGACCCCAGCAGCCGCAACTGCAGCCCGCCCACC
ACCTTCTTGCTGTTCAGCCAGAAATCTGCCATCAGTCGGATGATCCCGGACGA
CCAGCACAGCCCGGATCTCATCCTGCCCCTGCATGGACTGAGGAACGTCAAAG
CCATCGACTATGACCCACTGGACAAGTTCATCTACTGGGTGGATGGGCGCCA
GAACATCAAGCGAGCCAAGGACGACGGGACCCAGCCCTTTGTTTTGACCTCT
CTGAGCCAAGGCCAAAACCCAGACAGGCAGCCCCACGACCTCAGCATCGACA
TCTACAGCCGGACACTGTTCTGGACGTGCGAGGCCACCAATACCATCAACGTC
CACAGGCTGAGCGGGGAAGCCATGGGGGTGGTGCTGCGTGGGGACCGCGACAA
GCCCAGGGCCATCGTCGTCAACGCGGAGCGAGGGTACCTGTACTTCACCAACA
TGCAGGACCGGGCAGCCAAGATCGAACGCGCAGCCCTGGACGGCACCGAGCGC
GAGGTCCTCTTCACCACCGGCCTCATCCGCCCTGTGGCCCTGGTGGTAGACAAC
ACACTGGGCAAGCTGTTCTGGGTGGACGCGGACCTGAAGCGCATTGAGAGCT
GTGACCTGTCAGGGGCCAACCGCCTGACCCTGGAGGACGCCAACATCGTGCAG
CCTCTGGGCCTGACCATCCTTGGCAAGCATCTCTACTGGATCGACCGCCAGCAG
CAGATGATCGAGCGTGTGGAGAAGACCACCGGGGACAAGCGGACTCGCATCC
AGGGCCGTGTCGCCCACCTCACTGGCATCCATGCAGTGGAGGAAGTCAGCCTG
GAGGAGTTCTCAGCCCACCCATGTGCCCGTGACAATGGTGGCTGCTCCCACAT
CTGTATTGCCAAGGGTGATGGGACACCACGGTGCTCATGCCCAGTCCACCTCG
TGCTCCTGCAGAACCTGCTGACCTGTGGAGAGCCGCCCACCTGCTCCCCGGACC
AGTTTGCATGTGCCACAGGGGAGATCGACTGTATCCCCGGGGCCTGGCGCTGT
GACGGCTTTCCCGAGTGCGATGACCAGAGCGACGAGGAGGGCTGCCCCGTGTG
CTCCGCCGCCCAGTTCCCCTGCGCGCGCGGGGTCAGTGTGTGGACCTGCGCCTGCGCT
GCGACGGCGAGGCAGACTGTCAGGACCGCTCAGACGAGGCGGACTGTGACGCC
ATCTGCCTGCCCAACCAGTTCCGGTGTGCGAGCGGCCAGTGTGTCCTCATCAA
ACAGCAGTGCGACTCCTTCCCCGACTGTATCGACGGCTCCGACGAGCTCATGT

Figure 11 (a), Ctd.

GTGAAATCACCAAGCCGCCCTCAGACGACAGCCCGGCCCACAGCAGTGCCATC
GGGCCCGTCATTGGCATCATCCTCTCTCTCTTCGTCATGGGTGGTGTCTATTTT
GTGTGCCAGCGCGTGGTGTGCCAGCGCTATGCGGGGGCCAACGGGCCCTTCCCGC
ACGAGTATGTCAGCGGGACCCCGCACGTGCCCCTCAATTTCATAGCCCCGGGCG
GTTCCCAGCATGGCCCCTTCACAGGCATCGCATGCGGAAAGTCCATGATGAGC
TCCGTGAGCCTGATGGGGGGCCGGGGCGGGGTGCCCCTCTACGACCGGAACCAC
GTCACAGGGGCCTCGTCCAGCAGCTCGTCCAGCACGAAGGCCACGCTGTACCCG
CCGATCCTGAACCCGCCGCCCTCCCCGGCCACGGACCCCTCCCTGTACAACATGG
ACATGTTCTACTCTTCAAACATTCCGGCCACTGTGAGACCGTACAGGCCCTAC
ATCATTCGAGGAATGGCGCCCCCGACGACGCCCTGCAGCACCGACGTGTGTGA
CAGCGACTACAGCGCCAGCCGCTGGAAGGCCAGCAAGTACTACCTGGATTTG
AACTCGGACTCAGACCCCTATCCACCCCCACCCACGCCCCACAGCCAGTACCTG
TCGGCGGAGGACAGCTGCCCGCCCTCGCCCGCCACCGAGAGGAGCTACTTCCAT
CTCTTCCCGCCCCCTCCGTCCCCCTGCACGGACTCATCCTGACCTCGGCCGGGCCA
CTCTGGCTTCTCTGTGCCCCTGTAAATAGTTTTAAATATGAACAAAGAAAA
AAATATATTTTATGATTTAAAAAATAAATATAATTGGGATTTTAAAA
ACATGAGAAATGTGAACTGTGATGGGGTGGGCAGGGCTGGGAGAACTTTGT
ACAGTGGAACAAATATTTATAAACTTAATTT

Figure I I(b)

ATGTACTGGACAGACTGGGGTGAGACGCCCCGGATTGAGCGGGCAGGGATGG
ATGGCAGCACCCGGAAGATCATTGTGGACTCGGACATTTACTGGCCCAATGG
ACTGACCATCGACCTGGAGGAGCAGAAGCTCTACTGGGCTGACGCCAAGCTC
AGCTTCATCCACCGTGCCAACCTGGACGGCTCGTTCCGGCAGAAGGTGGTGGA
GGGCAGCCTGACGCACCCCTTCGCCCTGACGCTCTCCGGGGACACTCTGTACTGG
ACAGACTGGCAGACCCGCTCCATCCATGCCTGCAACAAGCGCACTGGGGGGA
AGAGGAAGGAGATCCTGAGTGCCCTCTACTCACCCATGGACATCCAGGTGCT
GAGCCAGGAGCGGCAGCCTTTCTTCCACACTCGCTGTGAGGAGGACAATGGCG
GCTGCTCCCACCTGTGCCTGCTGTCCCCAAGCGAGCCTTTCTACACATGCGCCTG
CCCCACGGGTGTGCAGCTGCAGGACAACGGCAGGACGTGTAAGGCAGGAGCC
GAGGAGGTGCTGCTGCTGGCCCGGCGGACGGACCTACGGAGGATCTCGCTGGA
CACGCCGGACTTTACCGACATCGTGCTGCAGGTGGACGACATCCGGCACGCCA
TTGCCATCGACTACGACCCGCTAGAGGGCTATGTCTACTGGACAGATGACGA
GGTGCGGGCCATCCGCAGGGCGTACCTGGACGGGTCTGGGGCGCAGACGCTGGT
CAACACCGAGATCAACGACCCCGATGGCATCGCGGTCGACTGGGTGGCCCGAA
ACCTCTACTGGACCGACACGGGCACGGACCGCATCGAGGTGACGCGCCTCAAC
GGCACCTCCCGCAAGATCCTGGTGTCGGAGGACCTGGACGAGCCCCGAGCCATC
GCACTGCACCCCGTGATGGGCCTCATGTACTGGACAGACTGGGGAGAGAACCC
TAAAATCGAGTGTGCCAACTTGGATGGGCAGGAGCGGCGTGTGCTGGTCAAT
GCCTCCCTCGGGTGGCCCAACGGCCTGGCCCTGGACCTGCAGGAGGGGAAGCTC
TACTGGGGAGACGCCAAGACAGACAAGATCGAGGTGATCAATGTTGATGGG
ACGAAGAGGCGGACCCTCCTGGAGGACAAGCTCCCGCACATTTTCGGGTTCAC
GCTGCTGGGGGACTTCATCTACTGGACTGACTGGCAGCGCCGCAGCATCGAGC
GGGTGCACAAGGTCAAGGCCAGCCGGGACGTCATCATTGACCAGCTGCCCGAC
CTGATGGGGCTCAAAGCTGTGAATGTGGCCAAGGTCGTCGGAACCAACCCGT
GTGCGGACAGGAACGGGGGGTGCAGCCACCTGTGCTTCTTCACACCCCACGCA
ACCCGGTGTGGCTGCCCCATCGGCCTGGAGCTGCTGAGTGACATGAAGACCTG
CATCGTGCCTGAGGCCTTCTTGGTCTTCACCAGCAGAGCCGCCATCCACAGGA
TCTCCCTCGAGACCAATAACAACGACGTGGCCATCCCGCTCACGGGCGTCAAG
GAGGCCTCAGCCCTGGACTTTGATGTGTCCAACAACCACATCTACTGGACAG
ACGTCAGCCTGAAGACCATCAGCCGCGCCTTCATGAACGGGAGCTCGGTGGAG
CACGTGGTGGAGTTTGGCCTTGACTACCCCGAGGGCATGGCCGTTGACTGGAT
GGGCAAGAACCTCTACTGGGCCGACACTGGGACCAACAGAATCGAAGTGGCG
CGGCTGGACGGGCAGTTCCGGCAAGTCCTCGTGTGGAGGGACTTGGACAACCC
GAGGTCGCTGGCCCTGGATCCCACCAAGGGCTACATCTACTGGACCGAGTGGG
GCGGCAAGCCGAGGATCGTGCGGGCCTTCATGGACGGGACCAACTGCATGACG
CTGGTGGACAAGGTGGGCCGGGCCAACGACCTCACCATTGACTACGCTGACCA
GCGCCTCTACTGGACCGACCTGGACACCAACATGATCGAGTCGTCCAACATGC
TGGGTCAGGAGCGGGTCGTGATTGCCGACGATCTCCCCGCACCCGTTCGGTCTGA

Figure 11(b), ctd,

CGCAGTACAGCGATTATATCTACTGGACAGACTGGAATCTGCACAGCATTG
AGCGGGCCGACAAGACTAGCGGCCGGAACCGCACCCTCATCCAGGGCCACCTG
GACTTCGTGATGGACATCCTGGTGTTCCACTCCTCCCGCCAGGATGGCCTCAAT
GACTGTATGCACAACAACGGGCAGTGTGGGCAGCTGTGCCTTGCCATCCCCGG
CGGCCACCGCTGCGGCTGCGCCTCACACTACACCCTGGACCCCAGCAGCCGCAA
CTGCAGCCCGCCCACCACCTTCTTGCTGTTCAGCCAGAAATCTGCCATCAGTCG
GATGATCCCGGACGACCAGCACAGCCCGGATCTCATCCTGCCCCTGCATGGAC
TGAGGAACGTCAAAGCCATCGACTATGACCCACTGGACAAGTTCATCTACT
GGGTGGATGGGCGCCAGAACATCAAGCGAGCCAAGGACGACGGGACCCAGCC
CTTTGTTTTGACCTCTCTGAGCCAAGGCCAAAACCCAGACAGGCAGCCCCACG
ACCTCAGCATCGACATCTACAGCCGGACACTGTTCTGGACGTGCGAGGCCACC
AATACCATCAACGTCCACAGGCTGAGCGGGGAAGCCATGGGGGTGGTGCTGC
GTGGGGACCGCGACAAGCCCAGGGCCATCGTCGTCAACGCGGAGCGAGGGTAC
CTGTACTTCACCAACATGCAGGACCGGGCAGCCAAGATCGAACGCGCAGCCC
TGGACGGCACCGAGCGCGAGGTCCTCTTCACCACCGGCCTCATCCGCCCTGTGGC
CCTGGTGGTAGACAACACACTGGGCAAGCTGTTCTGGGTGGACGCGGACCTG
AAGCGCATTGAGAGCTGTGACCTGTCAGGGGCCAACCGCCTGACCCTGGAGG
ACGCCAACATCGTGCAGCCTCTGGGCCTGACCATCCTTGGCAAGCATCTCTAC
TGGATCGACCGCCAGCAGCAGATGATCGAGCGTGTGGAGAAGACCACCGGGG
ACAAGCGGACTCGCATCCAGGGCCGTGTCGCCCACCTCACTGGCATCCATGCA
GTGGAGGAAGTCAGCCTGGAGGAGTTCTCAGCCCACCCATGTGCCCGTGACA
ATGGTGGCTGCTCCCACATCTGTATTGCCAAGGGGTGATGGGACACCACGGTGC
TCATGCCCAGTCCACCTCGTGCTCCTGCAGAACCTGCTGACCTGTGGAGAGCCG
CCCACCTGCTCCCCGGACCAGTTTGCATGTGCCACAGGGGAGATCGACTGTAT
CCCCGGGGCCTGGCGCTGTGACGGCTTTCCCGAGTGCGATGACCAGAGCGACGA
GGAGGGCTGCCCCGTGTGCTCCGCCGCCCAGTTCCCCTGCGCGCGGGGTCAGTGT
GTGGACCTGCGCCTGCGCTGCGACGGCGAGGCAGACTGTCAGGACCGCTCAGAC
GAGGCGGACTGTGACGCCATCTGCCTGCCCAACCAGTTCCGGTGTGCGAGCGGC
GAGTGTGTCCTCATCAAACAGCAGTGCGACTCCTTCCCCGACTGTATCGACGG
CTCCGACGAGCTCATGTGTGAAATCACCAAGCCGCCCTCAGACGACAGCCCGG
CCCACAGCAGTGCCATCGGGCCCGTCATTGGCATCATCCTCTCTCTCTTCGTCA
TGGGTGGTGTCTATTTTGTGTGCCAGCGCGTGGTGTGCCAGCGCTATGCGGGGG
CCAACGGGCCCTTCCCGCACGAGTATGTCAGCGGGACCCCGCACGTGCCCCTCA
ATTTCATAGCCCCGGGCGGTTCCCAGCATGGCCCCTTCACAGGCATCGCATGCG
GAAAGTCCATGATGAGCTCCGTGAGCCTGATGGGGGGCCGGGGCGGGGTGCCC
CTCTACGACCGGAACCACGTCACAGGGGCCTCGTCCAGCAGCTCGTCCAGCACG
AAGGCCACGCTGTACCCGCCGATCCTGAACCCGCCGCCCTCCCCGGCCACGGACC
CCTCCCTGTACAACATGGACATGTTCTACTCTTCAAACATTCCGGCCACTGTG
AGACCGTACAGGCCCTACATCATTCGAGGAATGGCGCCCCCGACGACGCCCTG
CAGCACCGACGTGTGTGACAGCGACTACAGCGCCAGCCGCTGGAAGGCCAGCA

Figure 11(b), Ctd,

AGTACTACCTGGATTTGAACTCGGACTCAGACCCCTATCCACCCCCACCCACG
CCCCACAGCCAGTACCTGTCGGCGGAGGACAGCTGCCCCGCCCTCGCCCGCCACCG
AGAGGAGCTACTTCCATCTCTTCCCGCCCCCTCCGTCCCCCTGCACGGACTCATC
C

Figure 11(c)

MYWTDWGETPRIERAGMDGSTRKIIVDSDIYWPNGLTIDLEEQKLYWADAK
LSFIHRANLDGSFRQKVVEGSLTHPFALTLSGDTLYWTDWQTRSIHACNKRT
GGKRKEILSALYSPMDIQVLSQERQPFFHTRCEEDNGGCSHLCLLSPSEPFYTCA
CPTGVQLQDNGRTCKAGAEEVLLLARRTDLRRISLDTPDFTDIVLQVDDIRHA
IAIDYDPLEGYVYWTDDEVRAIRRAYLDGSGAQTLVNTEINDPDGIAVDWV
ARNLYWTDTGTDRIEVTRLNGTSRKILVSEDLDEPRAIALHPVMGLMYWTD
WGENPKIECANLDGQERRVLVNASLGWPNGLALDLQEGKLYWGDAKTDKIE
VINVDGTKRRTLLEDKLPHIFGFTLLGDFTYWTDWQRRSIERVHKVKASRDVI
IDQLPDLMGLKAVNVAKVVGTNPCADRNGGCSHLCFFTPHATRCGCPIGLEL
LSDMKTCIVPEAFLVFTSRAAIHRISLETNNNDVAIPLTGVKEASALDFDVS
NNHIYWTDVSLKTISRAFMNGSSVEHVVEFGLDYPEGMAVDWMGKNLYW
ADTGTNRIEVARLDGQFRQVLVWRDLDNPRSLALDPTKGYTYWTEWGGKPR
IVRAFMDGTNCMTLVDKVGRANDLTIDYADQRLYWTDLDTNMIESSNMLG
QERVVIADDLPHPFGLTQYSDYTYWTDWNLHSIERADKTSGRNRTLIQGHLDF
VMDILVFHSSRQDGLNDCMHNNGQCGQLCLAIPGGHRCGCASHYTLDPSSRNC
SPPTTFLLFSQKSAISRMIPDDQHSPDLILPLHGLRNVKAIDYDPLDKFTYWV
DGRQNIKRAKDDGTQPFVLTSLSQGQNPDRQPHDLSIDIYSRTLFWTCEATNTI
NVHRLSGEAMGVVLRGDRDKPRAIVVNAERGYLYFTNMQDRAAKIERAAL
DGTEREVLFTTGLIRPVALVVDNTLGKLFWVDADLKRIESCDLSGANRLTLE
DANTVQPLGLTILGKHLYWIDRQQQMIERVEKTTGDKRTRIQGRVAHLTGIH
AVEEVSLEEFSAHPCARDNGGCSHICIAKGDGTPRCSCPVHLVLLQNLLTCGE
PPTCSPDQFACATGEIDCIPGAWRCDGFPECDDQSDEEGCPVCSAAQFPCARGQ
CVDLRLRCDGEADCQDRSDEADCDAICLPNQFRCASGQCVLIKQQCDSFPDCIDG
SDELMCEITKPPSDDSPAHSSAIGPVIGIILSLFVMGGVYFVCQRVVCQRYAG
ANGPFPHEYVSGTPHVPLNFIAPGGSQHGPFTGIACGKSMMSSVSLMGGRGG
VPLYDRNHVTGASSSSSSSTKATLYPPILNPPPSPATDPSLYNMDMFYSSNIP
ATVRPYRPYIIRGMAPPTTPCSTDVCDSDYSASRWKASKYYLDLNSDSDPYP
PPPTPHSQYLSAEDSCPPSPATERSYFHLFPPPPSPCTDSS

Figure 12(a)

```
TAAATGGCTTGGCAAAGGGAGTTCATTCCTTTTAGCGCTTCCATCTTCTGCA
GTGAGAGGACACCGCATTCTTCTTCTCCAGAGGATGCAGCAGCAAGGCGCCA
TCTTGAAACCAGAGACCAAACCAACCAGCAACTTCGTCTTGAACTTCCCAGC
CTCCACAACTCCTCGCCGCTCCTGCTATTTGCCAACCGCCGGGACGTACGGCTG
GTGGACGCCGGCGGAGTCAAGCTGGAGTCCACCATCGTGGTCAGCGGCCTGGA
GGATGCGGCCGCAGTGGACTTCCAGTTTTCCAAGGGAGCCGTGTACTGGACAG
ACGTGAGCGAGGAGGCCATCAAGCAGACCTACCTGAACCAGACGGGGGCCGC
CGTGCAGAACGTGGTCATCTCCGGCCTGGTCTCTCCCGACGGCCTCGCCTGCGAC
TGGGTGGGCAAGAAGCTGTACTGGACGGACTCAGAGACCAACCGCATCGAGG
TGGCCAACCTCAATGGCACATCCCGGAAGGTGCTCTTCTGGCAGGACCTTGAC
CAGCCGAGGGCCATCGCCTTGGACCCCGCTCACGGGTACATGTACTGGACAGA
CTGGGGTGAGACGCCCCGGATTGAGCGGGCAGGGATGGATGGCAGCACCCGGA
AGATCATTGTGGACTCGGACATTTACTGGCCCAATGGACTGACCATCGACCT
GGAGGAGCAGAAGCTCTACTGGGCTGACGCCAAGCTCAGCTTCATCCACCGTG
CCAACCTGGACGGCTCGTTCCGGCAGAAGGTGGTGGAGGGCAGCCTGACGCAC
CCCTTCGCCCTGACGCTCTCCGGGGACACTCTGTACTGGACAGACTGGCAGACC
CGCTCCATCCATGCCTGCAACAAGCGCACTGGGGGGAAGAGGAAGGAGATCC
TGAGTGCCCTCTACTCACCCATGGACATCCAGGTGCTGAGCCAGGAGCGGCAG
CCTTTCTTCCACACTCGCTGTGAGGAGGACAATGGCGGCTGCTCCCACCTGTGC
CTGCTGTCCCCAAGCGAGCCTTTCTACACATGCGCCTGCCCCACGGGTGTGCAG
CTGCAGGACAACGGCAGGACGTGTAAGGCAGGAGCCGAGGAGGTGCTGCTGC
TGGCCCGGCGGACGGACCTACGGAGGATCTCGCTGGACACGCCGGACTTTACCG
ACATCGTGCTGCAGGTGGACGACATCCGGCACGCCATTGCCATCGACTACGAC
CCGCTAGAGGGCTATGTCTACTGGACAGATGACGAGGTGCGGGCCATCCGCAG
GGCGTACCTGGACGGGTCTGGGGCGCAGACGCTGGTCAACACCGAGATCAACG
ACCCCGATGGCATCGCGGTCGACTGGGTGGCCCGAAACCTCTACTGGACCGAC
ACGGGCACGGACCGCATCGAGGTGACGCGCCTCAACGGCACCTCCCGCAAGAT
CCTGGTGTCGGAGGACCTGGACGAGCCCCGAGCCATCGCACTGCACCCCGTGAT
GGGCCTCATGTACTGGACAGACTGGGGAGAGAACCCTAAAATCGAGTGTGCC
AACTTGGATGGGCAGGAGCGGCGTGTGCTGGTCAATGCCTCCCTCGGGTGGCCC
AACGGCCTGGCCCTGGACCTGCAGGAGGGGAAGCTCTACTGGGGAGACGCCA
AGACAGACAAGATCGAGGTGATCAATGTTGATGGGACGAAGAGGCGGACC
CTCCTGGAGGACAAGCTCCCGCACATTTTCGGGTTCACGCTGCTGGGGGACTTC
ATCTACTGGACTGACTGGCAGCGCCGCAGCATCGAGCGGGTGCACAAGGTCA
AGGCCAGCCGGGACGTCATCATTGACCAGCTGCCCGACCTGATGGGGCTCAAA
GCTGTGAATGTGGCCAAGGTCGTCGGAACCAACCCGTGTGCGGACAGGAACG
GGGGGTGCAGCCACCTGTGCTTCTTCACACCCCACGCAACCCGGTGTGGCTGCCC
CATCGGCCTGGAGCTGCTGAGTGACATGAAGACCTGCATCGTGCCTGAGGCCT
TCTTGGTCTTCACCAGCAGAGCCGCCATCCACAGGATCTCCCTCGAGACCAAT
```

Figure 12(a), Ctd.

AACAACGACGTGGCCATCCCGCTCACGGGCGTCAAGGAGGCCTCAGCCCTGGA
CTTTGATGTGTCCAACAACCACATCTACTGGACAGACGTCAGCCTGAAGACC
ATCAGCCGCGCCTTCATGAACGGGAGCTCGGTGGAGCACGTGGTGGAGTTTGG
CCTTGACTACCCCGAGGGCATGGCCGTTGACTGGATGGGCAAGAACCTCTACT
GGGCCGACACTGGGACCAACAGAATCGAAGTGGCGCGGCTGGACGGGCAGTT
CCGGCAAGTCCTCGTGTGGAGGGACTTGGACAACCCGAGGTCGCTGGCCCTGG
ATCCCACCAAGGGCTACATCTACTGGACCGAGTGGGGCGGCAAGCCGAGGAT
CGTGCGGGCCTTCATGGACGGGACCAACTGCATGACGCTGGTGGACAAGGTGG
GCCGGGCCAACGACCTCACCATTGACTACGCTGACCAGCGCCTCTACTGGACCG
ACCTGGACACCAACATGATCGAGTCGTCCAACATGCTGGGTCAGGAGCGGGT
CGTGATTGCCGACGATCTCCCGCACCCGTTCGGTCTGACGCAGTACAGCGATT
ATATCTACTGGACAGACTGGAATCTGCACAGCATTGAGCGGGCCGACAAGA
CTAGCGGCCGGAACCGCACCCTCATCCAGGGCCACCTGGACTTCGTGATGGAC
ATCCTGGTGTTCCACTCCTCCCGCCAGGATGGCCTCAATGACTGTATGCACAA
CAACGGGCAGTGTGGGCAGCTGTGCCTTGCCATCCCCGGCGGCCACCGCTGCGGC
TGCGCCTCACACTACACCCTGGACCCCAGCAGCCGCAACTGCAGCCCGCCCACC
ACCTTCTTGCTGTTCAGCCAGAAATCTGCCATCAGTCGGATGATCCCGGACGA
CCAGCACAGCCCGGATCTCATCCTGCCCCTGCATGGACTGAGGAACGTCAAAG
CCATCGACTATGACCCACTGGACAAGTTCATCTACTGGGTGGATGGGCGCCA
GAACATCAAGCGAGCCAAGGACGACGGGACCCAGCCCTTTGTTTTGACCTCT
CTGAGCCAAGGCCAAAACCCAGACAGGCAGCCCCACGACCTCAGCATCGACA
TCTACAGCCGGACACTGTTCTGGACGTGCGAGGCCACCAATACCATCAACGTC
CACAGGCTGAGCGGGGAAGCCATGGGGGTGGTGCTGCGTGGGGACCGCGACAA
GCCCAGGGCCATCGTCGTCAACGCGGAGCGAGGGTACCTGTACTTCACCAACA
TGCAGGACCGGGCAGCCAAGATCGAACGCGCAGCCCTGGACGGCACCGAGCGC
GAGGTCCTCTTCACCACCGGCCTCATCCGCCCTGTGGCCCTGGTGGTAGACAAC
ACACTGGGCAAGCTGTTCTGGGTGGACGCGGACCTGAAGCGCATTGAGAGCT
GTGACCTGTCAGGGGCCAACCGCCTGACCCTGGAGGACGCCAACATCGTGCAG
CCTCTGGGCCTGACCATCCTTGGCAAGCATCTCTACTGGATCGACCGCCAGCAG
CAGATGATCGAGCGTGTGGAGAAGACCACCGGGGACAAGCGGACTCGCATCC
AGGGCCGTGTCGCCCACCTCACTGGCATCCATGCAGTGGAGGAAGTCAGCCTG
GAGGAGTTCTCAGCCCACCCATGTGCCCGTGACAATGGTGGCTGCTCCCACAT
CTGTATTGCCAAGGGTGATGGGACACCACGGTGCTCATGCCCAGTCCACCTCG
TGCTCCTGCAGAACCTGCTGACCTGTGGAGAGCCGCCCACCTGCTCCCCGGACC
AGTTTGCATGTGCCACAGGGGAGATCGACTGTATCCCCGGGGCCTGGCGCTGT
GACGGCTTTCCCGAGTGCGATGACCAGAGCGACGAGGAGGGCTGCCCCGTGTG
CTCCGCCGCCCAGTTCCCCTGCGCGCGCGGGGTCAGTGTGTGGACCTGCGCCTGCGCT
GCGACGGCGAGGCAGACTGTCAGGACCGCTCAGACGAGGCGGACTGTGACGCC
ATCTGCCTGCCCAACCAGTTCCGGTGTGCGAGCGGCCAGTGTGTCCTCATCAA
ACAGCAGTGCGACTCCTTCCCCGACTGTATCGACGGCTCCGACGAGCTCATGT
GTGAAATCACCAAGCCGCCCTCAGACGACAGCCCGGCCCACAGCAGTGCCATC

Figure 12(a), Ctd,

GGGCCCGTCATTGGCATCATCCTCTCTCTCTTCGTCATGGGTGGTGTCTATTTT
GTGTGCCAGCGCGTGGTGTGCCAGCGCTATGCGGGGGCCAACGGGCCCTTCCCGC
ACGAGTATGTCAGCGGGACCCCGCACGTGCCCCTCAATTTCATAGCCCCGGGCG
GTTCCCAGCATGGCCCCTTCACAGGCATCGCATGCGGAAAGTCCATGATGAGC
TCCGTGAGCCTGATGGGGGGCCGGGGCGGGGTGCCCCTCTACGACCGGAACCAC
GTCACAGGGGCCTCGTCCAGCAGCTCGTCCAGCACGAAGGCCACGCTGTACCCG
CCGATCCTGAACCCGCCGCCCTCCCCGGCCACGGACCCCTCCCTGTACAACATGG
ACATGTTCTACTCTTCAAACATTCCGGCCACTGCGAGACCGTACAGGCCCTAC
ATCATTCGAGGAATGGCGCCCCCGACGACGCCCTGCAGCACCGACGTGTGTGA
CAGCGACTACAGCGCCAGCCGCTGGAAGGCCAGCAAGTACTACCTGGATTTG
AACTCGGACTCAGACCCCTATCCACCCCCACCCACGCCCCACAGCCAGTACCTG
TCGGCGGAGGACAGCTGCCCGCCCTCGCCCGCCACCGAGAGGAGCTACTTCCAT
CTCTTCCCGCCCCCTCCGTCCCCCTGCACGGACTCATCCTGACCTCGGCCGGGCCA
CTCTGGCTTCTCTGTGCCCCTGTAAATAGTTTTAAATATGAACAAAGAAAA
AAATATATTTTATGATTTAAAAAATAAATATAATTGGGATTTTAAAA
ACATGAGAAATGTGAACTGTGATGGGGTGGGCAGGGCTGGGAGAACTTTGT
ACAGTGGAACAAATATTTATAAACTTAATTTTGTAAAACAG

Figure 12(b)

TAAAATGGCTTGGCAAAGGGAGTTCATTCCTTTTAGCGCTTCCATCTTCTGC
AGTGAGAGGACACCGCATTCTTCTTCTCCAGAGGATGCAGCAGCAAGGCGCC
ATCTTGAAACCAGAGACCAAACCAACCAGCAACTTCGTCTTGAACTTCCCA
GCCTCCACAACT

Figure 12(c)

ATGGCTTGGCAAAGGGAGTTCATTCCTTTTAGCGCTTCCATCTTCTGCAGTGA
GAGGACACCGCATTCTTCTTCTCCAGAGGATGCAGCAGCAAGGCGCCATCTTG
AAACCAGAGACCAAACCAACCAGCAACTTCGTCTTGAACTTCCCAGCCTCCA
CAACTCCTCGCCGCTCCTGCTATTTGCCAACCGCCGGGACGTACGGCTGGTGGA
CGCCGGCGGAGTCAAGCTGGAGTCCACCATCGTGGTCAGCGGCCTGGAGGATG
CGGCCGCAGTGGACTTCCAGTTTTCCAAGGGAGCCGTGTACTGGACAGACGTG
AGCGAGGAGGCCATCAAGCAGACCTACCTGAACCAGACGGGGGCCGCCGTGC
AGAACGTGGTCATCTCCGGCCTGGTCTCTCCCGACGGCCTCGCCTGCGACTGGGT
GGGCAAGAAGCTGTACTGGACGGACTCAGAGACCAACCGCATCGAGGTGGCC
AACCTCAATGGCACATCCCGGAAGGTGCTCTTCTGGCAGGACCTTGACCAGCC
GAGGGCCATCGCCTTGGACCCCGCTCACGGGTACATGTACTGGACAGACTGGG
GTGAGACGCCCCGGATTGAGCGGGCAGGGATGGATGGCAGCACCCGGAAGAT
CATTGTGGACTCGGACATTTACTGGCCCAATGGACTGACCATCGACCTGGAG
GAGCAGAAGCTCTACTGGGCTGACGCCAAGCTCAGCTTCATCCACCGTGCCAA
CCTGGACGGCTCGTTCCGGCAGAAGGTGGTGGAGGGCAGCCTGACGCACCCCTT
CGCCCTGACGCTCTCCGGGGACACTCTGTACTGGACAGACTGGCAGACCCGCTC
CATCCATGCCTGCAACAAGCGCACTGGGGGGAAGAGGAAGGAGATCCTGAG
TGCCCTCTACTCACCCATGGACATCCAGGTGCTGAGCCAGGAGCGGCAGCCTT
TCTTCCACACTCGCTGTGAGGAGGACAATGGCGGCTGCTCCCACCTGTGCCTGC
TGTCCCCAAGCGAGCCTTTCTACACATGCGCCTGCCCCACGGGTGTGCAGCTGC
AGGACAACGGCAGGACGTGTAAGGCAGGAGCCGAGGAGGTGCTGCTGCTGGC
CCGGCGGACGGACCTACGGAGGATCTCGCTGGACACGCCGGACTTTACCGACA
TCGTGCTGCAGGTGGACGACATCCGGCACGCCATTGCCATCGACTACGACCCGC
TAGAGGGCTATGTCTACTGGACAGATGACGAGGTGCGGGCCATCCGCAGGGC
GTACCTGGACGGGTCTGGGGCGCAGACGCTGGTCAACACCGAGATCAACGACC
CCGATGGCATCGCGGTCGACTGGGTGGCCCGAAACCTCTACTGGACCGACACG
GGCACGGACCGCATCGAGGTGACGCGCCTCAACGGCACCTCCCGCAAGATCCT
GGTGTCGGAGGACCTGGACGAGCCCCGAGCCATCGCACTGCACCCCGTGATGGG
CCTCATGTACTGGACAGACTGGGGAGAGAACCCTAAAATCGAGTGTGCCAA
CTTGGATGGGCAGGAGCGGCGTGTGCTGGTCAATGCCTCCCTCGGGTGGCCCAA
CGGCCTGGCCCTGGACCTGCAGGAGGGGAAGCTCTACTGGGGAGACGCCAAGA
CAGACAAGATCGAGGTGATCAATGTTGATGGGACGAAGAGGCGGACCCTCC
TGGAGGACAAGCTCCCGCACATTTTCGGGTTCACGCTGCTGGGGGACTTCATC
TACTGGACTGACTGGCAGCGCCGCAGCATCGAGCGGGTGCACAAGGTCAAGG
CCAGCCGGGACGTCATCATTGACCAGCTGCCCGACCTGATGGGGCTCAAAGCT
GTGAATGTGGCCAAGGTCGTCGGAACCAACCCGTGTGCGGACAGGAACGGGG
GGTGCAGCCACCTGTGCTTCTTCACACCCCACGCAACCCGGTGTGGCTGCCCCA
TCGGCCTGGAGCTGCTGAGTGACATGAAGACCTGCATCGTGCCTGAGGCCTTC
TTGGTCTTCACCAGCAGAGCCGCCATCCACAGGATCTCCCTCGAGACCAATA

Figure 12(c), Ctd.

```
ACAACGACGTGGCCATCCCGCTCACGGGCGTCAAGGAGGCCTCAGCCCTGGAC
TTTGATGTGTCCAACAACCACATCTACTGGACAGACGTCAGCCTGAAGACC
ATCAGCCGCGCCTTCATGAACGGGAGCTCGGTGGAGCACGTGGTGGAGTTTGG
CCTTGACTACCCCGAGGGCATGGCCGTTGACTGGATGGGCAAGAACCTCTACT
GGGCCGACACTGGGACCAACAGAATCGAAGTGGCGCGGCTGGACGGGCAGTT
CCGGCAAGTCCTCGTGTGGAGGGACTTGGACAACCCGAGGTCGCTGGCCCTGG
ATCCCACCAAGGGCTACATCTACTGGACCGAGTGGGGCGGCAAGCCGAGGAT
CGTGCGGGCCTTCATGGACGGGACCAACTGCATGACGCTGGTGGACAAGGTGG
GCCGGGCCAACGACCTCACCATTGACTACGCTGACCAGCGCCTCTACTGGACCG
ACCTGGACACCAACATGATCGAGTCGTCCAACATGCTGGGTCAGGAGCGGGT
CGTGATTGCCGACGATCTCCCGCACCCGTTCGGTCTGACGCAGTACAGCGATT
ATATCTACTGGACAGACTGGAATCTGCACAGCATTGAGCGGGCCGACAAGA
CTAGCGGCCGGAACCGCACCCTCATCCAGGGCCACCTGGACTTCGTGATGGAC
ATCCTGGTGTTCCACTCCTCCCGCCAGGATGGCCTCAATGACTGTATGCACAA
CAACGGGCAGTGTGGGCAGCTGTGCCTTGCCATCCCCGGCGGCCACCGCTGCGGC
TGCGCCTCACACTACACCCTGGACCCCAGCAGCCGCAACTGCAGCCCGCCCACC
ACCTTCTTGCTGTTCAGCCAGAAATCTGCCATCAGTCGGATGATCCCGGACGA
CCAGCACAGCCCGGATCTCATCCTGCCCCTGCATGGACTGAGGAACGTCAAAG
CCATCGACTATGACCCACTGGACAAGTTCATCTACTGGGTGGATGGGCGCCA
GAACATCAAGCGAGCCAAGGACGACGGGACCCAGCCCTTTGTTTTGACCTCT
CTGAGCCAAGGCCAAAACCCAGACAGGCAGCCCCACGACCTCAGCATCGACA
TCTACAGCCGGACACTGTTCTGGACGTGCGAGGCCACCAATACCATCAACGTC
CACAGGCTGAGCGGGGAAGCCATGGGGGTGGTGCTGCGTGGGGACCGCGACAA
GCCCAGGGCCATCGTCGTCAACGCGGAGCGAGGGTACCTGTACTTCACCAACA
TGCAGGACCGGGCAGCCAAGATCGAACGCGCAGCCCTGGACGGCACCGAGCGC
GAGGTCCTCTTCACCACCGGCCTCATCCGCCCTGTGGCCCTGGTGGTAGACAAC
ACACTGGGCAAGCTGTTCTGGGTGGACGCGGACCTGAAGCGCATTGAGAGCT
GTGACCTGTCAGGGGCCAACCGCCTGACCCTGGAGGACGCCAACATCGTGCAG
CCTCTGGGCCTGACCATCCTTGGCAAGCATCTCTACTGGATCGACCGCCAGCAG
CAGATGATCGAGCGTGTGGAGAAGACCACCGGGGACAAGCGGACTCGCATCC
AGGGCCGTGTCGCCCACCTCACTGGCATCCATGCAGTGGAGGAAGTCAGCCTG
GAGGAGTTCTCAGCCCACCCATGTGCCCGTGACAATGGTGGCTGCTCCCACAT
CTGTATTGCCAAGGGTGATGGGACACCACGGTGCTCATGCCCAGTCCACCTCG
TGCTCCTGCAGAACCTGCTGACCTGTGGAGAGCCGCCCACCTGCTCCCCGGACC
AGTTTGCATGTGCCACAGGGGAGATCGACTGTATCCCCGGGGCCTGGCGCTGT
GACGGCTTTCCCGAGTGCGATGACCAGAGCGACGAGGAGGGCTGCCCCGTGTG
CTCCGCCGCCCAGTTCCCCTGCGCGCGGGGGTCAGTGTGTGGACCTGCGCCTGCGCT
GCGACGGCGAGGCAGACTGTCAGGACCGCTCAGACGAGGCGGACTGTGACGCC
ATCTGCCTGCCCAACCAGTTCCGGTGTGCGAGCGGCCAGTGTGTCCTCATCAA
ACAGCAGTGCGACTCCTTCCCCGACTGTATCGACGGCTCCGACGAGCTCATGT
GTGAAATCACCAAGCCGCCCTCAGACGACAGCCCGGCCCACAGCAGTGCCATC
```

Figure 12(C), Ctd.

GGGCCCGTCATTGGCATCATCCTCTCTCTCTCTTCGTCATGGGTGGTGTCTATTTT
GTGTGCCAGCGCGTGGTGTGCCAGCGCTATGCGGGGGGCCAACGGGCCCTTCCCGC
ACGAGTATGTCAGCGGGACCCCGCACGTGCCCCTCAATTTCATAGCCCCGGGCG
GTTCCCAGCATGGCCCCTTCACAGGCATCGCATGCGGAAAGTCCATGATGAGC
TCCGTGAGCCTGATGGGGGGCCGGGGCGGGGTGCCCCTCTACGACCGGAACCAC
GTCACAGGGGCCTCGTCCAGCAGCTCGTCCAGCACGAAGGCCACGCTGTACCCG
CCGATCCTGAACCCGCCGCCCTCCCCGGCCACGGACCCCTCCCTGTACAACATGG
ACATGTTCTACTCTTCAAAACATTCCGGCCACTGCGAGACCGTACAGGCCCTAC
ATCATTCGAGGAATGGCGCCCCCGACGACGCCCTGCAGCACCGACGTGTGTGA
CAGCGACTACAGCGCCAGCCGCTGGAAGGCCAGCAAGTACTACCTGGATTTG
AACTCGGACTCAGACCCCTATCCACCCCCACCCACGCCCCACAGCCAGTACCTG
TCGGCGGAGGACAGCTGCCCGCCCTCGCCCGCCACCGAGAGGAGCTACTTCCAT
CTCTTCCCGCCCCCTCCGTCCCCCTGCACGGACTCATCC

Figure   12(d)

MAWQREFIPFSASIFCSERTPHSSSPEDAAARRHLETRDQTNQQLRLELPSLH
NSSPLLLFANRRDVRLVDAGGVKLESTIVVSGLEDAAAVDFQFSKGAVYWT
DVSEEAIKQTYLNQTGAAVQNVVISGLVSPDGLACDWVGKKLYWTDSETNR
IEVANLNGTSRKVLFWQDLDQPRAIALDPAHGYMYWTDWGETPRIERAGM
DGSTRKIIVDSDIYWPNGLTIDLEEQKLYWADAKLSFIHRANLDGSFRQKVVE
GSLTHPFALTLSGDTLYWTDWQTRSIHACNKRTGGKRKEILSALYSPMDIQV
LSQERQPFFHTRCEEDNGGCSHLCLLSPSEPFYTCACPTGVQLQDNGRTCKAGA
EEVLLLARRTDLRRISLDTPDFTDIVLQVDDIRHAIAIDYDPLEGYVYWTDDE
VRAIRRAYLDGSGAQTLVNTEINDPDGIAVDWVARNLYWTDTGTDRIEVTR
LNGTSRKILVSEDLDEPRAIALHPVMGLMYWTDWGENPKIECANLDGQERR
VLVNASLGWPNGLALDLQEGKLYWGDAKTDKIEVINVDGTKRRTLLEDKLP
HIFGFTLLGDFIYWTDWQRRSIERVHKVKASRDVIIDQLPDLMGLKAVNVA
KVVGTNPCADRNGGCSHLCFFTPHATRCGCPIGLELLSDMKTCIVPEAFLVFT
SRAAIHRISLETNNNDVAIPLTGVKEASALDFDVSNNHIYWTDVSLKTISRA
FMNGSSVEHVVEFGLDYPEGMAVDWMGKNLYWADTGTNRIEVARLDGQFR
QVLVWRDLDNPRSLALDPTKGYIYWTEWGGKPRIVRAFMDGTNCMTLVDK
VGRANDLTIDYADQRLYWTDLDTNMIESSNMLGQERVVIADDLPHPFGLTQ
YSDYIYWTDWNLHSIERADKTSGRNRTLIQGHLDFVMDILVFHSSRQDGLND
CMHNNGQCGQLCLAIPGGHRCGCASHYTLDPSSRNCSPPTTFLLFSQKSAISR
MIPDDQHSPDLILPLHGLRNVKAIDYDPLDKFIYWVDGRQNIKRAKDDGTQPF
VLTSLSQGQNPDRQPHDLSIDIYSRTLFWTCEATNTINVHRLSGEAMGVVLR
GDRDKPRAIVVNAERGYLYFTNMQDRAAKIERAALDGTEREVLFTTGLIRPV
ALVVDNTLGKLFWVDADLKRIESCDLSGANRLTLEDANTVQPLGLTILGKHL
YWIDRQQQMIERVEKTTGDKRTRIQGRVAHLTGIHAVEEVSLEEFSAHPCAR
DNGGCSHICIAKGDGTPRCSCPVHLVLLQNLLTCGEPPTCSPDQFACATGEIDCI
PGAWRCDGFPECDDQSDEEGCPVCSAAQFPCARGQCVDLRLRCDGEADCQDRS
DEADCDAICLPNQFRCASGQCVLIKQQCDSFPDCIDGSDELMCEITKPPSDDSPA
HSSAIGPVIGIILSLFVMGGVYFVCQRVVCQRYAGANGPFPHEYVSGTPHVP
LNFIAPGGSQHGPFTGIACGKSMMSSVSLMGGRGGVPLYDRNHVTGASSSSS
SSTKATLYPPILNPPPSPATDPSLYNMDMFYSSNIPATARPYRPYIIRGMAP
PTTPCSTDVCDSDYSASRWKASKYYLDLNSDSDPYPPPPTPHSQYLSAEDSCP
PSPATERSYFHLFPPPPSPCTDSS

Figure   12(e)

TATAAAATGGCTTGGCAAAGGGAGTTCATTCCTTTTAGCGCTTCCATCTTCT
GCAGTGAGAGGACACCGCATTCTTCTTCTCCAGAGGATG

Figure 13

TAAGAGTATAAAGGGCTCCTGAGACCAAAAAGGTTGAGAACCAGTGCTTT
AAAGCTTGATGTTTCTCAGGGTTTCATCCTTTGTGGATTAATGCCCATTATA
AAATGGCTTGGCAAAGGGAGTTCATTCCTTTTAGCGCTTCCATCTTCTGCAG
TGAGAGGACACCGCATTCTTCTTCTCCAGAGGATGCAGCAGCAAGGCGCCAT
CTTGAAACCAGAGACCAAACCAACCAGCAACTTCGTCTTGAACTTCCCAGCC
TCCACAACTCAGCAGTCTGTGCAGGACCCTGTGAGCAGAGCCGCAGCCTCGCC
GCTCCTGCTATTTGCCAACCGCCGGGACGTACGGCTGGTGGACGCCGGCGGAGT
CAAGCTGGAGTCCACCATCGTGGTCAGCGGCCTGGAGGATGCGGCCGCAGTGG
ACTTCCAGTTTTCCAAGGGAGCCGTGTACTGGACAGACGTGAGCGAGGAGGC
CATCAAGCAGACCTACCTGAACCAGACGGGGGGCCGCCGTGCAGAACGTGGTC
ATCTCCGGCCTGGTCTCTCCCGACGGCCTCGCCTGCGACTGGGTGGGCAAGAAG
CTGTACTGGACGGACTCAGAGACCAACCGCATCGAGGTGGCCAACCTCAATG
GCACATCCCGGAAGGTGCTCTTCTGGCAGGACCTTGACCAGCCGAGGGCCATC
GCCTTGGACCCCGCTCACGGGTACATGTACTGGACAGACTGGGGTGAGACGCC
CCGGATTGAGCGGGCAGGGATGGATGGCAGCACCCGGAAGATCATTGTGGAC
TCGGACATTTACTGGCCCAATGGACTGACCATCGACCTGGAGGAGCAGAAGC
TCTACTGGGCTGACGCCAAGCTCAGCTTCATCCACCGTGCCAACCTGGACGGCT
CGTTCCGGCAGAAGGTGGTGGAGGGCAGCCTGACGCACCCCTTCGCCCTGACGC
TCTCCGGGGACACTCTGTACTGGACAGACTGGCAGACCCGCTCCATCCATGCCT
GCAACAAGCGCACTGGGGGGAAGAGGAAGGAGATCCTGAGTGCCCTCTACTC
ACCCATGGACATCCAGGTGCTGAGCCAGGAGCGGCAGCCTTTCTTCCACACTC
GCTGTGAGGAGGACAATGGCGGCTGCTCCCACCTGTGCCTGCTGTCCCCAAGCG
AGCCTTTCTACACATGCGCCTGCCCCACGGGTGTGCAGCTGCAGGACAACGGC
AGGACGTGTAAGGCAGGAGCCGAGGAGGTGCTGCTGCTGGCCCGGCGGACGGA
CCTACGGAGGATCTCGCTGGACACGCCGGACTTTACCGACATCGTGCTGCAGG
TGGACGACATCCGGCACGCCATTGCCATCGACTACGACCCGCTAGAGGGCTAT
GTCTACTGGACAGATGACGAGGTGCGGGCCATCCGCAGGGCGTACCTGGACGG
GTCTGGGGCGCAGACGCTGGTCAACACCGAGATCAACGACCCCGATGGCATCG
CGGTCGACTGGGTGGCCCGAAACCTCTACTGGACCGACACGGGCACGGACCGC
ATCGAGGTGACGCGCCTCAACGGCACCTCCCGCAAGATCCTGGTGTCGGAGGA
CCTGGACGAGCCCCGAGCCATCGCACTGCACCCCGTGATGGGCCTCATGTACTG
GACAGACTGGGGAGAGAACCCTAAAAATCGAGTGTGCCAACTTGGATGGGCA
GGAGCGGCGTGTGCTGGTCAATGCCTCCCTCGGGTGGCCCAACGGCCTGGCCCTG
GACCTGCAGGAGGGGAAGCTCTACTGGGGAGACGCCAAGACAGACAAGATC
GAGGTGATCAATGTTGATGGGACGAAGAGGCGGACCCTCCTGGAGGACAAG
CTCCCGCACATTTTCGGGTTCACGCTGCTGGGGGACTTCATCTACTGGACTGAC
TGGCAGCGCCGCAGCATCGAGCGGGTGCACAAGGTCAAGGCCAGCCGGGACGT
CATCATTGACCAGCTGCCCGACCTGATGGGGCTCAAAGCTGTGAATGTGGCC
AAGGTCGTCGGAACCAACCCGTGTGCGGACAGGAACGGGGGGGTGCAGCCACC

Figure 13, ctd

TGTGCTTCTTCACACCCCACGCAACCCGGTGTGGCTGCCCCATCGGCCTGGAGCT
GCTGAGTGACATGAAGACCTGCATCGTGCCTGAGGCCTTCTTGGTCTTCACCA
GCAGAGCCGCCATCCACAGGATCTCCCTCGAGACCAATAACAACGACGTGGC
CATCCCGCTCACGGGCGTCAAGGAGGCCTCAGCCCTGGACTTTGATGTGTCCA
ACAACCACATCTACTGGACAGACGTCAGCCTGAAGACCATCAGCCGCGCCTT
CATGAACGGGAGCTCGGTGGAGCACGTGGTGGAGTTTGGCCTTGACTACCCCG
AGGGCATGGCCGTTGACTGGATGGGCAAGAACCTCTACTGGGCCGACACTGG
GACCAACAGAATCGAAGTGGCGCGGCTGGACGGGCAGTTCCGGCAAGTCCTC
GTGTGGAGGGACTTGGACAACCCGAGGTCGCTGGCCCTGGATCCCACCAAGGG
CTACATCTACTGGACCGAGTGGGGCGGCAAGCCGAGGATCGTGCGGGCCTTCA
TGGACGGGACCAACTGCATGACGCTGGTGGACAAGGTGGGCCGGGCCAACGA
CCTCACCATTGACTACGCTGACCAGCGCCTCTACTGGACCGACCTGGACACCA
ACATGATCGAGTCGTCCAACATGCTGGGTCAGGAGCGGGTCGTGATTGCCGA
CGATCTCCCGCACCCGTTCGGTCTGACGCAGTACAGCGATTATATCTACTGGA
CAGACTGGAATCTGCACAGCATTGAGCGGGCCGACAAGACTAGCGGCCGGAA
CCGCACCCTCATCCAGGGCCACCTGGACTTCGTGATGGACATCCTGGTGTTCCA
CTCCTCCCGCCAGGATGGCCTCAATGACTGTATGCACAACAACGGGCAGTGTG
GGCAGCTGTGCCTTGCCATCCCCGGCGGCCACCGCTGCGGCTGCGCCTCACACTA
CACCCTGGACCCCAGCAGCCGCAACTGCAGCCCGCCCACCACCTTCTTGCTGTTC
AGCCAGAAATCTGCCATCAGTCGGATGATCCCGGACGACCAGCACAGCCCGG
ATCTCATCCTGCCCCTGCATGGACTGAGGAACGTCAAAGCCATCGACTATGA
CCCACTGGACAAGTTCATCTACTGGGTGGATGGGCGCCAGAACATCAAGCGA
GCCAAGGACGACGGGACCCAGCCCTTTGTTTTGACCTCTCTGAGCCAAGGCCA
AAACCCAGACAGGCAGCCCCACGACCTCAGCATCGACATCTACAGCCGGACA
CTGTTCTGGACGTGCGAGGCCACCAATACCATCAACGTCCACAGGCTGAGCGG
GGAAGCCATGGGGGTGGTGCTGCGTGGGGACCGCGACAAGCCCAGGGCCATCG
TCGTCAACGCGGAGCGAGGGTACCTGTACTTCACCAACATGCAGGACCGGGC
AGCCAAGATCGAACGCGCAGCCCTGGACGGCACCGAGCGCGAGGTCCTCTTCA
CCACCGGCCTCATCCGCCCTGTGGCCCTGGTGGTAGACAACACACTGGGCAAGC
TGTTCTGGGTGGACGCGGACCTGAAGCGCATTGAGAGCTGTGACCTGTCAGGG
GCCAACCGCCTGACCCTGGAGGACGCCAACATCGTGCAGCCTCTGGGCCTGACC
ATCCTTGGCAAGCATCTCTACTGGATCGACCGCCAGCAGCAGATGATCGAGC
GTGTGGAGAAGACCACCGGGGACAAGCGGACTCGCATCCAGGGCCGTGTCGCC
CACCTCACTGGCATCCATGCAGTGGAGGAAGTCAGCCTGGAGGAGTTCTCAG
CCCACCCATGTGCCCGTGACAATGGTGGCTGCTCCCACATCTGTATTGCCAAG
GGTGATGGGACACCACGGTGCTCATGCCCAGTCCACCTCGTGCTCCTGCAGAA
CCTGCTGACCTGTGGAGAGCCGCCCACCTGCTCCCCGGACCAGTTTGCATGTGCC
ACAGGGGAGATCGACTGTATCCCCGGGGCCTGGCGCTGTGACGGCTTTCCCGAG
TGCGATGACCAGAGCGACGAGGAGGGCTGCCCCGTGTGCTCCGCCGCCCAGTTC
CCCTGCGCGCGGGGTCAGTGTGTGGACCTGCGCCTGCGCTGCGACGGCGAGGCAG
ACTGTCAGGACCGCTCAGACGAGGCGGACTGTGACGCCATCTGCCTGCCCAAC

Figure 13, ctd.

CAGTTCCGGTGTGCGAGCGGCCAGTGTGTCCTCATCAAACAGCAGTGCGACTC
CTTCCCCGACTGTATCGACGGCTCCGACGAGCTCATGTGTGAAATCACCAAGC
CGCCCTCAGACGACAGCCCGGCCCACAGCAGTGCCATCGGGCCCGTCATTGGCA
TCATCCTCTCTCTCTTCGTCATGGGTGGTGTCTATTTTGTGTGCCAGCGCGTGG
TGTGCCAGCGCTATGCGGGGGGCCAACGGGCCCTTCCCGCACGAGTATGTCAGCG
GGACCCCGCACGTGCCCCTCAATTTCATAGCCCCGGGCGGTTCCCAGCATGGCCC
CTTCACAGGCATCGCATGCGGAAAGTCCATGATGAGCTCCGTGAGCCTGATG
GGGGGCCGGGGCGGGGTGCCCCTCTACGACCGGAACCACGTCACAGGGGCCTCG
TCCAGCAGCTCGTCCAGCACGAAGGCCACGCTGTACCCGCGGATCCTGAACCCG
CCGCCCTCCCCGGCCACGGACCCCTCCCTGTACAACATGGACATGTTCTACTCT
TCAAACATTCCGGCCACTGCGAGACCGTACAGGCCCTACATCATTCGAGGAA
TGGCGCCCCCGACGACGCCCTGCAGCACCGACGTGTGTGACAGCGACTACAGCG
CCAGCCGCTGGAAGGCCAGCAAGTACTACCTGGATTTGAACTCGGACTCAGA
CCCCTATCCACCCCCACCCACGCCCCACAGCCAGTACCTGTCGGCGGAGGACAG
CTGCCCGCCCTCGCCCGCCACCGAGAGGAGCTACTTCCATCTCTTCCCGCCCCCTC
CGTCCCCCTGCACGGACTCATCCTGACCTCGGCCGGGCCACTCTGGCTTCTCTGT
GCCCCTGTAAATAGTTTTAAATATGAACAAAGAAAAAATATATTTTA
TGATTTAAAAAATAAATATAATTGGGATTTTAAAAACATGAGAAATGT
GAACTGTGATGGGGTGGGCAGGGCTGGGAGAACTTTGTACAGTGGAACAAA
TATTTATAAACTTAATTTTGTAAAACAG

Figure 14

GGCTGGTCTTGAACTCCTGGCCTGAGATGATCCTCTCTCCTCGGAAAGTGCTG
GGATTATAGCCTCGCCGCTCCTGCTATTTGCCAACCGCCGGGACGTACGGCTGG
TGGACGCCGGCGGAGTCAAGCTGGAGTCCACCATCGTGGTCAGCGGCCTGGAG
GATGCGGCCGCAGTGGACTTCCAGTTTTCCAAGGGAGCCGTGTACTGGACAG
ACGGAGCGAGGAGGCCATCAAGCAGACCTACCTGAACCAGACGGGGGCCGCC
GTGCAGAACGTGGTCATCTCCGGCCTGGTCTCTCCCGACGGCCTCGCCTGCGACT
GGGTGGGCAAGAAGCTGTACTGGACGGACTCAGAGACCAACCGCATCGAGGT
GGCCAACCTCAATGGCACATCCCGGAAGGTGCTCTTCTGGCAGGACCTTGACC
AGCCGAGGGCCATCGCCTTGGACCCCGCTCACGGGTACATGTACTGGACAGAC
TGGGGTGAGACGCCCCGGATTGAGCGGGCAGGGATGGATGGCAGCACCCGGA
AGATCATTGTGGACTCGGACATTTACTGGCCCAATGGACTGACCATCGACCT
GGAGGAGCAGAAGCTCTACTGGGCTGACGCCAAGCTCAGCTTCATCCACCGTG
CCAACCTGGACGGCTCGTTCCGGCAGAAGGTGGTGGAGGGCAGCCTGACGCAC
CCCTTCGCCCTGACGCTCTCCGGGGACACTCTGTACTGGACAGACTGGCAGACC
CGCTCCATCCATGCCTGCAACAAGCGCACTGGGGGGGAAGAGGAAGGAGATCC
TGAGTGCCCTCTACTCACCCATGGACATCCAGGTGCTGAGCCAGGAGCGGCAG
CCTTTCTTCCACACTCGCTGTGAGGAGGACAATGGCGGCTGCTCCCACCTGTGC
CTGCTGTCCCCAAGCGAGCCTTTCTACACATGCGCCTGCCCCACGGGTGTGCAG
CTGCAGGACAACGGCAGGACGTGTAAGGCAGGAGCCGAGGAGGTGCTGCTGC
TGGCCCGGCGGACGGACCTACGGAGGATCTCGCTGGACACGCCGGACTTTACCG
ACATCGTGCTGCAGGTGGACGACATCCGGCACGCCATTGCCATCGACTACGAC
CCGCTAGAGGGCTATGTCTACTGGACAGATGACGAGGTGCGGGCCATCCGCAG
GGCGTACCTGGACGGGTCTGGGGCGCAGACGCTGGTCAACACCGAGATCAACG
ACCCCGATGGCATCGCGGTCGACTGGGTGGCCCGAAACCTCTACTGGACCGAC
ACGGGCACGGACCGCATCGAGGTGACGCGCCTCAACGGCACCTCCCGCAAGAT
CCTGGTGTCGGAGGACCTGGACGAGCCCCGAGCCATCGCACTGCACCCCGTGAT
GGGCCTCATGTACTGGACAGACTGGGGAGAGAACCCTAAAATCGAGTGTGCC
AACTTGGATGGGCAGGAGCGGCGTGTGCTGGTCAATGCCTCCCTCGGGTGGCCC
AACGGCCTGGCCCTGGACCTGCAGGAGGGGAAGCTCTACTGGGGAGACGCCA
AGACAGACAAGATCGAGGTGATCAATGTTGATGGGACGAAGAGGCGGACC
CTCCTGGAGGACAAGCTCCCGCACATTTTCGGGTTCACGCTGCTGGGGGACTTC
ATCTACTGGACTGACTGGCAGCGCCGCAGCATCGAGCGGGTGCACAAGGTCA
AGGCCAGCCGGGACGTCATCATTGACCAGCTGCCCGACCTGATGGGGCTCAAA
GCTGTGAATGTGGCCAAGGTCGTCGGAACCAACCCGTGTGCGGACAGGAACG
GGGGGTGCAGCCACCTGTGCTTCTTCACACCCCACGCAACCCGGTGTGGCTGCCC
CATCGGCCT
GGAGCTGCTGAGTGACATGAAGACCTGCATCGTGCCTGAGGCCTTCTTGGTCT
TCACCAGCAGAGCCGCCATCCACAGGATCTCCCTCGAGACCAATAACAACGA
CGTGGCCATCCCGCTCACGGGCGTCAAGGAGGCCTCAGCCCTGGACTTTGATGT

Figure 14, ctd,

GTCCAACAACCACATCTACTGGACAGACGTCAGCCTGAAGACCATCAGCCGC
GCCTTCATGAACGGGAGCTCGGTGGAGCACGTGGTGGAGTTTGGCCTTGACTA
CCCCGAGGGCATGGCCGTTGACTGGATGGGCAAGAACCTCTACTGGGCCGACA
CTGGGACCAACAGAATCGAAGTGGCGCGGCTGGACGGGCAGTTCCGGCAAGT
CCTCGTGTGGAGGGACTTGGACAACCCGAGGTCGCTGGCCCTGGATCCCACCA
AGGGCTACATCTACTGGACCGAGTGGGGCGGCAAGCCGAGGATCGTGCGGGCC
TTCATGGACGGGACCAACTGCATGA
CGCTGGTGGACAAGGTGGGCCGGGCCAACGACCTCACCATTGACTACGCTGAC
CAGCGCCTCTACTGGACCGACCTGGACACCAACATGATCGAGTCGTCCAACA
TGCTGGGTCAGGAGCGGGTCGTGATTGCCGACGATCTCCCGCACCCGTTCGGTC
TGACGCAGTACAGCGATTATATCTACTGGACAGACTGGAATCTGCACAGCA
TTGAGCGGGCCGACAAGACTAGCGGCCGGAACCGCACCCTCATCCAGGGCCAC
CTGGACTTCGTGATGGACATCCTGGTGTTCCACTCCTCCCGCCAGGATGGCCTC
AATGACTGTATGCACAACAACGGGCAGTGTGGGCAGCTGTGCCTTGCCATCC
CCGGCGGCCACCGCTGCGGCTGCGCCTCACACTACACCCTGGACCCCAGCAGCCG
CAACTGCAGCCCGCCCACCACCTTCTTGCTGTTCAGCCAGAAATCTGCCATCA
GTCGGATGATCCCGGACGACCAGCACAGCCCGGATCTCATCCTGCCCCTGCATG
GACTGAGGAACGTCAAAGCCATCGACTATGACCCACTGGACAAGTTCATCT
ACTGGGTGGATGGGCGCCAGAACATCAAGCGAGCCAAGGACGACGGGACCCA
GCCCTTTGTTTTGACCTCTCTGAGCCAAGGCCAAAACCCAGACAGGCAGCCCC
ACGACCTCAGCATCGACATCTACAGCCGGACACTGTTCTGGACGTGCGAGGCC
ACCAATACCATCAACGTCCACAGGCTGAGCGGGGAAGCCATGGGGGTGGTGC
TGCGTGGGGACCGCGACAAGCCCAGGGCCATCGTCGTCAACGCGGAGCGAGGG
TACCTGTACTTCACCAACATGCAGGACCGGGCAGCCAAGATCGAACGCGCAG
CCCTGGACGGCACCGAGCGCGAGGTCCTCTTCACCACCGGCCTCATCCGCCCTGT
GGCCCTGGTGGTAGACAACACACTGGGCAAGCTGTTCTGGGTGGACGC
GGACCTGAAGCGCATTGAGAGCTGTGACCTGTCAGGGGCCAACCGCCTGACCC
TGGAGGACGCCAACATCGTGCAGCCTCTGGGCCTGACCATCCTTGGCAAGCAT
CTCTACTGGATCGACCGCCAGCAGCAGATGATCGAGCGTGTGGAGAAGACCA
CCGGGGACAAGCGGACTCGCATCCAGGGCCGTGTCGCCCACCTCACTGGCATCC
ATGCAGTGGAGGAAGTCAGCCTGGAGGAGTTCTCAGCCCACCCATGTGCCCGT
GACAATGGTGGCTGCTCCCACATCTGTATTGCCAAGGGTGATGGGACACCAC
GGTGCTCATGCCCAGTCCACCTCGTGCTCCTGCAGAACCTGCTGACCTGTGGAG
AGCCGCCCACCTGCTCCCCGGACCAGTTTGCATGTGCCACAGGGGAGATCGACT
GTATCCCCGGGGCCTGGCGCTGTGACGGCTTTCCCGAGTGCGATGACCAGAGCG
ACGAGGAGGGCTGCCCCGTGTGCTCCGCCGCCCAGTTCCCCTGCGCGCGGGGTCA
GTGTGTGGACCTGCGCCTGCGCTGCGACGGCGAGGCAGACTGTCAGGACCGCTC
AGACGAGGCGGACTGTGACGCCATCTGCCTGCCCAACCAGTTCCGGTGTGCGA
GCGGCCAGTGTGTCCTCATCAAACAGCAGTGCGACTCCTTCCCCGACTGTATC
GACGGCTCCGACGAGCTCATGTGTGTGAAATCACCAAGCCGCCCTCAGACGACA
GCCCCGGCCCACAGCAGTGCCATCGGGCCCCGTCATTGGCATCATCCTCTCTCTCTT

Figure 14, ctd.

CGTCATGGGTGGTGTCTATTTTGTGTGCCAGCGCGTGGTGTGCCAGCGCTATGC
GGGGGCCAACGGGCCCTTCCCGCACGAGTATGTCAGCGGGACCCCGCACGTGCC
CCTCAATTTCATAGCCCCGGGCGGTTCCCAGCATGGCCCCTTCACAGGCATCGC
ATGCGGAAAGTCCATGATGAGCTCCGTGAGCCTGATGGGGGGCCGGGGCGGGG
TGCCCCTCTACGACCGGAACCACGTCACAGGGGCCTCGTCCAGCAGCTCGTCCA
GCACGAAGGCCACGCTGTACCCGCCGATCCTGAACCCGCCGCCCTCCCCGGCCAC
GGACCCCTCCCTGTACAACATGGACATGTTCTACTCTTCAAACATTCCGGCCA
CTGTGAGACCGTACAGGCCCTACATCATTCGAGGAATGGCGCCCCCGACGACG
CCCTGCAGCACCGACGTGTGTGACAGCGACTACAGCGCCAGCCGCTGGAAGGC
CAGCAAGTACTACCTGGATTTGAACTCGGACTCAGACCCCTATCCACCCCCAC
CCACGCCCCACAGCCAGTACCTGTCGGCGGAGGACAGCTGCCCGCCCTCGCCCGC
CACCGAGAGGAGCTACTTCCATCTCTTCCCGCCCCCTCCGTCCCCCTGCACGGAC
TCATCCTGACCTCGGCCGGGCCACTCTGGCTTCTCTGTGCCCCTGTAAATAGTT
TTAAATATGAACAAAGAAAAAAATATATTTTATGATTTAAAAAATAA
ATATAATTGGGATTTTAAAAACATGAGAAATGTGAACTGTGATGGGGTG
GGCAGGGCTGGGAGAACTTTGTACAGTGGAACAAATATTTATAAACTTAA
TTTTGTAAAACAG

Figure 15(a)

AGGCTGGTCTCAAACTCCTGGCCTTAAGTGATCTGCCCGCCTCGGCCTCCCAAA
GTGCTGAGATGACAGGTGTGAGCCACCGTGCCCGGCCCAGAACTCTTTAATTC
CCACCTGAAACTTGCCGCCTTAAGCAGGTCCCCAGTCTCCCTCCCCTAGTCCCT
GGTCCCACCATTCTGCTTTCTGTCTCAATGAATTTGCCTACCCCTCGCCGCTCCT
GCTATTTGCCAACCGCCGGGACGTACGGCTGGTGGACGCCGGCGGAGTCAAGC
TGGAGTCCACCATCGTGGTCAGCGGCCTGGAGGATGCGGCCGCAGTGGACTTCC
AGTTTTCCAAGGGAGCCGTGTACTGGACAGACGTGAGCGAGGAGGCCATCA
AGCAGACCTACCTGAACCAGACGGGGGCCGCCGTGCAGAACGTGGTCATCTCC
GGCCTGGTCTCTCCCGACGGCCTCGCCTGCGACTGGGTGGGCAAGAAGCTGTAC
TGGACGGACTCAGAGACCAACCGCATCGAGGTGGCCAACCTCAATGGCACAT
CCCGGAAGGTGCTCTTCTGGCAGGACCTTGACCAGCCGAGGGCCATCGCCTTGG
ACCCCGCTCACGGGTACATGTACTGGACAGACTGGGGTGAGACGCCCCGGATT
GAGCGGGCAGGGATGGATGGCAGCACCCGGAAGATCATTGTGGACTCGGACA
TTTACTGGCCCAATGGACTGACCATCGACCTGGAGGAGCAGAAGCTCTACTG
GGCTGACGCCAAGCTCAGCTTCATCCACCGTGCCAACCTGGACGGCTCGTTCCG
GCAGAAGGTGGTGGAGGGCAGCCTGACGCACCCCTTCGCCCTGACGCTCTCCGG
GGACACTCTGTACTGGACAGACTGGCAGACCCGCTCCATCCATGCCTGCAACA
AGCGCACTGGGGGGAAGAGGAAGGAGATCCTGAGTGCCCTCTACTCACCCAT
GGACATCCAGGTGCTGAGCCAGGAGCGGCAGCCTTTCTTCCACACTCGCTGTG
AGGAGGACAATGGCGGCTGCTCCCACCTGTGCCTGCTGTCCCCAAGCGAGCCTT
TCTACACATGCCGCCTGCCCCACGGGTGTGCAGCTGCAGGACAACGGCAGGACG
TGTAAGGCAGGAGCCGAGGAGGTGCTGCTGCTGGCCCGGCGGACGGACCTACG
GAGGATCTCGCTGGACACGCCGGACTTTACCGACATCGTGCTGCAGGTGGACG
ACATCCGGCACGCCATTGCCATCGACTACGACCCGCTAGAGGGCTATGTCTAC
TGGACAGATGACGAGGTGCGGGCCATCCGCAGGGCGTACCTGGACGGGTCTGG
GGCGCAGACGCTGGTCAACACCGAGATCAACGACCCCGATGGCATCGCGGTCG
ACTGGGTGGCCCGAAACCTCTACTGGACCGACACGGGCACGGACCGCATCGAG
GTGACGCGCCTCAACGGCACCTCCCGCAAGATCCTGGTGTCGGAGGACCTGGA
CGAGCCCCGAGCCATCGCACTGCACCCCGTGATGGGCCTCATGTACTGGACAG
ACTGGGGAGAGAACCCTAAAATCGAGTGTGCCAACTTGGATGGGCAGGAGC
GGCGTGTGCTGGTCAATGCCTCCCTCGGGTGGCCCAACGGCCTGGCCCTGGACCT
GCAGGAGGGGAAGCTCTACTGGGGAGACGCCAAGACAGACAAGATCGAGGT
GATCAATGTTGATGGGACGAAGAGGCGGACCCTCCTGGAGGACAAGCTCCCG
CACATTTTCGGGTTCACGCTGCTGGGGGACTTCATCTACTGGACTGACTGGCA
GCGCCGCAGCATCGAGCGGGTGCACAAGGTCAAGGCCAGCCGGGACGTCATCA
TTGACCAGCTGCCCGACCTGATGGGGCTCAAAGCTGTGAATGTGGCCAAGGTC
GTCGGAACCAACCCCGTGTGCGGACAGGAACGGGGGGTGCAGCCACCTGTGCTT
CTTCACACCCCACGCAACCCGGTGTGGCTGCCCCATCGGCCTGGAGCTGCTGAG
TGACATGAAGACCTGCATCGTGCCTGAGGCCTTCTTGGTCTTCACCAGCAGAG

Figure 15(a), ctd.

CCGCCATCCACAGGATCTCCCTCGAGACCAATAACAACGACGTGGCCATCCCG
CTCACGGGCGTCAAGGAGGCCTCAGCCCTGGACTTTGATGTGTCCAACAACCA
CATCTACTGGACAGACGTCAGCCTGAAGACCATCAGCCGCGCCTTCATGAAC
GGGAGCTCGGTGGAGCACGTGGTGGAGTTTGGCCTTGACTACCCCGAGGGCAT
GGCCGTTGACTGGATGGGCAAGAACCTCTACTGGGCCGACACTGGGACCAAC
AGAATCGAAGTGGCGCGGCTGGACGGGCAGTTCCGGCAAGTCCTCGTGTGGAG
GGACTTGGACAACCCGAGGTCGCTGGCCCTGGATCCCACCAAGGGCTACATCT
ACTGGACCGAGTGGGGCGGCAAGCCGAGGATCGTGCGGGCCTTCATGGACGGG
ACCAACTGCATGACGCTGGTGGACAAGGTGGGCCGGGCCAACGACCTCACCA
TTGACTACGCTGACCAGCGCCTCTACTGGACCGACCTGGACACCAACATGATC
GAGTCGTCCAACATGCTGGGTCAGGAGCGGGTCGTGATTGCCGACGATCTCCC
GCACCCGTTCGGTCTGACGCAGTACAGCGATTATATCTACTGGACAGACTGG
AATCTGCACAGCATTGAGCGGGCCGACAAGACTAGCGGCCGGAACCGCACCC
TCATCCAGGGCCACCTGGACTTCGTGATGGACATCCTGGTGTTCCACTCCTCCC
GCCAGGATGGCCTCAATGACTGTATGCACAACAACGGGCAGTGTGGGCAGCT
GTGCCTTGCCATCCCCGGCGGCCACCGCTGCGGCTGCGCCTCACACTACACCCTG
GACCCCAGCAGCCGCAACTGCAGCCCGCCCACCACCTTCTTGCTGTTCAGCCAG
AAATCTGCCATCAGTCGGATGATCCCGGACGACCAGCACAGCCCGGATCTCA
TCCTGCCCCTGCATGGACTGAGGAACGTCAAAGCCATCGACTATGACCCACTG
GACAAGTTCATCTACTGGGTGGATGGGCGCCAGAACATCAAGCGAGCCAAG
GACGACGGGACCCAGCCCTTTGTTTTGACCTCTCTGAGCCAAGGCCAAAACCC
AGACAGGCAGCCCCACGACCTCAGCATCGACATCTACAGCCGGACACTGTTCT
GGACGTGCGAGGCCACCAATACCATCAACGTCCACAGGCTGAGCGGGGAAGC
CATGGGGGTGGTGCTGCGTGGGGACCGCGACAAGCCCAGGGCCATCGTCGTCA
ACGCGGAGCGAGGGTACCTGTACTTCACCAACATGCAGGACCGGGCAGCCAA
GATCGAACGCGCAGCCCTGGACGGCACCGAGCGCGAGGTCCTCTTCACCACCGG
CCTCATCCGCCCTGTGGCCCTGGTGGTAGACAACACACTGGGCAAGCTGTTCTG
GGTGGACGCGGACCTGAAGCGCATTGAGAGCTGTGACCTGTCAGGGGCCAACC
GCCTGACCCTGGAGGACGCCAACATCGTGCAGCCTCTGGGCCTGACCATCCTTG
GCAAGCATCTCTACTGGATCGACCGCCAGCAGCAGATGATCGAGCGTGTGGA
GAAGACCACCGGGGACAAGCGGACTCGCATCCAGGGCCGTGTCGCCCACCTCA
CTGGCATCCATGCAGTGGAGGAAGTCAGCCTGGAGGAGTTCTCAGCCCACCCA
TGTGCCCGTGACAATGGTGGCTGCTCCCACATCTGTATTGCCAAGGGTGATGG
GACACCACGGTGCTCATGCCCAGTCCACCTCGTGCTCCTGCAGAACCTGCTGAC
CTGTGGAGAGCCGCCCACCTGCTCCCCGGACCAGTTTGCATGTGCCACAGGGGA
GATCGACTGTATCCCCGGGGCCTGGCGCTGTGACGGCTTTCCCGAGTGCGATGA
CCAGAGCGACGAGGAGGGCTGCCCCCGTGTGCTCCGCCGCCCAGTTCCCCTGCGCG
CGGGGTCAGTGTGTGTGGACCTGCCGCCTGCGCTGCGACGGCGAGGCAGACTGTCAG
GACCGCTCAGACGAGGCGGACTGTGACGCCATCTGCCTGCCCAACCAGTTCCGG
TGTGCGAGCGGCCAGTGTGTGTCCTCATCAAACAGCAGTGCGACTCCTTCCCCGA
CTGTATCGACGGCTCCGACGAGCTCATGTGTGTGAAATCACCAAGCCGCCCTCAG

Figure 15(a), ctd.

ACGACAGCCCGGCCCACAGCAGTGCCATCGGGCCCGTCATTGGCATCATCCTCT
CTCTCTTCGTCATGGGTGGTGTCTATTTTGTGTGCCAGCGCGTGGTGTGCCAGC
GCTATGCGGGGGCCAACGGGCCCTTCCCGCACGAGTATGTCAGCGGGACCCCGC
ACGTGCCCCTCAATTTCATAGCCCCGGGCGGTTCCCAGCATGGCCCCTTCACAG
GCATCGCATGCGGAAAGTCCATGATGAGCTCCGTGAGCCTGATGGGGGGCCGG
GGCGGGGTGCCCCTCTACGACCGGAACCACGTCACAGGGGCCTCGTCCAGCAGC
TCGTCCAGCACGAAGGCCACGCTGTACCCGCGGATCCTGAACCCGCCGCCCTCCC
CGGCCACGGACCCCTCCCTGTACAACATGGACATGTTCTACTCTTCAAACATT
CCGGCCACTGCGAGACCGTACAGGCCCTACATCATTCGAGGAATGGCGCCCCC
GACGACGCCCTGCAGCACCGACGTGTGTGACAGCGACTACAGCGCCAGCCGCT
GGAAGGCCAGCAAGTACTACCTGGATTTGAACTCGGACTCAGACCCCTATCC
ACCCCCACCCACGCCCCACAGCCAGTACCTGTCGGCGGAGGACAGCTGCCCGCCC
TCGCCCGCCACCGAGAGGAGCTACTTCCATCTCTTCCCGCCCCCTCCGTCCCCCTG
CACGGACTCATCCTGACCTCGGCCGGGCCACTCTGGCTTCTCTGTGCCCCTGTAA
ATAGTTTTAAATATGAACAAAGAAAAAAATATATTTTATGATTTAAA
AAATAAATATAATTGGGATTTTAAAAACATGAGAAATGTGAACTGTGA
TGGGGTGGGCAGGGCTGGGAGAACTTTGTACAGTGGAACAAATATTTATA
AACTTAATTTTGTAAAACAG

Figure 15(b)

CAATGTCCAGTTCCGCTGCAGTTATAACATCCCATTTTTTTGATTTCTTTTTA
TTTTTTCCTTTTTTCTTTTTTGAGATGGAGTCTCGCTCTGTCACCCAGGCTGGAGT
GCAATGGG

Figure 16(a)

GCCGCGGCGCCCGAGGCGGGAGCAAGAGGCGCCGGGAGCCGCGAGGATCCACCG
CCGCCGCGCGCGCCATGGAGCCCGAGTGAGCGCGCGGCGCTCCCGGCCGCCGGAC
GACATGGAAACGGCGCCGACCCGGGCCCCTCCGCCGCCGCGCCGCCGCTGCTGC
TGCTGGTGCTGTACTGCAGCTTGGTCCCCGCCGCGGCCTCACCGCTCCTGTTGTT
TGCCAACCGCCGGGATGTGCGGCTAGTGGATGCCGGCGGAGTGAAGCTGGAGT
CCACCATTGTGGCCAGTGGCCTGGAGGATGCAGCTGCTGTAGACTTCCAGTTC
TCCAAGGGTGCTGTGTACTGGACAGATGTGAGCGAGGAGGCCATCAAACAG
ACCTACCTGAACCAGACTGGAGGTGCTGCACAGAACATTGTCATCTCGGGCC
TCGTGTCACCTGATGGCCTGGCCTGTGACTGGGTTGGCAAGAAGCTGTACTGG
ACGGACTCCGAGACCAACCGCATTGAGGTTGCCAACCTCAATGGGACGTCCCG
TAAGGTTCTCTTCTGGCAGGACCTGGACCAGCCAAGGGCCATTGCCCTGGATC
CTGCACATGGGTACATGTACTGGACTGACTGGGGGGAAGCACCCCGGATCGA
GCGGGCAGGGATGGATGGCAGTACCCGGAAGATCATTGTAGACTCCGACATT
TACTGGCCCAATGGGCTGACCATCGACCTGGAGGAACAGAAGCTGTACTGGG
CCGATGCCAAGCTCAGCTTCATCCACCGTGCCAACCTGGACGGCTCCTTCCGGC
AGAAGGTGGTGGAGGGCAGCCTCACTCACCCTTTTGCCCTGACACTCTCTGGG
GACACACTCTACTGGACAGACTGGCAGACCCGCTCCATCCACGCCTGCAACA
AGTGGACAGGGGAGCAGAGGAAGGAGATCCTTAGTGCTCTGTACTCACCCA
TGGACATCCAAGTGCTGAGCCAGGAGCGGCAGCCTCCCTTCCACACACCATGC
GAGGAGGACAACGGTGGCTGTTCCCACCTGTGCCTGCTGTCCCCGAGGGAGCCT
TTCTACTCCTGTGCCTGCCCCACTGGTGTGCAGTTGCAGGACAATGGCAAGAC
GTGCAAGACAGGGGCTGAGGAAGTGCTGCTGCTGGCTCGGAGGACAGACCTG
AGGAGGATCTCTCTGGACACCCCTGACTTCACAGACATAGTGCTGCAGGTGG
GCGACATCCGGCATGCCATTGCCATTGACTACGATCCCCTGGAGGGCTACGTG
TACTGGACCGATGATGAGGTGCGGGCTATCCGCAGGGCGTACCTAGATGGCTC
AGGTGCGCAGACACTTGTGAACACTGAGATCAATGACCCCGATGGCATTGCT
GTGGACTGGGTCGCCCGGAACCTCTACTGGACAGATACAGGCACTGACAGAA
TTGAGGTGACTCGCCTCAACGGCACCTCCCGAAAGATCCTGGTATCTGAGGAC
CTGGACGAACCGCGAGCCATTGTGTTGCACCCTGTGATGGGCCTCATGTACTG
GACAGACTGGGGGGAGAACCCCAAAATCGAATGCGCCAACCTAGATGGGAG
AGATCGGCATGTCCTGGTGAACACCTCCCTTGGGTGGCCCAATGGACTGGCCC
TGGACCTGCAGGAGGGCAAGCTGTACTGGGGGGATGCCAAAACTGATAAAA
TCGAGGTGATCAACATAGACGGG

Figure   16(b)

GCCGCGGCGCCCGAGGCGGGAGCAAGAGGCGCCGGGAGCCGCGAGGATCCACCG
CCGCCGCGCGCGCCATGGAGCCCGAGTGAGCGCGCGGCGCTCCCGGCCGCCGGAC
GACATGGAAACGGCGCCGACCCGGGCCCCTCCGCCGCCGCCGCCGCCGCTGCTGC
TGCTGGTGCTGTACTGCAGCTTGGTCCCCGCCGCGG

Figure 16(c)

ATGGAAACGGCGCCGACCCGGGCCCCTCCGCCGCCGCCGCCGCCGCTGCTGCTGCT
GGTGCTGTACTGCAGCTTGGTCCCCGCCGCGGCCTCACCGCTCCTGTTGTTTGCC.
AACCGCCGGGATGTGCGGCTAGTGGATGCCGGCGGAGTGAAGCTGGAGTCCAC
CATTGTGGCCAGTGGCCTGGAGGATGCAGCTGCTGTAGACTTCCAGTTCTCCA
AGGGTGCTGTGTACTGGACAGATGTGAGCGAGGAGGCCATCAAACAGACCT
ACCTGAACCAGACTGGAGGTGCTGCACAGAACATTGTCATCTCGGGCCTCGT
GTCACCTGATGGCCTGGCCTGTGACTGGGTTGGCAAGAAGCTGTACTGGACGG
ACTCCGAGACCAACCGCATTGAGGTTGCCAACCTCAATGGGACGTCCCGTAA
GGTTCTCTTCTGGCAGGACCTGGACCAGCCAAGGGCCATTGCCCTGGATCCTGC
ACATGGGTACATGTACTGGACTGACTGGGGGGAAGCACCCCGGATCGAGCGG
GCAGGGATGGATGGCAGTACCCGGAAGATCATTGTAGACTCCGACATTTAC
TGGCCCAATGGGCTGACCATCGACCTGGAGGAACAGAAGCTGTACTGGGCCG
ATGCCAAGCTCAGCTTCATCCACCGTGCCAACCTGGACGGCTCCTTCCGGCAG
AAGGTGGTGGAGGGCAGCCTCACTCACCCTTTTGCCCTGACACTCTCTGGGGA
CACACTCTACTGGACAGACTGGCAGACCCGCTCCATCCACGCCTGCAACAAGT
GGACAGGGGAGCAGAGGAAGGAGATCCTTAGTGCTCTGTACTCACCCATGG
ACATCCAAGTGCTGAGCCAGGAGCGGCAGCCTCCCTTCCACACACCATGCGAG
GAGGACAACGGTGGCTGTTCCCACCTGTGCCTGCTGTCCCCGAGGGAGCCTTTC
TACTCCTGTGCCTGCCCCACTGGTGTGCAGTTGCAGGACAATGGCAAGACGTG
CAAGACAGGGGCTGAGGAAGTGCTGCTGCTGGCTCGGAGGACAGACCTGAGG
AGGATCTCTCTGGACACCCCTGACTTCACAGACATAGTGCTGCAGGTGGGCG
ACATCCGGCATGCCATTGCCATTGACTACGATCCCCTGGAGGGCTACGTGTAC
TGGACCGATGATGAGGTGCGGGCTATCCGCAGGGCGTACCTAGATGGCTCAGG
TGCGCAGACACTTGTGAACACTGAGATCAATGACCCCGATGGCATTGCTGTG
GACTGGGTCGCCCGGAACCTCTACTGGACAGATACAGGCACTGACAGAATTG
AGGTGACTCGCCTCAACGGCACCTCCCGAAAGATCCTGGTATCTGAGGACCTG
GACGAACCGCGAGCCATTGTGTTGCACCCTGTGATGGGCCTCATGTACTGGAC
AGACTGGGGGGAGAACCCCAAAATCGAATGCGCCAACCTAGATGGGAGAG
ATCGGCATGTCCTGGTGAACACCTCCCTTGGGTGGCCCAATGGACTGGCCCTGG
ACCTGCAGGAGGGCAAGCTGTACTGGGGGGATGCCAAAACTGATAAAATCG
AGGTGATCAACATAGACGGG

Figure 16(d)

METAPTRAPPPPPPPLLLLVLYCSLVPAAASPLLLFANRRDVRLVDAGGVK
LESTIVASGLEDAAAVDFQFSKGAVYWTDVSEEAIKQTYLNQTGGAAQNIVI
SGLVSPDGLACDWVGKKLYWTDSETNRIEVANLNGTSRKVLFWQDLDQPRA
IALDPAHGYMYWTDWGEAPRIERAGMDGSTRKIIVDSDIYWPNGLTIDLEE
QKLYWADAKLSFIHRANLDGSFRQKVVEGSLTHPFALTLSGDTLYWTDWQT
RSIHACNKWTGEQRKEILSALYSPMDIQVLSQERQPPFHTPCEEDNGGCSHLCL
LSPREPFYSCACPTGVQLQDNGKTCKTGAEEVLLLARRTDLRRISLDTPDFTDI
VLQVGDIRHAIAIDYDPLEGYVYWTDDEVRAIRRAYLDGSGAQTLVNTEIND
PDGIAVDWVARNLYWTDTGTDRIEVTRLNGTSRKILVSEDLDEPRAIVLHP
VMGLMYWTDWGENPKIECANLDGRDRHVLVNTSLGWPNGLALDLQEGKLY
WGDAKTDKIEVINIDG

Figure 17(a)

CCTCGCCGCTCCTGCTATTTGCCAACCGCCGGGACGTACGGCTGGTGGACGCCGG
CGGAGTCAAGCTGGAGTCCACCATCGTGGTCAGCGGCCTGGAGGATGCGGCCG
CAGTGGACTTCCAGTTTTCCAAGGGAGCCGTGTACTGGACAGACGTGAGCGA
GGAGGCCATCAAGCAGACCTACCTGAACCAGACGGGGGCCGCCGTGCAGAAC
GTGGTCATCTCCGGCCTGGTCTCTCCCGACGGCCTCGCCTGCGACTGGGTGGGCA
AGAAGCTGTACTGGACGGACTCAGAGACCAACCGCATCGAGGTGGCCAACCT
CAATGGCACATCCCGGAAGGTGCTCTTCTGGCAGGACCTTGACCAGCCGAGGG
CCATCGCCTTGGACCCCGCTCACGGGTACATGTACTGGACAGACTGGGGTGAG
ACGCCCCGGATTGAGCGGGCAGGGATGGATGGCAGCACCCGGAAGATCATTG
TGGACTCGGACATTTACTGGCCCAATGGACTGACCATCGACCTGGAGGAGCA
GAAGCTCTACTGGGCTGACGCCAAGCTCAGCTTCATCCACCGTGCCAACCTGG
ACGGCTCGTTCCGGCAGAAGGTGGTGGAGGGCAGCCTGACGCACCCCTTCGCCC
TGACGCTCTCCGGGGACACTCTGTACTGGACAGACTGGCAGACCCGCTCCATCC
ATGCCTGCAACAAGCGCACTGGGGGGAAGAGGAAGGAGATCCTGAGTGCCCT
CTACTCACCCATGGACATCCAGGTGCTGAGCCAGGAGCGGCAGCCTTTCTTCC
ACACTCGCTGTGAGGAGGACAATGGCGGCTGCTCCCACCTGTGCCTGCTGTCCC
CAAGCGAGCCTTTCTACACATGCGCCTGCCCCACGGGTGTGCAGCTGCAGGAC
AACGGCAGGACGTGTAAGGCAGGAGCCGAGGAGGTGCTGCTGCTGGCCCGGCG
GACGGACCTACGGAGGATCTCGCTGGACACGCCGGACTTTACCGACATCGTGC
TGCAGGTGGACGACATCCGGCACGCCATTGCCATCGACTACGACCCGCTAGAG
GGCTATGTCTACTGGACAGATGACGAGGTGCGGGCCATCCGCAGGGCGTACCT
GGACGGGTCTGGGGCGCAGACGCTGGTCAACACCGAGATCAACGACCCCGATG
GCATCGCGGTCGACTGGGTGGCCCGAAACCTCTACTGGACCGACACGGGCACG
GACCGCATCGAGGTGACGCGCCTCAACGGCACCTCCCGCAAGATCCTGGTGTCG
GAGGACCTGGACGAGCCCCGAGCCATCGCACTGCACCCCGTGATGGGCCTCATG
TACTGGACAGACTGGGGAGAGAACCCTAAAATCGAGTGTGCCAACTTGGAT
GGGCAGGAGCGGCGTGTGCTGGTCAATGCCTCCCTCGGGTGGCCCAACGGCCTG
GCCCTGGACCTGCAGGAGGGGAAGCTCTACTGGGGAGACGCCAAGACAGACA
AGATCGAGGTGATCAATGTTGATGGGACGAAGAGGCGGACCCTCCTGGAGG
ACAAGCTCCCGCACATTTTCGGGTTCACGCTGCTGGGGGACTTCATCTACTGG
ACTGACTGGCAGCGCCGCAGCATCGAGCGGGTGCACAAGGTCAAGGCCAGCCG
GGACGTCATCATTGACCAGCTGCCCGACCTGATGGGGCTCAAAGCTGTGAAT
GTGGCCAAGGTCGTCGGAACCAACCCGTGTGCGGACAGGAACGGGGGGTGCA
GCCACCTGTGCTTCTTCACACCCCACGCAACCCGGTGTGGCTGCCCATCGGCCT
GGAGCTGCTGAGTGACATGAAGACCTGCATCGTGCCTGAGGCCTTCTTGGTCT
TCACCAGCAGAGCCGCCATCCACAGGATCTCCCTCGAGACCAATAACAACGA
CGTGGCCATCCCGCTCACGGGCGTCAAGGAGGCCTCAGCCCTGGACTTTGATGT
GTCCAACAACCACATCTACTGGACAGACGTCAGCCTGAAGACCATCAGCCGC
GCCTTCATGAACGGGAGCTCGGTGGAGCACGTGGTGGAGTTTGGCCTTGACTA
CCCCGAGGGGCATGGCCGTTGACTGGATGGGCAAGAACCTCTACTGGGCCGACA

Figure 17(a) Continued

CTGGGACCAACAGAATCGAAGTGGCGCGGCTGGACGGGCAGTTCCGGCAAGT
CCTCGTGTGGAGGGACTTGGACAACCCGAGGTCGCTGGCCCTGGATCCCACCA
AGGGCTACATCTACTGGACCGAGTGGGGCGGCAAGCCGAGGATCGTGCGGGCC
TTCATGGACGGGACCAACTGCATGACGCTGGTGGACAAGGTGGGCCGGGCCA
ACGACCTCACCATTGACTACGCTGACCAGCGCCTCTACTGGACCGACCTGGAC
ACCAACATGATCGAGTCGTCCAACATGCTGGGTCAGGAGCGGGTCGTGATTG
CCGACGATCTCCCGCACCCGTTCGGTCTGACGCAGTACAGCGATTATATCTAC
TGGACAGACTGGAATCTGCACAGCATTGAGCGGGCCGACAAGACTAGCGGCC
GGAACCGCACCCTCATCCAGGGCCACCTGGACTTCGTGATGGACATCCTGGTG
TTCCACTCCTCCCGCCAGGATGGCCTCAATGACTGTATGCACAACAACGGGCA
GTGTGGGCAGCTGTGCCTTGCCATCCCCGGCGGCCACCGCTGCGGCTGCGCCTCA
CACTACACCCTGGACCCCAGCAGCCGCAACTGCAGCCCGCCCACCACCTTCTTG
CTGTTCAGCCAGAAATCTGCCATCAGTCGGATGATCCCGGACGACCAGCACA
GCCCGGATCTCATCCTGCCCCTGCATGGACTGAGGAACGTCAAAGCCATCGAC
TATGACCCACTGGACAAGTTCATCTACTGGGTGGATGGGCGCCAGAACATCA
AGCGAGCCAAGGACGACGGGACCCAGCCCTTTGTTTTGACCTCTCTGAGCCAA
GGCCAAAACCCAGACAGGCAGCCCCACGACCTCAGCATCGACATCTACAGCC
GGACACTGTTCTGGACGTGCGAGGCCACCAATACCATCAACGTCCACAGGCT
GAGCGGGGAAGCCATGGGGGTGGTGCTGCGTGGGGACCGCGACAAGCCCAGGG
CCATCGTCGTCAACGCGGAGCGAGGGTACCTGTACTTCACCAACATGCAGGA
CCGGGCAGCCAAGATCGAACGCGCAGCCCTGGACGGCACCGAGCGCGAGGTCC
TCTTCACCACCGGCCTCATCCGCCCTGTGGCCCTGGTGGTAGACAACACACTGG
GCAAGCTGTTCTGGGTGGACGCGGACCTGAAGCGCATTGAGAGCTGTGACCTG
TCAGGGGCCAACCGCCTGACCCTGGAGGACGCCAACATCGTGCAGCCTCTGGGC
CTGACCATCCTTGGCAAGCATCTCTACTGGATCGACCGCCAGCAGCAGATGA
TCGAGCGTGTGGAGAAGACCACCGGGGACAAGCGGACTCGCATCCAGGGCCG
TGTCGCCCACCTCACTGGCATCCATGCAGTGGAGGAAGTCAGCCTGGAGGAGT
TCTCAGCCCACCCATGTGCCCGTGACAATGGTGGCTGCTCCCACATCTGTATTG
CCAAGGGTGATGGGACACCACGGTGCTCATGCCCAGTCCACCTCGTGCTCCTGC
AGAACCTGCTGACCTGTGGAGAGCCGCCCACCTGCTCCCCGGACCAGTTTGCAT
GTGCCACAGGGGAGATCGACTGTATCCCCGGGGCCTGGCGCTGTGACGGCTTTC
CCGAGTGCGATGACCAGAGCGACGAGGAGGGCTGCCCCGTGTGCTCCGCCGCCC
AGTTCCCCTGCGCGCGGGGTCAGTGTGTGGACCTGCGCCTGCGCTGCGACGGCGA
GGCAGACTGTCAGGACCGCTCAGACGAGGCGGACTGTGACGCCATCTGCCTGC
CCAACCAGTTCCGGTGTGCGAGCGGCCAGTGTGTCCTCATCAAACAGCAGTGC
GACTCCTTCCCCGACTGTATCGACGGCTCCGACGAGCTCATGTGTGAAATCAC
CAAGCCGCCCTCAGACGACAGCCCGGCCCACAGCAGTGCCATCGGGCCCGTCAT
TGGCATCATCCTCTCTCTCTTCGTCATGGGTGGTGTCTATTTTGTGTGCCAGCG
CGTGGTGTGCCAGCGCTATGCGGGGGGCCAACGGGCCCTTCCCGCACGAGTATGT
CAGCGGGACCCCCGCACGTGCCCCCTCAATTTCATAGCCCCGGGCCGGTTCCCAGCA
TGGCCCCTTCACAGGCATCGCATGCGGAAAGTCCATGATGAGCTCCGTGAGCC

Figure 17(a) Continued

TGATGGGGGGCCGGGGCGGGGTGCCCCTCTACGACCGGAACCACGTCACAGGGG
CCTCGTCCAGCAGCTCGTCCAGCACGAAGGCCACGCTGTACCCGCCGATCCTGA
ACCCGCCGCCCTCCCCGGCCACGGACCCCTCCCTGTACAACATGGACATGTTCT
ACTCTTCAAACATTCCGGCCACTGTGAGACCGTACAGGCCCTACATCATTCG
AGGAATGGCGCCCCCGACGACGCCCTGCAGCACCGACGTGTGTGACAGCGACT
ACAGCGCCAGCCGCTGGAAGGCCAGCAAGTACTACCTGGATTTGAACTCGGA
CTCAGACCCCTATCCACCCCCACCCACGCCCCACAGCCAGTACCTGTCGGCGGA
GGACAGCTGCCCGCCCTCGCCCGCCACCGAGAGGAGCTACTTCCATCTCTTCCCG
CCCCCTCCGTCCCCCTGCACGGACTCATCCTGACCTCGGCCGGGCCACTCTGGCTT
CTCTGTGCCCCTGTAAATAGTTTTAAATATGAACAAAGAAAAAAATATA
TTTTATGATTTAAAAAAATAAATATAATTGGGATTTTAAAAACATGAGA
AATGTGAACTGTGATGGGGTGGGCAGGGCTGGGAGAACTTTGTACAGTGGA
ACAAATATTTATAAACTTAATTTTGTAAAACAG

Figure 17(b)

                                                      SPLLLFANRRDVRLVDAGGVKLESTIV
VSGLEDAAAVDFQFSKGAVYWTDVSEEAIKQTYLNQTGAAVQNVVISGLVSPDGLAC
DWVGKKLYWTDSETNRIEVANLNGTSRKVLFWQDLDQPRAIALDPAHGYMYWTDW
GETPRIERAGMDGSTRKIIVDSDIYWPNGLTIDLEEQKLYWADAKLSFIHRANLDGSFR
QKVVEGSLTHPFALTLSGDTLYWTDWQTRSIHACNKRTGGKRKEILSALYSPMDIQVLS
QERQPFFHTRCEEDNGGCSHLCLLSPSEPFYTCACPTGVQLQDNGRTCKAGAEEVLLL
ARRTDLRRISLDTPDFTDIVLQVDDIRHAIAIDYDPLEGYVYWTDDEVRAIRRAYLDGS
GAQTLVNTEINDPDGIAVDWVARNLYWTDTGTDRIEVTRLNGTSRKILVSEDLDEPRAI
ALHPVMGLMYWTDWGENPKIECANLDGQERRVLVNASLGWPNGLALDLQEGKLYW
GDAKTDKIEVINVDGTKRRTLLEDKLPHIFGFTLLGDFIYWTDWQRRSIERVHKVKASR
DVIIDQLPDLMGLKAVNVAKVVGTNPCADRNGGCSHLCFFTPHATRCGCPIGLELLSD
MKTCIVPEAFLVFTSRAAIHRISLETNNNDVAIPLTGVKEASALDFDVSNNHIYWTDVSL
KTISRAFMNGSSVEHVVEFGLDYPEGMAVDWMGKNLYWADTGTNRIEVARLDGQFR
QVLVWRDLDNPRSLALDPTKGYIYWTEWGGKPRIVRAFMDGTNCMTLVDKVGRAND
LTIDYADQRLYWTDLDTNMIESSNMLGQERVVIADDLPHPFGLTQYSDYIYWTDWNL
HSIERADKTSGRNRTLIQGHLDFVMDILVFHSSRQDGLNDCMHNNGQCGQLCLAIPGG
HRCGCASHYTLDPSSRNCSPPTTFLLFSQKSAISRMIPDDQHSPDLILPLHGLRNVKAIDY
DPLDKFIYWVDGRQNIKRAKDDGTQPFVLTSLSQGQNPDRQPHDLSIDIYSRTLFWTCE
ATNTINVHRLSGEAMGVVLRGDRDKPRAIVVNAERGYLYFTNMQDRAAKIERAALDG
TEREVLFTTGLIRPVALVVDNTLGKLFWVDADLKRIESCDLSGANRLTLEDANIVQPLG
LTILGKHLYWIDRQQQMIERVEKTTGDKRTRIQGRVAHLTGIHAVEEVSLEEFSAHPCA
RDNGGCSHICIAKGDGTPRCSCPVHLVLLQNLLTCGEPPTCSPDQFACATGEIDCIPGA
WRCDGFPECDDQSDEEGCPVCSAAQFPCARGQCVDLRLRCDGEADCQDRSDEADCD
AICLPNQFRCASGQCVLIKQQCDSFPDCIDGSDELMCEITKPPSDDSPAHSSAIGPVIGIIL
SLFVMGGVYFVCQRVVCQRYAGANGPFPHEYVSGTPHVPLNFIAPGGSQHGPFTGIAC
GKSMMSSVSLMGGRGGVPLYDRNHVTGASSSSSSSTKATLYPPILNPPPSPATDPSLYN
MDMFYSSNIPATVRPYRPYIIRGMAPPTTPCSTDVCDSDYSASRWKASKYYLDLNSDSD
PYPPPPTPHSQYLSAEDSCPPSPATERSYFHLFPPPPSPCTDSS

Figure 18(a)

GCCGCGGCGCCCGAGGCGGGAGCAAGAGGCGCCGGGAGCCGCGAGGATCCACCGCCGCCG
CGCGCGCCATGGAGCCCGAGTGAGCGCGCGGCGCTCCCGGCCGCCGGACGACATGGAAAC
GGCGCCGACCCGGGCCCCTCCGCCGCCGCCGCCGCCGCTGCTGCTGCTGGTGCTGTACTG
CAGCTTGGTCCCCGCCGCGGCCTCACCGCTCCTGTTGTTTGCCAACCGCCGGGATGTGCG
GCTAGTGGATGCCGGCGGAGTGAAGCTGGAGTCCACCATTGTGGCCAGTGGCCTGGAGGA
TGCAGCTGCTGTAGACTTCCAGTTCTCCAAGGGTGCTGTGTACTGGACAGATGTGAGCGA
GGAGGCCATCAAACAGACCTACCTGAACCAGACTGGAGCTGCTGCACAGAACATTGTCAT
CTCGGGCCTCGTGTCACCTGATGGCCTGGCCTGTGACTGGGTTGGCAAGAAGCTGTACTG
GACGGACTCCGAGACCAACCGCATTGAGGTTGCCAACCTCAATGGGACGTCCCGTAAGGT
TCTCTTCTGGCAGGACCTGGACCAGCCAAGGGCCATTGCCCTGGATCCTGCACATGGGTA
CATGTACTGGACTGACTGGGGGGAAGCACCCCGGATCGAGCGGGCAGGGATGGATGGCAG
TACCCGGAAGATCATTGTAGACTCCGACATTTACTGGCCCAATGGGCTGACCATCGACCT
GGAGGAACAGAAGCTGTACTGGGCCGATGCCAAGCTCAGCTTCATCCACCGTGCCAACCT
GGACGGCTCCTTCCGGCAGAAGGTGGTGGAGGGCAGCCTCACTCACCCTTTTGCCCTGAC
ACTCTCTGGGGACACACTCTACTGGACAGACTGGCAGACCCGCTCCATCCACGCCTGCAA
CAAGTGGACAGGGGAGCAGAGGAAGGAGATCCTTAGTGCTCTGTACTCACCCATGGACAT
CCAAGTGCTGAGCCAGGAGCGGCAGCCTCCCTTCCACACACCATGCGAGGAGGACAACGG
TGGCTGTTCCCACCTGTGCCTGCTGTCCCCGAGGGAGCCTTTCTACTCCTGTGCCTGCCC
CACTGGTGTGCAGTTGCAGGACAATGGCAAGACGTGCAAGACAGGGGCTGAGGAAGTGCT
GCTGCTGGCTCGGAGGACAGACCTGAGGAGGATCTCTCTGGACACCCCTGACTTCACAGA
CATAGTGCTGCAGGTGGGCGACATCCGGCATGCCATTGCCATTGACTACGATCCCCTGGA
GGGCTACGTGTACTGGACCGATGATGAGGTGCGGGCTATCCGCAGGGCGTACCTAGATGG
CTCAGGTGCGCAGACACTTGTGAACACTGAGATCAATGACCCCGATGGCATTGCTGTGGA
CTGGGTCGCCCGGAACCTCTACTGGACAGATACAGGCACTGACAGAATTGAGGTGACTCG
CCTCAACGGCACCTCCCGAAAGATCCTGGTATCTGAGGACCTGGACGAACCGCGAGCCAT
TGTGTTGCACCCTGTGATGGGCCTCATGTACTGGACAGACTGGGGGGAGAACCCCAAAAT
CGAATGCGCCAACCTAGATGGGAGAGATCGGCATGTCCTGGTGAACACCTCCCTTGGGTG
GCCCAATGGACTGGCCCTGGACCTGCAGGAGGGCAAGCTGTACTGGGGGGATGCCAAAAC
TGATAAAATCGAGGTGATCAACATAGACGGGACAAAGCGGAAGACCCTGCTTGAGGACAA
GCTCCCACACATTTTTGGGTTCACACTGCTGGGGGACTTCATCTACTGGACCGACTGGCA
GAGACGCAGTATTGAAAGGGTCCACAAGGTCAAGGCCAGCCGGGATGTCATCATTGATCA
ACTCCCCGACCTGATGGGACTCAAAGCCGTGAATGTGGCCAAGGTTGTCGGAACCAACCC
ATGTGCGGATGGAAATGGAGGGTGCAGCCATCTGTGCTTCTTCACCCCACGTGCCACCAA
GTGTGGCTGCCCCATTGGCCTGGAGCTGTTGAGTGACATGAAGACCTGCATAATCCCCGA
GGCCTTCCGGTATTCACCAGCAGAGCCACCATCCACAGGATCTCCCTGGAGACTAACAAC
AACGATGTGGCTATCCCACTCACGGGTGTCAAAGAGGCCTCTGCACTGGACTTTGATGTG
TCCAACAATCACATCTACTGGACTGATGTTAGCCTCAAGACGATCAGCCGAGCCTTCATG
AATGGGAGCTCAGTGGAGCACGTGATTGAGTTTGGCCTCGACTACCCTGAAGGAATGGCT
GTGGACTGGATGGGCAAGAACCTCTATTGGGCGGACACAGGGACCAACAGGATTGAGGTG
GCCCGGCTGGATGGGCAGTTCCGGCAGGTGCTTGTGTGGAGAGACCTTGACAACCCCAGG
TCTCTGGCTCTGGATCCTACTAAAGGCTACATCTACTGGACTGAGTGGGGTGGCAAGCCA
AGGATTGTGCGGGCCTTCATGGATGGGACCAATTGTATGACACTGGTAGACAAGGTGGGC
CGGGCCAACGACCTCACCATTGATTATGCCGACCAGCGACTGTACTGGACTGACCTGGAC
ACCAACATGATTGAGTCTTCCAACATGCTGGGTCAGGAGCGCATGGTGATAGCTGACGAT
CTGCCCTACCCGTTTGGCCTGACTCAATATAGCGATTACATCTACTGGACTGACTGGAAC
CTGCATAGCATTGAACGGGCGGACAAGACCAGTGGGCGGAACCGCACCCTCATCCAGGGT
CACCTGGACTTCGTCATGGACATCCTGGTGTTCCACTCCTCCCGTCAGGATGGCCTCAAC

Figure 18(a) Continued

GACTGCGTGCACAGCAATGGCCAGTGTGGGCAGCTGTGCCTCGCCATCCCCGGAGGCCAC
CGCTGTGGCTGTGCTTCACACTACACGCTGGACCCCAGCAGCCGCAACTGCAGCCCGCCC
TCCACCTTCTTGCTGTTCAGCCAGAAATTTGCCATCAGCCGGATGATCCCCGATGACCAG
CTCAGCCCGGACCTTGTCCTACCCCTTCATGGGCTGAGGAACGTCAAAGCCATCAACTAT
GACCCGCTGGACAAGTTCATCTACTGGGTGGACGGGCGCCAGAACATCAAGAGGGCCAAG
GACGACGGTACCCAGCCCTCCATGCTGACCTCTCCCAGCCAAAGCCTGAGCCCAGACAGA
CAGCCACACGACCTCAGCATTGACATCTACAGCCGGACACTGTTCTGGACCTGTGAGGCC
ACCAACACTATCAATGTCCACCGGCTGGATGGGGATGCCATGGGAGTGGTGCTTCGAGGG
GACCGTGACAAGCCAAGGGCCATTGCTGTCAATGCTGAGCGAGGGTACATGTACTTTACC
AACATGCAGGACCATGCTGCCAAGATCGAGCGAGCCTCCCTGGATGGCACAGAGCGGGAG
GTCCTCTTCACCACAGGCCTCATCCGTCCCGTGGCCCTTGTGGTGGACAATGCTCTGGGC
AAGCTCTTCTGGGTGGATGCCGACCTAAAGCGAATCGAAAGCTGTGACCTCTCTGGGGCC
AACCGCCTGACCCTGGAAGATGCCAACATCGTACAGCCAGTAGGTCTGACAGTGCTGGGC
AGGCACCTCTACTGGATCGACCGCCAGCAGCAGATGATCGAGCGCGTGGAGAAGACCACT
GGGGACAAGCGGACTAGGGTTCAGGGCCGTGTCACCCACCTGACAGGCATCCATGCCGTG
GAGGAAGTCAGCCTGGAGGAGTTCTCAGCCCATCCTTGTGCCCGAGACAATGGCGGCTGC
TCCCACATCTGTATCGCCAAGGGTGATGGAACACCGCGCTGCTCGTGCCCTGTCCACCTG
GTGCTCCTGCAGAACCTGCTGACTTGTGGTGAGCCTCCTACCTGCTCCCCTGATCAGTTT
GCATGTACCACTGGTGAGATCGACTGCATCCCCGGAGCCTGGCGCTGTGACGGCTTCCCT
GAGTGTGCTGACCAGAGTGATGAAGAAGGCTGCCCAGTGTGCTCCGCCTCTCAGTTCCCC
TGCGCTCGAGGCCAGTGTGTGGACCTGCGGTTACGCTGCGACGGTGAGGCCGACTGCCAG
GATCGCTCTGATGAAGTAACTGCGATGCTGTCTGTCTGCCCAATCAGTTCCGGTGCACCA
GCGGCCAGTGTGTCCTCATCAAGCAACAGTGTGACTCCTTCCCCGACTGTGCTGATGGGT
CTGATGAGCTCATGTGTGAAATCAACAAGCCACCCTCTGATGACATCCCAGCCCACAGCA
GTGCCATTGGGCCCGTCATTGGTATCATCCTCTCCCTCTTCGTCATGGGCGGGGTCTACT
TTGTCTGCCAGCGTGTGATGTGCCAGCGCTACACAGGGGCCAGTGGGCCCTTTCCCCACG
AGTATGTTGGTGGAGCCCCTCATGTGCCTCTCAACTTCATAGCCCCAGGTGGCTCACAGC
ACGGTCCCTTCCCAGGCATCCCGTGCAGCAAGTCCGTGATGAGCTCCATGAGCCTGGTGG
GGGGGCGCGGCAGCGTGCCCCTCTATGACCGGAATCACGTCACTGGGGCCTCATCCAGCA
GCTCGTCCAGCACAAAGGCCACACTATATCCGCCGATCCTGAACCCACCCCCGTCCCCGG
CCACAGACCCCTCTCTCTACAACGTGGACGTGTTTATTCTTCAGGCATCCCGGCCACCG
CTAGACCATACAGGCCCTACGTCATTCGAGGTATGGCACCCCCAACAACACCGTGCAGCA
CAGATGTGTGTGACAGTGACTACAGCATCAGTCGCTGGAAGAGCAGCAAATACTACCTGG
ACTTGAATTCGGACTCAGACCCCTACCCCCCCCCGCCCACCCCCCACAGCCAGTACCTAT
CTGCAGAGGACAGCTGCCCACCCTCACCAGGCACTGAGAGGAGTTACTGCCACCTCTTCC
CGCCCCCACCGTCCCCCTGCACGGACTCGTCCTGACCTCGGCCGTCCACCCGGCCCTGCT
GCCTCCCTGTAAATATTTTTAAATATGAACAAAGGAAAAATATATTTTATGATTTAAAAA
ATAAATATAATTGGGGTTTTTAACAAGTGAGAAATGTGAGCGGTGAAGGGGTGGGCAGGG
CTGGGAAACTTTTCTAG

Figure 18(b)

ATGGAAACGGCGCCGACCCGGGCCCCTCCGCCGCCGCCGCCGCCGCTGCTGCTGCTGGTG
CTGTACTGCAGCTTGGTCCCCGCCGCGGCCTCACCGCTCCTGTTGTTTGCCAACCGCCGG
GATGTGCGGCTAGTGGATGCCGGCGGAGTGAAGCTGGAGTCCACCATTGTGGCCAGTGGC
CTGGAGGATGCAGCTGCTGTAGACTTCCAGTTCTCCAAGGGTGCTGTGTACTGGACAGAT
GTGAGCGAGGAGGCCATCAAACAGACCTACCTGAACCAGACTGGAGCTGCTGCACAGAAC
ATTGTCATCTCGGGCCTCGTGTCACCTGATGGCCTGGCCTGTGACTGGGTTGGCAAGAAG
CTGTACTGGACGGACTCCGAGACCAACCGCATTGAGGTTGCCAACCTCAATGGGACGTCC
CGTAAGGTTCTCTTCTGGCAGGACCTGGACCAGCCAAGGGCCATTGCCCTGGATCCTGCA
CATGGGTACATGTACTGGACTGACTGGGGGGAAGCACCCCGGATCGAGCGGGCAGGGATG
GATGGCAGTACCCGGAAGATCATTGTAGACTCCGACATTTACTGGCCCAATGGGCTGACC
ATCGACCTGGAGGAACAGAAGCTGTACTGGGCCGATGCCAAGCTCAGCTTCATCCACCGT
GCCAACCTGGACGGCTCCTTCCGGCAGAAGGTGGTGGAGGGCAGCCTCACTCACCCTTTT
GCCCTGACACTCTCTGGGGACACACTCTACTGGACAGACTGGCAGACCCGCTCCATCCAC
GCCTGCAACAAGTGGACAGGGGAGCAGAGGAAGGAGATCCTTAGTGCTCTGTACTCACCC
ATGGACATCCAAGTGCTGAGCCAGGAGCGGCAGCCTCCCTTCCACACACCATGCGAGGAG
GACAACGGTGGCTGTTCCCACCTGTGCCTGCTGTCCCCGAGGGAGCCTTTCTACTCCTGT
GCCTGCCCCACTGGTGTGCAGTTGCAGGACAATGGCAAGACGTGCAAGACAGGGGCTGAG
GAAGTGCTGCTGCTGGCTCGGAGGACAGACCTGAGGAGGATCTCTCTGGACACCCCTGAC
TTCACAGACATAGTGCTGCAGGTGGGCGACATCCGGCATGCCATTGCCATTGACTACGAT
CCCCTGGAGGGCTACGTGTACTGGACCGATGATGAGGTGCGGGCTATCCGCAGGGCGTAC
CTAGATGGCTCAGGTGCGCAGACACTTGTGAACACTGAGATCAATGACCCCGATGGCATT
GCTGTGGACTGGGTCGCCCGGAACCTCTACTGGACAGATACAGGCACTGACAGAATTGAG
GTGACTCGCCTCAACGGCACCTCCCGAAAGATCCTGGTATCTGAGGACCTGGACGAACCG
CGAGCCATTGTGTTGCACCCTGTGATGGGCCTCATGTACTGGACAGACTGGGGGGAGAAC
CCCAAAATCGAATGCGCCAACCTAGATGGGAGAGATCGGCATGTCCTGGTGAACACCTCC
CTTGGGTGGCCCAATGGACTGGCCCTGGACCTGCAGGAGGGCAAGCTGTACTGGGGGGAT
GCCAAAACTGATAAAATCGAGGTGATCAACATAGACGGGACAAAGCGGAAGACCCTGCTT
GAGGACAAGCTCCCACACATTTTTGGGTTCACACTGCTGGGGGACTTCATCTACTGGACC
GACTGGCAGAGACGCAGTATTGAAAGGGTCCACAAGGTCAAGGCCAGCCGGGATGTCATC
ATTGATCAACTCCCCGACCTGATGGGACTCAAAGCCGTGAATGTGGCCAAGGTTGTCGGA
ACCAACCCATGTGCGGATGGAAATGGAGGGTGCAGCCATCTGTGCTTCTTCACCCCACGT
GCCACCAAGTGTGGCTGCCCCATTGGCCTGGAGCTGTTGAGTGACATGAAGACCTGCATA
ATCCCCGAGGCCTTCCTGGTATTCACCAGCAGAGCCACCATCCACAGGATCTCCCTGGAG
ACTAACAACAACGATGTGGCTATCCCACTCACGGGTGTCAAAGAGGCCTCTGCACTGGAC
TTTGATGTTCCAACAATCACATCTACTGGACTGATGTTAGCCTCAAGACGATCAGCCGAG
CCTTCATGAATGGGAGCTCAGTGGAGCACGTGATTGAGTTTGGCCTCGACTACCCTGAAG
GAATGGCTGTGGACTGGATGGGCAAGAACCTCTATTGGGCGGACACAGGGACCAACAGGA
TTGAGGTGGCCCGGCTGGATGGGCAGTTCCGGCAGGTGCTTGTGTGGAGAGACCTTGACA
ACCCCAGGTCTCTGGCTCTGGATCCTACTAAAGGCTACATCTACTGGACTGAGTGGGGTG
GCAAGCCAAGGATTGTGCGGGCCTTCATGGATGGGACCAATTGTATGACACTGGTAGACA
AGGTGGGCCGGGCCAACGACCTCACCATTGATTATGCCGACCAGCGACTGTACTGGACTG
ACCTGGACACCAACATGATTGAGTCTTCCAACATGCTGGGTCAGGAGCGCATGGTGATAG
CTGACGATCTGCCCTACCCGTTTGGCCTGACTCAATATAGCGATTACATCTACTGGACTG
ACTGGAACCTGCATAGCATTGAACGGGCGGACAAGACCAGTGGGCGGAACCGCACCCTCA
TCCAGGGTCACCTGGACTTCGTCATGGACATCCTGGTGTTCCACTCCTCCCGTCAGGATG
GCCTCAACGACTGCGTGCACAGCAATGGCCAGTGTGGGCAGCTGTGCCTCGCCATCCCCG
GAGGCCACCGCTGTGGCTGTGCTTCACACTACACGCTGGACCCCAGCAGCCGCAACTGCA

Figure 18(b) Continued

GCCCGCCCTCCACCTTCTTGCTGTTCAGCCAGAAATTTGCCATCAGCCGGATGATCCCCG
ATGACCAGCTCAGCCCGGACCTTGTCCTACCCCTTCATGGGCTGAGGAACGTCAAAGCCA
TCAACTATGACCCGCTGGACAAGTTCATCTACTGGGTGGACGGGCGCCAGAACATCAAGA
GGGCCAAGGACGACGGTACCCAGCCCTCCATGCTGACCTCTCCCAGCCAAAGCCTGAGCC
CAGACAGACAGCCACACGACCTCAGCATTGACATCTACAGCCGGACACTGTTCTGGACCT
GTGAGGCCACCAACACTATCAATGTCCACCGGCTGGATGGGGATGCCATGGGAGTGGTGC
TTCGAGGGGACCGTGACAAGCCAAGGGCCATTGCTGTCAATGCTGAGCGAGGGTACATGT
ACTTTACCAACATGCAGGACCATGCTGCCAAGATCGAGCGAGCCTCCCTGGATGGCACAG
AGCGGGAGGTCCTCTTCACCACAGGCCTCATCCGTCCCGTGGCCCTTGTGGTGGACAATG
CTCTGGGCAAGCTCTTCTGGGTGGATGCCGACCTAAAGCGAATCGAAAGCTGTGACCTCT
CTGGGGCCAACCGCCTGACCCTGGAAGATGCCAACATCGTACAGCCAGTAGGTCTGACAG
TGCTGGGCAGGCACCTCTACTGGATCGACCGCCAGCAGCAGATGATCGAGCGCGTGGAGA
AGACCACTGGGGACAAGCGGACTAGGGTTCAGGGCCGTGTCACCCACCTGACAGGCATCC
ATGCCGTGGAGGAAGTCAGCCTGGAGGAGTTCTCAGCCCATCCTTGTGCCCGAGACAATG
GCGGCTGCTCCCACATCTGTATCGCCAAGGGTGATGGAACACCGCGCTGCTCGTGCCCTG
TCCACCTGGTGCTCCTGCAGAACCTGCTGACTTGTGGTGAGCCTCCTACCTGCTCCCCTG
ATCAGTTTGCATGTACCACTGGTGAGATCGACTGCATCCCCGGAGCCTGGCGCTGTGACG
GCTTCCCTGAGTGTGCTGACCAGAGTGATGAAGAAGGCTGCCCAGTGTGCTCCGCCTCTC
AGTTCCCCTGCGCTCGAGGCCAGTGTGTGGACCTGCGGTTACGCTGCGACGGTGAGGCCG
ACTGCCAGGATCGCTCTGATGAAGCTAACTGCGATGCTGTCTGTCTGCCCAATCAGTTCC
GGTGCACCAGCGGCCAGTGTGTCCTCATCAAGCAACAGTGTGACTCCTTCCCCGACTGTG
CTGATGGGTCTGATGACTCATGTGTGAAATCAACAAGCCACCCTCTGATGACATCCCAGC
CCACAGCAGTGCCATTGGGCCCGTCATTGGTATCATCCTCTCCCTCTTCGTCATGGGCGG
GGTCTACTTTGTCTGCCAGCGTGTGATGTGCCAGCGCTACACAGGGGCCAGTGGGCCCTT
TCCCCACGAGTATGTTGGTGGAGCCCCTCATGTGCCTCTCAACTTCATAGCCCCAGGTGG
CTCACAGCACGGTCCCTTCCCAGGCATCCCGTGCAGCAAGTCCGTGATGAGCTCCATGAG
CCTGGTGGGGGGGCGCGGCAGCGTGCCCCTCTATGACCGGAATCACGTCACTGGGGCCTC
ATCCAGCAGCTCGTCCAGCACAAAGGCCACACTATATCCGCCGATCCTGAACCCACCCCC
GTCCCCGGCCACAGACCCCTCTCTCTACAACGTGGACGTGTTTTATTCTTCAGGCATCCC
GGCCACCGCTAGACCATACAGGCCCTACGTCATTCGAGGTATGGCACCCCCAACAACACC
GTGCAGCACAGATGTGTGTGACAGTGACTACAGCATCAGTCGCTGGAAGAGCAGCAAATA
CTACCTGGACTTGAATTCGGACTCAGACCCCTACCCCCCCCCGCCCACCCCCCACAGCCA
GTACCTATCTGCAGAGGACAGCTGCCCACCCTCACCAGGCACTGAGAGGAGTTACTGCCA
CCTCTTCCCGCCCCCACCGTCCCCCTGCACGGACTCGTCCTGA

Figure 18(c)

```
   1   METAPTRAPP  PPPPPLLLLV  LYCSLVPAAA  SPLLLFANRR  DVRLVDAGGV
  51   KLESTIVASG  LEDAAAVDFQ  FSKGAVYWTD  VSEEAIKQTY  LNQTGAAAQN
 101   IVISGLVSPD  GLACDWVGKK  LYWTDSETNR  IEVANLNGTS  RKVLFWQDLD
 151   QPRAIALDPA  HGYMYWTDWG  EAPRIERAGM  DGSTRKIIVD  SDIYWPNGLT
 201   IDLEEQKLYW  ADAKLSFIHR  ANLDGSFRQK  VVEGSLTHPF  ALTLSGDTLY
 251   WTDWQTRSIH  ACNKWTGEQR  KEILSALYSP  MDIQVLSQER  QPPFHTPCEE
 301   DNGGCSHLCL  LSPREPFYSC  ACPTGVQLQD  NGKTCKTGAE  EVLLLARRTD
 351   LRRISLDTPD  FTDIVLQVGD  IRHAIAIDYD  PLEGYVYWTD  DEVRAIRRAY
 401   LDGSGAQTLV  NTEINDPDGI  AVDWVARNLY  WTDTGTDRIE  VTRLNGTSRK
 451   ILVSEDLDEP  RAIVLHPVMG  LMYWTDWGEN  PKIECANLDG  RDRHVLVNTS
 501   LGWPNGLALD  LQEGKLYWGD  AKTDKIEVIN  IDGTKRKTLL  EDKLPHIFGF
 551   TLLGDFIYWT  DWQRRSIERV  HKVKASRDVI  IDQLPDLMGL  KAVNVAKVVG
 601   TNPCADGNGG  CSHLCFFTPR  ATKCGCPIGL  ELLSDMKTCI  IPEAFLVFTS
 651   RATIHRISLE  TNNNDVAIPL  TGVKEASALD  FDVSNNHIYW  TDVSLKTISR
 701   AFMNGSSVEH  VIEFGLDYPE  GMAVDWMGKN  LYWADTGTNR  IEVARLDGQF
 751   RQVLVWRDLD  NPRSLALDPT  KGYIYWTEWG  GKPRIVRAFM  DGTNCMTLVD
 801   KVGRANDLTI  DYADQRLYWT  DLDTNMIESS  NMLGQERMVI  ADDLPYPFGL
 851   TQYSDYIYWT  DWNLHSIERA  DKTSGRNRTL  IQGHLDFVMD  ILVFHSSRQD
 901   GLNDCVHSNG  QCGQLCLAIP  GGHRCGCASH  YTLDPSSRNC  SPPSTFLLFS
 951   QKFAISRMIP  DDQLSPDLVL  PLHGLRNVKA  INYDPLDKFI  YWVDGRQNIK
1001   RAKDDGTQPS  MLTSPSQSLS  PDRQPHDLSI  DIYSRTLFWT  CEATNTINVH
1051   RLDGDAMGVV  LRGDRDKPRA  IAVNAERGYM  YFTNMQDHAA  KIERASLDGT
1101   EREVLFTTGL  IRPVALVVDN  ALGKLFWVDA  DLKRIESCDL  SGANRLTLED
1151   ANIVQPVGLT  VLGRHLYWID  RQQQMIERVE  KTTGDKRTRV  QGRVTHLTGI
1201   HAVEEVSLEE  FSAHPCARDN  GGCSHICIAK  GDGTPRCSCP  VHLVLLQNLL
1251   TCGEPPTCSP  DQFACTTGEI  DCIPGAWRCD  GFPECADQSD  EEGCPVCSAS
1301   QFPCARGQCV  DLRLRCDGEA  DCQDRSDEAN  CDAVCLPNQF  RCTSGQCVLI
1351   KQQCDSFPDC  ADGSDELMCE  INKPPSDDIP  AHSSAIGPVI  GIILSLFVMG
1401   GVYFVCQRVM  CQRYTGASGP  FPHEYVGGAP  HVPLNFIAPG  GSQHGPFPGI
1451   PCSKSVMSSM  SLVGGRGSVP  LYDRNHVTGA  SSSSSSSTKA  TLYPPILNPP
1501   PSPATDPSLY  NVDVFYSSGI  PATARPYRPY  VIRGMAPPTT  PCSTDVCDSD
1551   YSISRWKSSK  YYLDLNSDSD  PYPPPPTPHS  QYLSAEDSCP  PSPGTERSYC
1601   HLFPPPPSPC  TDSS
```

Figure 18(d)


```
1   METAPTRAPPPPPPPLLLLLVLYCSL.VPAAASPLLLFANRRDVRLVDAGG 49
    || ||    |   |    ||||.   |     |||||||||||||||||||||||||
1   MEAAPPGPPWPLLLLLLLLLALCGCPAPAAASPLLLFANRRDVRLVDAGG 50

50  VKLESTIVASGLEDAAAVDFQFSKGAVYWTDVSEEAIKQTYLNQTGAAAQ 99
    ||||||||  ||||||||||||||||||||||||||||||||||||||| |
51  VKLESTIVVSGLEDAAAVDFQFSKGAVYWTDVSEEAIKQTYLNQTGAAVQ 100

100 NIVISGLVSPDGLACDWVGKKLYWTDSETNRIEVANLNGTSRKVLFWQDL 149
    |:||||||||||||||||||||||||||||||||||||||||||||||||
101 NVVISGLVSPDGLACDWVGKKLYWTDSETNRIEVANLNGTSRKVLFWQDL 150

150 DQPRAIALDPAHGYMYWTDWGEAPRIERAGMDGSTRKIIVDSDIYWPNGL 199
    ||||||||||||||||||||||| |||||||||||||||||||||||||
151 DQPRAIALDPAHGYMYWTDWGETPRIERAGMDGSTRKIIVDSDIYWPNGL 200

200 TIDLEEQKLYWADAKLSFIHRANLDGSFRQKVVEGSLTHPFALTLSGDTL 249
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 TIDLEEQKLYWADAKLSFIHRANLDGSFRQKVVEGSLTHPFALTLSGDTL 250

250 YWTDWQTRSIHACNKWTGEQRKEILSALYSPMDIQVLSQERQPPFHTPCE 299
    ||||||||||||| ||  .|||||||||||||||||||||| |||  ||
251 YWTDWQTRSIHACNKRTGGKRKEILSALYSPMDIQVLSQERQPFFHTRCE 300

300 EDNGGCSHLCLLSPREPFYSCACPTGVQLQDNGKTCKTGAEEVLLLARRT 349
    |||||||||||||| ||||.|||||||||||||||:||| ||||||||||
301 EDNGGCSHLCLLSPSEPFYTCACPTGVQLQDNGRTCKAGAEEVLLLARRT 350

350 DLRRISLDTPDFTDIVLQVGDIRHAIAIDYDPLEGYVYWTDDEVRAIRRA 399
    ||||||||||||||||||| ||||||||||||||||||||||||||||||
351 DLRRISLDTPDFTDIVLQVDDIRHAIAIDYDPLEGYVYWTDDEVRAIRRA 400

400 YLDGSGAQTLVNTEINDPDGIAVDWVARNLYWTDTGTDRIEVTRLNGTSR 449
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 YLDGSGAQTLVNTEINDPDGIAVDWVARNLYWTDTGTDRIEVTRLNGTSR 450

450 KILVSEDLDEPRAIVLHPVMGLMYWTDWGENPKIECANLDGRDRHVLVNT 499
    ||||||||||||||| ||||||||||||||||||||||||||.:| ||||
451 KILVSEDLDEPRAIALHPVMGLMYWTDWGENPKIECANLDGQERRVLVNA 500

500 SLGWPNGLALDLQEGKLYWGDAKTDKIEVINIDGTKRKTLLEDKLPHIFG 549
    |||||||||||||||||||||||||||||||:||||||:|||||||||||
501 SLGWPNGLALDLQEGKLYWGDAKTDKIEVINVDGTKRRTLLEDKLPHIFG 550

550 FTLLGDFIYWTDWQRRSIERVHKVKASRDVIIDQLPDLMGLKAVNVAKVV 599
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 FTLLGDFIYWTDWQRRSIERVHKVKASRDVIIDQLPDLMGLKAVNVAKVV 600
```

Figure 18(d) Continued

```
600  GTNPCADGNGGCSHLCFFTPRATKCGCPIGLELLSDMKTCIIPEAFLVFT  649
     |||||||  ||||||||||||  ||:|||||||||||||||||||:|||||||||
601  GTNPCADRNGGCSHLCFFTPHATRCGCPIGLELLSDMKTCIVPEAFLVFT  650

650  SRATIHRISLETNNNDVAIPLTGVKEASALDFDVSNNHIYWTDVSLKTIS  699
     |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
651  SRAAIHRISLETNNNDVAIPLTGVKEASALDFDVSNNHIYWTDVSLKTIS  700

700  RAFMNGSSVEHVIEFGLDYPEGMAVDWMGKNLYWADTGTNRIEVARLDGQ  749
     ||||||||||||:||||||||||||||||||||||||||||||||||||||||
701  RAFMNGSSVEHVVEFGLDYPEGMAVDWMGKNLYWADTGTNRIEVARLDGQ  750

750  FRQVLVWRDLDNPRSLALDPTKGYIYWTEWGGKPRIVRAFMDGTNCMTLV  799
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
751  FRQVLVWRDLDNPRSLALDPTKGYIYWTEWGGKPRIVRAFMDGTNCMTLV  800

800  DKVGRANDLTIDYADQRLYWTDLDTNMIESSNMLGQERMVIADDLPYPFG  849
     ||||||||||||||||||||||||||||||||||||||.||||||:||||
801  DKVGRANDLTIDYADQRLYWTDLDTNMIESSNMLGQERVVIADDLPHPFG  850

850  LTQYSDYIYWTDWNLHSIERADKTSGRNRTLIQGHLDFVMDILVFHSSRQ  899
     ||||||||||||||||||||||||||||||||||||||||||||||||||
851  LTQYSDYIYWTDWNLHSIERADKTSGRNRTLIQGHLDFVMDILVFHSSRQ  900

900  DGLNDCVHSNGQCGQLCLAIPGGHRCGCASHYTLDPSSRNCSPPSTFLLF  949
     ||||||.|.|||||||||||||||||||||||||||||||||||.|||||
901  DGLNDCMHNNGQCGQLCLAIPGGHRCGCASHYTLDPSSRNCSPPTTFLLF  950

950  SQKFAISRMIPDDQLSPDLVLPLHGLRNVKAINYDPLDKFIYWVDGRQNI  999
     |||  |||||||||  ||||:|||||||||||.|||||||||||||||||
951  SQKSAISRMIPDDQHSPDLILPLHGLRNVKAIDYDPLDKFIYWVDGRQNI  1000

1000 KRAKDDGTQPSMLTSPSQSLSPDRQPHDLSIDIYSRTLFWTCEATNTINV  1049
     |||||||||| .||| ||  .|||||||||||||||||||||||||||||
1001 KRAKDDGTQPFVLTSLSQGQNPDRQPHDLSIDIYSRTLFWTCEATNTINV  1050

1050 HRLDGDAMGVVLRGDRDKPRAIAVNAERGYMYFTNMQDHAAKIERASLDG  1099
     |||  |:|||||||||||||||  |||||||:|||||| |||||||.|||
1051 HRLSGEAMGVVLRGDRDKPRAIVVNAERGYLYFTNMQDRAAKIERAALDG  1100

1100 TEREVLFTTGLIRPVALVVDNALGKLFWVDADLKRIESCDLSGANRLTLE  1149
     |||||||||||||||||||||||  ||||||||||||||||||||||||||
1101 TEREVLFTTGLIRPVALVVDNTLGKLFWVDADLKRIESCDLSGANRLTLE  1150

1150 DANIVQPVGLTVLGRHLYWIDRQQQMIERVEKTTGDKRTRVQGRVTHLTG  1199
     ||||||.|||:||:||||||||||||||||||||||||:||||  ||||
1151 DANIVQPLGLTILGKHLYWIDRQQQMIERVEKTTGDKRTRIQGRVAHLTG  1200
```

Figure 18(d) Continued


```
1200 IHAVEEVSLEEFSAHPCARDNGGCSHICIAKGDGTPRCSCPVHLVLLQNL 1249
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 IHAVEEVSLEEFSAHPCARDNGGCSHICIAKGDGTPRCSCPVHLVLLQNL 1250

1250 LTCGEPPTCSPDQFACTTGEIDCIPGAWRCDGFPECADQSDEEGCPVCSA 1299
     ||||||||||||||| ||||||||||||||||||| ||||||||||||||
1251 LTCGEPPTCSPDQFACATGEIDCIPGAWRCDGFPECDDQSDEEGCPVCSA 1300

1300 SQFPCARGQCVDLRLRCDGEADCQDRSDEANCDAVCLPNQFRCTSGQCVL 1349
     .||||||||||||||||||||||||||||||.|||:|||||||| ||||||
1301 AQFPCARGQCVDLRLRCDGEADCQDRSDEADCDAICLPNQFRCASGQCVL 1350

1350 IKQQCDSFPDCADGSDELMCEINKPPSDDIPAHSSAIGPVIGIILSLFVM 1399
     |||||||||||| |||||||||| |||||| |||||||||||||||||||
1351 IKQQCDSFPDCIDGSDELMCEITKPPSDDSPAHSSAIGPVIGIILSLFVM 1400

1400 GGVYFVCQRVMCQRYTGASGPFPHEYVGGAPHVPLNFIAPGGSQHGPFPG 1449
     ||||||||||.||||| ||.||||||||| |||||||||||||||||||| |
1401 GGVYFVCQRVVCQRYAGANGPFPHEYVSGTPHVPLNFIAPGGSQHGPFTG 1450

1450 IPCSKSVMSSMSLVGGRGSVPLYDRNHVTGASSSSSSSTKATLYPPILNP 1499
     | | ||.|||.||.|||| |||||||||||||||||||||||||||||||
1451 IACGKSMMSSVSLMGGRGGVPLYDRNHVTGASSSSSSSTKATLYPPILNP 1500

1500 PPSPATDPSLYNVDVFYSSGIPATARPYRPYVIRGMAPPTTPCSTDVCDS 1549
     ||||||||||||.|.|||| |||| ||||||:|||||||||||||||||
1501 PPSPATDPSLYNMDMFYSSNIPATVRPYRPYIIRGMAPPTTPCSTDVCDS 1550

1550 DYSISRWKSSKYYLDLNSDSDPYPPPPTPHSQYLSAEDSCPPSPGTERSY 1599
     ||| ||||.||||||||||||||||||||||||||||||||||| |||||
1551 DYSASRWKASKYYLDLNSDSDPYPPPPTPHSQYLSAEDSCPPSPATERSY 1600

1600 CHLFPPPPSPCTDSS 1614
     |||||||||||||||
1601 FHLFPPPPSPCTDSS 1615
```

Figure 18(e)

```
 25 CPAPAAASPLLLFANRRDVRLVDAGGVKLESTIVVSGLEDAAAVDFQFSK 74
       |||||||||||||||||||||||||||||||  ||||||||||||||||
 29 .....AASPLLLFANRRDVRLVDAGGVKLESTIVASGLEDAAAVDFQFSK 73

 75 GAVYWTDVSEEAIKQTYLNQTGAAVQNVVISGLVSPDGLACDWVGKKLYW 124
    |||||||||||||||||||||||||  ||:|||||||||||||||||||
 74 GAVYWTDVSEEAIKQTYLNQTGAAAQNIVISGLVSPDGLACDWVGKKLYW 123

125 TDSETNRIEVANLNGTSRKVLFWQDLDQPRAIALDPAHGYMYWTDWGETP 174
    |||||||||||||||||||||||||||||||||||||||||||||||| |
124 TDSETNRIEVANLNGTSRKVLFWQDLDQPRAIALDPAHGYMYWTDWGEAP 173

175 RIERAGMDGSTRKIIVDSDIYWPNGLTIDLEEQKLYWADAKLSFIHRANL 224
    |||||||||||||||||||||||||||||||||||||||||||||||||
174 RIERAGMDGSTRKIIVDSDIYWPNGLTIDLEEQKLYWADAKLSFIHRANL 223

225 DGSFRQKVVEGSLTHPFALTLSGDTLYWTDWQTRSIHACNKRTGGKRKEI 274
    ||||||||||||||||||||||||||||||||||||||||||| || .||||
224 DGSFRQKVVEGSLTHPFALTLSGDTLYWTDWQTRSIHACNKWTGEQRKEI 273

275 LSALYSPMDIQVLSQERQPFFHTRCEEDNGGCSHLCLLSPSEPFYTCACP 324
    |||||||||||||||||||| |||  |||||||||||||||| ||||.||||
274 LSALYSPMDIQVLSQERQPPFHTPCEEDNGGCSHLCLLSPREPFYSCACP 323

325 TGVQLQDNGRTCKAGAEEVLLLARRTDLRRISLDTPDFTDIVLQVDDIRH 374
    |||||||||:|||  ||||||||||||||||||||||||||||||| ||||
324 TGVQLQDNGKTCKTGAEEVLLLARRTDLRRISLDTPDFTDIVLQVGDIRH 373

375 AIAIDYDPLEGYVYWTDDEVRAIRRAYLDGSGAQTLVNTEINDPDGIAVD 424
    |||||||||||||||||||||||||||||||||||||||||||||||||
374 AIAIDYDPLEGYVYWTDDEVRAIRRAYLDGSGAQTLVNTEINDPDGIAVD 423

425 WVARNLYWTDTGTDRIEVTRLNGTSRKILVSEDLDEPRAIALHPVMGLMY 474
    ||||||||||||||||||||||||||||||||||||||| |||||||||
424 WVARNLYWTDTGTDRIEVTRLNGTSRKILVSEDLDEPRAIVLHPVMGLMY 473

475 WTDWGENPKIECANLDGQERRVLVNASLGWPNGLALDLQEGKLYWGDAKT 524
    ||||||||||||||||||.:|  ||||  |||||||||||||||||||||
474 WTDWGENPKIECANLDGRDRHVLVNTSLGWPNGLALDLQEGKLYWGDAKT 523

525 DKIEVINVDGTKRRTLLEDKLPHIFGFTLLGDFIYWTDWQRRSIERVHKV 574
    |||||||:||||||:||||||||||||||||||||||||||||||||||||
524 DKIEVINIDGTKRKTLLEDKLPHIFGFTLLGDFIYWTDWQRRSIERVHKV 573

575 KASRDVIIDQLPDLMGLKAVNVAKVVGTNPCADRNGGCSHLCFFTPHATR 624
    ||||||||||||||||||||||||||||||||  |||||||||||| ||:
574 KASRDVIIDQLPDLMGLKAVNVAKVVGTNPCADGNGGCSHLCFFTPRATK 623
```

Figure 18(e) Continued

625 CGCPIGLELLSDMKTCIVPEAFLVFTSRAAIHRISLETNNNDVAIPLTGV 674
    ||||||||||||||||||||:||||||||||| |||||||||||||||||
624 CGCPIGLELLSDMKTCIIPEAFLVFTSRATIHRISLETNNNDVAIPLTGV 673

675 KEASALDFDVSNNHIYWTDVSLKTISRAFMNGSSVEHVVEFGLDYPEGMA 724
    |||||||||||||||||||||||||||||||||||||||||:||||||||
674 KEASALDFDVSNNHIYWTDVSLKTISRAFMNGSSVEHVIEFGLDYPEGMA 723

725 VDWMGKNLYWADTGTNRIEVARLDGQFRQVLVWRDLDNPRSLALDPTKGY 774
    ||||||||||||||||||||||||||||||||||||||||||||||||||
724 VDWMGKNLYWADTGTNRIEVARLDGQFRQVLVWRDLDNPRSLALDPTKGY 773

775 IYWTEWGGKPRIVRAFMDGTNCMTLVDKVGRANDLTIDYADQRLYWTDLD 824
    ||||||||||||||||||||||||||||||||||||||||||||||||||
774 IYWTEWGGKPRIVRAFMDGTNCMTLVDKVGRANDLTIDYADQRLYWTDLD 823

825 TNMIESSNMLGQERVVIADDLPHPFGLTQYSDYIYWTDWNLHSIERADKT 874
    |||||||||||||.|||||||:||||||||||||||||||||||||||||
824 TNMIESSNMLGQERMVIADDLPYPFGLTQYSDYIYWTDWNLHSIERADKT 873

875 SGRNRTLIQGHLDFVMDILVFHSSRQDGLNDCMHNNGQCGQLCLAIPGGH 924
    |||||||||||||||||||||||||||||.|.||||||||||||||||||
874 SGRNRTLIQGHLDFVMDILVFHSSRQDGLNDCVHSNGQCGQLCLAIPGGH 923

925 RCGCASHYTLDPSSRNCSPPTTFLLFSQKSAISRMIPDDQHSPDLILPLH 974
    ||||||||||||||||||||.||||||||||.|||||||||||||:||||
924 RCGCASHYTLDPSSRNCSPPSTFLLFSQKFAISRMIPDDQLSPDLVLPLH 973

975 GLRNVKAIDYDPLDKFIYWVDGRQNIKRAKDDGTQPFVLTSLSQGQNPDR 1024
    ||||||||.||||||||||||||||||||||||||| .||| || .|||
974 GLRNVKAINYDPLDKFIYWVDGRQNIKRAKDDGTQPSMLTSPSQSLSPDR 1023

1025 QPHDLSIDIYSRTLFWTCEATNTINVHRLSGEAMGVVLRGDRDKPRAIVV 1074
     |||||||||||||||||||||||||||||||| |:||||||||||||||| |
1024 QPHDLSIDIYSRTLFWTCEATNTINVHRLDGDAMGVVLRGDRDKPRAIAV 1073

1075 NAERGYLYFTNMQDRAAKIERAALDGTEREVLFTTGLIRPVALVVDNTLG 1124
     |||||||:||||||| |||||||.|||||||||||||||||||||||| ||
1074 NAERGYMYFTNMQDHAAKIERASLDGTEREVLFTTGLIRPVALVVDNALG 1123

1125 KLFWVDADLKRIESCDLSGANRLTLEDANIVQPLGLTILGKHLYWIDRQQ 1174
     ||||||||||||||||||||||||||||||||||||.|||:||:||||||||
1124 KLFWVDADLKRIESCDLSGANRLTLEDANIVQPVGLTVLGRHLYWIDRQQ 1173

1175 QMIERVEKTTGDKRTRIQGRVAHLTGIHAVEEVSLEEFSAHPCARDNGGC 1224
     ||||||||||||||||:|||| ||||||||||||||||||||||||||||||
1174 QMIERVEKTTGDKRTRVQGRVTHLTGIHAVEEVSLEEFSAHPCARDNGGC 1223

Figure 18(e) Continued

1225 SHICIAKGDGTPRCSCPVHLVLLQNLLTCGEPPTCSPDQFACATGEIDCI 1274
     |||||||||||||||||||||||||||||||||||||||||| |||||||
1224 SHICIAKGDGTPRCSCPVHLVLLQNLLTCGEPPTCSPDQFACTTGEIDCI 1273

1275 PGAWRCDGFPECDDQSDEEGCPVCSAAQFPCARGQCVDLRLRCDGEADCQ 1324
     ||||||||||||| |||||||||||||·||||||||||||||||||||||
1274 PGAWRCDGFPECADQSDEEGCPVCSASQFPCARGQCVDLRLRCDGEADCQ 1323

1325 DRSDEADCDAICLPNQFRCASGQCVLIKQQCDSFPDCIDGSDELMCEITK 1374
     |||||||·|||:||||||||| |||||||||||||||| ||||||||| |
1324 DRSDEANCDAVCLPNQFRCTSGQCVLIKQQCDSFPDCADGSDELMCEINK 1373

1375 PPSDDSPAHSSAIGPVIGIILSLFVMGGVYFVCQRVVCQRYAGANGPFPH 1424
     ||||| |||||||||||||||||||||||||||||||·|||| ||·|||||
1374 PPSDDIPAHSSAIGPVIGIILSLFVMGGVYFVCQRVMCQRYTGASGPFPH 1423

                 .   .   .   .   .
1425 EYVSGTPHVPLNFIAPGGSQHGPFTGIACGKSMMSSVSLMGGRGGVPLYD 1474
     ||| | ||||||||||||||| || | ||·|||·||·|||| |||||
1424 EYVGGAPHVPLNFIAPGGSQHGPFPGIPCSKSVMSSMSLVGGRGSVPLYD 1473

1475 RNHVTGASSSSSSSTKATLYPPILNPPPSPATDPSLYNMDMFYSSNIPAT 1524
     |||||||||||||||||||||||||||||||||||||||·|·|||| ||||
1474 RNHVTGASSSSSSSTKATLYPPILNPPPSPATDPSLYNVDVFYSSGIPAT 1523

1525 VRPYRPYIIRGMAPPTTPCSTDVCDSDYSASRWKASKYYLDLNSDSDPYP 1574
     ||·|||:|||||||||||||||||||||| ||||·||||||||||||||
1524 ARPYRPYVIRGMAPPTTPCSTDVCDSDYSISRWKSSKYYLDLNSDSDPYP 1573

1575 PPPTPHSQYLSAEDSCPPSPATERSYFHLFPPPPSPCTDSS 1615
     |||||||||||||||||||| ||||| |||||||||||||
1574 PPPTPHSQYLSAEDSCPPSPGTERSYCHLFPPPPSPCTDSS 1614

Figure 19

Figure 20

25kb

Minimum region of IBD between
the four susceptible haplotypes,
B,C,D and E

☒ Mouse homology
☒ Putative promotor