(54) RIEG COMPOSITIONS AND THERAPEUTIC AND DIAGNOSTIC USES THEREOF

(75) Inventors: Jeffrey C. Murray, Iowa City, IA (US); Elena Semina, Coralville, IA (US)

(73) Assignee: University of Iowa Research Foundation, Iowa City, IA (US)

(40) Abstract


Marn, M et al. GENBANK Accession No. W59350, Jun. 6, 1996.


* cited by examiner

Primary Examiner—Jeffrey Siew
Assistant Examiner—Brandon Fetterolf
(74) Attorney, Agent, or Firm—Foley Hoag LLP

(57) ABSTRACT

The present invention relates to the discovery of novel genes encoding RIEG polypeptides. Therapeutics, diagnostics and screening assays based on these molecules are also disclosed.

28 Claims, 12 Drawing Sheets
(A)

1  CGAGAACCT  GAGAGGAGC  AGAAAGAAC  TCAGGGTGGC  GGGCAGATT
51  CGGAGGCCCC  AGTTGACCGG  TGAGCTCTT  CGGACTTGG  CACCTGAG
101  AGGCCGTCAGG  TCTCTCAGG  CGCGGCTTCC  GGGCCTTCC  CGTGCAAGGG
151  GACGCGTGGC  TCGCTGAAAA  TGCCCGGCTG  AACAGGTGGG  AGCGGGAGAC
201  AGCAGCTTCC  TCCCGGTAGC  CCGtaagttgg  .......cgc  ttctctcagc

ttcctttct  tctgtttttgc  agATAACGGG  GAATAG  GAG  ACC  AAC  TGC
Met  Glu  Thr  Asn  Cys

242  CGC  AAA  CTG  GTG  TCG  GGG  TGT  CTG  CAA  TTA  Ggtaagaaccccccctctcc
Arg  Lys  Leu  Val  Ser  Ala  Cys  Val  Gln  Leu

286  CAG  CAG  GGG  AAG  AAT  GAG  GAC  GTG  CCC  GAG  GAC  CCC  TCT  AAG
Gln  Gln  Gly  Lys  Asn  Glu  Asp  Val  Gly  Ala  Glu  Asp  Pro  Ser  Lys

331  AAG  AAG  CGG  CAA  AGG  CGG  CAG  CGG  ACT  CAC  TTT  ACC  AGC  CAG  CAG
Lys  Lys  Arg  Glu  Arg  Arg  Glu  Arg  His  Thr  Ser  Ser  Gln  Gln

374  CTC  CAG  CAG  CTG  GAG  CCC  ACT  TTC  CAG  AGG  AAG  CCC  TAC  CGG  GAC
Leu  Gln  Gln  Leu  Glu  Ala  Thr  Phe  Gln  Arg  Asn  Arg  Tyr  Pro  Asp

413  ATG  TCC  ACA  CGG  GAA  GAA  ATC  GCT  TGG  ACC  AAC  CTT  AGC  GAA
Met  Ser  Thr  Arg  Glu  Glu  Ile  Ala  Val  Thr  Asn  Leu  Thr  Glu

464  GCC  CGA  GTC  CGGgtgacgcaagcagggagtctggggag...ctgtttgctcctttgc
Ala  Arg  Val  Arg

C  cccaaaccggccccag  gtt  TGG  TTC  AAG  AAT  AAT  CCG  GCC  AAA  TGG  AGG
Val  Trp  Phe  Lys  Asn  Arg  Arg  Ala  Lys  Trp  Arg

509  AAG  AAG  CGG  CCC  AAG  CAG  CAG  CCC  GAG  CTA  TGC  AAG  AAT  GGC  TTC
Lys  Arg  Glu  Arg  Gln  Gln  Ala  Glu  Leu  Cys  Lys  Asn  Gly  Phe

554  GGG  CGG  CAG  TCC  AAT  GGG  CAC  AGT  CAG  CCC  TAC  GAC  GAC  ATG  TAC
Gly  Pro  Gln  Phe  Asn  Gly  Leu  Met  Gln  Pro  Tyr  Asp  Asp  Met  Tyr

FIGURE 1A
(B)  
1  CACCGAGCGG CTCCCCTAAAG AGGCTCGCAG CGGTTCCCTC CTGCCCAACGT CCCCACGCTCT  
61  GCCTGGCCCC CCCCCTCTTT CGGCTGCGCA ACTCTTTTCT GCTGGAGCTCT GAAGCTAGAG  
121  GAGACAGGGG TGGAGGATTG GCCAGTTTGC GTTCCCTGGC TCTTTCAGAT CTCGGCTAAC  
181  ACGGGGACAC TTTGGCCGCTA GAGGAAAGCC GGCACCCGGG GCCACCGCAG GCCACCTGCG  
241  CGGCTAGCCCT GGACTCATAG GGCTCCGCAC TCCCTGCTCACG TAC TGC ATG AAA GGC  

    Met Asn Cys Met Lys Gly  

CCG CTG CCC TTG GAG CAC CGA GCA GCC GGG ACT AAG CTG TCG GCC GCC TCC  
Pro Leu Pro Leu Glu His Arg Ala Ala Gly Thr Lys Leu Ser Ala Ala Ser  

TCA CCC TTC TGT CAC CAT CCC CAG GCG TTA GCC ATG GCT TCG GTC CTA GCT  
Ser Pro Pro Cys His His Pro Gln Ala Leu Ala Met Ala Ser Val Leu Ala  

CCT GGC CAG CCC CGC TCC TTG GAC TCC TCC AAA CAT AGA CTG GAG GTG CAT  
Pro Gly Gln Pro Arg Ser Leu Asp Ser Ser Lys His Arg Leu Glu Val His  

    #  

AGA ATC TCC GAT ACT TCC AGC CCT GAA GTC GCA  
Thr Ile Ser Asp Thr Ser Ser Ser Pro Glu Val Ala  

(C)  

ACCGGGCCCA GGGCGGGGGG ATCCGAGGAC TGTCCAGGTG GGCAACTCTG CCCCGAAGAA  
ACTGAGAATT GTCTAGAAG TTGGTGCCCA CTATGGGAAG GAAGAGGGGG GAAAAAGAGAT  
CAGAGGAAA GAAACCACTG AATCCAAGAG GAGACGGCCT TTTGATCAGA AGGAAATGGCC  
CCAAGTGCT ACCTGGTTCG CTAAGAGTAT TCCCAACAGT TGGAGAGGCC GTACGCCAC  
AAATGTTTTGA CTCGATATGA CATTTTTACA TTACTATAAG TTGTATATT TTTAGATTAA  
GCATGTTTTAA CATTAAAAAG ACTGAAGAGA TGGATATATG TCAGAAGTTT AATTTTTT  
TATAAAAAGCA GTTGTATGATT ATATCAGACG ATGTGTATTA AAAGCCTAGGC TTAAAAATAA  
AGCATGTTAT ACAAATCACG TAAGATTTTT TCGGTGCGGA GCAAAGGAAG GTATATACTA  
AATGCCCACAC TGTAAGTTTC TAAACATTAT TTATT

FIGURE 1B
<table>
<thead>
<tr>
<th>gene name</th>
<th>homeo domain sequence</th>
<th>identity</th>
<th>SEQ ID NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solurshin</td>
<td>QRRQRTHTFS QQLQBLEARF</td>
<td>97%</td>
<td>SEQ ID NO. 7</td>
</tr>
<tr>
<td>ptx-1</td>
<td>--------- ---------</td>
<td>83%</td>
<td>SEQ ID NO. 8</td>
</tr>
<tr>
<td>unc-30</td>
<td>--------- ---------</td>
<td>62%</td>
<td>SEQ ID NO. 9</td>
</tr>
<tr>
<td>otx-1</td>
<td>B-T--R S-DV---L-</td>
<td>62%</td>
<td>SEQ ID NO. 20</td>
</tr>
<tr>
<td>otx-2</td>
<td>B-T--R A-DV---L-</td>
<td>62%</td>
<td>SEQ ID NO. 11</td>
</tr>
<tr>
<td>drd-ode</td>
<td>B-T--K A-DV---L-</td>
<td>62%</td>
<td>SEQ ID NO. 12</td>
</tr>
<tr>
<td>goosecoid</td>
<td>K--H--I--D E--SA--NL-</td>
<td>53%</td>
<td>SEQ ID NO. 13</td>
</tr>
<tr>
<td>pax-7</td>
<td>S--T--A E--B--KA--</td>
<td>65%</td>
<td>SEQ ID NO. 14</td>
</tr>
<tr>
<td>pax-3</td>
<td>S--T--A E--B--KA--</td>
<td>60%</td>
<td>SEQ ID NO. 15</td>
</tr>
<tr>
<td>paired</td>
<td>C--T--SA S--D--RA--</td>
<td>58%</td>
<td>SEQ ID NO. 16</td>
</tr>
<tr>
<td>rat drg11</td>
<td>N--T--AL E--A--V--</td>
<td>67%</td>
<td>SEQ ID NO. 17</td>
</tr>
<tr>
<td>prd-like</td>
<td>Y--T--AL F--E--KA--</td>
<td>67%</td>
<td>SEQ ID NO. 18</td>
</tr>
<tr>
<td>dm-1al</td>
<td>Y--T--AL F--E--KA--</td>
<td>67%</td>
<td>SEQ ID NO. 19</td>
</tr>
<tr>
<td>cart-1</td>
<td>K--H--I--D L--B--KV--</td>
<td>65%</td>
<td>SEQ ID NO. 20</td>
</tr>
<tr>
<td>prx-1</td>
<td>N--T--N S--A--RV--</td>
<td>62%</td>
<td>SEQ ID NO. 21</td>
</tr>
<tr>
<td>prx-2</td>
<td>N--T--N S--A--RV--</td>
<td>62%</td>
<td>SEQ ID NO. 22</td>
</tr>
<tr>
<td>phox-2</td>
<td>I--T--I-- A--K--RV--</td>
<td>52%</td>
<td>SEQ ID NO. 23</td>
</tr>
<tr>
<td>murine hmk2</td>
<td>N--T--N S--A--RV--</td>
<td>62%</td>
<td>SEQ ID NO. 24</td>
</tr>
<tr>
<td>murine chx10</td>
<td>K--H--I--Y--E--KA--</td>
<td>60%</td>
<td>SEQ ID NO. 25</td>
</tr>
<tr>
<td>murine otp</td>
<td>K--H--R--F A--N--RS--</td>
<td>58%</td>
<td>SEQ ID NO. 26</td>
</tr>
<tr>
<td>Gene Name</td>
<td>Sequence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solushin</td>
<td>SSLASLRAKQHS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>murine ptx-1</td>
<td>I--I-N-I--I-EY-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chick prx-2</td>
<td>N-I-A--M--RE--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chick prx-1</td>
<td>N-I-V-M--E-T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>human Cart-1</td>
<td>T-I-R--R--LE-T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dm-aristaless</td>
<td>N-I-A--M--RE--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>murine hmk2</td>
<td>A-V-A--M--RE--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rat dreg11</td>
<td>A-V-A--M--RE--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>murine chx10</td>
<td>A-V-A--M--RE--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>murine otp</td>
<td>A-V-A--M--RE--</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Genomic Structure of the Human Gene

Figure 3

- 14 amino acid conserved region
- Homeobox
- 270 bp 3' UTR Conserved
- cDNA Sequence
- CpG Islands

MS
- 5.15 kb
- 49 bp
- 222 bp

- 3R
- 3F
- 2F
- 1R
- 1F

TGA
- 730 bp
- 206 bp
- 4F
- 3R

cont
FIGURE 4E

Family 5

Unaffected

Affected

SEQ ID NO. 47

SEQ ID NO. 48
FIGURE 4F

(F)

Family 6

Unaffected

Affected

CA ACTGGCCGCCA

ACA ACTGAGCCGCCA

SEQ ID NO. 49

SEQ ID NO. 50
RIEG COMPOSITIONS AND THERAPEUTIC
AND DIAGNOSTIC USES THEREFOR

CROSS-REFERENCE TO RELATED
APPLICATIONS

This application is a continuation of U.S. patent application
Ser. No. 08/754,477, now U.S. Pat. No. 6,518,411, filed
on Nov. 22, 1996, the contents of which are specifically
incorporated by reference herein.

BACKGROUND OF THE INVENTION

Rieger syndrome (RS) is an autosomal dominant disorder
of morphogenesis characterized by abnormalities of the
anterior segment of the eye and dental hypoplasia. Other
systemic anomalies have been reported in association with
the syndrome, the most common of which are mild cranio-
facial dysmorphism, including maxillary hypoplasia and
failure of involution of the periamphibial skin and enopa-
locele. Gastrointestinal defects reported include Meckels
diverticulum (Kreps, Y. P. and D. Pertschulidis (1979) Am.
J. Gastroenterol. 71:608–610) colon atresia (Rogers, R. C.
51:438–440) The ocular features can include a prominent,
anteri orly displaced Schwalbe line, iris processes which
insert into the Schwalbe line, hypoplasia of the anterior iris
stroma or corectopia and pseudopolycoria. Roughly one-
half of the patients will develop glaucoma, usually by young
adulthood. The dental features associated with the condition
include oligodontia and microdontia.

There is general agreement that patients with both the
ocular features and systemic abnormalities (dental, umbili-
cal, craniofacial) should be designated Rieger syndrome.
The nomenclature of patients with only the ocular features,
however, is confusing and disputed. For example, the con-
dition of only ocular abnormalities is sometimes referred to
as Rieger syndrome and other times referred to as Axenfeld-
Rieger anomaly.

The underlying genetic defect and the chromosomal local-
ization of RS is uncertain. Cases with Axenfeld-Rieger
anomaly have been reported with various aberrations of
chromosomes 4, 6, 10, 13, 16, and 22 (Wilcox, L. M. et al.,
developed RS, including the dental and umbilical anomalies,
have been associated with deletions in the region of chro-
mosome 4q23–26. One case of a deletion in the region of
4q26 which did not have RS has also been reported. These
cases indicate that a gene for RS lies in the region of 4q25.

In RS, a high risk of developing increased intraocular
pressures and subsequent glaucomatous damage necessitats
vigorous screening procedures beginning in infancy. Current
screening for the ocular abnormalities of RS requires slit
lamp examination, tonometry, gonioscopy and a dilated
examination of the optic nerve head.

SUMMARY OF THE INVENTION

The present invention is based on the discovery of novel
molecules, referred to herein as RIEG nucleic acids and
RIEG or “Solushrin” polypeptide molecules. The human
RIEG gene, which is approximately 18 kb, consists of four
exons, 222, 49, 206 and 1306 nucleotides in length. The
ATG initiation codon is located in the second exon and the
homeobox region in the third and fourth exons. Translation
of the open reading frame yields a protein of about 271
amino acids.

The gene is expressed in oral epithelium, the umbilicus,
and periorcular mesoderm, consistent with the phenotypic
abnormalities seen in Rieger syndrome patients. In addition,
the gene is expressed in Rathke’s pouch and the vitelline
artery.

In one aspect, the invention features isolated vertebrate
RIEG nucleic acid molecules. The disclosed molecules can
be non-coding, (e.g. probe, antisense or ribozyme mole-
cules) or can encode a functional RIEG polypeptide (e.g. a
polypeptide which specifically modulates, e.g., by acting as
either an agonist or antagonist, at least one bioactivity of
the human RIEG polypeptide). In one embodiment, the nucleic
acids of the present invention can hybridize to a vertebrate
RIEG gene or to the complement of a vertebrate RIEG gene.
In a further embodiment, the claimed nucleic acid hybridizes
with the coding sequence designated in at least one of SEQ
ID Nos: 1 or 3 or to the complement to the coding sequence
designated in at least one of SEQ ID Nos: 1 or 3. In a
preferred embodiment, the hybridization is conducted under
mildly stringent or stringent conditions.

In a further embodiments, the nucleic acid molecule is an
RIEG nucleic acid that is at least 70%, preferably 80%, more
preferably 85%, and even more preferably at least 95%
homologous in sequence to the nucleic acids shown as SEQ
ID Nos: 1 or 3 or to the complement of the nucleic acids
shown as SEQ ID Nos: 1 or 3. In another embodiment, the
RIEG nucleic acid molecule encodes a polypeptide that is at
least 90% and more preferably at least 94% similar in
sequence to the polypeptide shown in SEQ ID No: 2.

The invention also provides probes and primers compris-
ing substantially purified oligonucleotides, which corre-
spond to a region of nucleotide sequence which hybridizes
to at least 6 consecutive nucleotides of the sequences set
forth as SEQ ID Nos: 1 or 3 or complements of the
sequences set forth as SEQ ID Nos: 1 or 3, or naturally
occurring mutants thereof. In preferred embodiments, the
probe/primer further includes a label group attached thereto,
which is capable of being detected.

For expression, the subject RIEG nucleic acids can include
a transcriptional regulatory sequence, e.g. at least one of a
transcriptional promoter (e.g., for constitutive expression
or inducible expression) or transcriptional enhancer sequence,
which regulatory sequence is operably linked to the RIEG
genome sequence. Such regulatory sequences in conjunction
with a RIEG nucleic acid molecule can be useful vectors for
gene expression. This invention also describes host cells transfect ed
with said expression vector whether prokaryotic or eukaryotic
and in vitro (e.g. cell culture) and in vivo (e.g. transgenic)
methods for producing RIEG proteins by employing said expression
vectors.

In another aspect, the invention features isolated RIEG
or Solushrin polypeptides, preferably substantially pure prep-
nations e.g. of plasma purified or recombinantly produced
RIEG polypeptides. In preferred embodiments, the polypep-
tide is a functional transcription factor.

In one embodiment, the polypeptide is identical to or
similar to a RIEG protein represented in SEQ ID No: 2.
Related members of the vertebrate and particularly the
mammalian RIEG family are also within the scope of the
invention. Preferably, a RIEG polypeptide has an amino acid sequence at least 60% homologous and preferably at least 80% homologous to the polypeptide represented in SEQ ID No: 2. The subject RIEG proteins also include modified protein, which are resistant to post-translation modification, as for example, due to mutations which alter modification sites (such as tyrosine, threonine, serine or asparagine residues), or which prevent glycosylation of the protein, or which prevent interaction of the protein with intracellular proteins involved in signal transduction.

The RIEG polypeptide can comprise a full length protein, such as represented in SEQ ID No: 2, or it can comprise a fragment corresponding to one or more particular motifs/domains, or to arbitrary sizes, e.g., at least 5, 10, 25, 50, 100, 150, 175, 200, 225, 250 or 260 amino acids in length.

Another aspect of the invention features chimeric molecules (e.g. fusion proteins) comprised of a RIEG protein. For instance, the RIEG protein can be provided as a recombinant fusion protein which includes a second polypeptide portion, e.g., a second polypeptide having an amino acid sequence unrelated (heterologous) to the RIEG polypeptide (e.g. the second polypeptide portion is glutathione-S-transferase, an enzymatic activity such as alkaline phosphatase or an epitope tag).

Yet another aspect of the present invention concerns an immunogen comprising a RIEG polypeptide in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for a RIEG polypeptide; e.g., a humoral response, an antibody response and/or cellular response. In preferred embodiments, the immunogen comprises an antigenic determinant, e.g. a unique determinant, from the protein represented in SEQ ID No: 2.

A still further aspect of the present invention features antibodies and antibody preparations specifically reactive with an epitope of the RIEG protein. In preferred embodiments the antibody specifically binds to an epitope represented in SEQ ID No: 2.

The invention also features transgenic non-human animals which include (and preferably express) a heterologous form of a RIEG gene described herein, or which misexpress an endogenous RIEG gene (e.g., an animal in which expression of one or more of the subject RIEG proteins is disrupted). Such a transgenic animal can serve as an animal model for studying cellular and tissue disorders comprising mutated or mis-expressed RIEG alleles or for use in drug screening. Alternatively, such a transgenic animal can be useful for expressing recombinant RIEG polypeptides.

In yet another aspect, the invention provides assays, e.g., for screening test compounds to identify inhibitors, or alternatively, potentiation of an interaction between a RIEG protein and, for example, a virus, an extracellular ligand of the RIEG protein, or an intracellular protein which binds to the RIEG protein. An exemplary method includes the steps of (i) combining a RIEG polypeptide or bioactive fragments thereof, a RIEG target molecule (such as a RIEG ligand or a RIEG substrate), and a test compound, e.g., under conditions wherein, but for the test compound, the RIEG protein and target molecule are able to interact; and (ii) detecting the formation of a complex which includes the RIEG protein and the target polypeptide either by directly quantitating the complex, by measuring inductive effects of the protein, or, in the instance of a substrate, measuring the conversion to product. A statistically significant change, such as a decrease, in the interaction of the RIEG and target molecule in the presence of a test compound (relative to what is detected in the absence of the test compound) is indicative of a modulation (e.g., inhibition or potentiation of the interaction between the RIEG protein and the target molecule).

Yet another aspect of the present invention concerns a method for modulating the transcription of certain genes in a cell by modulating RIEG bioactivity, (e.g., by potentiating or disrupting RIEG bioactivity). In general, whether carried out in vivo, in vitro, or in situ, the method comprises treating the cell with an effective amount of a RIEG therapeutic so as to alter, relative to the cell in the absence of treatment, the level of transcription of certain genes. Accordingly, the method can be carried out with RIEG therapeutics such as peptide and peptidomimetics or other molecules identified in the above-referenced drug screens which agonize or antagonize the effects of signaling from a RIEG protein or ligand binding of a RIEG protein. Other RIEG therapeutics include antisense constructs for inhibiting expression of RIEG proteins, and dominant negative mutants of RIEG proteins which competitively inhibit ligand interactions upstream and signal transduction downstream of the wild-type RIEG protein.

A further aspect of the present invention provides a method of determining if a subject is at risk for Riesger Syndrome or other disorder. The method includes detecting, in a tissue of the subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a RIEG protein, e.g., represented in one of SEQ ID Nos: 1 or 3 or a homolog thereof; or (ii) the mis-expression of a RIEG gene. In preferred embodiments, detecting the genetic lesion includes ascertaining the existence of at least one of: a deletion of one or more nucleotides from a RIEG gene; an addition of one or more nucleotides to the gene, a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene; an alteration in the level of a messenger RNA transcript of the gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; a non-wild type level of the protein; and/or an aberrant level of soluble RIEG protein.

For example, detecting the genetic lesion can include (i) providing a probe/primer comprised of an oligonucleotide which hybridizes to a sense or antisense sequence of a RIEG gene or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with the RIEG gene; (ii) contacting the probe/primer to an appropriate nucleic acid containing sample; and (iii) detecting, by hybridization of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion; e.g., wherein detecting the lesion comprises utilizing the probe/primer to determine the nucleotide sequence of the RIEG gene and, optionally, of the flanking nucleic acid sequences. For instance, the primer can be employed in a polymerase chain reaction (PCR) or in a ligation chain reaction (LCR). In alternate embodiments, the level of a RIEG protein is detected in an immunoassay using an antibody which is specifically immunoreactive with the RIEG protein.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1. a) Shows the nucleotide sequence of the human (h) RIEG gene including a portion of the 5' and 3' untranslated region (UTR) (SEQ. ID. NO.1). In addition, the deduced amino-acid sequence of the human RIEG gene is shown (SEQ. ID. NO.2) based on the coding sequence of h RIEG (SEQ. ID. NO.3). The intron sequence (shown in
lower case) was determined only at intron-exon junctions. The start (ATG) and stop (TGA) codons are in boldface, as are the conserved gt and ag sequences flanking each intron and the polyadenylation signals in the 3'UTR. Multiple stop codons 5' from the initiating methionine are underlined and the in-frame stop codon is boxed. The 5' ATG is shown in bold italic. The homeobox has a solid underline, the conserved 14-amino-acid sequence has a dashed underline, and the conserved region in the 3' UTR is italicized. The nucleotide sequence of the mouse (m)Rieg gene is shown in comparison including the 5' and 3' UTR (SEQ. ID. NO.4); and the deduced amino acid sequence of the mouse Rieg gene (SEQ. ID. NO. 5) based on the coding sequence of mRieg (SEQ. ID. NO.6) is shown in comparison to the human. Where mRieg and hRieg sequences differ, the mouse Rieg nucleotides appear above their corresponding Rieg nucleotides in the sequence. The homologous coding region, beginning within the third exon (start point marked with an arrow), is indicated by an initial square bracket. Two amino-acid changes in the protein sequence are shown italicized above the main sequence as well. Intron-exon junction sequences have been identified only for RIEG. Nucleotides below the main sequence indicate the positions of and nucleotide changes in the RGS families’ mutations; corresponding amino-acid changes are shown as well. The numbers at left and right indicate positions within the cdNA and protein sequences, respectively. The region spanning from nt 950–1760 had significant homology with EST sequences H95407, T64905, and H13139. b) cdNA sequence from the 3'UTR of Rieg. The region sharing 97% homology with the human RIEG 3'UTR is italicized.

FIG. 2. a) Alignment of the solushin homedomain region with the homedomains of related proteins. The conserved lysine at position 9 of the third helix is shown in bold. b) Alignment of the 14-amino-acid region in the C-terminal end of solushin with protein sequences of other homedomain-containing genes.

FIG. 3. Genomic organization of the human RIEG gene. The numbered boxes represent the 4 exons; the horizontal lines, introns, with intrinsic sizes indicated by the numbers above them. The line above the second exon shows the position of the 319N1 CpG-island genomic probe used for isolation of the RIEG cdNA. The lines 187,8 and 163 represents cosmid from the cosmid contig comprising the Rieg region. The translation start (ATG) and stop (TGA) signals are indicated. The boxed black box represents the homeobox; the grey box, the 14-amino acid conserved element; and hatched, the conserved region of the 3' UTR of the gene. The arrows indicated the amplification primers used for mutation screening.

FIG. 4 is a schematic of the results of mutation screening in RIEG families. (A), (B), (C), (D) represent the pedigree and the segregation of the mutated allele in the families 1, 2, 3, 4 respectively. (E), (F) represent proband with de novo RIEG mutations. Open squares, unaffected males; closed squares, affected males; open circles, unaffected females; closed circles, affected females. Corresponding sequences of the amplified genomic DNA from unaffected and affected individuals from the family are shown to the right of each pedigree and the position their mutation occurred is pointed on both pictures with arrows.

FIG. 5. Whole mount in situ hybridization on mouse day 11 embryos with sense (A) and anti-sense (B) digoxigen labeled riboprobes derived from the 3'UTR of mouse Rieg cdNA. Marked regions of embryo are: eye (e), maxillary (mx) and mandibular (md) and in the umbilical cord (u). Bars: 0.57 mm.

FIG. 6. Shows brightfield (A, B, C) and darkfield (D, E, F) micrographs of sections through the head (A and B) and midgut regions (C) of day 11 mouse embryo. Note the signal around the eye (e) and in Rathke’s pouch (r) in (D) (arrows); in the epithelium of the maxillary (mx) and mandibular (md) processes in (E) (arrows); and in (F) at the juncture of the limb (l) to the body, the dorsal mesentery (dm) the vitelline (v) and umbilical (u) vessels (arrows). Bars: d and E, 0.37 mm; F, 0.42 mm.

DETAILED DESCRIPTION OF THE INVENTION

4.1. General

The present invention is based on the discovery of a family of novel genes, referred to herein as “RIEG” gene. The RIEG genes or gene products provided by the present invention are exemplified by hRIEG and mRIEG. The mRNA transcripts including 5' and 3' untranslated regions as well as the coding region; and the deduced amino acid sequences of hRIEG is shown in FIG. 1 and in SEQ ID Nos: 1, 2 and 3. The mRNA transcripts including 5' and 3' untranslated regions as well as the coding region; and the deduced amino acid sequences of mRIEG is shown in FIG. 2 and in SEQ ID Nos: 4, 5 and 6.

Based on in situ hybridization, the RIEG gene has been detected in the following tissues: eye mesenchyme, maxillary and mandibular epithelium, Rathke’s pouch, limb-body junction, and umbilicus.

The gene itself has several interesting features. First, comparison of the mouse sequence shows there is strong amino acid conservation throughout the gene, particularly in the homedomain. In addition, there is also a region of strong nucleotide conservation located in the 3' untranslated region of the gene between human and mouse nucleotide sequences. The strongest region of homology extends over 270 nucleotides, beginning 200 base pairs 3' of the stop codon and spanning the polyadenylation signal where 261/270 bp are shared (97%). This strong region of conservation suggests that this is an important and conserved regulatory element.

Another interesting region of sequence conservation is a 14-amino acid sequence in the C-terminal of the protein that is identical to Pto 1 (Lamonerie T., et al. (1996) Genes Dev. 10:1284–1295) and similar to sequences in genes found in Drosophila, chicken, mouse, and human with craniofacial-specific expression. Such sequences may be targeting recognition sequences that have regional specificity.

As reported herein, six mutations were identified in six unrelated families that both cosegregate with the disorder and were not identified in screens of 200 unaffected individuals drawn from populations of similar ethnic and geographic origin. These mutations were identified in regions that either demonstrated conserved amino acid homology between the human and the mouse, generated stop codons within the gene, or were found immediately adjacent to splice sites, suggesting that a disruption of splicing could result from the mutation. This provides compelling evidence for the role of this gene in the etiology of this disorder.

Further, the expression pattern in the mouse strongly overlaps with the human phenotypic abnormalities (e.g. oral epithelium and dental hypoplasia, perioral mesenchyme and anterior chamber abnormalities and umbilical) providing further evidence for the role of the RIEG gene in the etiology of relevant disorders, as well as strong evidence of conservation between vertebrates. In addition, there is strong expression of the gene in Rathke’s pouch, the pituitary
anlage, and at least one reported Rieger family has had
growth hormone deficiency as part of the phenotype (Fetin-
with the discovered strong conservation in the homeobox
region suggest that RIEG is important in pituitary develop-
ment and gene expression.

Accordingly, certain aspects of the present invention
relate to nucleic acid molecules encoding vertebrate RIEG
proteins, the RIEG proteins, antibodies immunoreactive
with RIEG proteins, and preparations of such compositions.
In addition, drug discovery assays are provided for identi-
fying agents which can modulate the biological function of
RIEG proteins, such as by altering the interaction of verte-
brate RIEG molecules with either downstream or upstream
elements in the signal transduction pathway. Such agents
can be useful therapeutically as further described herein.
Moreover, the present invention provides diagnostic and ther-
apeutic assays and reagents for detecting and treating disor-
der involving, for example, aberrant expression (or loss thereto) of vertebrate RIEG genes. Other aspects of the
invention are described below or will be apparent to those
skilled in the art in light of the present disclosure.

4.2 Definitions

For convenience, the meaning of certain terms and
phrases employed in the specification, examples, and
appendix claims are provided below.

"Cells," "host cells" or "recombinant host cells" are terms
used interchangeably herein. It is understood that such terms
refer not only to the particular subject cell but to the progeny
or potential progeny of such a cell. Because certain modi-
fications may occur in succeeding generations due to either
mutation or environmental influences, such progeny may
not, in fact, be identical to the parent cell, but are still
included within the scope of the term as used herein.

A "chimeric protein" or "fusion protein" is a fusion of a
first amino acid sequence encoding one of the subject RIEG
polypeptides with a second amino acid sequence defining a
domain (e.g. polypeptide portion) foreign to and not sub-
stantially homologous with any domain of one of the RIEG
proteins. A chimeric protein may present a foreign domain
which is found (albeit in a different protein) in an organism
which also expresses the first protein, or it may be an
"interspecies", "intergeneric", etc. fusion of protein structures
expressed by different kinds of organisms. In general, a
fusion protein can be represented by the general formula
X-RIEG-Y, wherein RIEG represents a portion of the protein
which is derived from one of the RIEG proteins, and X and
Y are independently absent or represent amino acid
sequences which are not related to one of the RIEG
sequences in an organism, including naturally occurring
mutants.

"Complementary" sequences as used herein refer to
sequences which have sufficient complementarity to be able
to hybridize, forming a stable duplex.

A "delivery complex" shall mean a targeting means (e.g.
a molecule that results in higher affinity bonding of a gene,
protein, polypeptide or peptide to a target cell surface and/or
increased cellular uptake by a target cell). Examples of
targeting means include: steroids (e.g. cholesterol), lipids
(e.g. a cationic lipid, virosome or liposome), viruses (e.g.
adenovirus, adeno-associated virus, and retrovirus) or target
cell specific binding agents (e.g. ligands recognized by
target cell specific receptors). Preferred complexes are suf-
ficiently stable in vivo to prevent significant uncoupling
prior to internalization by the target cell. However, the
complex is cleavable under appropriate conditions within
the cell so that the gene, protein, polypeptide or peptide is
released in a functional form.

As is well known, genes for a particular polypeptide may
exist in single or multiple copies within the genome of an
individual. Such duplicate genes may be identical or may
have certain modifications, including nucleotide substitu-
tions, additions or deletions, which all still code for polypep-
tides having substantially the same activity. The term "DNA
sequence encoding a RIEG polypeptide" may thus refer to
one or more genes within a particular individual. Moreover,
certain differences in nucleotide sequences may exist
between individual organisms, which are called alleles. Such
allelic differences may or may not result in differences in
amino acid sequence of the encoded polypeptide yet still
encode a protein with the same biological activity.

As used herein, the term "gene" or "recombinant gene"
refers to a nucleic acid molecule comprising an open reading
frame encoding one of the RIEG polypeptides of the present
invention, including both exon and (optionally) intron
sequences. A "recombinant gene" refers to nucleic acid
encoding a RIEG polypeptide and comprising RIEG-encod-
ing exon sequences, though it may optionally include intron
sequences which are either derived from a chromosomal
RIEG gene or from an unrelated chromosomal gene. Exem-
plary recombinant genes encoding the subject RIEG
polypeptides are represented in the appended Sequence
Listing. The term "intron" refers to a DNA sequence present
in a given RIEG gene which is not translated into protein and
is generally found between exons.

"Homology" or "identity" or "similarity" refers to
sequence similarity between two peptides or between two
nucleic acid molecules. Homology can be determined by
comparing a position in each sequence which may be
aligned for purposes of comparison. When a position in the
compared sequence is occupied by the same base or amino
acid, then the molecules are homologous at that position. A
degree of homology between sequences is a function of the
number of matching or homologous positions shared by the
sequences. An "unrelated" or "non-homologous" sequence
shares less than 40% identity, though preferably less than
25% identity, with one of the RIEG sequences of the present
invention.

The term "interact" as used herein is meant to include
detectable interactions between molecules, such as can be
detected using, for example, a yeast two hybrid assay. The
term interact is also meant to include "binding" interactions
between molecules. Interactions may be protein—protein or
protein-nucleic acid in nature.

The term "isolated" as used herein with respect to nucleic
acids, such as DNA or RNA, refers to molecules separated
from other DNAs, or RNAs, respectively, that are present in
the natural source of the nucleic molecule. For example, an
isolated nucleic acid encoding one of the subject RIEG
polypeptides preferably includes no more than 10 kilobases
(kb) of nucleic acid sequence which naturally immediately
flanks the RIEG gene in genomic DNA, more preferably no
more than 5 kb of such naturally occurring flanking
sequences, and most preferably less than 1.5 kb of such
naturally occurring flanking sequence. The term isolated as
used herein also refers to a nucleic acid or peptide that is
substantially free of cellular material, viral material, or
culture medium when produced by recombinant DNA tech-
niques, or chemical precursors or other chemicals when
chemically synthesized. Moreover, an “isolated nucleic
acid” is meant to include nucleic acid fragments which are
not naturally occurring as fragments and would not be found
in the natural state. The term "isolated" is also used herein
to refer to polypeptides which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides.

The term "modulation" as used herein refers to both upregulation, i.e., stimulation, and downregulation, i.e., suppression, of a response.

The "non-human animals" of the invention include mammals such as rodents, non-human primates, sheep, dog, cow, chickens, amphibians, reptiles, etc. Preferred non-human animals are selected from the rodent family including rat and mouse, most preferably mouse, though transgenic amphibians such as members of the Xenopus genus, and transgenic chickens can also provide important tools for understanding and identifying agents which can affect, for example, embryogenesis and tissue formation. The term "chimeric animal" is used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that one of the recombinant RIEG genes is present and/or expressed or disrupted in some tissues but not others.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

As used herein, the term "promoter" means a DNA sequence that regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in cells. The term encompasses "tissue specific" promoters, i.e. promoters, which effect expression of the selected DNA sequence only in specific cells (e.g. cells of a specific tissue). The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well. The term also encompasses non-tissue specific promoters and promoters that constitutively express or that are inducible (i.e. expression levels can be controlled).

The terms "protein", "polypeptide" and "peptide" are used interchangeably herein when referring to a gene product.

The term "recombinant protein" refers to a polypeptide of the present invention which is produced by recombinant DNA techniques, wherein generally, DNA encoding a RIEG polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. Moreover, the phrase "derived from," with respect to a recombinant RIEG gene, is meant to include within the meaning of "recombinant protein" those proteins having an amino acid sequence of a native RIEG protein, or an amino acid sequence similar thereto which is generated by mutations including substitutions and deletions (including truncation) of a naturally occurring form of the protein.

As used herein, the term "specifically hybridizes" or "specifically detects" refers to the ability of a nucleic acid molecule of the invention to hybridize to at least approximately 6, 12, 20, 30, 50, 100, 150, 200, 300, 350, 400 or 425 consecutive nucleotides of a vertebrate, preferably RIEG gene, such as a RIEG sequence designated in one of SEQ ID Nos. 1 or 3, or a sequence complementary thereto, or naturally occurring mutants thereof, such that it shows more than 10 times more hybridization, preferably more than 100 times more hybridization, and even more preferably more than 100 times more hybridization than it does to a cellular nucleic acid (e.g., mRNA or genomic DNA) encoding a protein other than a vertebrate RIEG protein as defined herein.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of protein coding sequences with which they are operably linked. In preferred embodiments, transcription of one of the recombinant mammalian RIEG genes is under the control of a promoter sequence (or other transcriptional regulatory sequence) which controls the expression of the recombinant gene in a cell-type in which expression is intended. It will also be understood that the recombinant gene can be under the control of transcriptional regulatory sequences which are the same or which are different from those sequences which control transcription of the naturally-occurring forms of RIEG proteins.

As used herein, the term "transfection" means the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of a mammalian RIEG polypeptide or, in the case of anti-sense expression from the transferred gene, the expression of a naturally-occurring form of the RIEG protein is disrupted.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., one of the mammalian RIEG polypeptides, or a transcript thereof), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

A "transgenic animal" refers to any animal, preferably a non-human mammal, bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animals described herein, the transgene causes cells to express a recombinant form of one of the mammalian RIEG proteins, e.g. either agonistic or antagonistic forms. However, transgenic animals in which the recombinant RIEG gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. Moreover, "transgenic animal" also includes those recombinant animals in which gene disruption of one or more RIEG genes is caused by human intervention, including both recombination and antisense techniques.
As used herein, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomously replicating and expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred herein as “expression vectors”. In general, expression vectors of utility in recombinant DNA techniques are often in the form of “plasmids” which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, “plasmid” and “vector” are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

4.3 Nucleic Acids of the Present Invention

As described below, one aspect of the invention pertains to isolated nucleic acids comprising nucleotide sequences encoding RIEG polypeptides, and/or equivalents of such nucleic acids. The term equivalent is understood to include nucleotide sequences encoding functionally equivalent RIEG polypeptides or functionally equivalent peptides having an activity of a vertebrate RIEG protein such as described herein. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include sequences that differ from the nucleotide sequence of the RIEG gene shown in any of SEQ ID Nos: 1 or 3 due to the degeneracy of the genetic code.

Preferred nucleic acids are vertebrate RGS nucleic acids. Particularly preferred vertebrate RGS nucleic acids are mammalian. Regardless of species, particularly preferred RGS nucleic acids encode polypeptides that are at least 90% similar to an amino acid sequence of a vertebrate RGS. Preferred nucleic acids encode a RGS polypeptide comprising an amino acid sequence at least 92.5% homologous and more preferably 94% homologous with an amino acid sequence of a vertebrate RGS, e.g., such as a sequence shown in one of SEQ ID Nos: 1 or 3. Nucleic acids which encode polypeptides at least about 95%, and even more preferably at least about 98–99% similarity with an amino acid sequence represented in one of SEQ ID Nos: 2 or 5 are also within the scope of the invention. In a particularly preferred embodiment, the nucleic acid of the present invention encodes an amino acid sequence shown in one of SEQ ID No: 2 in one embodiment, the nucleic acid is a cDNA encoding a peptide having at least one bioactivity of the subject RGS polypeptide. Preferably, the nucleic acid includes all or a portion of the nucleotide sequence corresponding to the coding region of SEQ ID Nos: 1 or 3.

Still other preferred nucleic acids of the present invention encode a RGS polypeptide which includes a polypeptide sequence corresponding to all or a portion of amino acid residues of SEQ ID No: 2 or of SEQ ID No: 5, e.g., at least 2, 5, 10, 25, 50, 100, 150 or 200 amino acid residues of that region. For example, preferred nucleic acid molecules for use as probes/primer or antisense molecules (i.e. noncoding nucleic acid molecules) can comprise at least about 6, 12, 20, 30, 50, 100, 125, 150 or 200 base pairs in length, whereas coding nucleic acid molecules can comprise about 200, 250, 300, 350, 400, 410, 420, 430, 435 or 440 base pairs.

Another aspect of the invention provides a nucleic acid which hybridizes under stringent conditions to a nucleic acid represented by one of SEQ ID Nos: 1 or 3. Appropriate stringency conditions which promote DNA hybridization, for example, 60°C sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0×SSC at 50°C, are known to those skilled in the art or can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1–6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0×SSC at 50°C to a high stringency of about 0.2×SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Both temperature and salt may be varied, or temperature of salt concentration may be held constant while the other variable is changed. In a preferred embodiment, a RIEG nucleic acid of the present invention will bind to one of SEQ ID Nos 1 or 3 under moderately stringent conditions, for example at about 2.0×SSC and about 40°C. In a particularly preferred embodiment, a RIEG nucleic acid of the present invention will bind to one of SEQ ID Nos: 1 or 3 under high stringency conditions, but will not bind to the nucleic acid shown in SEQ ID No: 5.

Preferred nucleic acids have a sequence at least 75% homologous and more preferably 80% and even more preferably at least 85% homologous with nucleotide sequence of a mammalian RGS, e.g., such as a sequence shown in one of SEQ ID Nos: 1 and 3. Nucleic acids at least 90%, more preferably 95%, and most preferably at least about 98–99% homologous with a nucleic sequence represented in one of SEQ ID Nos: 1 and 3 are of course also within the scope of the invention. In preferred embodiments, the nucleic acid is a mammalian RGS gene and in particularly preferred embodiments, includes all or a portion of the nucleotide sequence corresponding to the coding region of one of SEQ ID Nos: 1 or 3.

Nucleic acids having a sequence that differs from the nucleotide sequences shown in one of SEQ ID Nos: 1 or 3 due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent peptides (i.e., a peptide having a biological activity of a mammalian RIEG polypeptide) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC each encode histidine) may result in “silent” mutations which do not affect the amino acid sequence of a mammalian RIEG polypeptide. However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject RIEG polypeptides will exist among mammals. One skilled in the art will appreciate that these variations in one or more nucleotides (e.g., up to about 3–5% of the nucleotides) of the nucleic acids encoding polypeptides having an activity of a mammalian RIEG polypeptide may exist among individuals of a given species due to natural allelic variation.

As indicated by the examples set out below, RIEG protein-encoding nucleic acids can be obtained from mRNA present in any of a number of eukaryotic cells. It should also be possible to obtain nucleic acids encoding mammalian RIEG polypeptides of the present invention from genomic DNA from both adults and embryos. For example, a gene encoding a RIEG protein can be cloned from either a cDNA or a genomic library in accordance with protocols described herein, as well as those generally known to persons skilled in the art. Examples of tissues and/or libraries suitable for isolation of the subject nucleic acids include breast, spleen,
thymus, prostate, testes, ovary, small intestine, colon, and peripheral blood cells, among others. A cDNA encoding a RIEG protein can be obtained by isolating total mRNA from a cell, e.g., a vertebrate cell, a mammalian cell, or a human cell, including embryonic cells. Double stranded cDNAs can then be prepared from the total mRNA, and subsequently inserted into a suitable plasmid or bacteriophage vector using any one of a number of known techniques. The gene encoding a mammalian RIEG protein can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acid of the invention can be DNA or RNA. A preferred nucleic acid is a cDNA represented by a sequence selected from the group consisting of SEQ ID Nos: 1 and 3.

4.3.1. Vectors.

This invention also provides expression vectors containing a nucleic acid encoding a RIEG polypeptide, operably linked to at least one transcriptional regulatory sequence. “Operably linked” is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and are selected to direct expression of the subject mammalian RIEG proteins. Accordingly, the term “transcriptional regulatory sequence” includes promoters, enhancers and other expression control elements. Such regulatory sequences are described in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). In one embodiment, the expression vector includes a recombinant gene encoding a peptide having an agonistic activity of a subject RIEG polypeptide, or alternatively, encoding a peptide which is an antagonistic form of the RIEG protein. Such expression vectors can be used to transfect cells and thereby produce polypeptides, including fusion proteins, encoded by nucleic acids as described herein. Moreover, the gene constructs of the present invention can also be used as a part of a gene therapy protocol to deliver nucleic acids encoding either an agonistic or antagonistic form of one of the subject mammalian MEG proteins. Thus, another aspect of the invention features expression vectors for in vivo or in vitro transfection and expression of a mammalian RIEG polypeptide in particular cell types so as to reconstitute the function of, or alternatively, abrogate the function of RIEG-induced signaling in a tissue. This could be desirable, for example, when the naturally-occurring form of the protein is misexpressed; or to deliver a form of the protein which alters differentiation of tissue. Expression vectors may also be employed to inhibit neoplastic transformation.

In addition to viral transfer methods, such as those illustrated above, non-viral methods can also be employed to cause expression of a subject RIEG polypeptide in the tissue of an animal. Most nonviral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral targeting means of the present invention rely on endocytic pathways for the uptake of the subject RIEG polypeptide gene by the targeted cell. Exemplary targeting means of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

4.3.2. Probes and Primers

Moreover, the nucleotide sequences determined from the cloning of RIEG genes from mammalian organisms will further allow for the generation of probes and primers designed for use in identifying and/or cloning RIEG homologs in other cell types, e.g. from other tissues, as well as RIEG homologs from other mammalian organisms. For instance, the present invention also provides a probe/primer comprising a substantially purified oligonucleotide, which oligonucleotide comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least approximately 12, preferably 25, more preferably 40, 50 or 75 consecutive nucleotides of sense or anti-sense sequence selected from the group consisting of SEQ ID No:1 and 3, or naturally occurring mutants thereof. For instance, primers based on the nucleic acid represented in SEQ ID Nos:1 and 3 can be used in PCR reactions to clone RIEG homologs. Likewise, probes based on the subject RIEG sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto and able to be detected, e.g. the label group is selected from amongst radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors.

As discussed in more detail below, such probes can also be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a RIEG protein, such as by measuring a level of a RIEG-encoding nucleic acid in a sample of cells from a patient; e.g. detecting RIEG mRNA levels or determining whether a genomic RIEG gene has been mutated or deleted. Briefly, nucleotide probes can be generated from the subject RIEG genes which facilitate histological screening of intact tissue and tissue samples for the presence (or absence) of RIEG-encoding transcripts. Similar to the diagnostic uses of anti-RIEG antibodies, the use of probes directed to RIEG messages, or to genomic RIEG sequences, can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, neoplastic or hyperplastic disorders (e.g. unwanted cell growth) or abnormal differentiation of tissue. Used in conjunction with immunostains as described herein, the oligonucleotide probes can help facilitate the determination of the molecular basis for the development disorder which may involve some abnormality associated with expression (or lack thereof) of a RIEG protein. For instance, variation in polypeptide synthesis can be differentiated from a mutation in a coding sequence.

4.3.3. Antisense, Ribozyme and Triplex techniques

One aspect of the invention relates to the use of the isolated nucleic acid in “antisense” therapy. As used herein, “antisense” therapy refers to administration or in situ generation of oligonucleotide molecules or their derivatives which specifically hybridize (e.g. bind) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding one or more of the subject RIEG proteins so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, “antisense” therapy refers to the range of techniques generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a mammalian RIEG protein. Alternatively, the antisense construct is an oligonucleotide probe which is generated ex vivo and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences of a mammalian RIEG gene. Such oligonucleotide probes are preferably
modified oligonucleotides which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and are therefore stable in vivo. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoromidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Pat. Nos. 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by Van der Krol et al. (1988) Biotechniques 6:958–976; and Stein et al. (1988) Cancer Res 48:2659–2668. With respect to antisense DNA, oligodeoxyribonucleotides derived from the translation initiation site, e.g., between the –10 and +10 regions of the RIEG nucleotide sequence of interest, are preferred.

Antisense approaches involve the design of oligonucleotides (either DNA or RNA) that are complementary to RIEG mRNA. The antisense oligonucleotides will bind to the RIEG mRNA transcripts and prevent translation. Absolute complementarity, although preferred, is not required. A sequence “complementary” to a portion of an RNA, as referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

Oligonucleotides that are complementary to the 5’ end of the message, e.g., the 5’ untranslated sequence up to and including the AUG initiation codon, should work most efficiently at inhibiting translation. However, sequences complementary to the 3’ untranslated sequences of mRNAs have recently been shown to be effective at inhibiting translation of mRNAs as well. (Wagner, R. 1994. Nature 372:333). Therefore, oligonucleotides complementary to either the 5’ or 3’ untranslated, non-coding regions of a RIEG gene could be used in an antisense approach to inhibit translation of endogenous RIEG mRNA. Oligonucleotides complementary to the 5’ untranslated region of the mRNA should include the complement of the AUG start codon. Antisense oligonucleotides complementary to mRNA coding regions are less efficient inhibitors of translation but could be used in accordance with the invention. Whether designed to hybridize to the 5’, 3’ or coding region of RIEG mRNA, antisense nucleic acids should be at least six nucleotides in length, and are preferably oligonucleotides ranging from 6 to about 50 nucleotides in length. In certain embodiments, the oligonucleotide is at least 10 nucleotides, at least 17 nucleotides, at least 25 nucleotides, or at least 50 nucleotides.

Regardless of the choice of target sequence, it is preferred that in vitro studies are first performed to quantify the ability of the antisense oligonucleotide to quantitatively achieve the ability of the antisense oligonucleotide to inhibit gene expression. It is preferred that these studies utilize controls that distinguish between antisense gene inhibition and non-specific biological effects of oligonucleotides. It is also preferred that these studies compare levels of the target RNA or protein with that of an internal control RNA or protein. Additionally, it is envisioned that results obtained using the antisense oligonucleotide are compared with those obtained using a control oligonucleotide. It is preferred that the control oligonucleotide is of approximately the same length as the test oligonucleotide and that the nucleotide sequence of the oligonucleotide differs from the antisense sequence no more than is necessary to prevent specific hybridization to the target sequence.

The oligonucleotides can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, etc. The oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors in vivo), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. 84:648-652; PCT Publication No. WO88/09810, published Dec. 15, 1988) or the blood-brain barrier (see, e.g., PCT Publication No. WO89/10134, published Apr. 25, 1989), hybridization-triggered cleavage agents. (See, e.g., Krol et al., 1988, BioTechniques 6:958–976) or intercalating agents. (See, e.g., Zon, 1988, Pharm. Res. 5:539–549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The antisense oligonucleotide may comprise at least one modified base moiety which is selected from the group including but not limited to 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-ace
telytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyl adenine, 2-methylguanine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannoslyqueosine, 5-methoxy carbonyl methyluracil, 5-methoxyuracil, 2-methylthio-N-6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiycytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylster, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acep)w, and 2,6-diaminopurine.

The antisense oligonucleotide may also comprise at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidate, a phosphorimidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof.

In yet another embodiment, the antisense oligonucleotide is an α-anomeric oligonucleotide. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier et al., 1987, Nucl. Acids Res. 15:6625–6641). The oligonucleotide is a 2’-0-methylribonucleotide (Inoue et al., 1987, Nucl. Acids Res. 15:6131–6148), or a chimeric RNA-DNA analogue (Inoue et al., 1987, FEBS Lett. 215:327-330).

Oligonucleotides of the invention may be synthesized by standard methods known in the art, e.g. by use of an
automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides may be synthesized by the method of Stein et al. (1988, *Nucleic Acids Res.* 16:3209), methylphosphonate oligonucleotides can be prepared by use of controlled pore glass polymer supports (Sarin et al., 1988, *Proc. Natl. Acad. Sci. U.S.A.* 85:7448–7451), etc.

While antisense nucleotides complementary to the RIEG coding region sequence could be used, those complementary to the transcribed untranslated region are most preferred. For example, an antisense oligonucleotide as set forth in SEQ ID No. 14 can be utilized in accordance with the invention.

The antisense molecules should be delivered to cells which express the RIEG in vivo. A number of methods have been developed for delivering antisense DNA or RNA to cells; e.g., antisense molecules can be injected directly into the tissue site, or modified antisense molecules, designed to target the desired cells (e.g., antisense linked to peptides or antibodies that specifically bind receptors or antigens expressed on the target cell surface) can be administered systematically.

However, it is often difficult to achieve intracellular concentrations of the antisense sufficient to suppress translation of endogenous mRNAs. Therefore a preferred approach utilizes a recombinant DNA construct in which the antisense oligonucleotide is placed under the control of a strong pol III or pol II promoter. The use of such a construct to transfect target cells in the patient will result in the transcription of sufficient amounts of single stranded RNAs that will form complementary base pairs with the endogenous RIEG transcripts and thereby prevent translation of the RIEG mRNA. For example, a vector can be introduced in vivo such that it is taken up by a cell and directs the transcription of an antisense RNA. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequence encoding the antisense RNA can be by any promoter known in the art to act in mammalian, preferably human cells. Such promoters can be inducible or constitutive. Such promoters include but are not limited to: the SV40 early promoter region (Bernoist and Chambon, 1981, *Nature* 290:304–310), the promoter contained in the 3′ long terminal repeat of Rous sarcoma virus (Yamamoto et al., 1980, *Cell* 22:787–797), the herpes thymidine kinase promoter (Wagner et al., 1981, *Proc. Natl. Acad. Sci. U.S.A.* 78:1441–1445), the regulatory sequences of the metallothio-
nein gene (Brüning et al., 1982, *Nature* 296:39–42), etc. Any type of plasmid, cosmid, YAC or viral vector can be used to prepare the recombinant DNA construct which can be introduced directly into the tissue site; e.g., the choroid plexus or hypothalamus. Alternatively, viral vectors can be used which selectively infect the desired tissue; (e.g., for brain, herpes virus vectors may be used), in which case administration may be accomplished by another route (e.g., systematically).

Ribozyine molecules designed to catalytically cleave RIEG mRNA transcripts can also be used to prevent translation of RIEG mRNA and expression of RIEG. (See, e.g., PCT International Publication WO90/11364, published Oct. 4, 1990; Sarver et al., 1990, *Science* 247:1222–1225). While ribozymes that cleave mRNA at site specific recognition sequences can be used to destroy RIEG mRNAs, the use of hammerhead ribozymes is preferred. Hammerhead ribozymes cleave mRNAs at locations dictated by flanking regions that form complementary base pairs with the target mRNA. The sole requirement is that the target mRNA have the following sequence of two bases: 5′-UG-3′. The construction and production of hammerhead ribozymes is well known in the art and is described more fully in Haseloff and Gerlach, 1988, *Nature*, 334:585–591. There are hundreds of potential hammerhead ribozyme cleavage sites within the nucleotide sequence of human RIEG cDNA (FIG. 3). Preferably, the ribozyme is engineered so that the cleavage recognition site is located near the 5′ end of the RIEG mRNA; i.e., to increase efficiency and minimize the intracellular accumulation of non-functional mRNA transcripts.

For example, ribozymes having the sequence set forth in SEQ ID NO 13 can be utilized in accordance with the invention. The ribozymes of the present invention also include RNA endoribonucleases (hereinafter “Cech-type ribozymes”) such as the one which occurs naturally in Tetrahymena Thermophilus (known as the IVS, or L-19 IVS RNA) and which has been extensively described by Thomas Cech and collaborators (Zaug, et al., 1984, *Science*, 224:574–578; Zaug and Cech, 1986, *Science*, 231:470–475; Zaug, et al., 1986, *Nature*, 324:429–433; published International patent application No. W088/04300 by University Patents Inc.; Been and Cech, 1986, *Cell*, 47:207–216). The Cech-type ribozymes have an eight base pair active site which hybridizes to a target RNA sequence whereas cleavage of the target RNA takes place. The invention encompasses those Cech-type ribozymes which target eight base-pair active site sequences that are present in RIEG.

As in the antisense approach, the ribozymes can be composed of modified oligonucleotides (e.g., for improved stability, targeting, etc.) and should be delivered to cells which express the RIEG in vivo e.g., hypothalamus and/or the choroid plexus. A preferred method of delivery involves using a DNA construct “encoding” the ribozyme under the control of a strong constitutive pol III or pol II promoter, so that transfected cells will produce sufficient quantities of the ribozyme to destroy endogenous RIEG messages and inhibit translation. Because ribozymes unlike antisense molecules, are catalytic, a lower intracellular concentration is required for efficiency.

Endogenous RIEG gene expression can also be reduced by inactivating or “knocking out” the RIEG gene or its promoter using targeted homologous recombination. (E.g., see Smithies et al., 1985, *Nature* 317:230–234; Thomas & Capecci, 1987, *Cell* 51:503–512; Thompson et al., 1989 *Cell* 5:313–321; each of which is incorporated by reference herein in its entirety.) For example, a mutant, non-functional RIEG (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous RIEG gene (either the coding regions or regulatory regions of the RIEG gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transflect cells that express RIEG in vivo. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the RIEG gene. Such approaches are particularly suited in the agricultural field where modifications to ES (embryonic stem) cells can be used to generate animal offspring with an inactive RIEG (e.g., see Thomas & Capecci 1987 and Thompson 1989, supra). However this approach can be adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors, e.g., herpes virus vectors for delivery to brain tissue, e.g., the hypothalamus and/or choroid plexus.

Likewise, the antisense constructs of the present invention, by antagonizing the normal biological activity of one of the RIEG proteins, can be used in the manipulation of tissue, e.g., tissue differentiation, both in vivo and for ex vivo tissue cultures.

Furthermore, the anti-sense techniques (e.g., microinjection of antisense molecules, or transfection with plasmids whose transcripts are anti-sense with regard to a RIEG mRNA or gene sequence) can be used to investigate role of RIEG in developmental events, as well as the normal cellular function of RIEG in adult tissue. Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals, as detailed below.

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by an endonucleolytic cleavage. The composition of ribozyme molecules must include one or more sequences complementary to the target gene mRNA, and must include the well known catalytic sequence responsible for mRNA cleavage. For this sequence, see U.S. Pat. No. 5,093,246, which is incorporated by reference herein in its entirety. As such within the scope of the invention are engineered hammerhead motif ribozyme molecules that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences encoding FCHD5944 transcription factors.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the molecule of interest for ribozyme cleavage sites which include the following sequences, GUU, GGU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features, such as secondary structure, that may render the oligonucleotide sequence unsuitable. The suitability of candidate sequences may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

Nucleic acid molecules to be used in triple helix formation for the inhibition of transcription are preferably single stranded and composed of deoxyribonucleotides. The base composition of these oligonucleotides should promote stable helix formation via Hoogsteen base pairing rules, which generally require sizable stretches of either purines or pyrimidines to be present on one strand of a duplex. Nucleotide sequences may be pyrimidine-based, which will result in TAT and CGC triplets across the three associated strands of the resulting triple helix. The pyrimidine-rich molecules provide base complementarity to a purine-rich region of a single strand of the duplex in a parallel orientation to that strand. In addition, nucleic acid molecules may be chosen that are purine-rich, for example, containing a stretch of G residues. These molecules will form a triple helix with a DNA duplex that is rich in GC pairs, in which the majority of the purine residues are located on a single strand of the targeted duplex, resulting in CGC triplets across the three strands in the triplex.

Alternatively, the potential sequences that can be targeted for triple helix formation may be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'3', 3'5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizable stretch of either purines or pyrimidines to be present on one strand of a duplex.

Antisense RNA and DNA, ribozyme, and triple helix molecules of the invention may be prepared by any method known in the art for the synthesis of DNA and RNA molecules. These include techniques for chemically synthesizing oligodeoxyribonucleotides and oligoribonucleotides well known in the art such as for example solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding the antisense RNA molecule. Such DNA sequences may be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, can be introduced stably into cell lines.

Moreover, various well-known modifications to nucleic acid molecules may be introduced as a means of increasing intracellular stability and half-life. Possible modifications include but are not limited to the addition of flanking sequences of ribonucleotides or deoxyribonucleotides to the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2'O-methyl rather than phosphodiester linkages within the oligodeoxyribonucleotide backbone.

4.4. Polypeptides of the Present Invention

The present invention also makes available isolated RIEG ("Soltushin") polypeptides which are isolated from, or otherwise substantially free of other cellular proteins, especially other signal transduction factors and/or transcription factors which may normally be associated with the RIEG polypeptide. The term "substantially free of other cellular proteins" (also referred to herein as "contaminating proteins") or "substantially pure or purified preparations" are defined as encompassing preparations of RIEG polypeptides having less than about 20% (by dry weight) contaminating protein, and preferably having less than about 5% contaminating protein. Functional forms of the subject polypeptides can be prepared, for the first time, as purified preparations by using a cloned gene as described herein. By "purified", it is meant, when referring to a peptide or DNA or RNA sequence, that the indicated molecule is present in the substantial absence of other biological macromolecules, such as other proteins. The term "purified" as used herein preferably means at least 80% by dry weight, more preferably in the range of 95-99% by weight, and most preferably at least 99.8% by weight, of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). The term "pure" as used herein preferably has the same numerical limits as "purified" immediately above. "Isolated" and "purified" do not encompass either natural materials in their native state or natural materials that have been separated into components (e.g., in an acrylamide gel) but not obtained either as pure (e.g., lacking contaminating proteins, or chromatography reagents such as denaturing agents and polymers, e.g. acrylamide or agarose) substances or solutions. In preferred embodiments, purified RIEG preparations will lack any contaminating proteins from the same animal from which RIEG is normally produced, as can
be accomplished by recombinant expression of, for example, a human RIEG protein in a non-human cell.

Full length proteins or fragments corresponding to one or more particular motifs and/or domains or to arbitrary sizes, for example, at least 5, 10, 25, 50, 75, 100, 125, 150 amino acids in length are within the scope of the present invention.

For example, isolated RIEG polypeptides can include all or a portion of an amino acid sequences corresponding to a RIEG polypeptide represented in one or more of SEQ ID No: 2 or 5. Isolated peptidyl portions of RIEG proteins can be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, a RIEG polypeptide of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of a wild-type (e.g., “authentic”) RIEG protein.

Another aspect of the present invention concerns recombinant forms of the RIEG proteins. Recombinant polypeptides preferred by the present invention, in addition to native RIEG proteins, are at least 92% homologous and more preferably 94% homologous and most preferably 95% homologous with an amino acid sequence represented by any of SEQ ID Nos: 2 or 5. Polypeptides which are at least about 98–99% homologous with a sequence selected from the group consisting of SEQ ID Nos: 2 or 5 are also within the scope of the invention. In a preferred embodiment, a RIEG protein of the present invention is a mammalian RIEG protein. In a particularly preferred embodiment a RIEG polypeptide comprises the coding sequence of one of SEQ ID Nos: 2 or 5. In a particularly preferred embodiment, a RIEG protein comprises the amino acid coding sequence of one of SEQ ID No: 2 or 5.

In general, polypeptides referred to herein as having an activity (e.g., are “bioactive”) of a mammalian RIEG protein are defined as polypeptides which include an amino acid sequence corresponding (e.g., identical or homologous) to all or a portion of the amino acid sequences of a mammalian RIEG proteins shown in any one or more of SEQ ID Nos: 2 or 5 and which mimic or antagonize all or a portion of the biological/biochemical activities of a naturally occurring RIEG protein. In preferred embodiments, the biochemical activities are related to gene expression, pituitary development, and abdominal development related to umbilical and vitelline artery expression.

Other biological activities of the subject RIEG proteins are described herein or will be reasonably apparent to those skilled in the art. According to the present invention, a polypeptide has biological activity if it is a specific agonist or antagonist of a naturally-occurring form of a mammalian RIEG protein.

The present invention further pertains to methods of producing the subject RIEG polypeptides. For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding the subject polypeptides can be cultured under appropriate conditions to allow expression of the peptide to occur. The cells may be harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. The recombinant RIEG polypeptide can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such peptide. In a preferred embodiment, the recombinant RIEG polypeptide is a fusion protein containing a domain which facilitates its purification, such as GST fusion protein or polyHis fusion protein.

Moreover, it will be generally appreciated that, under certain circumstances, it may be advantageous to provide homologs of one of the subject RIEG polypeptides which function in a limited capacity as one of either a RIEG agonist (mimetic) or a RIEG antagonist, in order to promote or inhibit only a subset of the biological activities of the naturally-occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homolog or limited function, and with fewer side effects relative to treatment with agonists or antagonists which are directed to all of the biological activities of naturally occurring forms of RIEG proteins.

Homologs of each of the subject RIEG proteins can be generated by mutagenesis, such as by discrete point mutation(s), or by truncation. For instance, mutation can give rise to homologs which retain substantially the same, or merely a subset, of the biological activity of the RIEG polypeptide from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally occurring form of the protein, such as by competitively binding to a downstream or upstream member of the biochemical pathway, which includes the RIEG protein. In addition, agonistic forms of the protein may be generated which are constitutively active. Thus, the mammalian RIEG protein and homologs thereof provided by the subject invention may be either positive or negative regulators of gene expression.
The recombinant Rieg polypeptides of the present invention also include homologs of the authentic Rieg proteins, such as versions of those protein which are resistant to proteolytic cleavage, as for example, due to mutations which alter ubiquitination or other enzymatic targeting associated with the protein. Rieg polypeptides may also be chemically modified to create Rieg derivatives by forming covalent or aggregate conjugates with other chemical moieties, such as glycosyl groups, lipids, phosphate, acetyl groups and the like. Covalent derivatives of Rieg proteins can be prepared by linking the chemical moieties to functional groups on amino acid sidechains of the protein or at the N-terminus or at the C-terminus of the polypeptide.

Modification of the structure of the subject mammalian Rieg polypeptides can be for such purposes as enhancing therapeutic or prophylactic efficacy, stability (e.g., ex vivo shelf life and resistance to proteolytic degradation in vivo), or post-translational modifications (e.g., to alter phosphorylation pattern of protein). Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, or to produce specific antagonists thereof, are considered functional equivalents of the Rieg polypeptides described in more detail herein. Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e., isosteric and/or isoelectric mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids can be divided into four families: (1) acidic—aspartate, glutamate; (2) basic—lysine, arginine, histidine; (3) nonpolar—alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar—glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. In similar fashion, the amino acid repertoire can be grouped as (1) acidic—aspartate, glutamate; (2) basic—lysine, arginine, histidine; (3) aliphatic—glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally grouped separately as aliphatic-hydroxy; (4) aromatic—phenylalanine, tyrosine, tryptophan; (5) amide—asparagine, glutamine; and (6) sulfur-containing—cysteine and methionine. (see, for example, Biochemistry, 2nd ed., Ed. by L. Stryer, WH Freeman and Co.: 1981). Whether a change in the amino acid sequence of a peptide results in a functional Rieg homolog (e.g., functional in the sense that the resulting polypeptide mimics or antagonizes the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type protein, or competitively inhibit such a response. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method for generating sets of combinatorial mutants of the subject Rieg proteins as well as truncation mutants, and is especially useful for identifying potential variant sequences (e.g., homologs) that are functional in modulating gene expression. The purpose of screening such combinatorial libraries is to generate, for example, novel Rieg homologs which can act as either agonists or antagonists, or alternatively, possess novel activities all together.

Likewise, Rieg homologs can be generated by the present combinatorial approach to selectively inhibit gene expression. For instance, mutagenesis can provide Rieg homologs which are able to bind other signal pathway proteins (or DNA) yet prevent propagation of the signal, e.g. the homologs can be dominant negative mutants. Moreover, manipulation of certain domains of Rieg by the present method can provide domains more suitable for use in fusion proteins.

In one embodiment, the variegated library of Rieg variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For instance, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential Rieg sequences are expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g. for phage display) containing the set of Rieg sequences therein.

There are many ways by which such libraries of potential Rieg homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential Rieg sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example, Narang, S A (1983) Tetrahehedron 39:3; Itakura et al. (1981) Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules, ed. A G Walton, Amsterdam: Elsevier pp273–289; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) Science 249:386–390; Roberts et al. (1992) PNAS 89:2429–2433; Devlin et al. (1990) Science 249: 404–406; Cwrla et al. (1990) PNAS 87: 6378–6382; as well as U.S. Pat. Nos. 5,223,409, 5,198,346, and 5,096,815).

Likewise, a library of coding sequence fragments can be provided for a Rieg clone in order to generate a variegated population of Rieg fragments for screening and subsequent selection of bioactive fragments. A variety of techniques are known in the art for generating such libraries, including chemical synthesis. In one embodiment, a library of coding sequence fragments can be generated by (i) treating a double stranded PCR fragment of a Rieg coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule; (ii) denaturing the double stranded DNA; (iii) renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products; (iv) removing single stranded portions from reformer duplexes by treatment with S1 nuclease; and (v) ligating the resulting fragment library into an expression vector. By this exemplary method, an expression library can be derived which codes for N-terminal, C-terminal and internal fragments of various sizes.

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of Rieg homologs. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of
vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the illustrative assays described below are amenable to high throughput analysis as necessary to screen large numbers of degenerate RIEG sequences created by combinatorial mutagenesis techniques.

Combinatorial mutagenesis has a potential to generate very large libraries of mutant proteins, e.g., in the order of $10^{20}$ molecules. Combinatorial libraries of this size may be technically challenging to screen even with high throughput screening assays. To overcome this problem, a new technique has been developed recently, recursive ensemble mutagenesis (REM), which allows one to avoid the very high proportion of non-functional proteins in a random library and simply enhances the frequency of functional proteins by decreasing the complexity required to achieve a useful sampling of sequence space. REM is an algorithm which enhances the frequency of functional mutants in a library when an appropriate selection or screening method is employed (Arkin and Yorvann, 1992, PNAS USA 89:7811-7815; Yorvann et al., 1992, Parallel Problem Solving from Nature 2. In Maender and Manderick, eds., Elsevier Publishing Co., Amsterdam, pp. 401-410; Delgrave et al., 1993, Protein Engineering 6(3):327-331).

The invention also provides for reduction of the mammalian RIEG proteins to generate mimetics, e.g., peptide or non-peptide agents, which are able to disrupt binding of a mammalian RIEG polypeptide of the present invention with either upstream or downstream components. Thus, such mutagenic techniques as described above are also useful to map the determinants of the RIEG proteins which participate in protein-protein interactions involved in, for example, binding of the subject RIEG polypeptide to proteins which may function upstream (including both activators and repressors of its activity) or to proteins or nucleic acids which may function downstream of the RIEG polypeptide, whether they are positively or negatively regulated by it. To illustrate, the critical residues of a subject RIEG polypeptide which are involved in molecular recognition of a component upstream or downstream of a RIEG can be determined and used to generate RIEG derived peptidomimetics which competitively inhibit binding of the authentic RIEG protein with that moiety. By employing, for example, scanning mutagenesis to map the amino acid residues of each of the subject RIEG proteins which are involved in binding other extracellular proteins, peptidomimetic compounds can be generated which mimic those residues of the RIEG protein which facilitate the interaction. Such mimetics may then be used to interfere with the normal function of a RIEG protein. For instance, non-hydrolysable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in Peptides: Chemistry and Biology, G. R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in Peptides: Chemistry and Biology, G. R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in Peptides: Chemistry and Biology, G. R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), ketone-methylene pseudopeptides (Ewenson et al. (1986) J Med Chem 29:295; and Ewenson et al. in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, Ill., 1985), beta-turn dipeptide cores (Nagai et al. (1985) Tifehen- dron Lett 26:647; and Sato et al. (1986) J Chem Soc Perkin Trans 1:1231), and beta-aminoalcohols (Gordon et al. (1985) Biochem Biophys Res Commun 126:419; and Dann et al. (1986) Biochem Biophys Res Commun 134:71).


This invention also pertains to a host cell transfected to express a recombinant form of the subject RIEG polypeptide. The host cell may be any prokaryotic or eukaryotic cell. Thus, a nucleotide sequence derived from the cloning of mammalian RIEG proteins, encoding all or a selected portion of the full-length protein, can be used to produce a recombinant form of a RIEG polypeptide via microbial or eukaryotic cellular processes. Ligating the polynucleotide sequence into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells), are standard procedures used in producing other well-known proteins, e.g., MAP kinase, p53, WT1, PTP phosphatases, SRC, and the like. Similar procedures, or modifications thereof, can be employed to prepare recombinant RIEG polypeptides by microbial means or tissue culture technology in accord with the subject invention.

The recombinant RIEG genes can be produced by ligating nucleic acid encoding a RIEG protein, or a portion thereof, into a vector suitable for expression in either prokaryotic cells, eukaryotic cells, or both. Expression vectors for production of recombinant forms of the subject RIEG polypeptides include plasmids and other vectors. For instance, suitable vectors for the expression of a RIEG polypeptide include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as E. coli.

A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into S. cerevisiae (see, for example, Brouch et al. (1983) in Experimental Manipulation of Gene Expression, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in E. coli due the presence of the pBR322 ori, and in S. cerevisiae due to the replication determinant of the yeast 2 micron plasmid. In addition, drug resistance markers such as ampicillin can be used. In an illustrative embodiment, a RIEG polypeptide is produced recombinantly utilizing an expression vector generated by sub-cloning the coding sequence of one of the RIEG genes represented in SEQ ID Nos: 1 and 3.

The preference mammalian expression vectors contain both prokaryotic sequences, to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pCDNAI/amp, pcDNA1neo, pCtCMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSV77, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the hovine papillomavirus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see Molecular Cloning A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, N.Y. 1989.
In some instances, it may be desirable to express the recombinant RIEG polypeptide by the use of a baculovirus expression system. Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcEUW-derived vectors (such as pAcEUW1), and pBacHac-derived vectors (such as the β-gal containing pBacBlueHac III).

When it is desirable to express only a portion of a RIEG protein, such as a form lacking a portion of the N-terminus, i.e. a truncation mutant which lacks the signal peptide, it may be necessary to add a start codon (ATG) to the oligonucleotide fragment containing the desired sequence to be expressed. It is well known in the art that a methionine at the N-terminal position can be enzymatically cleaved by the use of the enzyme methionine aminopeptidase (MAP). MAP has been cloned from E. coli (Ben-Bassat et al. (1987) J. Bacteriol. 169:751–757) and Salmonella typhimurium and its in vitro activity has been demonstrated on recombinant proteins (Miller et al. (1987) PNAS 84:2718–2722). Therefore, removal of an N-terminal methionine, if desired, can be achieved either in vivo by expressing RIEG-derived polypeptides in a host which produces MAP (e.g., E. coli or CM89 or S. cerevisiae), or in vitro by use of purified MAP (e.g., procedure of Miller et al., supra). In other embodiments transgenic animals, described in more detail below could be used to produce recombinant proteins.

4.4.2 Fusion Proteins and Immunogens

In another embodiment, the coding sequences for the polypeptide can be incorporated as a part of a fusion gene including a nucleotide sequence encoding a different polypeptide. This type of expression system can be useful under conditions where it is desirable to produce an immunogenic fragment of a RIEG protein. For example, the VP6 capsid protein of rotavirus can be used as an immunogenic carrier protein for portions of the RIEG polypeptide, either in the monomeric form or in the form of a viral particle. The nucleic acid sequences corresponding to the portion of a subject RIEG protein to which antibodies are to be raised can be incorporated into a fusion gene construct which includes coding sequences for a late vaccinia virus structural protein to produce a set of recombinant viruses expressing fusion proteins comprising RIEG epitopes as part of the virion. It has been demonstrated with the use of immunogenic fusion proteins utilizing the Hepatitis B surface antigen fusion proteins that recombinant Hepatitis B virions can be utilized in this role as well. Similarly, chimeric constructs coding for fusion proteins containing a portion of a RIEG protein and the poliovirus capsid protein can be created to enhance immunogenicity of the set of polypeptide antigens (see, for example, EP Publication No: 0259149; and Evans et al. (1989) Nature 339:385; Huang et al. (1988) J. Virol. 62:3855; and Schlienger et al. (1992) J. Virol. 66:2).

The Multiple Antigen Peptide system for peptide-based immunization can also be utilized to generate an immunogen, wherein a desired portion of a RIEG polypeptide is obtained directly from organo-chemical synthesis of the peptide onto an oligomeric branching lysine core (see, for example, Posnett et al. (1988) JBC 263:1719 and Nardelli et al. (1992) J. Immunol. 148:914). Antigenic determinants of RIEG proteins can also be expressed and presented by bacterial cells.

In addition to utilizing fusion proteins to enhance immunogenicity, it is widely appreciated that fusion proteins can also facilitate the expression of proteins, and accordingly, can be used in the expression of the mammalian RIEG polypeptides of the present invention. For example, RIEG polypeptides can be generated as glutathione-S-transferase (GST-fusion) proteins. Such GST-fusion proteins can enable easy purification of the RIEG polypeptide, as for example by the use of glutathione-derivatized matrices (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al. (N.Y.: John Wiley & Sons, 1991)).

In another embodiment, a fusion gene coding for a purification leader sequence, such as a poly-(His)-enterokinase cleavage site sequence at the N-terminus of the desired portion of the recombinant protein, can allow purification of the expressed fusion protein by affinity chromatography using a Ni2+ metal resin. The purification leader sequence can then be subsequently removed by treatment with enterokinase to provide the purified protein (e.g., see Hochuli et al. (1987) J. Chromatography 411:177; and Janknecht et al. PNAS 88:8972).

Techniques for making fusion genes are known to those skilled in the art. Essentially, the joining of various DNA fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or staggered-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overlaps between two consecutive gene fragments which can subsequently be annealed to generate a chimeric gene sequence (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al. John Wiley & Sons: 1992).

4.4.3. Antibodies

Another aspect of the invention pertains to an antibody specifically reactive with a RIEG protein. For example, by using immunogens derived from a RIEG protein, e.g. based on the cDNA sequences, anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, Antibodies: A Laboratory Manual ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal, such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide (e.g., a RIEG polypeptide or an antigenic fragment which is capable of eliciting an antibody response, or a fusion protein as described above). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of a RIEG protein can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunossays can be used with the immunogen as antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of a RIEG protein of a mammal, e.g. antigenic determinants of a protein represented by SEQ ID Nos:2 or 4 or closely related homologs (e.g. at least 92% homologous, and more preferably at least 94% homologous).

Following immunization of an animal with an antigenic preparation of a RIEG polypeptide, anti-RIEG antisera can be obtained and, if desired, polyclonal anti-RIEG antibodies isolated from the serum. To produce monoclonal antibodies, antibody-producing cells (lymphocytes) can be harvested from an immunized animal and fused by standard somatic
cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Such techniques are well known in the art, as included, for example, the hybridoma technique (originally developed by Kohler and Milstein, (1975) Nature, 256: 495-497), the human B cell hybridoma technique (Kozbor et al., (1983) Immuno-ology Today, 4: 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., (1985) Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. pp. 77–96). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with a RIEG polypeptide of the present invention and monoclonal antibodies isolated from a culture comprising such hybridoma cells.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with one of the subject mammalian RIEG polypeptides. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, F(ab)_2 fragments can be generated by treating antibody with pepsin. The resulting F(ab)_2 fragment can be treated to reduce disulfide bridges to produce Fab fragments. The antibody of the present invention is further intended to include bispecific and chimeric molecules having affinity for a RIEG protein conferred by at least one CDR region of the antibody.

Antibodies which specifically bind RIEG epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of each of the subject RIEG polypeptides. Anti-RIEG antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate RIEG protein levels in tissue as part of a clinical testing procedure. For instance, such measurements can be useful in predictive valuations of the onset or progression of proliferative disorders. Likewise, the ability to monitor RIEG protein levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. The level of RIEG polypeptides may be measured from cells in bodily fluid, such as in samples of cerebral spinal fluid or amniotic fluid, or can be measured in tissue, such as produced by biopsy. Diagnostic assays using anti-RIEG antibodies can include, for example, immunoassays designed to aid in early diagnosis of a degenerative disorder, particularly ones which are manifest at birth. Diagnostic assays using anti-RIEG polypeptide antibodies can also include immunoassays designed to aid in early diagnosis and phenotyping neoplastic or hyperplastic disorders.

Another application of anti-RIEG antibodies of the present invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λgt11, λgt18–23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λgt11 will produce fusion proteins whose amino termini consist of β-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a RIEG protein, e.g. other orthologs of a particular RIEG protein or other paralogs from the same species, can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-RIEG antibodies. Positive phage detected by this assay can then be isolated from the infected plate. Thus, the presence of RIEG homologs can be detected and cloned from other animals, as can alternate isoforms (including splicing variants) from humans.

4.5 Methods of Treating Disease

The RIEG therapeutics of the present invention can be used in treatment of Riegler Syndrome, or associated conditions including pituitary disorders and abdominal disorders related to unbalanced and vitelline artery expression.

A “RIEG therapeutic,” whether an antagonist or agonist of wild type RIEG, can be, as appropriate, any of the preparations described above, including isolated polypeptides, gene therapy constructs, antisense molecules, peptidomimetics or agents identified in the drug assays provided herein.

Since, in some cases, genes may be upregulated in a disease state and in other cases may be downregulated, it will be desirable to activate and/or potentiate or suppress and/or downmodulate RIEG bioactivity depending on the condition to be treated using the techniques compounds and methods described herein. Some genes may be underexpressed in certain disease states. Several genes are now known to be down-regulated in monocytes under disease conditions. For example, bcl-2 and glutathione peroxidase gene expression is down-regulated in the monocytes of patients exposed to a high lipid diet, e.g. cholesterol or fat, that leads to high serum LDL levels. The activity of RIEG gene products may be in some way impaired, leading to the development of cardiovascular disease symptoms. Such down-regulation of RIEG gene expression or decrease in the activity of a RIEG protein may have a causative or exacerbating effect on the disease state.

Among the approaches which may be used to ameliorate disease symptoms involving the misexpression of a RIEG gene are, for example, antisense, ribozyme, and triple helix molecules described above. Compounds that compete with an RIEG protein for binding to upstream or downstream element in a signaling cascade will antagonize a RIEG protein, thereby inducing a therapeutic effect. Examples of suitable compounds include the antagonists or homologues described in detail above. In other instances, the increased expression or bioactivity of a RIEG protein may be desirable and may be accomplished by, for example the use of the RIEG agonists or mimetics or by gene replacement therapy, as described herein.

Compounds identified as increasing or decreasing RIEG gene expression or protein activity can be administered to a subject at therapeutically effective dose to treat or ameliorate a disease or condition.

4.5.1. Effective Dose

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to unaffected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective...
dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

4.5.2. Formulation and Use

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by, for example, injection, inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral or rectal administration.

For such therapy, the oligomers of the invention can be formulated for a variety of loads of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remington's Pharmaceutical Sciences, Meade Publishing Co., Easton, Pa. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the oligomers of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

Preparations for oral administration may be suitably formulated to give controlled release of the active compound.

For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of e.g. gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqeous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g. sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art.

In clinical settings, the gene delivery systems for the therapeutic RIEG gene can be introduced into a patient by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Pat. No. 5,328,470) or by stereotactic injection (e.g. Chen et al. (1994) PNAS 91: 3054-3057). A RIEG gene, such as any one of the sequences represented in the group consisting of SEQ ID NO: 1 or 3, or a sequence homologous thereto can be delivered in a gene therapy construct by electroporation using techniques described, for example, by Dev et al. ((1994) Cancer Treat Rev 20:105–115). The pharmaceutical preparation of the gene therapy construct can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alterna-
tively, where the complete gene delivery system can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system. The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

4.6 Diagnostic and Prognostic Assays

In the diagnostic and prognostic assays described herein, in addition to the RIEG nucleic acid molecules and polypeptides described above, the present invention provides for the use of nucleic comprising at least a portion of the nucleic acid sequence shown in SEQ ID No: 1 or 3 or polypeptides comprising at least a portion of the amino acid sequence shown in SEQ ID No: 2.

The present method provides a method for determining if a subject is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In preferred embodiments, the methods can be characterized as comprising detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of (i) an alteration affecting the integrity of a gene encoding a RIEG-protein, or (ii) the mis-expression of the RIEG gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a RIEG-gene, (ii) an addition of one or more nucleotides to a RIEG-gene, (iii) a substitution of one or more nucleotides of a RIEG-gene, and (iv) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a RIEG-gene. As set out below, the present invention provides a large number of assay techniques for detecting lesions in RIEG genes, and importantly, provides the ability to discern between different molecular causes underlying RIEG-dependent aberrant cell growth, proliferation and/or differentiation.

In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) Science 241:1077–1080; and Nakazawa et al. (1994) PNAS 91:360–364), the latter of which can be particularly useful for detecting point mutations in the RIEG-gene (see Abravaya et al. (1995) Nuc Acid Res 23:675–682). In a merely illustrative embodiment, the method includes the steps of (i) collecting a sample of cells from a patient, (ii) isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, (iii) contacting the nucleic acid sample with one or more primers which specifically hybridize to a RIEG gene under conditions such that hybridization and amplification of the RIEG-gene (if present) occurs, and (iv) detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein. Alternative amplification methods include: self sustained sequence replication (Guaitelli, J. C. et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874–1878), transcriptional amplification system (Kwoh, D. Y. et al., 1989, Proc. Natl. Acad. Sci. USA 86:1173–1177), Q-Beta Replicase (Lizardi, P. M. et al., 1988, Bio/Technology 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In a preferred embodiment of the subject assay, mutations in a RIEG gene from a sample cell are identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Pat. No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the RIEG gene and detect mutations by comparing the sequence of the sample RIEG with the corresponding wild-type (control) sequence. Exemplary sequencing reactions include those based on techniques developed by Maxam and Gilbert (Proc. Natl. Acad Sci USA (1977) 74:560) or Sanger (Sanger et al (1977) Proc Natl Acad Sci 74:5463). It is also
contemplated that any of a variety of automated sequencing procedures may be utilized when performing the subject assays (Biotechniques (1995) 19:448), including by sequencing by mass spectrometry (see, for example PCT publication WO 94/16101; Cohen et al. (1996) Adv Chromatogr 36:127–162; and Griffin et al. (1993) Appl Biochem Biotechnol 38:147–159). It will be evident to one skilled in the art that, for certain embodiments, the occurrence of only one, two or three of the nucleic acid bases need be determined in the sequencing reaction. For instance, A-tract or the like, e.g., where only one nucleic acid is detected, can be carried out.

In a further embodiment, protection from cleavage agents (such as a nuclease, hydroxylamine or osmium tetroxide and with peridione) can be used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) Science 230:1242). In general, the art technique of “mismatch cleavage” starts by providing heteroduplexes of formed by hybridizing (labelled) RNA or DNA containing the wild-type RIEG sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nucleases to enzymatically digest the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with peridione in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton et al. (1988) Proc Natl Acad Sci USA 85:4397; Saleeba et al. (1992) Methods Enzymol. 217:286–295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called “DNA mismatch repair” enzymes) in defined systems for detecting and mapping point mutations in RIEG cDNAs obtained from samples of cells. For example, the mutS enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) Carcinogenesis 15:1657–1662). According to an exemplary embodiment, a probe based on a RIEG sequence, e.g., a wild-type RIEG sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in RIEG genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) Proc Natl Acad Sci USA 86:2766, see also Cotton (1993) Mutat Res 285:125–144; and Hayashi (1992) Genet Anal Tech Appl 9:73–79). Single-stranded DNA fragments of sample and control RIEG nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labelled or detected with labelled probes.

The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) Trends Genet 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al (1985) Nature 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing agent gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) Biophys Chem 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) Nature 324:163; Saiki et al (1989) Proc Natl Acad Sci USA 86:6230). Such allele specific oligonucleotide hybridization techniques may be used to test one mutation per reaction when oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labelled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al (1989) Nucleic Acids Res. 17:2437–2448) or at the extreme 3′ end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) Tibtech 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al (1992) Mol. Cell Probes 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) Proc Natl Acad Sci USA 88:189). In such cases, ligation will occur only if there is a perfect match at the 3′ end of the 5′ sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

For mutations that produce premature termination of protein translation, the protein truncation test (PTT) offers an efficient diagnostic approach (Roest, et. al., (1993) Hum Mol Genet 2:1719–21; van der Luijt, et. al., (1994) Genomics 20:14). For PTT, RNA is initially isolated from available tissue and reverse-transcribed, and the segment of interest is amplified by PCR. The products of reverse transcription PCR are then used as a template for nested PCR amplification with a primer that contains an RNA polymerase promoter and a sequence for initiating eukaryotic translation. After amplification of the region of interest, the unique motifs incorporated into the primer permits sequential in vitro transcription and translation of the PCR products. Upon
sodium dodecyl sulfate-polyacrylamide gel electrophoresis of translation products, the appearance of truncated polypeptides signals the presence of a mutation that causes premature termination of translation. In a variation of this technique, DNA (as opposed to RNA) is used as a PCR template when the target region of interest is derived from a single exon.

Another embodiment of the invention provides for a nucleic acid composition comprising a (purified) oligonucleotide probe including a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of a RIEG-gene, or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject RIEG-genes or naturally occurring mutants thereof. The nucleic acid of a cell is rendered accessible for hybridization, the probe is exposed to nucleic acid of the sample, and the hybridization of the probe to the sample nucleic acid is detected. Such techniques can be used to detect lesions at either the genomic or mRNA level, including deletions, substitutions, etc., as well as to determine mRNA transcript levels. Such oligonucleotide probes can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, neoplastic or hyperplastic disorders (e.g. aberrant cell growth).

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a RIEG gene.

Any cell type or tissue, preferably monocytes, endothelial cells, or smooth muscle cells, in which the RIEG is expressed may be utilized in the diagnostics described below. For example, a subject's bodily fluid (e.g. blood) can be obtained by known techniques (e.g. venipuncture). Alternatively, nucleic acid tests can be performed on dry samples (e.g. hair or skin). Fetal nucleic acid samples can be obtained from maternal blood as described in International Patent Application No. WO91/07660 to Bianchi. Alternatively, amniocentesis or chorionic villi may be obtained for performing prenatal testing.

Diagnostic procedures may also be performed in situ directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents may be used as probes and/or primers for such in situ procedures (see, for example, Nuovo, G. J., 1992, PCR in situ hybridization: protocols and applications, Raven Press, NY).

In addition to methods which focus primarily on the detection of one nucleic acid sequence, profiles may also be assessed in such detection schemes. Fingerprint profiles may be generated, for example, by utilizing a differential display procedure, Northern analysis and/or RT-PCR.

Antibodies directed against wild type or mutant RIEG proteins, which are discussed, above, may also be used in disease diagnostics and prognostics. Such diagnostic methods, may be used to detect abnormalities in the level of RIEG protein expression, or abnormalities in the structure and/or tissue, cellular, or subcellular location of RIEG protein. Structural differences may include, for example, differences in the size, electronegativity, or antigenicity of the mutant RIEG protein relative to the normal RIEG protein. Protein from the tissue or cell type to be analyzed may easily be detected or isolated using techniques which are well known to one of skill in the art, including but not limited to Western blot analysis. For a detailed explanation of methods for carrying out Western blot analysis, see Sambrook et al., 1989, supra, at Chapter 18. The protein detection and isolation methods employed herein may also be such as those described in Harlow and Lane, for example, (Harlow, E. and Lane, D., 1988, “Antibodies: A Laboratory Manual”, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.), which is incorporated herein by reference in its entirety.

This can be accomplished, for example, by immunofluorescence techniques employing a fluorescently labeled antibody (see below) coupled with light microscopic, flow cytometric, or fluorometric detection. The antibodies (or fragments thereof) useful in the present invention may, additionally, be employed histologically, as in immunofluorescence or immunoelectron microscopy, for in situ detection of RIEG proteins. In situ detection may be accomplished by removing a histological specimen from a patient, and applying thereto a labeled antibody of the present invention. The antibody (or fragment) is preferably applied by overlaying the labeled antibody (or fragment) onto a biological sample. Through the use of such a procedure, it is possible to determine not only the presence of the RIEG protein, but also its distribution in the examined tissue. Using the present invention, one of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such in situ detection.

Often a solid phase support or carrier is used as a support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amyloses, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antigen or antibody, or will be able to ascertain the same by use of routine experimentation.

One means for labeling an anti-RIEG protein specific antibody is via linkage to an enzyme and use in an enzyme immunoassay (EIA) (Voller, "The Enzyme Linked Immunosorbent Assay (ELISA)", Diagnostic Horizons 2:1-7, 1978; Microbiological Associates Quarterly Publication, Walkersville, Md.; Voller, et al., J. Clin. Pathol. 31:507-520 (1978); Butler, Meth. Enzymol. 73:482-523 (1981); Maggio, (ed.) Enzyme Immunoassay, CRC Press, Boca Raton, Fla., 1980; Ishikawa, et al., (eds.) Enzyme Immunoassay, Kagaku Shoin, Tokyo, 1981). The enzyme which is bound to the antibody will react with an appropriate substrate, preferably a chromogenic substrate, in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorometric or by visual means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, starch phosphorylase, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycoexophastrase, dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glu-
cose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect fingerprint gene wild type or mutant peptides through the use of a radioimmunoassay (RIA) (see, for example, Weintraub, B., Principles of Radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March, 1986, which is incorporated by reference herein). The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycoerythrin, allophycocyanin, o-phthaldehyde and fluorescamine.

The antibody can also be detectably labeled using fluorescence emitting metals such as 152Eu, or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylene-triaminepentaacetic acid (DTPA) or ethylenediaminetetraacetic acid (EDTA).

The antibody can also be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, thermatic acidinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

Moreover, it will be understood that any of the above methods for detecting alterations in a RIEG gene or gene product can be used to monitor the course of treatment or therapy. 4.7. Drug Screening Assays

In drug screening assays described herein, in addition to the RIEG nucleic acid molecules and polypeptides described above, the present invention also provides for the use of nucleic comprising at least a portion of the nucleic acid sequence shown in SEQ ID No:5 or polypeptides comprising at least a portion of the amino acid sequence shown in SEQ ID No:6.

Furthermore, by making available purified and recombinant RIEG polypeptides, the present invention facilitates the development of assays which can be used to screen for drugs, including RIEG homologs, which are either agonists or antagonists of the normal cellular function of the subject RIEG polypeptides, or of their role in the pathogenesis of cellular differentiation and/or proliferation and disorders related thereto. In one embodiment, the assay evaluates the ability of a compound to modulate binding between a RIEG polypeptide and a molecule, be it protein or DNA, that interacts either upstream or downstream of the RIEG polypeptide in the TGFβ signaling pathway. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by a skilled artisan.

4.7.1 Cell-Free Assays

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as “primary” screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the in vitro system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifested in an alteration of binding affinity with upstream or downstream elements. Accordingly, in an exemplary screening assay of the present invention, the compound of interest is contacted with proteins which may function upstream (including both activators and repressors of its activity) or to proteins or nucleic acids which may function downstream of the RIEG polypeptide, whether they are positively or negatively regulated by it. The mixture of the compound and the upstream or downstream element is then added a composition containing a RIEG polypeptide. Detection and quantification of complexes of RIEG with it’s upstream or downstream elements provide a means for determining a compound’s efficacy in inhibiting (or potentiating) complex formation between RIEG and the RIEG-binding elements. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, isolated and purified RIEG polypeptide is added to a composition containing the RIEG-binding element, and the formation of a complex is quantitated in the absence of the test compound.

Complex formation between the RIEG polypeptide and a RIEG binding element may be detected by a variety of techniques. Modulation of the formation of complexes can be quantitated using, for example, detectably labeled proteins such as radiolabeled, fluorescently labeled, or enzymatically labeled RIEG polypeptides, by immunocoupling, or by chromatographic detection. Typically, it will be desirable to immobilize either RIEG and/or its binding protein to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of RIEG to an upstream or downstream element, in the presence and absence of a candidate agent, can be accomplished in any vessel suitable for containing the reactants. Examples include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/RIEG (GST/RIEG) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtiter plates, which are then combined with the cell lysates, e.g., an 35S-labeled, and the test compound, and the mixture incubated under conditions conducive to complex formation, e.g., at physiological conditions for salt and pH, though slightly more stringent...
conditions may be desired. Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly (e.g. beads placed in scintillant), or in the supernatant after the complexes are subsequently dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of RIEG-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques such as described in the appended examples.

Other techniques for immobilizing proteins on matrices are also available for use in the subject assay. For instance, either RIEG or its cognate binding protein can be immobilized utilizing conjugation of biotin and streptavidin. For instance, biotinylated RIEG molecules can be prepared from biotin-NHS (N-hydroxysuccinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with RIEG but which do not interfere with binding of upstream or downstream elements can be derivatized to the wells of the plate, and RIEG trapped in the wells by antibody conjugation. As above, preparations of a RIEG-binding protein and a test compound are incubated in the RIEG-presenting wells of the plate, and the amount of complex trapped in the well can be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the RIEG binding element, or which are reactive with RIEG protein and compete with the binding element; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the binding element, either intrinsic or extrinsic activity. In the instance of the latter, the enzyme can be chemically conjugated or provided as a fusion protein with the RIEG-BP. To illustrate, the RIEG-BP can be chemically cross-linked or genetically fused with horseradish peroxidase, and the amount of polypeptide trapped in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. 3,3′-diaminobenzidine terahydrochloride or 4-chloro-1-naphthol. Likewise, a fusion protein comprising the polypeptide and glutathione-S-transferase can be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2.4-dinitrobenzene (Habig et al (1974) J Biol Chem 249:7130).

For processes which rely on immunodetection for quantitating one of the proteins trapped in the complex, antibodies against the protein, such as anti-RIEG antibodies, can be used. Alternatively, the protein to be detected in the complex can be “epitope tagged” in the form of a fusion protein which includes, in addition to the RIEG sequence, a second polypeptide for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include myc-epitopes (e.g., see Ellison et al. (1991) J Biol Chem 266:21150–21157) which includes a 10-residue sequence from c-myc, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharmacia, N.J.).

4.7.2. Cell Based Assays

In addition to cell-free assays, such as described above, the readily available source of mammalian RIEG proteins provided by the present invention also facilitates the generation of cell-based assays for identifying small molecule agonists/antagonists and the like. For example, cells which are sensitive to TGFβ signals can be caused to overexpress a recombinant RIEG protein in the presence and absence of a test agent of interest, with the assay scoring for modulation in RIEG responses by the target cell mediated by the test agent. As with the cell-free assays, agents which produce a statistically significant change in RIEG-dependent responses (either inhibition or potentiation) can be identified. In an illustrative embodiment, the expression or activity of a RIEG is modulated embryos or cells and the effects of compounds of interest on the readout of interest (such as tissue differentiation, proliferation, tumorigenesis) are measured. For example, the expression of genes which are up- or down-regulated in response to a RIEG-dependent signal cascade can be assayed. In preferred embodiments, the regulatory regions of such genes, e.g., the 5′ flanking promoter and enhancer regions, are operably linked to a detectable marker (such as luciferase) which encodes a gene product that can be readily detected.

Exemplary cell lines may include non-recombinant monocyte cell lines, such as U937 (ATCC® CRL-1593), THP-1 (ATCC® TIB-202), and P388D1 (ATCC® TIB-63) endothelial cells such as HUVEC’s and bovine aortic endothelial cells (BAEC’s); as well as generic mammalian cell lines such as HeLa cells and COS cells, e.g. COS-7 (ATCC® CRL-1651). Further, the transgenic animals discussed herein may be used to generate cell lines, containing one or more cell types involved in cardiovascular disease, that can be used as cell culture models for this disorder. While primary cultures derived from the cardiovascular disease transgenic animals of the invention may be utilized, the generation of continuous cell lines is preferred. For examples of techniques which may be used to derive a continuous cell line from the transgenic animals, see Small et al., 1985, Mol. Cell Biol. 5:642–648.

For example, the effect of a test compound on a variety of end points could be tested including rates of LDL uptake, adhesion to endothelial cells, transmigration, foam cell formation, fatty streak formation, and production by foam cells of growth factors such as bFGF, IGF-I, VEGF, IL-1, M-CSF, TGFβ, TGFα, TNFα, HB-EGF, PDGF, IFN-γ, and GM-CSF. Similarly, HUVEC’s can be treated with test compounds or transfected with genetically engineered RIEG genes. The HUVEC’s can then be examined for phenotypes associated with cardiovascular disease, including, but not limited to changes in cellular morphology, cell proliferation, cell migration, and mononuclear cell adhesion; or for the effects on production of other proteins involved in vascular disease such as ICAM, VCAM, PDGF-β, and E-selectin.

In the event that the RIEG proteins, or in complexes with other proteins, are capable of binding DNA and modifying transcription of a gene, a transcriptional based assay could be used, for example, in which a RIEG responsive regulatory sequence is operably linked to a detectable marker gene.

Monitoring the influence of compounds on cells may be applied not only in basic drug screening, but also in clinical trials. In such clinical trials, the expression of a panel of genes may be used as a “read out” of a particular drug’s therapeutic effect.

In yet another aspect of the invention, the subject RIEG polypeptides can be used to generate a “two hybrid” assay (see, for example, U.S. Pat. No. 5,283,317; Zverov et al. (1993) Cell 72:223–232; Madura et al. (1993) J Biol Chem 268:12046–12054; Bartel et al. (1993) Biotechniques 14:920–924; Iwabuchi et al. (1993) Oncogene 8:1693–1696; and Brent WO94/0500), for isolating coding sequences for
other cellular proteins which bind to or interact with RIEG ("RIEG-binding proteins" or "RIEG-bp"), such as FCHD 534, and the like. Such RIEG-binding proteins would likely also be involved in the propagation of TGFβ signals by the RIEG proteins as, for example, the upstream or downstream elements of the RIEG pathway or as collatoral regulators of signal bioactivity.

Briefly, the two hybrid assay relies on reconstituting in vivo a functional transcriptional activator protein from two separate fusion proteins. In particular, the method makes use of chimeric genes which express hybrid proteins. To illustrate, a first hybrid gene comprises the coding sequence for a DNA-binding domain of a transcriptional activator fused in frame to the coding sequence for a RIEG polypeptide. The second hybrid protein encodes a transcriptional activation domain fused in frame to a sample gene from a CDNA library. If the bait and sample hybrid proteins are able to interact, e.g., form a RIEG-dependent complex, they bring into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription of a reporter gene which is operably linked to a transcriptional regulatory site responsive to the transcriptional activator, and expression of the reporter gene can be detected and used to score for the interaction of the RIEG and sample proteins.

4.8 Transgenic Animals

These systems may be used in a variety of applications. For example, the cell- and animal-based model systems may be used to further characterize RIEG genes and proteins. In addition, such assays may be utilized as part of screening strategies designed to identify compounds which are capable of ameliorating disease symptoms. Thus, the animal- and cell-based models may be used to identify drugs, pharmaceuticals, therapies and interventions which may be effective in treating disease.

4.8.1. Animal-Based Systems

One aspect of the present invention concerns transgenic animals which are comprised of cells (or that animal) which contain a transgene of the present invention and which preferably (though optionally) express an exogenous RIEG protein in one or more cells in the animal. A RIEG transgene can encode the wild-type form of the protein, or can encode homologs thereof, including both agonists and antagonists, as well as antisense constructs. In preferred embodiments, the expression of the transgene is restricted to specific subsets of cells, tissues or developmental stages utilizing, for example, cis-acting sequences that control expression in the desired pattern. In the present invention, such mosaic expression of a RIEG protein can be essential for many forms of lineage analysis and can additionally provide a means to assess the effects of, for example, lack of RIEG expression which might generally alter development in small patches of tissue within an otherwise normal embryo. Toward this and, tissue-specific regulatory sequences and conditional regulatory sequences can be used to control expression of the transgene in certain spatial patterns. Moreover, temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences.

Genetic techniques which allow for the expression of transgenes can be regulated via site-specific genetic manipulation in vivo are known to those skilled in the art. For instance, genetic systems are available which allow for the regulated expression of a recombinase that catalyzes the genetic recombination a target sequence. As used herein, the phrase "target sequence" refers to a nucleotide sequence that is genetically recombined by a recombinase. The target sequence is flanked by recombinase recognition sequences and is generally either excised or inverted in cells expressing recombinase activity. Recombinase catalyzed recombination events can be designed such that recombination of the target sequence results in either the activation or repression of expression of one of the subject RIEG proteins. For example, excision of a target sequence which interferes with the expression of a recombinant RIEG gene, such as one which encodes an antagonistic homolog or an antisense transcript, can be designed to activate expression of that gene. This interference with expression of the protein can result from a variety of mechanisms, such as spatial separation of the RIEG gene from the promoter element or an internal stop codon. Moreover, the transgene can be made wherein the coding sequence of the gene is flanked by recombinase recognition sequences and is initially transcribed into cells in a 5' to 3' orientation with respect to the promoter element. In such an instance, inversion of the target sequence will reorient the subject gene by placing the 3' end of the coding sequence in an orientation with respect to the promoter element which allow for promoter driven transcriptional activation.

The transgenic animals of the present invention all include within a plurality of their cells a transgene of the present invention, which transgene alters the phenotype of the "host cell" with respect to regulation of cell growth, death and/or differentiation. Since it is possible to produce transgenic organisms of the invention utilizing one or more of the transgene constructs described herein, a general description will be given of the production of transgenic organisms by referring generally to exogenous genetic material. This general description can be adapted by those skilled in the art in order to incorporate specific transgene sequences into organisms utilizing the methods and materials described below.

In an illustrative embodiment, either the cre/loxP recombinase system of bacteriophage P1 (Lalson et al. (1992) PNAS 89:6232–6236; Orban et al. (1992) PNAS 89:6861–6865) or the FLP recombinase system of Saccharomyces cerevisiae (O'Gorman et al. (1991) Science 251: 1351–1355; PCT publication WO 92/15604) can be used to generate in vivo site-specific genetic recombination systems. Cre recombinase catalyzes the site-specific recombination of an intervening target sequence located between loxP sequences. loxP sequences are 34 base pair nucleotide repeat sequences to which the Cre recombinase binds and are required for Cre recombinase mediated genetic recombination. The orientation of loxP sequences determines whether the intervening target sequence is excised or inverted when Cre recombinase is present (Abremski et al. (1984) J. Biol. Chem. 259:1509–1514); catalyzing the excision of the target sequence when the loxP sequences are oriented as direct repeats and catalyzes inversion of the target sequence when loxP sequences are oriented as inverted repeats.

Accordingly, genetic recombination of the target sequence is dependent on expression of the Cre recombinase. Expression of the recombinase can be regulated by promoter elements which are subject to regulatory control, e.g., tissue-specific, developmental stage-specific, inducible or repressible by externally added agents. This regulated control will result in genetic recombination of the target sequence only in cells where recombinase expression is mediated by the promoter element. Thus, the activation expression of a recombinant RIEG protein can be regulated via control of recombinase expression.

Use of the cre/loxP recombinase system to regulate expression of a recombinant RIEG protein requires the
construction of a transgenic animal containing transgenes encoding both the Cre recombinase and the subject protein. Animals containing both the Cre recombinase and a recomb- inant RIEG gene can be provided through the construction of “double” transgenic animals. A convenient method for providing such animals is to mate two transgenic animals each containing a transgene, e.g., a RIEG gene and recom- binase gene.

One advantage derived from initially constructing transgenic animals containing a RIEG transgene in a recombi- nase-mediated expressible format derives from the likely- hood that the subject protein, whether agonistic or antagonistic, can be deleterious upon expression in the transgenic animal. In such an instance, a founder population, in which the subject transgene is silent in all tissues, can be propagated and maintained. Individuals of this founder population can be crossed with animals expressing the recombinase in, for example, one or more tissues and/or a desired temporal pattern. Thus, the creation of a founder population in which, for example, an antagonistic RIEG transgene is silent will allow the study of progeny from that founder in which disruption of RIEG mediated induction in a particular tissue or at certain developmental stages would result in, for example, a lethal phenotype.

Similar conditional transgenes can be provided using prokaryotic promoter sequences which require prokaryotic proteins to be simultaneously expressed in order to facilitate expression of the RIEG transgene. Exemplary promoters and the corresponding trans-activating prokaryotic proteins are given in U.S. Pat. No. 4,833,080.

Moreover, expression of the conditional transgenes can be induced by gene therapy-like methods wherein a gene encoding the trans-activating protein, e.g. a recombinase or a prokaryotic protein, is delivered to the tissue and caused to be expressed, such as in a cell-type specific manner. By this method, a RIEG transgene could remain silent into adult- hood until “turned on” by the introduction of the trans- activator.

In an exemplary embodiment, the “transgenic non-human animals” of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonal target cell. The specific line(s) of any animal used to practice this invention are selected for general good health, good embryo yields, good pronuclear visibility in the embryo, and good reproductive fitness. In addition, the haplotype is a significant factor. For example, when transgenic mice are to be produced, strains such as C57BL/6 or FVB lines are often used (Jackson Laboratory, Bar Harbor, Me.). Preferred strains are those with H-2k, H-2d or H-2a haplotypes such as C57BL/6 or DBA/1. The line(s) used to practice this invention may themselves be transgenics, and/or may be knock- outs (i.e., obtained from animals which have one or more genes partially or completely suppressed).

In one embodiment, the transgene construct is introduced into a single stage embryo. The zygote is the best target for micro-injection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter which allows reproducible injection of 1-2 pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster et al. (1985) PNAS 82:4438-4442). As a conse- quence, all cells of the transgenic animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene.

Normally, fertilized embryos are incubated in suitable media until the pronuclei appear. At about this time, the nucleotide sequence comprising the transgene is introduced into the female or male pronucleus as described below. In some species such as mice, the male pronucleus is preferred. It is most preferred that the exogenous genetic material be added to the male DNA complement of the zygote prior to its being processed by the ovum nucleus or the zygote female pronucleus. It is thought that the ovum nucleus or female pronucleus release molecules which affect the male DNA complement, perhaps by replacing the protamines of the male DNA with histones, thereby facilitating the combi- nation of the female and male DNA complements to form the diploid zygote.

Thus, it is preferred that the exogenous genetic material be added to the male complement of DNA or any other complement of DNA prior to its being affected by the female pronucleus. For example, the exogenous genetic material is added to the early male pronucleus, as soon as possible after the formation of the male pronucleus, which is when the male and female pronuclei are well separated and both are located close to the cell membrane. Alternatively, the exogenous genetic material could be added to the nucleus of the sperm after it has been induced to undergo decondensation. Sperm containing the exogenous genetic material can then be added to the ovum or the decondensed sperm could be added to the ovum with the transgene constructs being added as soon as possible thereafter.

Introduction of the transgene nucleotide sequence into the embryo may be accomplished by any means known in the art such as, for example, microinjection, electroporation, or lipofection. Following introduction of the transgene nucleo- tide sequence into the embryo, the embryo may be incubated in vitro for varying amounts of time, or reimplemented into the surrogate host, or both. In vitro incubation to maturity is within the scope of this invention. One common method is to incubate the embryos in vitro for about 1-7 days, depending on the species, and then reimplant them into the surrogate host.

For the purposes of this invention a zygote is essentially the formation of a diploid cell which is capable of develop- ing into a complete organism. Generally, the zygote will be comprised of an egg containing a nucleus formed, either naturally or artificially, by the fusion of two haploid nuclei from a gamete or gametotes. Thus, the gamete nuclei must be ones which are naturally compatible, i.e., ones which result in a viable zygote capable of undergoing differentiation and developing into a functioning organism. Generally, a euploid zygote is preferred. If an aneuploid zygote is obtained, then the number of chromosomes should not vary by more than one with respect to the euploid number of the organism from which either gamete originated.

In addition to similar biological considerations, physical ones also govern the amount (e.g., volume) of exogenous genetic material which can be added to the nucleus of the zygote or to the genetic material which forms a part of the zygote nucleus. If no genetic material is removed, then the amount of exogenous genetic material which can be added is limited by the amount which will be absorbed without being physically disruptive. Generally, the volume of exog- enous genetic material inserted will not exceed about 10 picoliters. The physical effects of addition must not be so great as to physically destroy the viability of the zygote. The biological limit of the number and variety of DNA
sequences will vary depending upon the particular zygote and functions of the exogenous genetic material and will be readily apparent to one skilled in the art, because the genetic material, including the exogenous genetic material, of the resulting zygote must be biologically capable of initiating and maintaining the differentiation and development of the zygote into a functional organism.

The number of copies of the transgene constructs which are added to the zygote is dependent upon the total amount of exogenous genetic material added and will be the amount which enables the genetic transformation to occur. Theoretically only one copy is required; however, generally, numerous copies are utilized, for example, 1,000–20,000 copies of the transgene construct, in order to insure that one copy is functional. As regards the present invention, there will often be an advantage to having more than one functioning copy of each of the inserted exogenous DNA sequences to enhance the phenotypic expression of the exogenous DNA sequences.

Any technique which allows for the addition of the exogenous genetic material into nucleic gene material can be utilized so long as it is not destructive to the cell, nuclear membrane or other existing cellular or genetic structures. The exogenous genetic material is preferentially inserted into the nuclear gene material by microinjection. Microinjection of cells and cellular structures is known and is used in the art.

Reimplantation is accomplished using standard methods. Usually, the surrogate host is anesthetized, and the embryos are inserted into the oviduct. The number of embryos implanted into a particular host will vary by species, but will usually be comparable to the number of offspring the species naturally produces.

Transgenic offspring of the surrogate host may be screened for the presence and/or expression of the transgene by any suitable method. Screening is often accomplished by Southern blot or Northern blot analysis, using a probe that is complementary to at least a portion of the transgene. Western blot analysis using an antibody against the protein encoded by the transgene may be employed as an alternative or additional method for screening for the presence of the transgene product. Typically, DNA is prepared from tail tissue and analyzed by Southern analysis or PCR for the transgene. Alternatively, the tissues or cells believed to express the transgene at the highest levels are tested for the presence and expression of the transgene using Southern analysis or PCR, although any tissues or cell types may be used for this analysis.

Alternative or additional methods for evaluating the presence of the transgene include, without limitation, suitable biochemical assays such as enzyme and/or immunological assays, histological stains for particular marker or enzyme activities, flow cytometric analysis, and the like. Analysis of the blood may also be useful to detect the presence of the transgene product in the blood, as well as to evaluate the effect of the transgene on the levels of various types of blood cells and other blood constituents.

Progeny of the transgenic animals may be obtained by mating the transgenic animal with a suitable partner, or by in vitro fertilization of eggs and/or sperm obtained from the transgenic animal. Where mating with a partner is to be performed, the partner may or may not be transgenic and/or a knockout; where it is transgenic, it may contain the same or a different transgene, or both. Alternatively, the partner may be a parental line. Where in vitro fertilization is used, the fertilized embryo may be implanted into a surrogate host or incubated in vitro, or both. Using either method, the progeny may be evaluated for the presence of the transgene using methods described above, or other appropriate methods.

The transgenic animals produced in accordance with the present invention will include exogenous genetic material. As set out above, the exogenous genetic material will, in certain embodiments, be a DNA sequence which results in the production of a RIEG protein (either agonistic or antagonistic), and antisense transcript, or a RIEG mutant. Further, in such embodiments the sequence will be attached to a transcription control element, e.g., a promoter, which preferably allows the expression of the transgene product in a specific type of cell.

Retroviral infection can also be used to introduce transgene into a non-human animal. The developing non-human embryo can be cultured in vitro to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenicke, R. (1976) PNAS 73:1260–1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Manipulating the Mouse Embryo, Hogan eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahnner et al. (1985) PNAS 82:6927–6931; Van der Putten et al. (1985) PNAS 82:6148–6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, supra; Stewart et al. (1987) EMBO J. 6:383–388). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahnner et al. (1982) Nature 298:623–628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic non-human animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrataminal retroviral infection of the midgestation embryo (Jahnner et al. (1982) supra).


In one embodiment, gene targeting, which is a method of using homologous recombination to modify an animal’s genome, can be used to introduce changes into cultured embryonic stem cells. By targeting a RIEG gene of interest in ES cells, these changes can be introduced into the germplines of animals to generate chimeras. The gene targeting procedure is accomplished by introducing into tissue culture cells a DNA targeting construct that includes a segment homologous to a target RIEG locus, and which also includes an intended sequence modification to the RIEG genomic sequence (e.g., insertion, deletion, point mutation). The treated cells are then screened for accurate targeting to identify and isolate those which have been properly targeted.
Gene targeting in embryonic stem cells is in fact a scheme contemplated by the present invention as a means for disrupting a RIEG gene function through the use of a targeting transgene construct designed to undergo homologous recombination with one or more RIEG genomic sequences. The targeting construct can be arranged so that, upon recombination with an element of a RIEG gene, a positive selection marker is inserted into (or replaces) coding sequences of the targeted signalin gene. The inserted sequence functionally disrupts the RIEG gene, while also providing a positive selection trait. Exemplary RIEG targeting constructs are described in more detail below.

Generally, the embryonic stem cells (ES cells) used to produce the knockout animals will be of the same species as the knockout animal to be generated. Thus for example, mouse embryonic stem cells will usually be used for generation of knockout mice.

Embryonic stem cells are generated and maintained using methods well known to the skilled artisan such as those described by Doetschman et al. (1985) J. Embryol. Exp. Morphol. 87:27-45. Any line of ES cells can be used; however, the line chosen is typically selected for the ability of the cells to integrate into and become part of the germ line of a developing embryo so as to generate germ line transmission of the knockout construct. Thus, any ES cell line that is believed to have this capability is suitable for use herein. One mouse strain that is typically used for production of ES cells, is the 129J strain. Another ES cell line is murine cell line D3 (American Type Culture Collection, catalog no. CLK 1934) Still another preferred ES cell line is the HW6 cell line (Iofe et al. (1995) PNAS 92:7357-7361). The cells are cultured and prepared for knockout construct insertion using methods well known to the skilled artisan, such as those set forth by Robertson in: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, E. J. Robertson, ed. IRL Press, Washington, D.C. [1987]); by Bradley et al. (1986) Current Topics in Devol. Biol. 20:357-371); and by Hogan et al. (Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [1986]).

Insertion of the knockout construct into the ES cells can be accomplished using a variety of methods well known in the art including for example, electroporation, microinjection, and calcium phosphate treatment. A preferred method of insertion is electroporation.

Each knockout construct to be inserted into the cell must first be in the linear form. Therefore, if the knockout construct has been inserted into a vector (described infra), linearization is accomplished by digesting the DNA with a suitable restriction endonuclease selected to cut only within the vector sequence and not within the knockout construct sequence.

For insertion, the knockout construct is added to the ES cells under appropriate conditions for the insertion method chosen, as is known to the skilled artisan. Where more than one construct is to be introduced into the ES cell, each knockout construct can be introduced simultaneously or one at a time.

If the ES cells are to be electroporated, the ES cells and knockout construct DNA are exposed to an electric pulse using an electroporation machine and following the manufacturer’s guidelines for use. After electroporation, the ES cells are typically allowed to recover under suitable incubation conditions. The cells are then screened for the presence of the knockout construct.

Screening can be accomplished using a variety of methods. Where the marker gene is an antibiotic resistance gene, for example, the ES cells may be cultured in the presence of an otherwise lethal concentration of antibiotic. Those ES cells that survive have presumably integrated the knockout construct. If the marker gene is other than an antibiotic resistance gene, a Southern blot of the ES cell genomic DNA can be probed with a sequence of DNA designed to hybridize only to the marker sequence. Alternatively, PCR can be used. Finally, if the marker gene is a gene that encodes an enzyme whose activity can be detected (e.g., β-galactosidase), the enzyme substrate can be added to the cells under suitable conditions, and the enzymatic activity can be analyzed. One skilled in the art will be familiar with other useful markers and the means for detecting their presence in a given cell. All such markers are contemplated as being included within the scope of the teaching of this invention.

The knockout construct may integrate into several locations in the ES cell genome, and may integrate into a different location in each ES cell’s genome due to the occurrence of random insertion events. The desired location of insertion is in a complementary position to the DNA sequence to be knocked out (e.g., the RIEG coding sequence, transcriptional regulatory sequence, etc. Typically, less than about 1–5% of the ES cells that take up the knockout construct will actually integrate the knockout construct in the desired location. To identify those ES cells with proper integration of the knockout construct, total DNA can be extracted from the ES cells using standard methods. The DNA can then be probed on a Southern blot with a probe or probes designed to hybridize to a specific pattern to genomic DNA digested with particular restriction enzyme(s). Alternatively, or additionally, the genomic DNA can be amplified by PCR with probes specifically designed to amplify DNA fragments of a particular size and sequence (i.e., only those cells containing the knockout construct in the proper position will generate DNA fragments of the proper size).

After suitable ES cells containing the knockout construct in the proper location have been identified, the cells can be inserted into an embryo. Insertion may be accomplished in a variety of ways known to the skilled artisan, however a preferred method is by microinjection. For microinjection, about 10–30 cells are collected into a micropipet and injected into embryos that are at the proper stage of development to permit integration of the foreign ES cell containing the knockout construct into the developing embryo. For instance, as the appended Examples describe, the transformed ES cells can be microinjected into blastocysts.

The suitable stage of development for the embryo used for insertion of ES cells is very species dependent, however for mice it is about 3.5 days. The embryos are obtained by perfusing the uterus of pregnant females. Suitable methods for accomplishing this are known to the skilled artisan, and are set forth by, e.g., Bradley et al. (supra).

While any embryo of the right stage of development is suitable for use, preferred embryos are male. In mice, the preferred embryos also have genes coding for a coat color that is different from the coat color encoded by the ES cell genes. In this way, the offspring can be screened easily for the presence of the knockout construct by looking for mosaic coat color (indicating that the ES cell was incorporated into the developing embryo). Thus, for example, if the ES cell line carries the genes for white fur, the embryo selected will carry genes for black or brown fur.

After the ES cell has been introduced into the embryo, the embryo may be implanted into the uterus of a pseudopregnant foster mother for gestation. While any foster mother may be used, the foster mother is typically selected for her ability to breed and reproduce well, and for her ability to
care for the young. Such foster mothers are typically prepared by mating with vasectomized males of the same species. The stage of the pseudopregnant foster mother is important for successful implantation, and it is species dependent. For mice, this stage is about 2–3 days pseudopregnant.

Offspring that are born to the foster mother may be screened initially for mosaic coat color where the coat color selection strategy (as described above, and in the appended examples) has been employed. In addition, or as an alternative, DNA from tail tissue of the offspring may be screened for the presence of the knockout construct using Southern blots and/or PCR as described above. Offspring that appear to be mosaics may then be crossed to each other, if they are believed to carry the knockout construct in their germ line, in order to generate homozygous knockout animals. Homozygotes may be identified by Southern blotting of equivalent amounts of genomic DNA from mice that are the product of this cross, as well as mice that are known heterozygotes and wild type mice.

Other means of identifying and characterizing the knockout offspring are available. For example, Northern blots can be used to probe the mRNA for the presence or absence of transcripts encoding either the gene knocked out, the marker gene, or both. In addition, Western blots can be used to assess the level of expression of the RIEG gene knocked out in various tissues of the offspring by probing the Western blot with an antibody against the particular RIEG protein, or an antibody against the marker gene product, where this gene is expressed. Finally, in situ analysis (such as fixing the cells and labeling with antibody) and/or FACS (fluorescence activated cell sorting) analysis of various cells from the offspring can be conducted using suitable antibodies to look for the presence or absence of the knockout construct gene product.

Yet other methods of making knock-out or disruption transgenic animals are also generally known. See, for example, Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Recombinase dependent knockouts can also be generated, e.g. by homologous recombination to insert target sequences, such that tissue specific and/or temporal control of inactivation of a RIEG-gene can be controlled by recombinase sequences (described infra).

Animals containing more than one knockout construct and/or more than one transgene expression construct are prepared in any of several ways. The preferred manner of preparation is to generate a series of mammals, each containing one of the desired transgenic phenotypes. Such animals are bred together through a series of crosses, backcrosses and selections, to ultimately generate a single animal containing all desired knockout constructs and/or expression constructs, where the animal is otherwise congenic (genetically identical) to the wild type except for the presence of the knockout construct(s) and/or transgene(s).

The present invention is further illustrated by the following examples which should not be construed as limiting in any way. The contents of all cited references (including literature references, issued patents, published patent applications as cited throughout this application are hereby expressly incorporated by reference. The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press; 1989); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Pat. No. 4,683,195; Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds., 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Methods In Enzymology, Vols. 154 and 155 (Wu et al. eds.), Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I–IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

The present invention is further illustrated by the following examples which should not be construed as limiting in any way. The contents of all cited references (including literature references, issued patents, published patent applications, and co-pending patent applications) cited throughout this application are hereby expressly incorporated by reference.

EXAMPLE 1

Mapping of the Rieger Syndrome Causing Gene and Exclusion of the Epidermal Growth Factor (EGF) Gene as a Candidate Gene for Rieger Syndrome

Materials and Methods

Patients and Cell Lines.

Seven families were identified with characteristic RGS. In all families at least one individual within the pedigree manifested all three cardinal signs of RGS: abnormalities of the anterior segment of the eye, hypodontia, and failure of normal involution of the periumbilical skin. Individuals were considered to be affected within families if they had any of these features.

The MS cell line was established from a RGS patient with karyotype 46,XX,t(4;11)q26q22. MS and her daughter HS have the same balanced translocation and ocular features of Rieger syndrome as well as dental hypoplasia.

The LR cell line was established from a patient with an apparent de novo balanced translocation and karyotype 46,XX,t(4;11)q27;121. Both of LR’s parents have normal karyotypes, and LR has anterior segment abnormalities, dental hypoplasia, and umbilical outpouching (all consistent with Rieger). In addition, LR has mild development delay and polydactyly.

The AR cell line was established from a patient with the karyotype 46,XY,?del(4)inv(4)(q26q28). AR had an omphalocele repaired at birth, anterior segment abnormalities, dental hypoplasia, and moderate developmental delay. AR is adopted, and family history data are unavailable.

Single-Strand Conformation Polymorphism (SSCP) Analysis and Sequencing.

SSCP analysis was carried out using primers selected from each of the introns of EGF and chosen to flank the appropriate exon and to generate fragments of 100 to 220 base pairs in length. The sequence of primers for each of the
individual exons is shown in Table 1. SSCP analysis was carried out using a modification of a technique of Orita et al. (Orita M. et al., (1989) *Genomics* 5:874-879) with $^{35}$S-labeled PCR products run on 6% polyacrylamide gels with 10% glycerol at room temperature. Following autoradiography, films were inspected for potential allelic variants.

**Table 1**

<table>
<thead>
<tr>
<th>EGF PRIMERS</th>
<th>SIZE</th>
<th>TM</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGF1 FWD: GAGAGTCCAATCTACATTCAAG (SEQ ID NO.51)</td>
<td>150</td>
<td>55</td>
</tr>
<tr>
<td>EGF1 REV: GAGTACCTACCCACACAGTAA (SEQ ID NO.52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF2 FWD: GGTCTCTTCTCTCCACCCA (SEQ ID NO.53)</td>
<td>239</td>
<td>51</td>
</tr>
<tr>
<td>EGF2 REV: TGTTAGTTAAGCATTATTAC (SEQ ID NO.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF3 FWD: TAAACATTGATTCTGCCAGGAGGAG (SEQ ID NO.55)</td>
<td>230</td>
<td>55</td>
</tr>
<tr>
<td>EGF3 REV: CTGCAATAGCCAGAATTACC (SEQ ID NO.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF4 FWD: TGAAGACCTGGTTGGCCTCA (SEQ ID NO.57)</td>
<td>267</td>
<td>51</td>
</tr>
<tr>
<td>EGF4 REV: ATACAGAAGGTAAGATTTAC (SEQ ID NO.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF5 FWD: TTTCTTTCTGCTCTTCTTTAG (SEQ ID NO.59)</td>
<td>243</td>
<td>53</td>
</tr>
<tr>
<td>EGF5 REV: CAAGCTCAACGAGTCACTCAC (SEQ ID NO.60)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF6 FWD: CCCACTCCAAATAAGGATC (SEQ ID NO.61)</td>
<td>102</td>
<td>53</td>
</tr>
<tr>
<td>EGF6 REV: CATTACCCCATGACGTTC (SEQ ID NO.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF7 FWD: TTGAAATGTATTGCTGGTGCTTG (SEQ ID NO.63)</td>
<td>187</td>
<td>53</td>
</tr>
<tr>
<td>EGF7 REV: TAAAAGCAGTAAATACTGGTGCC (SEQ ID NO.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF8 FWD: AACACAGATTCTGCCGTCTCT (SEQ ID NO.65)</td>
<td>186</td>
<td>54</td>
</tr>
<tr>
<td>EGF8 REV: CCAACAAAAGTTAAAGAGCA (SEQ ID NO.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF9 FWD: GAGCAGCATTACATTACGAAT (SEQ ID NO.67)</td>
<td>188</td>
<td>54</td>
</tr>
<tr>
<td>EGF9 REV: GAGCACTTCCAGCAACACCA (SEQ ID NO.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF10 FWD: TAAATCCAGCATTATTTGAAATT (SEQ ID NO.69)</td>
<td>200</td>
<td>53</td>
</tr>
<tr>
<td>EGF10 REV: TTCCACATCGTTCAACTAATAC (SEQ ID NO.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF11 FWD: AAAAGGCAGTCTGCCTGTCT (SEQ ID NO.71)</td>
<td>209</td>
<td>55</td>
</tr>
<tr>
<td>EGF11 REV: ACAGGTTAAAACTGAAGCAG (SEQ ID NO.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGF12 FWD: TATTTGCCGCTGCTGGCTCG (SEQ ID NO.73)</td>
<td>277</td>
<td>53</td>
</tr>
</tbody>
</table>

DNA sequence analysis on individual exons demonstrating SSCP variants was carried out using a technique of Sheffield (Sheffield, V. C. et al. (1993) *Genomics* 16:325-332). All sequences were compared to previously published cDNA sequences for EGF.
Somatic Cell Hybrid Creation and Breakpoint Mapping

Somatic cell hybrids were obtained from a polyethylene glycol-mediated fusion between lymphoblastoid cells established from the RGS patients carrying the translocations and hamster UCWS56 cells (provided by John Wasmuth). Hybrid cell clones that retained the der(4) and had lost all other chromosome 4 material were identified by PCR screening with several chromosome 4 STSs. From the total 50 hybrids, three independent cell lines derived from MS were isolated which carried 4q26-ter: M 9-1, MS 22-2, MS 19-2. From the fusion between LR and USW56, 46 hybrids were obtained, and four independent cell lines carrying 4pter-cen4q26 (LR 11-2, LR 10-1, LR 22-2 and LR 19-2) and 2 hybrids carrying 4q26-4tel (LR 22-2, LR 23-1) were selected. Hybrids AR22-2 and AR24-1 were created from AR lymphoblastoid cell line and contained different chromosome 4 counterparts (Table 2).

### Table 2

<table>
<thead>
<tr>
<th>Number</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS9-1</td>
<td>4q26-gcr:16pter-q22</td>
</tr>
<tr>
<td>MS22-2</td>
<td>same as above</td>
</tr>
<tr>
<td>L11-2</td>
<td>4pter-q26:13q21-ter</td>
</tr>
<tr>
<td>L10-1</td>
<td>same as above</td>
</tr>
<tr>
<td>L19-2</td>
<td>same as above</td>
</tr>
<tr>
<td>L22-2</td>
<td>4q26-ter:1pter-q21</td>
</tr>
<tr>
<td>L23-1</td>
<td>same as above</td>
</tr>
<tr>
<td>AR24-2</td>
<td>?del or inv(4) (q26g28)</td>
</tr>
<tr>
<td>AR22-1</td>
<td>normal 4</td>
</tr>
</tbody>
</table>

The table of somatic cell hybrids obtained from the RGS patients carrying translocations were tested with 82 unknown STSs derived from chromosome 4, with the most dense distribution in the region between D4S1564–D4S1573 loci as this region showed linkage with RGS (Murray, J. D. et al., (1992) Nature Genet. 2:46–49).

Screening of YAC Libraries

The YAC library used in this study was prepared at CEPH. The YAC library was screened by PCR assays. The YACs for the Génétion markers were identified through the CEPH-Génétion information service on the World Wide Web (http://www.cephp.genethon-map.html).

Construction of the YAC Contig

YAC clones were grown in AHC medium (deficient in uracil and tryptophan), and their DNA was isolated according to the method of Green and Olson (Green, E. D. and M. V. Olson (1990) Proc. Nat’l Acad. Sci. USA, 87:1213–1217). Total-yeast DNA containing YAC(s) was prepared in agarose plugs (Dixon, M. J. (1994) Am. J. Hum. Genetics 55:372–378), and the YAC and yeast chromosomes were separated and sized by pulsed-field gel electrophoresis using a Biorad CEO buffer and DR1. Electrophoresis was performed in 0.5 TBE at 12°C and 240V and with a switch interval of 20–50 sec for 20–25 h. They were stained with ethidium bromide and prepared for DNA transfer to Hybond N+ filters. To visualize the YACs the filters were hybridized with random-prime-labeled total human genomic DNA and autoradiographed MegaBlot 11 DNA standards from Biorad were used as size markers.

DNA extracted from the YAC clones was used for amplification with known STSs in the critical region identified by translocation breakpoint mapping as described above. The amplification was carried out in 10 μl total volume, with 20 ng of DNAs, 1 μl of 10X Boehringer PCR buffer, 2.5 μM of each primer, 2 mM of each dNTPs, and 0.25 units of Taq polymerase (Boehringer). The YAC contig was constructed based on the STS content. Additional STSs from the YACs were isolated by Alu-PCR method (Nelson, D. L. et al., Proc. Nat’l Acad. Sci. USA, 86:6686–6690) and tested on the somatic cell hybrid panel as well as with the other YACs to verify the contig integrity. STS49 amplification used primers F-AAGAATGAGCGTGAGATG (SEQ ID NO. 97) and R-AGCAGCTTCCACATTTCCCTG (SEQ ID NO. 98) at T_m=55°C, with a product size of 146 bp.

Cosmid Subcloning Preparations and Contig Construction

Cosmid sublibraries were made from three YACs covering the breakpoints in LR and MS: 928_e_5, 747_e_8 (1130 kb), 747_e_8 (1760 kb), and 896_d_3 (1500 kb). Ten milligrams of yeast DNA containing YAC(s) DNA were digested with 0.001 units of Sau3A in a total volume of 100 μl for 10 min. at 37°C to provide partially digested DNA. The digestion was checked on 0.6% agarose gel. Ultracentrifugation in a sucrose gradient was applied to the reaction, and fractions containing 30- to 500 kb fragments of DNA were selected for further experiments. Preparation of vector (SinI) DNA and genomic DNA, ligation, packaging, and preparation of plating cultures were carried out according to an instruction manual from Stratagene. Cosmid libraries from each YAC were plated on 150 mm LB with ampicillin (100 μg/ml). Master and replica lifts were made from each plate using Biodyne 0.45 micromembranes. The membranes were denatured and neutralized using manufacturer-specified conditions and UV cross-linked. Cosmid clones containing human DNA inserts were selected by colony hybridization using total human DNA labeled by random prime reaction (Boehringer) as a probe. Human clones were picked into 96-well microtiter plates for manipulation and storage.

From the YAC 928_e_5, 80 human clones were obtained from the total 2000 clones; from YAC 747_e_8, 1000 clones were selected out of 10,000; and, from YAC 896_d_3, 6,000 from a total of 7,000. Altogether that was equivalent to nearly twenty genomes. First cosmid sublibraries with known STS markers were screened. The original YACs were shown to be positive for D4S1571, D4S1564, D4S193, D4S406, D4S407, D4S1651, D4S448, D4S2350, GATA62A12, and GATA 65D06 by colony hybridization. From three to ten positive clones were identified for each STS. Then cosmid DNA from 370 cosmids, equivalent to nearly 3 genomes was extracted using Qiagen’s mini prep protocol. One hundred nanograms of DNA from each individual clone were digested with EcoRI and completion of digestion confirmed on 0.8% agarose gel containing ethidium bromide. Two nanograms of each sample were then electrophoresed in 0.8% agarose gel at 50V for 3 hours following Southern transfer to Hybond-n membranes. In order to identify overlapping clones, hybridization was carried out using whole cosmids DNA as a probe. Ten nanograms of whole cosmids DNA were labeled with 32P-dCTP using a random prime labeling kit from Boehringer, prehybridized for 1 h with human placental DNA (200 μg/ml) at 65°C, and then hybridized with the membranes in a hybridization buffer containing 7% SDS, 10% PEG-8000, 1 mM EDTA, and 200 mg/ml of human placental DNA, at a probe concentration of 10 cpm/ml. The membranes were prehybridized in the same buffer at 65°C for no less than 4 hours before hybridization (Murata, Y., Hum Mol. Genet. (1994) 3:1341–1344).

Cosmids were preliminarily grouped according to the digestion pattern with EcoRI. (Cosmids sharing three or more bands suggested contiguity). Three clones at the end of each contig were selected as probes for Southern hybridiza-
tion to confirm grouping and to detect clones that bridged contigs (Murai, Y. (1994) Hum. Mo. Genet. 3:1341–1344). Cosmid walks were initiated first from two groups of cosmids; the groups being positive for markers D4S2350, EGF, D4S193 and D4S406, as they were considered to be the closest to the breakpoints according to the physical map of the region, and the groups with present NotI restriction site, as was shown to be strongly associated with CpG islands surrounding the genes. Multiple contigs were assembled, and the end points of each contig were sequenced and mapped back on the somatic cell hybrid panel. The cosmid contig crossing both breakpoints was identified. Amplification primers for STS 24013 were F-GCTGTGCTCCCAAAAAATTGC (SEQ ID NO. 99) and R-CATTATAAGTGACTATAAGTG (SEQ ID NO. 100), yielding a PCR product of 148 bp; STS 24017, F-GAATTG-TTGTGTTTGGTGGTGGT (SEQ ID NO. 101) and R-CCCCACTTTTGTTAGTTAGTAC (SEQ ID NO. 102), with a 137-bp product; STS 2223, F-AAACCTCTGACCAACTGTGA (SEQ ID NO. 103) and R-TTGGCCATCAGCCGCAAACT (SEQ ID NO. 104); 160 bp: STS 2227, F-TCTATCTCTTCCTGGTAGTCA (SEQ ID NO. 105) and R-AATTGTGAGAATGTGAGGGTCAAAAG (SEQ ID NO. 106), 190 bp; STS 1871, F-CCAGTTTAAATCATGAC (SEQ ID NO. 107) and R-TTGGGATTTGGCAACATC (SEQ ID NO. 108), 120 bp; STS 1877, F-CTGTCAGCCATCTGACTCCAG (SEQ ID NO. 109) and R-GTCAGTCGTCTGCAAGG (SEQ ID NO. 110), 134 bp.

Sequencing

Cosmid DNA was extracted using QiaGen Midi prep and digested with NotI to linearize (Dugaszczyk, A. et al. (1992) Nucleic Acids Res. 20:6421–6422). Aliquots were checked on 0.8% agarose gel to assure complete digestion. Then the reaction was run through the Micronex column according to the manufacturer’s protocol or cleaned with phenol-chloroform and precipitated with ethanol. One to 1.5 μg of cosmid DNA were used for dye terminator sequencing with conditions according to the manufacturer’s (Applied Biosystems) protocol and loaded in Applied Biosystems Model 373A DNA sequencer. Sequences were analyzed using Blast and Grail software available on the WWW.

RESULTS

Mutation Search in the Epidermal Growth Factor Gene

SSCP analysis was performed on PCR amplification products generated from genomic DNA of seven unrelated patients with appropriate oligonucleotides spanning each of 24 exons of the epidermal growth factor (EGF) gene (Table 1). Four abnormal variants were identified as shown in Fig. 2. In all cases the variants were also identified in individuals unaffected with Rieger syndrome and in unrelated family members. The variants then underwent DNA sequence analysis; these sequence differences are shown in Fig. 3 and the specificities of the variants are outlined in Table 3. As can be seen, in three cases single base-pair changes resulted in alterations of the amino acid sequence in exons 14, 15 and 19. In exon 7, differences were also located within coding sequences but resulted in no change in the amino acid sequence.

A control group of 50 unrelated individuals selected from a combination of CEPH parents and other Caucasian individuals available from populations not segregating from Rieger syndrome were studied. The frequencies of the individual allelic variants were calculated from this material and are shown in Table 3. These allelic variants were also compared and haplotyped with previously described variants for the EGF gene as reported in the studies of Murray et al. (Murray, J. D. et al. (1992) Nature Genet. 2:46–49; and Murray, J. C. et al. (1986) Nucleic Acids Research 14:5117) and Ritty et al., (Ritty T. M. et al. and J. C. Murray (1989) Nucleic Acids Res. 17:5870). When haplotypes were combined, a level of heterozygosity of 0.95 in 100 unrelated Caucasians was obtained.

Construction of the High-Resolution Map of 4q25-26

Genetic maps of the region have been obtained through the Genethon (http://www.cceb.fr/ceph-genethon-map.html) and the CHLC (http://www.chlc.org/) servers on the World Wide Web. Identified markers were then tested on the somatic cell hybrid panel derived from the Rieger syndrome patients with translocations (Table 2). The panel consisted often cell lines: hybrids MS9-1, MS22-2, and MS19-2, containing the human translocation chromosome 4q26-ter:16pter-q22; hybrids LR11-2, LR10-1, and LR-19-2, containing the human translocation chromosome 4pter-q26:11q21-pter; the hybrids LR22-2 and LR23-1, containing the human translocation chromosome 4q26-ter:11pter-q21, without the normal human chromosome 4 and the reciprocal translocation chromosomes; and hybrids AR22-2 and AR24-1, carrying normal and deleted chromosomes 4. Chromosome 4 markers present in MS9-1, 22-2, and 19-2 must be located distally to both breakpoints (i.e., between the breakpoints and the end of the arm of chromosome 4), whereas markers missing from the MS9 hybrid cell lines and present in LR11-2, 10-1, and 19-2 must be from the LR22-2 and LR23-1 cell lines (or vice versa) must be located in between breakpoints. From this analysis it was deduced that the EGF-D4S1571 markers were the closest proximal markers to the breakpoint, and the microsatellite markers D4S1651-D4S406-D4S193 were then closest flanking distal markers (Fig. 3). The genetic distance between D4S1571 and D4S193 was identified as 3 cm.

Localization of the Translocation Breakpoints

The markers flanking the translocation breakpoints were used to screen the CEPH YAC libraries. (The YACs for the Genethon markers were identified through the CEPH-Genethon information service on the WWW). The YAC contig was constructed based on STS content (Fig. 3). Probes from the ends of the YACs were isolated and tested on the somatic cell hybrid panel as well as with the other YACs to verify the contig integrity. The YACs 928..3..5, 747..3..8, and 896..d..3 contained STSs that were localized proximally as well as distally form both breakpoints; thus, all these YACs crossed both breakpoints, but the STS content of YAC 928..e..5 was different from the STS content of YACs 747..3..8 and 896..d..3. Cosmid sublibraries were made from all three YACs to assure complete representation of the region of interest.

Cosmids positive for the STSs present in the original YACs were identified and used as starting points for the cosmids walks. Multiple contigs were assembled. End points of each contig were sequenced and mapped back on to the somatic cell hybrid panel. A large contig, spanning approximately 500 kb and consisting of about 70 cosmids was identified as encompassing both breakpoints. This region was narrowed down to a cosmid consisting of two cosmids: 240 (overlapping LR’s breakpoint) and 222 (overlapping MS’s breakpoint; Fig. 3). The distance between breakpoints is approximately 50 kb, as determined by mapping of STSs derived from the cosmids ends.
Cloning and Characterization of RIEG

Clinical Status of the Rieger Syndrome Pedigrees

Seventeen families were identified with characteristic RGS. In seven of them the patient also had cytogenetic abnormalities. From the remaining ten families at least one of the individuals in the pedigree manifested all normal involution of the periurnbival skin (Table 3).

<table>
<thead>
<tr>
<th>Family no.</th>
<th>Diagnosed by</th>
<th>Features within family</th>
<th>Mutation (region and nature)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>JCM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>WLA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>KWS</td>
<td>homeobox</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>RSB</td>
<td>homeobox</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DBN</td>
<td>iatros</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PB</td>
<td>homeobox</td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>BUZ</td>
<td>translocation</td>
<td>occurred at ~5-15 kb from 5' region of gene</td>
</tr>
<tr>
<td>LR</td>
<td>JCC</td>
<td>translocation</td>
<td>occurred at ~55-65 kb from 5' region of gene</td>
</tr>
</tbody>
</table>

Identification of CpG Islands

CpG islands were identified by restriction analysis of the cosmids from the contig spanning the cytogenetic breakpoints with EcoRI, HindIII, PstI, and the rare-cutter enzymes NotI, HPalI, Smal, Eall, and Sacl. Significant overrepresentation of CG cut sites in 203 kb DNA fragment relative to an average distribution (Lindsay, S. and Bird, A. P. (1987) Nature 327:336-338) indicated likely CpG-island presence. Such fragments were subcloned into pBlueScript plasmids and sequenced in both directions to ascertain CG-rich content. Sequences were analyzed using Blastn and Grail search engines available on the World-Wide Web.

cDNA Library Screening

Several cDNA libraries were used in this study: human fetal brain (Stratagene), human craniofacial constructed from mRNA derived from the craniofacial region of human embryos ranging from 42 to 53 days' gestation (Padanilam, B. J., Studier H. S., Mills, K. A., McLeod, J. B., Sol turin, M., Lee, B., Ramirez, F., Buetow, K. H., and Murray, J. C., 1992. Characterization of the human HOX7 cDNA and identification of polymorphic markers. Hum. Mol. Genet. 1, 407-410.), mouse embryonic carcinoma (Stratagene), and mouse fetal 15-day (Novagen). The 3.5-kb 319not1.896 and the 2-kb Secor1.747 GC-rich fragments were random-prime labeled with P-according to instructions from the kit's manufacturer (Boehringer) and were used to screen the craniofacial cDNA library in hybridization. The hybridization was carried out at 55°C, over 18 hours and was followed by several washes at room temperature and by a 2-hour wash at 60°C in 0.1% SSC and 0.5% SDS. cDNA fragments isolated in this initial screening were then used in subsequent screenings of the same (human craniofacial) and of the other human and mouse cDNA libraries to identify homologous and overlapping clones.

Clone Isolation and Sequencing

Plasmids from positive plaques were isolated from their L-Zap II or Uni-Zap XR hosts by in-vivo excision with R408 helper phage as described by the manufacturer (Stratagene). Bluescript plasmids containing cDNA inserts were sequenced and analyzed using the Blastn and Grail engines.

Genomic Structure and Identification

Exon-intron boundaries were identified by primer walking between the cDNA and genomic cosmids clones. Introns were sized by PCR with exon-specific primers and cosmid-end derived primers (FIG. 3), the cosmids ends had been previously mapped by restriction and hybridization, with a total of 24 overlapping cosmids identified in the region (Semina, E. et al., (1996) Am. J. Hum. Genet.).

SSCP and DNA Analysis

Oligonucleotide primers for the PCR amplification of the various RIEG fragments (See FIG. 3) were as follows: primer 1 forward GATAAAGGCCACGAGGGGAA (SEQ ID NO. 111) and reverse GAGGGAACTGTAATCTCGCA (SEQ ID NO.112); primer 2 forward GTAATCCTGCAT-GTTCGAC (SEQ ID NO.113) and reverse CAAGTGTTTTGTTGAGGAA (SEQ ID NO.114); primer 3 forward AGTTCATGCGCTACAGTGCAG (SEQ ID NO.115) and reverse GGAGAAACAGCTCTTGTTAG (SEQ ID NO.116). Genomic DNA from the 10 unrelated probands diagnosed with Rieger syndrome, was amplified in a PCR thermocycler (Perkin-Elmer Cetus), with a single four-minute, 94°C cycle followed by 30 cycles comprising 94°C, 55°C, and 72°C steps of 30 seconds each, in 10 μl total volume: 1 μl Boehringer PCR buffer, 2.5 pmol of each primer, 2 mM in each dNTP, and 0.25 units of Taq polymerase (Boehringer). PCR products were heated for four minutes at 95°C and electrophoresed for 4 hours at 20 W through a fan-cooled gel composed of 3.3 ml of 10 TBE, 1.37 ml of glycerol, 13.75 ml of MDE mix (from FMC), 36.6 ml of water, 220 μl of 10% APS, and 22 μl of TEMED. After silver staining the gels were visually inspected for potential abnormal variants.

Potential variant fragments were then extracted from the gel, reamplified, sequenced in both directions together with genomic PCR products from the corresponding individual, and compared with the normal sequence. Genomic DNA from family members was analyzed via SSCP to verify cosegregation of the relevant fragment with the phenotype.

In-Situ Hybridization

NIH Swiss mice from Harlan (Indianapolis, Ind.) were used in this study. Developmental staging was done according to Theiler, K. (1989) The Mouse House. Atlas of Embryonic development.

35S-uridine 5’-triphosphate (NEN) and digoxigenin (DIG) 5’-triphosphate (Boehringer) were used as labeling materials. The 535 bp DNA fragment (SL I-Degl) of 3’UTR region of the mouse rgs cDNA clone was cloned into pBlueScript vector and used as a probe for in-situ hybridization. Sense and antisense RNA probes were synthesized with labeling material using either T3 or T7 RNA polymerase (Stratagene) linearized DNA. Labeled RNA probes were purified with Bio-spin chromatography columns #30 (Bio-Rad). Probes were hydrolysed before use in hybridization. For 35S-labeled probes 3.5 x 106 cpm/kb per 50 μl of hybridization mixture was used, and 300-500 ng of probe per ml of hybridization mixture was used for digoxigenin-labeled probes.

Embryonic mice for radioactive in-situ hybridization were fixed overnight at 4°C in 4% paraformaldehyde in PBS, dehydrated, cleared in Histoclear (National Diagnostics), and embed in Paraplast Plus (Oxford). Seven-micron sec-
tions were cut and mounted onto Superfrost plus slides (Fisher) with DEPC-treated water, dried overnight at 40°C, and then stored at room temperature. Before use the slides were baked at 60°C overnight and then processed for in-situ hybridization according to the methods described by Sassone, D. and Rosenthal, N. (1993) *Meth. Enzymol.* 225:384–404. Slides were hybridized overnight at 50°C, washed in 5×SSC at 50°C to remove coverslips, and then washed in 2×SSC, 50% formamide at 60°C. Autoradiography was done with Kodak NTB-2 emulsion. A two-week exposure time at 4°C was determined to be optimal. Developed slides were stained for 30 minutes in hematoxylin, dehydrated, cleared in Histosol, and mounted with Permount (Fisher).

Embryonic mice for whole-mount in-situ hybridization were fixed and processed following a modification of the procedures described by Harland and Richard M. H. (1991) *Meth. Cell Bio.* 36:685–695. Briefly, embryos were fixed overnight at 4°C in 4% paraformaldehyde in M buffer (100 mM MOPS, pH 7.4, 2 mM EGTA, 1 mM MgSO₄) and then bleached in 3% hydrogen peroxide in M buffer, dehydrated in a graded series of M buffer and methanol, and stored at −20°C. Hybridization was performed at 60°C overnight with rocking. Post-hybridization washes were done at 60°C in 2×SSC.

In preparation for antibody incubation the embryos were washed in 10% heat-inactivated sheep serum in PBS+0.5% Tween 20 (PBSTw). A 1:2000 dilution of a-digoxigenin antibody (Boehringer) was added, and the embryos were incubated overnight at 4°C with gentle rocking. Embryos were rinsed for 24 hours in PBSTw with at least 10 changes of buffer and then washed in color reaction buffer (100 mM Tris, pH 9.0, 50 mM MgCl₂, 100 mM NaCl, 0.5% Tween 20, 1 mM levamisole) 3× for 20 minutes each. Color reagents (4.5 μl/ml NBT and 3.5 μl/ml BCIP) were then added. The reaction was monitored under a dissecting scope and stopped with 2 washes of PBS. The embryos were again fixed in Bouin’s buffer to help preserve the color and dehydrated in a graded series of PBS/glycerol and stored in 80% glycerol: 20% PBS, 0.2% sodium azide.

**Results**

**Cloning of the Gene**

Example 1 reports on fine physical mapping of translocation breakpoints on 4q25 from two patients with RGS. By restriction analysis of the cosmids crossing both breakpoints with CG-cutting enzymes six potentially C+G-rich regions ranging from one to four kilobases were identified, subcloned into plasmids and sequenced in both directions. Five fragments were shown to be C+G-rich by sequence analysis.


Sequence analysis of the 5′ end of the cDNA revealed high G+C content (66% for bases 1 through 222) and five restriction sites for Hpal, one for Ssh1, and one for BstUI—commonly associated with the CpG islands found to surround the transcription sites of vertebrate genes (Lindsay, S. and Bird, A. P. (1987) *Nature* 327:356–338). An ATG initiation codon was found at position 226, followed by a G at 229 (−4 from the beginning of the ATG codon) and preceded by a G at 223 (−3 from the ATG), in agreement with Kozak’s rules (Kozak, M. (1987) *Nucleic Acids Res.* 15:115–115), and represents the only in-frame ATG codon in the 5′ region. Two potential CTG codons were identified upstream of the 5′ region which might play a role in the regulation of tissue-specific transcription of the gene (Kozak, M. (1987) *J. Mol. Biol.* 196:947–950). The 3′ region of the cDNA was found to contain a polyadenylation signal at positions 1393–1399 and 1458–1463 with poly-A tail of 25 nucleotides. Translocation of the open reading frame yielded a protein of 271 amino acids.

Sequence comparison using GenBank indicated that it is a novel gene that contains a homeobox, and the gene and encoded protein were named RIEG. The protein alone is also referred to as “Solushrin” (Syndrome of orofacial and umbilical abnormalities of the Rieger syndrome homeobox). Comparison of the RIEG homeodomain with others identified the greatest homology with the recently published murine Ptx-1 and P-OTX proteins which themselves appear to be identical (Lamonerie T., et al. (1996) *Genes Dev.* 10:1284–1292; Szeto, D. P. et al. (1996) *Proc. Natl. Acad. Sci.* 93:7706–7710). The RIEG homeodomain sequence is different from Ptx-1 or P-OTX by just two residues: RIEG has Thr at position 3 of the second helix and an Ala residue at the position 2 of the third helix instead of the Met and Pro residues, respectively, at the corresponding positions in the Ptx-1/P-OTX homeodomain. Both homeodomains, however, share the Lys at position 9 of the predicted third helix (position 50 of the homeodomain) which is characteristic for the bicoid-related proteins as found in *C. elegans* unc-30 gene (Jin Y. et al., (1994) *Nature* 372, 780–783), the murine OtX-1 and OtX-2, and with the *Drosophila* Otd homeodomains (Simeone, A. et al., *Enbo. J.* 12:2735–2747) (FIG. 2).

The amino acid at position 9 of the recognition helix has been shown to be essential in some other homeodomains to determine the specificity of binding to the two bases following the TAAAT core (Hanes, S. D. and R. Brent (1989) *Cell* 57:1275–1283; Hanes, S. D. et al., (1994) *Mol. Cell. Biol.* 14:356–375). Other classes of homeodomain proteins show identity of amino acid at position 9 in the majority of their members: the ant-class proteins usually have a Gln, the paired proteins a Ser, proteins such as cut products usually contain a His, and POU-class proteins a Cys (Hanes, S. D. and R. Brent (1989)) Cell 14:1275–1283).

less (Schneitz, K. et al., (1993) Genes Dev. 7:114–129), rat
13:81–101) proteins. Prx-1/Prtx-1, prx-2, and prx-2, Cart-1,
arietofine, cxh10 and otp mRNAs are expressed at high
levels in the craniofacial region in different species; three of
these genes (prx-1 and prx-2, Cart-1) are also expressed in
the limb mesenchyme, which correlates with the expression
pattern of murine solorushin (see below). This 14-a.a
sequence might represent an element important for target
specificity of those transcription factors. The fact that pro-
tins sharing this conserved 14-a.a sequence also have correlations in their expression patterns suggest that this
may be important for region-specific function of those
proteins.

The gene was found to comprise about 18 kb in the human
and to consist of four exons of 572, 57, 206, and 1290
nucleotides in length. The initiation codon is located in the
second exon with the homeobox region in the third and
fourth exons. The genomic structure of the gene is illustrated
in FIG. 3. The approximate distances of the most 5’ end of
RIEG to the translocation breakpoints were approximated by
analysis of genomic clones containing both breakpoints in
one contig (FIG. 3) and were estimated at 5 to 15 kb for MS
and 55 to 65 kb for LR, suggesting position effect as the
cause of RGS in those patients. “Position effect” mutations
(usually translocations, inversions, deletions, duplications or
retroviral integrations) alter gene expression through a long-
range effect on chromatin structure (position effect variega-
tion) or by disrupting distal elements of a gene and have
been described in Droshilla, mouse, and human (Bedell, M.

Mutation Analysis

Ten families with dominant Rieger syndrome have been
identified (Table 3). In those families at least one of the
affected individuals in the pedigree manifested all three
cardinal features of the syndrome: abnormalities of the
anterior segment of the eye, hypodontia, and failure of normal
involution of the periumbilical skin (FIG. 4). Parts of the
exon and adjacent intron sequences were subsequently
amplified from the DNA of one affected individual from
each of ten Rieger syndrome families and screened for
mutations using single-strand conformational polymor-
phism (SSCP) analysis. The mutation screen was initiated at
the homeobox region of the gene using three overlapping
sets of primers (FIG. 3) and then expanded to cover the full
coding sequence.

Amplification with primer set 1 revealed a variant band in
each of three probands (FIGS. 5a, b, c); two familial cases
which showed cosegregation of SSCP variants and pheno-
type, and a third case that was a de novo Rieger phenotype
mutation. In family 1 sequencing identified a missense
mutation T394A, that changes the CTG codon for Leu into
cAG codon for Gln in the first helix of the homeodomain.
In family 2 a G-to-C point mutation was found at position +5
of the 5’ splice site of the third intron separating parts of the
homeobox element. This position is occupied by G in 82%
of 5’ splice sites, which indicates its importance in the
mechanism of splicing (Shapiro, M. B. (1987) Nucleic Acid
Res. 15:7155–7174). In family 3 the band shift was not
obvious, but the band structure consistently looked different
from the other bands, and sequencing revealed a point
mutation A435C, changing an ACA codon for a Thr residue
to a CCA coding for Pro in the second helix of the home-
domain. Amplification with primer set 2 revealed bands
with altered mobility in two additional families (FIGS.
5d,e). The mutation in family 4 changes A to G at position
–11 in the third intron, which creates an additional acceptor
site, AG, 5’ from the current one. The nearest 3’ acceptor site
usually becomes the major site used by splicing machinery
(Reed, R. (1989) Genes Dev. 3:2113–2123), so that, in this
case, an insertion of three additional amino acids and a
frame shift would occur in the third helix of the home-
domain. In family 5, where the proband has apparent de
novo Rieger syndrome, sequencing identified a missense
mutation G505C changing a CCG codon for Arg to a CCG
for Pro in the third helix of the homeobox. Amplification
with primer set 3 revealed an abnormal band in family 6
(FIG. 2a) that cosegregated with the disease phenotype.
Sequencing of DNA from affected individuals in this family
identified a point mutation G632A that results in a TGA stop
codon in place of a TGG for a Thr residue, resulting in
premature termination of the protein 34-a.a C-terminal region of the
homeodomain (FIG. 1).

All the SSCP variants for mutations were found to cose-
gregate with the disease phenotype in families with more than
one affected (FIG. 3), and the SSCP variants were not
detected in any of the chromosomes from ~200 control
individuals (~400 chromosomes) with the same ethnic back-
ground as the corresponding Rieger syndrome families
(Caucasian for families 1, 2, 3, 4 and 6 and Filipino for
family 5). No SSCP variants were detected in the remaining
four classic RIEG families with primer pairs spanning the
whole RIEG cDNA sequence.

A mouse embryonic carcinoma cDNA library (Strat-
agene) was screened with the PCR product containing the
human RIEG gene sequence form nt 880–1120. Positive
clones were identified from secondary screens and
sequenced. The longest clone was found to represent partial
cDNA sequences homologous to the human clone starting in
exon 3 (FIG. 1). The cDNA sequences of the human and
gene sequences shared 91% of their nucleotide sequence
through the coding region and averaged 85% nucleotide
sequence identity in the 3’ UTR, with the highest homology
~97% seen over 270 nucleotides of (A+T) rich-region of the
3’ UTR. Homology at the protein level was found to be
99.2% with 100% homology through the homeodomain
region (FIG. 1). Considering the strong nucleotide and
amino acid homolog of the isolated mouse cDNA and the
human RIEG cDNA this sequence was designated as a
partial cDNA sequence of the murine homolog of the human
RIEG gene, RIEG.

A 355-bp fragment from the mouse RIEG cDNA spanning
nt 740–1182 and containing mostly the 3’ UTR region (the
stop codon is located at position 782) was subcloned and
used for preparing sense and anti-sense riboprobes for
in-situ hybridization experiments. Whole-mount in-situ
hybridization in day-11 mouse embryos revealed signal
around the eye, the maxillary and mandibular epithelia, at
the base of the limb, and in the umbilical cord (FIG. 4).
Hybridization on sections through the head and the midgut
region of the mouse day-11 embryo detected strong signal in
the mesenchyme around the eye, in Rathke’s pouch, in the
dental lamina, the limb mesenchyme, in the dorsal mesen-
tery, and the vitelline umbilical vessels (FIG. 5). Expression
seen in the eye mesenchyme, dental lamina and umbilical
cord is consistent with the role of RIEG gene in the
pathogenesis or Rieger syndrome. Expression in the limb
mesenchyme and Rathke’s pouch is stron in day-11 mouse
embryos and suggests this gene may be playing in some
hand an CNS abnormalities reported in association with
RIEG.

EQUIVALENTS

Those skilled in the art will recognize, or be able to
ascertain using no more than routine experimentation, many
equivalents of the specific embodiments of the invention
described herein. Such equivalents are intended to be
encumbered by the following claims.

<table>
<thead>
<tr>
<th>Sequence Listing</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) GENERAL INFORMATION:</td>
</tr>
<tr>
<td>(iii) NUMBER OF SEQUENCES: 139</td>
</tr>
<tr>
<td>(2) INFORMATION FOR SEQ ID NO: 1:</td>
</tr>
<tr>
<td>(i) SEQUENCE CHARACTERISTICS:</td>
</tr>
<tr>
<td>(A) LENGTH: 1775 base pairs</td>
</tr>
<tr>
<td>(B) TYPE: nucleic acid</td>
</tr>
<tr>
<td>(C) STRANDEDNESS: single</td>
</tr>
<tr>
<td>(D) TOPOLOGY: linear</td>
</tr>
<tr>
<td>(ii) MOLECULE TYPE: cDNA</td>
</tr>
<tr>
<td>(ix) FEATURE:</td>
</tr>
<tr>
<td>(A) NAME/KEY: CDS</td>
</tr>
<tr>
<td>(B) LOCATION: 234..1047</td>
</tr>
<tr>
<td>(x) SEQUENCE DESCRIPTION: SEQ ID NO: 1:</td>
</tr>
<tr>
<td>CGAGAAAGCT GGAAGAGGAC AGAAGAGAAG TGGCACTGSGC GGGTAGAATTT CGAGGAGCCC</td>
</tr>
<tr>
<td>ACGGAGCCCG TGGACTCTTT CGGAACCTTGCGACCTCGAGT CTCCTGAGG</td>
</tr>
<tr>
<td>CCGGCTTCTC CGGAGCCTGCG CGGAGCTCGGC TGCTGGAAA TGGCCCGGG</td>
</tr>
<tr>
<td>ACGGAGCTGG ACGGAGAAG ACAGCCTCTC TCCCGTACGC CGAATACGG GAA ATG</td>
</tr>
<tr>
<td>Met</td>
</tr>
<tr>
<td>GAG ACC TGC CGC AAA CTG GTG TCG CGG CCG TUT CTG CAA TTA GAG AAA</td>
</tr>
<tr>
<td>Glu Thr Asn Cys Arg Lys Leu Val Ser Ala Cys Leu Gln Leu Glu Lys</td>
</tr>
<tr>
<td>5 10 15</td>
</tr>
<tr>
<td>GAT AAA AGC CAG CGG AGG AAG GAT GAC GTG GGC GCC GAG GAC CGG</td>
</tr>
<tr>
<td>Asp Lys Ser Gln Gln Gly Lys Asn Gln Asp Val Gly Ala Gln Asp Pro</td>
</tr>
<tr>
<td>20 25 30</td>
</tr>
<tr>
<td>TCT AAG AAG CGG CAA AGG CGG ACT CAG TTT ACC AGC CAG</td>
</tr>
<tr>
<td>Ser Lys Lys Arg Gln Arg Arg Gln Arg Thr His Phe Thr Ser Gln</td>
</tr>
<tr>
<td>35 40 45</td>
</tr>
<tr>
<td>CAG CTC CAG CGT GAG GCC ACT TTC CAG AGG AAC CGG TAC CGG GAC</td>
</tr>
<tr>
<td>Gln Leu Gln Leu Ala Thr Ala Phe Arg Asn Arg Tyr Pro Asp</td>
</tr>
<tr>
<td>50 55 60 65</td>
</tr>
<tr>
<td>AGT TCC ACA CGG GAA ATG GCT GTG TGG ACC AAG CTT AGG GAA GGC</td>
</tr>
<tr>
<td>Met Ser Thr Arg Gln Glu Ile Ala Val Trp Thr Asn Leu Thr Glu Ala</td>
</tr>
<tr>
<td>70 75 80</td>
</tr>
<tr>
<td>CGA GTC CGG GTT TGC TGC AAG ATG GCT CGG CCC AAA TGG AGA AAG AGG</td>
</tr>
<tr>
<td>Arg Val Arg Val Trp Phe Lys Asn Arg Asp Ala Lys Trp Arg Lys Arg</td>
</tr>
<tr>
<td>95 100 105</td>
</tr>
<tr>
<td>GAG GCC AAC CAG CGC GAG CTA TGC AAG AAT GCC TTC GGG CCC GAG</td>
</tr>
<tr>
<td>Glu Arg Asn Gln Gln Ala Leu Cys Lys Asn Gly Phe Gly Pro Gln</td>
</tr>
<tr>
<td>120</td>
</tr>
<tr>
<td>TTC AAT GGG CTC ATG CAG CCC TAC GAC ATG TAC CCA GCC TAT TCC</td>
</tr>
<tr>
<td>Phe Asn Gly Leu Met Gln Pro Tyr Asp Met Tyr Pro Gly Tyr Ser</td>
</tr>
<tr>
<td>115 120 125</td>
</tr>
</tbody>
</table>
2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Glu Thr Asn Cys Arg Lys Leu Val Ser Ala Cys Leu Gln Leu Glu
1  5  10  15

Lys Asp Lys Ser Glu Gln Gln Gly Lys Asn Gly Asp Val Gly Ala Glu
20  25  30
Pro Ser Lys Lys Arg Gln Arg Arg Gln Arg Thr His Phe Thr Ser
35  40
Gln Gln Leu Gln Gln Leu Glu Ala Thr Phe Gln Arg Asn Arg Tyr Pro
50  55  60
Asp Met Ser Thr Arg Glu Ile Ala Val Trp Thr Asn Leu Thr Glu
65  70  75  80
Ala Arg Val Arg Val Trp Phe Lys Asn Arg Asp Ala Lys Trp Arg Lys
85  90  95
Arg Glu Arg Asn Gln Gln Ala Glu Leu Cys Lys Asn Gly Phe Gly Pro
100 105 110
Gln Phe Asn Gly Leu Met Gln Pro Tyr Asp Asp Met Tyr Pro Gly Tyr
115 120 125
Ser Tyr Asn Asn Trp Ala Ala Lys Gly Leu Thr Ser Ala Ser Leu Ser
130 135 140
Thr Lys Ser Phe Pro Phe Phe Asn Ser Met Asn Val Asn Pro Leu Ser
145 150 155 160
Ser Glu Ser Met Phe Ser Pro Pro Asn Ser Ile Ser Ser Met Ser Met
165 170 175
Ser Ser Ser Met Val Pro Ser Ala Val Gly Leu Pro Gly Ser Ser
180 185 190
Leu Asn Ser Leu Asn Leu Asn Leu Ser Ser Pro Ser Leu Asn
195 200 205
Ser Ala Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Pro Thr Pro
210 215 220
Tyr Val Tyr Arg Asp Thr Cys Asn Ser Ser Leu Ala Ser Leu Arg Leu
225 230 235 240
Lys Ala Lys Glu His Ser Ser Ser Gly Tyr Ala Ser Val Gln Lys Pro
245 250 255
Ala Ser Asn Leu Ser Ala Cys Gln Tyr Ala Val Asp Arg Pro Val
260 265 270

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 816 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..814

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATG GAG ACC AAC TGC GCG AAA CTG GTG TCG GCG TGT CTG CAA TTA GAG
Met Glu Thr Aen Cys Arg Lys Leu Val Ser Ala Cys Leu Gln Leu Glu
1  5  10  15
AAA GAT AAA ACC CAG CAG GGG ARG AAT GAG GAC GTG GGC GCC GAG GAC
Lys Asp Lys Ser Gln Gln Gly Lys Aen Glu Aep Val Gly Ala Gln Asp
20  25
CCG TCT AAG AAG CGG CAA AGG CGG CAG ACT CAC TTT ACC AGC
Pro Ser Lys Lys Arg Arg Glu Arg Arg Gln Arg Arg Thr His Phe Thr Ser
35  40  45
CAG CAC CTC CAG CAG CTG GAG GCC ACT TTC CAG AGG AAC GCG TAC CCG
Gln Gln Leu Gln Gln Leu Glu Ala Thr Phe Gln Arg Aen Arg Tyr Pro
50  55  60
GAC ATG TCC ACA CGC GAA GAA ATC GCT TGG TGG ACC AAC CTT AGC GAA
Asp Met Ser Thr Arg Glu Glu Ile Ala Val Trp Thr Aen Leu Thr Glu
65 70 75 80
GCC CCA GTC CGG GTG TGG TCC AAC AAT CCT CGG GCC AAA TGG AGA
Ala Arg Val Arg Val Trp Phe Aan Arg Arg Ala Lys Trp Arg Lys
85 90 95
AGG GAG CGC AAC CAG CAC GCC GAG CTA TGC AAG AAT GCC TTC GSG CCG
Arg Glu Arg Aen Gln Glu Ala Glu Leu Cys Lys Aen Gln Gly Phe Gly Pro
100 105 110
CAG TTC AAG GCG TCT ATG CAG CCC TAC GAC GAC AGT TCC CTA GCC
Gln Phe Aen Gly Leu Met Gln Pro Tyr Asp Met Tyr Pro Gly Tyr
115 120 125
TCC TAC AAC AGC AAC TGC GCC GCA AAG GCC CTT ACA TCC GCC CTA TCC
Ser Tyr Aen Aen Trp Ala Aen Tyr Thr Glu Thr Ser Ala Ser Leu Ser
130 135 140
ACC AAG AGC TTC CCC TTC TCC AAT CAA GAC CCC TCT TGG AAG
Thr Lys Ser Phe Pro Phe Aen Ser Met Aen Val Aen Pro Leu Ser
145 150 155 160
TCA CAG AGC AGT TTT TCC CCA CCC TAC TGT AAC TCC AAG ATG
Ser Glu Ser Met Phe Ser Pro Aen Ser Ile Ser Ser Met Ser Met
165 170 175
TGC TCC AAC AGT GTG CCC TCA GCA GTC ACA GGC GTG GCC GCC TCC AGT
Ser Ser Met Val Pro Ser Ala Val Thr Gly Val Pro Gly Ser Ser
180 185 190
CTC AAC AGC CTT AAT AAC TGT ACG AAC CTG AGT AGC CCG TCG CTA ATG
Leu Aen Ser Leu Aen Leu Leu Ser Pro Ser Leu Aen
195 200 205
TCC CGG GTG CGG ACG CCT GCC TCT CCT TAC GCG CGG ACT CCT CGG
Ser Ala Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Pro Pro
210 215 220
TAT GGT TAT AGG GAC AGC TGT AAC TCG CAC GGC AAC CTC AGA CTG
Tyr Val Tyr Arg Aep Thr Cys Aen Ser Ser Leu Ala Ser Leu Arg Leu
225 230 235 240
AAA GCA AAG CAG CAC TCC AGC TTC GGC TAC GCC AGC GCT GAG
Lys Ala Lys Glu His Ser Ser Phe Gly Tyr Ala Ser Val Gln Lys Pro
245 250 255
GCC TCC AAC CTG AGT GCT TGC CAG TAT GCA GTG GAC CGG CCC GTG T
Ala Ser Aen Leu Ser Ala Cys Gly Thr Ala Val Asp Arg Pro Val
260 265 270

GA
816

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1775 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 234..1047

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CGAGAAACTG CGAGAAGACC AAGAAGAAC TCACAGTGCC GGCAGATTTG CGGAGGCGCCC
60
AGTGCCAGGC TGAGACTCTT CGAACTCTGG GACCTCAGG ACGCTGTGCA ATCTGTCAGG
120
CCGCGCTTGC GCGCGTCTCC CGTGCACCGG GAGGCTCGGC TGGCTGAAA TGGCCCCGGG
180
AAACGACTGGG ACAGGAGAC AGACGGCTTC TCCGGTGAC CGTACAGGG GAA ATG
236
---continued---

GAG ACC AAC TCC CCC AAA CTG GTC TCG GCC TGT TCC CAA TTA GAG AAA 284
Glu Thr Aen Cys Arg Lys Leu Val Ser Ala Cys Leu Glu Leu Lys
5 10 15

GAT AAG GGC CGC CAA GGA GAG AAT GAG GAT GTC GGC GGC GAC CCG 332
Asp Lys Gly Gln Gly Lys Lys Gln Glu Asp Val Gly Ala Glu Asp Pro
20 25 30

TCC AAG AAG AGG CGG CCA CCC CCG CAG AGG ACT CAT TAC ACT ACC CAG 380
Ser Lys Lys Lys Arg Gln Arg Arg Glu His Phe Thr Ser Glu
35 40 45

CAG CTG CAG CTG GAA GCC ACT TCC CAG AGA ACC CAC TAC CCA GAC 428
Gln Leu Gln Leu Glu Ala Thr Phe Glu Arg Arg Tyr Pro Asp
50 55 60 65

ATG TCC ACT CGC GAA GAA ATC GCT GTC TGG ACC AAG ACT CGG GAA GCC 476
Met Ser Thr Arg Glu Ile Ala Val Thr Ser Ala Leu Thr Glu Ala
70 75 80

CGA GTC CGG GTT TCG TTC AAG AAT CGG AGG AAC TCA TGG AGA AAG CCG 524
Arg Val Arg Val Trp Phe Lys Ala Val Arg Arg Ala Lys Trp Arg Lys Arg
90 95

GAA GCC ACC CAG CCC GCC GCC ACC GCC GCC TGG AAG GCC GTG TTC AAG 572
Glu Arg Aen Glu Glu Glu Ala Glu Leu Glu Gly Phe Gly Pro Gln
100 105 110

TTC ACC GGG CTC ATG CAG CCC TAC GAC GAC ATG TAC CCC GGC TAC TCG 620
Phe Aen Gly Leu Met Gln Ala Pro Tyr Asp Asp Met Tyr Pro Gly Tyr Ser
115 120 125

TAC AAC ACC TGG CCT GCC AAG GCC CTC ACC ACC TCA GCC TCT GCT TCC ACC 668
Tyr Aen Aen Trp Ala Ala Lys Gln Leu Thr Ser Ala Ser Leu Ser Thr
130 135 140 145

AAG ACC TTC CCC TTC TTC AAC TCC ATG ACC GTC AAT CCC TCT TCG TCT 716
Lys Ser Phe Phe Phe Ser Met Aen Val Aen Pro Leu Ser Ser
150 155 160

CAG AGT AGT TTT TCC CGG CCC ACC TCC ATC TCA TCT ATG AGT TCG 764
Gln Ser Met Phe Ser Pro Aen Ser Ser Leu Val Met Ser Ser Leu Leu
165 170 175

TCC AGG ACG GTG CCC TCC GCC GCC ACC GCC GCC CCC GCC ACC CTC 812
Ser Met Val Pro Ser Ala Val Val Gly Pro Gly Ser Ser Leu
180 185 190

AAT ACC AGG AAT AAC TTC AAC ACC AGG TGC CTG CAA ATT TCC 860
Aen Ser Aen Aen Aen Leu Aen Leu Leu Ser Pro Ser Aen Ser Leu
195 200 205

GCG GTG GCC ACC CCC GGC TCT CAT GCG CGG ACC ACT CCT CGG AAC 908
Ala Val Pro Pro Ala Cys Pro Tyr Tyr Pro Tyr Pro Tyr Pro Tyr
210 215 220 225

GTT TAT AGG GCC AAG TCT ACC AAC CGG AGG ACT CAT AAG AAC 956
Val Tyr Arg Asp Thr Cys Aen Ser Ser Ala Leu Ala Ser Leu Arg Lys
230 235 240

GCA AAG GAC TCC ACC TCC GCC GCC AGC GTC CAG ACG CGC GCC 1004
Ala Lys Glu His Ser Phe Gly Tyr Ala Ser Val Glu Aen Pro Ala
245 250 255

TCC AAC CAG AGT GCT TCC GAG TAT GCA GTG GAC CGG CCG GGT T 1047
Ser Aen Leu Ser Ala Cys Gly Thr Ala Val Asp Pro Pro Val
260 265 270

GAGCGGACCC CACAGGGCGG GGGCTCAGGG ACCTGGCGGG ACGGCGAACC TCCGCCTCGG 1107

AAGACTGGG AATTATCTAT GAAGTGCCTG GCCATTAAG AAGGGAGAAG AAGAGAAACC 1167

TATGAGA AAGGAAACC ACTGAATCAA AGAGAGAGCT CTTTGATTT CAAAGGGGATG 1227

TCTGAGCAGT CGCCAGCTT TCAGGCTAA GATGCAAGT ACACATACC 1287

AAGAANATG TTTGACCTGG TATGACAAT TATCATGAC ATPAGCGTAT TATTTTTAA 1347
GTGAGCATT GTGACATTAA AAAGACTGA AAGGATGAT ATATATGCAA ATGTCAATT
AAATTAATTA AACAGGTTG TAGTAAATATCA ACAAACGTTG TTTTAAGGG TAGCTTTTAA
AAATTAACGAT GTTACAGCA AGGATTAGG ATTTTCCGCT TOGG7CAAAG GAGCTTTATA
TACTAAATGC CACACTGTAT GTTCTCAACA TATTATTATT ATTATATAAAA ATGTGTAAT
ATCAGTTTA GATAGCTTG TGGGCTGAT GGAAGATGT TCTCGAAACT GCTAGTACA
ACCTACCCCTG TGATTAACGT TTCGTCACAT ATTATGTGTT TTCTTTCAGA CAAATGGA
ACCAAAGGTG TTTATTTTCA TGCAAGTAAA ATATACTGCA TACAAGAAAA AAAA********
A********

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Met Glu Thr Aas Cys Arg Lys Val Ser Ala Cys Leu Gln Leu Glu
1 5 10 15
Lys Asp Lys Gly Gln Gln Gly Lys Aan Glu Asp Val Gly Ala Glu Asp
20 25 30
Pro Ser Lys Lys Arg Gln Gln Arg Gln Arg Thr His Phe Thr Ser
35 40 45
Gln Gln Leu Gln Gln Leu Gln Ala Thr Phe Gln Gln Aen Aen Tyr Pro
50 55 60
Asp Met Ser Thr Arg Glu Glu Ile Ala Val Trp Thr Aen Leu Thr Glu
65 70 75 80
Ala Arg Val Arg Val Trp Phe Lys Aen Arg Arg Ala Lys Trp Arg Lys
85 90 95
Arg Glu Arg Aen Gln Gln Ala Glu Leu Cys Lys Aen Gly Phe Gly Pro
100 105 110
Gln Phe Aen Gly Leu Met Gln Pro Tyr Aen Asp Met Tyr Pro Gly Tyr
115 120 125
Ser Tyr Aen Aen Trp Ala Ala Lys Gly Leu Thr Ser Aen Ser Leu Ser
130 135 140
Thr Lys Ser Phe Pro Phe Aen Ser Met Aan Val Aen Pro Leu Ser
145 150 155 160
Ser Gln Ser Met Phe Ser Pro Aen Ser Ile Ser Ser Met Ser Met
165 170 175
Ser Ser Ser Met Val Pro Ser Ala Val Thr Gly Val Pro Gly Ser Ser
180 185 190
Leu Aan Ser Leu Lys Gln Ser Leu Ser Leu Ser Aen Ser Leu
195 200 205
Ser Ala Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Pro Thr Pro Pro
210 215 220
Tyr Val Tyr Arg Asp Thr Cys Ser Ser Leu Ala Ser Leu Arg Leu
225 230 235 240
Lys Ala Lys Glu His Ser Ser Phe Gly Tyr Ala Ser Val Gln Aen Pro
245 250 255
Aa Ser Aen Leu Ser Ala Cys Gln Tyr Ala Val Asp Pro Pro Val
260 265 270
(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 816 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..814

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

ATG GAG ACC AAC TGC CCG AAA CTG GTG TCG GCG TGT CTC CAA TTA GAG
Met Glu Thr Aaa Cys Arg Lys Leu Val Ser Ala Cys Leu Gin Leu Glu
1  5 10  15

AAA GAT AAG GGC CAG CAA CCA GGA AAG AAT GAG GAT GTC GGC GAG GAC
Lys Asp Lys Gly Gln Gln Lys Ala Glu Asp Val Gly Ala Glu Asp
20  25 30

CCG TCC AAG AAG CCG CAA CCG CAG AGG ACT CAT TTC ACT AGG
Pro Ser Lys Lys Arg Gln Arg Arg Arg Arg Thr His Phe Thr Ser
35  40  45

CAG CAT CTG CAG CAG CCG CAA CCG ACT TTC CAG AGA AAG GGC TAC CCA
Gln Glu Leu Gln Leu Ala Thr Gin Gin Gin Gin Gin Gin Arg Arg Asp Pro
50  55  60

GAC ATG TCC ACT CCG GAA GAA ATC GCT GTG TCG ACC AAG TTY AGG GAA
Asp Met Ser Thr Arg Gin Ile Ala Val Trp Thr Asn Leu Thr Glu
65  70  75  80

GCC CGA GTC CGG GTT TGG TTC AAG AAT CCG CGG GCC AAA TGG AGA
Ala Arg Val Arg Val Trp Phe Lys Asn Arg Arg Glu Arg Lys Ala Glu
85  90  95

CGG GAA CGC CAG CAG GCC GCG CTG TGC AAG AAT GGC TTT GGG CGG
Arg Glu Arg Aaa Cys Gin Gln Ala Leu Cys Lys Asp Asp Gin Gin Gin
100 105 110

CAG TTC AAC GGG CTC ATG CAG CCC TAT GAC GAC ATG TCC CCC TCC
Gln Phe Aaa Gin Thr Aaa Thr Aaa Thr Aaa Thr Aaa Thr Aaa Thr Aaa
115 120 125

TCG TAG AAC TAC TGG ACT GAC CCC TCC CCG TCC GTC TCT TCC
Ser Tyr Aaa Aaa Aaa Cys Aaa Ser Aaa Aaa Aaa Aaa Aaa Aaa Aaa
130 135 140

ACC AAG GGC TTC CCC TCC TCC TTC AAG ACG GTC AAT CCG CTG
Thr Lys Ser Phe Phe Phe Ser Phe Ser Met Aaa Val Aaa Val Aaa Val
145 150 155

TCT CAG AGT TTT TTC CCG CCC AAC TCC ATC TGA CTT AGT AGT
Ser Gin Ser Phe Ser Pro Phe Ser Pro Met Ser Ser Ser Ser Ser Ser Ser
160 165 170 175

TCG TCC AAG GTC CCC TCC CCG TTC GCC ACC GGC GCC GTC CCG GCC TCC AGC
Ser Ser Met Val Pro Ser Ala Val Thr Gly Val Pro Gly Ser Ser
180 185 190

CTC AAT GGC CAG TCG AAC ATG AAC GCC TCG CAG
Leu Aaa Ser Leu Aaa Ser Leu Aaa Ser Leu Aaa Ser Leu Aaa Ser Leu Aaa
195 200 205

TCC GCG GTG CCC ACG GCC GCC TCC CCT TAC GCG CGC ACT CCT CGG
Ser Ala Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Thr Pro Thr Pro
210 215 220

TAC GTC TAT AGG GAC TAT ACG TCC AGC GAC CCC ACG CTG AGC
Tyr Val Tyr Arg Asp Thr Cys Asn Ser Ser Ala Ser Les Ser Les Les Ser
225 230 235 240

AAA GCA AGG CAG CAC TCG TTC GGC TAC GCC AGC GCG CAG AAC CGG
768

(7) SEQ ID NO: 6

ATG GAG ACC AAC TGC CCG AAA CTG GTG TCG GCG TGT CTC CAA TTA GAG
Met Glu Thr Aaa Cys Arg Lys Leu Val Ser Ala Cys Leu Gin Leu Glu
1  5 10  15

AAA GAT AAG GGC CAG CAA CCA GGA AAG AAT GAG GAT GTC GGC GAG GAC
Lys Asp Lys Gly Gln Gln Lys Ala Glu Asp Val Gly Ala Glu Asp
20  25 30

CCG TCC AAG AAG CCG CAA CCG CAG AGG ACT CAT TTC ACT AGG
Pro Ser Lys Lys Arg Gln Arg Arg Arg Arg Thr His Phe Thr Ser
35  40  45

CAG CAT CTG CAG CAG CCG CAA CCG ACT TTC CAG AGA AAG GGC TAC CCA
Gln Glu Leu Gln Leu Ala Thr Gin Gin Gin Gin Gin Gin Arg Arg Asp Pro
50  55  60

GAC ATG TCC ACT CCG GAA GAA ATC GCT GTG TCG ACC AAG TTY AGG GAA
Asp Met Ser Thr Arg Gin Ile Ala Val Trp Thr Asn Leu Thr Glu
65  70  75  80

GCC CGA GTC CGG GTT TGG TTC AAG AAT CCG CGG GCC AAA TGG AGA
Ala Arg Val Arg Val Trp Phe Lys Asn Arg Arg Glu Arg Lys Ala Glu
85  90  95

CGG GAA CGC CAG CAG GCC GCG CTG TGC AAG AAT GGC TTT GGG CGG
Arg Glu Arg Aaa Cys Gin Gln Ala Leu Cys Lys Asp Asp Gin Gin Gin
100 105 110

CAG TTC AAC GGG CTC ATG CAG CCC TAT GAC GAC ATG TCC CCC TCC
Gln Phe Aaa Gin Thr Aaa Thr Aaa Thr Aaa Thr Aaa Thr Aaa Thr Aaa
115 120 125

TCG TAG AAC TAC TGG ACT GAC CCC TCC CCG TCC GTC TCT TCC
Ser Tyr Aaa Aaa Aaa Cys Aaa Ser Aaa Aaa Aaa Aaa Aaa Aaa Aaa
130 135 140

ACC AAG GGC TTC CCC TCC TCC TTC AAG ACG GTC AAT CCG CTG
Thr Lys Ser Phe Phe Phe Ser Phe Ser Met Aaa Val Aaa Val Aaa Val
145 150 155

TCT CAG AGT TTT TTC CCG CCC AAC TCC ATC TGA CTT AGT AGT
Ser Gin Ser Phe Ser Pro Phe Ser Pro Met Ser Ser Ser Ser Ser Ser Ser
160 165 170 175

TCG TCC AAG GTC CCC TCC CCG TTC GCC ACC GGC GCC GTC CCG GCC TCC AGC
Ser Ser Met Val Pro Ser Ala Val Thr Gly Val Pro Gly Ser Ser
180 185 190

CTC AAT GGC CAG TCG AAC ATG AAC GCC TCG CAG
Leu Aaa Ser Leu Aaa Ser Leu Aaa Ser Leu Aaa Ser Leu Aaa Ser Leu Aaa
195 200 205

TCC GCG GTG CCC ACG GCC GCC TCC CCT TAC GCG CGC ACT CCT CGG
Ser Ala Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Thr Pro Thr Pro
210 215 220

TAC GTC TAT AGG GAC TAT ACG TCC AGC GAC CCC ACG CTG AGC
Tyr Val Tyr Arg Asp Thr Cys Asn Ser Ser Ala Ser Les Ser Les Les Ser
225 230 235 240

AAA GCA AGG CAG CAC TCG TTC GGC TAC GCC AGC GCG CAG AAC CGG
768
Lys Ala Lys Gln His Ser Ser Phe Gly Tyr Ala Ser Val Gln Asn Pro
245
250
255
GCC TCC AAC CGG AGT GCT TGC CAG TAT GCA GTG GAC CCG CCC GTG T
814
Ala Ser Asn Leu Ser Ala Cys Gln Tyr Ala Val Asp Pro Pro Val
260
265
270
GA
816

(2) INFORMATION FOR SEQ ID NO: 7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:
Gln Arg Arg Gln Arg Thr His Phe Thr Ser Gin Gin Leu Gin Glu Leu
1       5       10       15
Glu Ala Thr Phe Gin Arg Gin Tyr Pro Asp Met Ser Thr Arg Glu
20      25      30
Glu Ile Ala Val Trp Thr Asn Leu Thr Glu Ala Arg Val Arg Val Trp
35      40      45
Phe Lys Asn Arg Arg Ala Lys Trp Arg Gin Gin Glu
50      55      60

(2) INFORMATION FOR SEQ ID NO: 8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
Gln Arg Arg Gln Arg Thr His Phe Thr Ser Gin Gin Leu Gin Glu Leu
1       5       10       15
Glu Ala Thr Phe Gin Arg Gin Tyr Pro Asp Met Ser Met Arg Glu
20      25      30
Glu Ile Ala Val Trp Thr Asn Leu Thr Glu Pro Arg Val Arg Val Trp
35      40      45
Phe Lys Asn Arg Arg Ala Lys Trp Arg Gin Gin Glu
50      55      60

(2) INFORMATION FOR SEQ ID NO: 9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
Pro Arg Arg Gin Arg Thr His Phe Thr Ser His Gin Leu Thr Glu Leu
1       5       10       15
Glu Asn Trp Phe Ser Arg Asn Arg Tyr Pro Asp Met Ala Cys Arg Glu
(2) INFORMATION FOR SEQ ID NO: 10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
Glu Arg Arg Glu Arg Thr Thr Arg Ser Glu Leu Asp Val Leu
1 5 10 15
Glu Ala Leu Phe Ala Lys Thr Arg Tyr Pro Asp Ile Phe Thr Arg Glu
20 25 30
Glu Val Ala Leu Lys Ile Asn Leu Pro Glu Ser Arg Val Glu Val Trp
35 40 45
Phe Lys Asn Arg Arg Ala Lys Cys Arg Glu Gln Gln Gln
50 55 60

(2) INFORMATION FOR SEQ ID NO: 11:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
Gln Arg Arg Glu Arg Thr Thr Arg Ala Gln Leu Asp Val Leu
1 5 10 15
Glu Ala Leu Phe Ala Lys Thr Arg Tyr Pro Asp Ile Phe Thr Arg Glu
20 25 30
Glu Val Ala Leu Lys Ile Asn Leu Pro Glu Ser Arg Val Glu Val Trp
35 40 45
Phe Lys Asn Arg Arg Ala Lys Cys Arg Glu Gln Gln Gln
50 55 60

(2) INFORMATION FOR SEQ ID NO: 12:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
Gln Arg Arg Glu Arg Thr Thr Arg Ala Gln Leu Asp Val Leu
1 5 10 15
Glu Ala Leu Phe Gly Lys Thr Arg Tyr Pro Asp Ile Phe Thr Arg Glu
20 25 30
(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Lys Arg Arg His Arg Thr Ile Phe Thr Asp Glu Gln Leu Glu Ala Leu
1  5  10  15
Glu Asn Leu Phe Glu Gln Glu Thr Tyr Pro Asp Val Gly Thr Arg Glu
20  25  30
Gln Leu Ala Arg Lys Val Leu Arg Glu Gln Lys Val Glu Val Trp
35  40  45
Phe Lys Asn Arg Arg Ala Lys Trp Arg Arg Glu Lys
50  55  60

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Gln Arg Arg Ser Arg Thr Thr Ala Glu Gln Leu Glu Glu Leu
1  5  10  15
Glu Lys Ala Phe Glu Arg Thr His Tyr Pro Asp Ile Tyr Thr Arg Glu
20  25  30
Glu Leu Ala Gln Arg Thr Lys Thr Leu Thr Gln Ala Arg Val Gln Val Trp
35  40  45
Phe Ser Asn Arg Arg Ala Trp Arg Lys Glu Ala
50  55  60

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Gln Arg Arg Ser Arg Thr Thr Ala Glu Gln Leu Glu Glu Leu
1  5  10  15
Glu Arg Ala Phe Glu Arg Thr His Tyr Pro Asp Ile Tyr Thr Arg Glu
20  25  30
Glu Leu Ala Gln Arg Ala Lys Leu Thr Glu Ala Arg Val Gln Val Trp
35  40
Fhe Ser Asn Arg Ala Arg Trp Arg Lys Gln Ala
50  55  60

(2) INFORMATION FOR SEQ ID NO: 16:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:
Gln Arg Arg Cys Arg Thr Thr Phe Ser Ala Ser Gln Leu Asp Glu Leu
1  5  10  15
Glu Arg Ala Phe Glu Arg Thr Gln Tyr Pro Asp Ile Tyr Thr Arg Glu
20  25  30
Glu Leu Ala Gln Arg Thr Asn Leu Thr Glu Ala Arg Ile Gln Val Trp
35  40  45
Fhe Ser Asn Arg Ala Arg Leu Arg Lys Gln His
50  55  60

(2) INFORMATION FOR SEQ ID NO: 17:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:
Gln Arg Arg Asn Arg Thr Thr Phe Ala Leu Gln Gln Leu Glu Ala Leu
1  5  10  15
Glu Ala Val Phe Ala Gln Thr His Tyr Pro Asp Val Phe Thr Arg Glu
20  25  30
Glu Leu Ala Met Lys Ile Asn Leu Thr Glu Ala Arg Val Gln Val Trp
35  40  45
Fhe Gln Asn Arg Arg Ala Lys Trp Arg Lys Thr Glu
50  55  60

(2) INFORMATION FOR SEQ ID NO: 18:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:
Gln Arg Arg Tyr Arg Thr Thr Ser Phe Gln Leu Glu Glu Leu
1  5  10  15
Glu Lys Ala Phe Ser Arg Thr His Tyr Pro Asp Val Phe Thr Arg Glu
20  25  30
Glu Leu Ala Met Lys Ile Gly Leu Thr Glu Ala Arg Ile Gln Val Trp
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>Phe</td>
<td>Gln</td>
</tr>
<tr>
<td>40</td>
<td>Asn</td>
<td>Arg</td>
</tr>
<tr>
<td>45</td>
<td>Arg</td>
<td>Ala</td>
</tr>
<tr>
<td>50</td>
<td>Trp</td>
<td>Arg</td>
</tr>
<tr>
<td>55</td>
<td>Gln</td>
<td>Gln</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gln</td>
<td>Arg</td>
</tr>
<tr>
<td>5</td>
<td>Arg</td>
<td>Tyr</td>
</tr>
<tr>
<td>10</td>
<td>Arg</td>
<td>Thr</td>
</tr>
<tr>
<td>15</td>
<td>Thr</td>
<td>Thr</td>
</tr>
<tr>
<td>20</td>
<td>Phe</td>
<td>Thr</td>
</tr>
<tr>
<td>25</td>
<td>Ser</td>
<td>Phe</td>
</tr>
<tr>
<td>30</td>
<td>Gln</td>
<td>Leu</td>
</tr>
<tr>
<td>35</td>
<td>Glu</td>
<td>Leu</td>
</tr>
<tr>
<td>40</td>
<td>Glu</td>
<td>Leu</td>
</tr>
<tr>
<td>45</td>
<td>Glu</td>
<td>Leu</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lys</td>
<td>Arg</td>
</tr>
<tr>
<td>5</td>
<td>Arg</td>
<td>His</td>
</tr>
<tr>
<td>10</td>
<td>Thr</td>
<td>Thr</td>
</tr>
<tr>
<td>15</td>
<td>Thr</td>
<td>Ser</td>
</tr>
<tr>
<td>20</td>
<td>Leu</td>
<td>Gln</td>
</tr>
<tr>
<td>25</td>
<td>Leu</td>
<td>Gln</td>
</tr>
<tr>
<td>30</td>
<td>Leu</td>
<td>Gln</td>
</tr>
<tr>
<td>35</td>
<td>Leu</td>
<td>Gln</td>
</tr>
<tr>
<td>40</td>
<td>Leu</td>
<td>Gln</td>
</tr>
<tr>
<td>45</td>
<td>Leu</td>
<td>Gln</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gln</td>
<td>Arg</td>
</tr>
<tr>
<td>5</td>
<td>Arg</td>
<td>Asn</td>
</tr>
<tr>
<td>10</td>
<td>Arg</td>
<td>Thr</td>
</tr>
<tr>
<td>15</td>
<td>Thr</td>
<td>Thr</td>
</tr>
<tr>
<td>20</td>
<td>Phe</td>
<td>Thr</td>
</tr>
<tr>
<td>25</td>
<td>Ser</td>
<td>Phe</td>
</tr>
<tr>
<td>30</td>
<td>Ser</td>
<td>Gln</td>
</tr>
<tr>
<td>35</td>
<td>Leu</td>
<td>Gln</td>
</tr>
<tr>
<td>40</td>
<td>Leu</td>
<td>Gln</td>
</tr>
<tr>
<td>45</td>
<td>Leu</td>
<td>Gln</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Asp</td>
<td>Leu</td>
</tr>
<tr>
<td>5</td>
<td>Ala</td>
<td>Arg</td>
</tr>
<tr>
<td>10</td>
<td>Val</td>
<td>Arg</td>
</tr>
<tr>
<td>15</td>
<td>Val</td>
<td>Arg</td>
</tr>
<tr>
<td>20</td>
<td>Gln</td>
<td>Val</td>
</tr>
<tr>
<td>25</td>
<td>Val</td>
<td>Val</td>
</tr>
<tr>
<td>30</td>
<td>Val</td>
<td>Val</td>
</tr>
<tr>
<td>35</td>
<td>Val</td>
<td>Val</td>
</tr>
<tr>
<td>40</td>
<td>Val</td>
<td>Val</td>
</tr>
<tr>
<td>45</td>
<td>Val</td>
<td>Val</td>
</tr>
<tr>
<td>Phe</td>
<td>Gln</td>
<td>Asn</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>50</td>
<td>55</td>
<td>60</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

| Gln | Arg | Arg | Asn | Arg | Thr | Thr | Phe | Asn | Ser | Ser | Gln | Leu | Gln | Ala | Leu |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1   | 5   | 10  | 15  |

<table>
<thead>
<tr>
<th>Glu</th>
<th>Arg</th>
<th>Val</th>
<th>Phe</th>
<th>Glu</th>
<th>Arg</th>
<th>Thr</th>
<th>His</th>
<th>Tyr</th>
<th>Pro</th>
<th>Asp</th>
<th>Ala</th>
<th>Phe</th>
<th>Val</th>
<th>Arg</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Glu</th>
<th>Leu</th>
<th>Ala</th>
<th>Arg</th>
<th>Val</th>
<th>Leu</th>
<th>Ser</th>
<th>Gln</th>
<th>Ala</th>
<th>Val</th>
<th>Glu</th>
<th>Val</th>
<th>Trp</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phe</th>
<th>Gln</th>
<th>Asn</th>
<th>Arg</th>
<th>Arg</th>
<th>Ala</th>
<th>Lys</th>
<th>Phe</th>
<th>Arg</th>
<th>Arg</th>
<th>Asn</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

<table>
<thead>
<tr>
<th>Gln</th>
<th>Arg</th>
<th>Arg</th>
<th>Ile</th>
<th>Arg</th>
<th>Thr</th>
<th>Thr</th>
<th>Phe</th>
<th>Thr</th>
<th>Ser</th>
<th>Ala</th>
<th>Gln</th>
<th>Leu</th>
<th>Lys</th>
<th>Glu</th>
<th>Leu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Glu</th>
<th>Arg</th>
<th>Val</th>
<th>Phe</th>
<th>Ala</th>
<th>Glu</th>
<th>Thr</th>
<th>His</th>
<th>Tyr</th>
<th>Pro</th>
<th>Asp</th>
<th>Ile</th>
<th>Tyr</th>
<th>Thr</th>
<th>Arg</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Glu</th>
<th>Leu</th>
<th>Ala</th>
<th>Leu</th>
<th>Lys</th>
<th>Ile</th>
<th>Asp</th>
<th>Leu</th>
<th>Thr</th>
<th>Gln</th>
<th>Ala</th>
<th>Arg</th>
<th>Val</th>
<th>Glu</th>
<th>Val</th>
<th>Trp</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phe</th>
<th>Gln</th>
<th>Asn</th>
<th>Arg</th>
<th>Arg</th>
<th>Ala</th>
<th>Lys</th>
<th>Phe</th>
<th>Arg</th>
<th>Arg</th>
<th>Lys</th>
<th>Gin</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

<table>
<thead>
<tr>
<th>Gln</th>
<th>Arg</th>
<th>Arg</th>
<th>Asn</th>
<th>Arg</th>
<th>Thr</th>
<th>Thr</th>
<th>Phe</th>
<th>Asn</th>
<th>Ser</th>
<th>Ser</th>
<th>Gln</th>
<th>Leu</th>
<th>Gln</th>
<th>Ala</th>
<th>Leu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Glu</th>
<th>Arg</th>
<th>Val</th>
<th>Phe</th>
<th>Glu</th>
<th>Arg</th>
<th>Thr</th>
<th>His</th>
<th>Tyr</th>
<th>Pro</th>
<th>Asp</th>
<th>Ala</th>
<th>Phe</th>
<th>Val</th>
<th>Arg</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Asp | Leu | Ala | Arg | Val | Asn | Leu | Thr | Gln | Ala | Arg | Val | Glu | Val | Trp |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 35  | 40  | 45  |
(2) INFORMATION FOR SEQ ID NO: 25:
   (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 60 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: peptide
   (v) FRAGMENT TYPE: internal
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

   Lys Arg Arg His Arg Thr Ile Phe Thr Ser Tyr Gln Leu Glu Glu Leu
   1    5    10   15
   Glu Lys Ala Phe Asn Glu Ala His Tyr Pro Asp Val Tyr Ala Arg Glu
   20   25   30
   Met Leu Ala Met Lys Thr Glu Leu Pro Glu Asp Arg Ile Gln Val Trp
   35   40   45
   Phe Glu Asn Arg Arg Ala Lys Trp Arg Lys Arg Glu
   50   55   60

(2) INFORMATION FOR SEQ ID NO: 26:
   (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 60 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: peptide
   (v) FRAGMENT TYPE: internal
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

   Gln Lys Arg Arg His Arg Thr Arg Phe Thr Pro Ala Gln Leu Asn Glu Leu
   1    5    10   15
   Glu Arg Ser Phe Ala Lys Thr Tyr Pro Asp Ile Phe Met Arg Glu
   20   25   30
   Glu Leu Ala Leu Arg Ile Gln Leu Thr Glu Ser Arg Val Gln Val Trp
   35   40   45
   Phe Glu Asn Arg Arg Ala Lys Trp Lys Lys Arg Lys
   50   55   60

(2) INFORMATION FOR SEQ ID NO: 27:
   (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 14 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: peptide
   (v) FRAGMENT TYPE: internal
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

   Ser Ser Leu Ala Ser Leu Arg Leu Lys Ala Lys Gln His Ser
   1    5    10

(2) INFORMATION FOR SEQ ID NO: 28:
   (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 14 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:
Ser Ser Leu Ala Ser Leu Arg Leu Lys Ala Lys Gln His Ser
1  5  10

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:
Asn Ser Ile Ala Ser Leu Arg Leu Lys Ala Lys Glu Phe Ser
1  5  10

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:
Asn Ser Ile Ala Asn Leu Arg Leu Lys Ala Lys Glu Tyr Ser
1  5  10

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:
Ser Ser Ile Ala Val Leu Arg Met Lys Ala Lys Glu His Thr
1  5  10

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:
Ser Ser Ile Ala Ala Leu Arg Leu Lys Ala Arg Glu His Glu
1  5  10
(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 14 amino acids
   (B) TYPE: amino acid
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

Asn Ser Ile Ala Asn Leu Arg Leu Lys Ala Lys Glu Tyr Ser
1   5   10

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 14 amino acids
   (B) TYPE: amino acid
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

Ala Ser Val Ala Leu Arg Met Lys Ala Arg Glu His Ser
1   5   10

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 14 amino acids
   (B) TYPE: amino acid
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

Asn Ser Ile Ala Ala Leu Arg Ala Lys Ala Gln Glu His Ser
1   5   10

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 14 amino acids
   (B) TYPE: amino acid
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

Thr Ser Ile Ala Ser Leu Arg Arg Lys Ala Leu Glu His Thr
1   5   10

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 16 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:
GCGCTCCACG TCTCGG

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 16 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GCGCTCCACG TCTCGG

GCGCTCCACG TCTCGG

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 15 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CTGCTCTCCG CCGG

CTGCTCTCCG CCGG

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 15 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CTGCTCTCCG CCGG

CTGCTCTCCG CCGG

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 15 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

TTTCTCCCG TGCG

TTTCTCCCG TGCG

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 15 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GGACATGCC ACAACG
(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

TTGCTCCGCTG

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGGACATGTCGG

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GCGGGCCGTTGGG

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GCGGGCCGNTGGG

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

TTGGGCOCGA CGATT

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(3) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 48:
TTTGGCCGGA CCATT

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xiii) SEQUENCE DESCRIPTION: SEQ ID NO: 49:
CAACTGGCC GCCA

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xiv) SEQUENCE DESCRIPTION: SEQ ID NO: 50:
ACAACGTAC CCAC

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xv) SEQUENCE DESCRIPTION: SEQ ID NO: 51:
GAAAGTCCAA ACTCATCAAC

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xvi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:
GACTACTAC CCAAGACTA

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

GCTCTTTC TCTCCACCA

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

TUTAGGTTAG GCTATTTTAC

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

TAAAATTTCA TTTCGAGAG AG

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

CTGAATACAG CAGAATTTAC C

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

TGAAGTGTG TGGCCTCA

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:
ATACCAAGAG TAAAAGCTAC

(2) INFORMATION FOR SEQ ID NO: 59:
   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 20 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: cDNA
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

TTGTCTGTT GCCTTTTAGG

(2) INFORMATION FOR SEQ ID NO: 60:
   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 20 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: cDNA
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

CAAGGTCAC GTGACTCAC

(2) INFORMATION FOR SEQ ID NO: 61:
   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 20 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: cDNA
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

CCCAGTCTCA ATGAAAGCAGC

(2) INFORMATION FOR SEQ ID NO: 62:
   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 20 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: cDNA
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

CATTACCTC ACGACTCAC

(2) INFORMATION FOR SEQ ID NO: 63:
   (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 21 base pairs
       (B) TYPE: nucleic acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear
   (ii) MOLECULE TYPE: cDNA
   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

TTTGAATGTA TGCTGTCTGTT G

(2) INFORMATION FOR SEQ ID NO: 64:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:
TAATGCTAA AATACCTGTT GC

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:
AACACTAATC TGAACCTGCT

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:
CAACAATAAG GTAATAAAG CAA

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:
GTTAGCATC TACTAGCAAG TT

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:
GACACAGTC CAGACAACCA

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

TAATTCAGC CATATTGAA ATT

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

TCCACATCTG TTCAAGTAAC

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

AAATCACAGT GCTGTGTYCT

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

ACAGCTTAAA ACTGAAACAG

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

TATTTGTGTY GYGGTTCCTG

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 74:
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

AAATCGCC ATCCCAAG

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

GGGTATTTA ACAAGCTTGA AT

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

GAGAGCTG GYACATCTC

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

CCCACTCTG AATTTGTTC

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

TTAGGTAAAT GCCACAATT TT

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

AAATTTAAT GCATCTTACG AC
(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

AGGATATGGGAGAACACTG

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

ATAACTGCACGAGATTCTG

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

TGCACAATAATGAATTCG

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

CCTTAATATTGCTACTGTGATAT

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

CCTCTCTCCCCAGCACATG

(2) INFORMATION FOR SEQ ID NO: 85:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

TCTCTCCGFT ACTCTGCTTF

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TCCCTTCAC TGTGACCAAC

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

GAAAGTAAA GTAAGTCTTF GG

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

ATCTCTCGTF CCACAGCAAA

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

CTGATGACCT CTGGTGTGF

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

AAAGAGTTT CCAGTACCA

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

TGGAGATTTC TCAAATTTTG G

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

TCCTGCTTC ATGCTGTTG

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

AATACTCTCT CTOCTCTTT

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

ACCGAGGAC CAGGATAAAA

(2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

GAAATTGAA ATTCTTTTG TC

(2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

TTCTTCTGA CTTCCTTTT A

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

AAGAATGGGAC GTGCACTG

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

AGCAGCTTCC ACGTTCCTG

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

GCTGCTTCC CAAAATAGC

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

CATTATAAGT CGACTATAAG TG
(2) INFORMATION FOR SEQ ID NO: 101:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:
GAGTTTAG TTTGCTTOT G

(2) INFORMATION FOR SEQ ID NO: 102:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:
CCCACTTTG GATAGTATC C

(2) INFORMATION FOR SEQ ID NO: 103:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:
TAACTCTTAG CCAACTGAY C

(2) INFORMATION FOR SEQ ID NO: 104:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:
TTGACCACAT CAGCCCAAC

(2) INFORMATION FOR SEQ ID NO: 105:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:
TCTATCTCT CTGGATTCA G

(2) INFORMATION FOR SEQ ID NO: 106:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:
AATGGAGAA TTAGGGGCAAG

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:
CCACCGTCCA AATCGACTC

(2) INFORMATION FOR SEQ ID NO: 108:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:
TTGCTATAT GGCAACACTC

(2) INFORMATION FOR SEQ ID NO: 109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:
CTGCTCCCCA ATTCCTCACT

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:
GTCAGCAGTC TTTGACAAGG

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

GATAAAGCC AGCAAGGGA  

20

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

GAGGCACTG TAATCTGCA  

20

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

GTAATCTGCA CAGGCGCATC  

20

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

CCAGTTGTTG TAGGAATGCC C  

21

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

AGTCCAATGG GCACATCGAC  

20

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:
GGGAAAAGCA TGCTCTGTGA

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1929 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

CGAGGAACTG GGAGAGAGG AGAAAGAAAG TGCCCATGGG GCTGAGATT CCGAGAGCCC 60
AGTGACCCCG TGGACCTCTT CGGAACCTTG CACCCCTGAG AGCCCTGCGG TCTCTTACGG 120
CCGGCCCTTC GGGGCTTCC CGTGCCAGCG CAGGCTCCGG TGCTCTGAAA TGCCGCGGGG 180
AACAACTGGG AGGCGGAGAC AGGACCTGCT TGGCGCTGAC CGTGAAGTGG AGGCCCTTCT 240
TACACCTCCT CTCTCCTCTG TTCTCGACAG ATGGGAAAGA TGAGACCCAA CTCGCGGAAA 300
CTGSGTGTCGG CTGTGTCTGA ATATGGTAA GACCCCCCTC TCTGGCACC GGCTACGCGG 360
TGTTTTTACG TTTTACAGA AGATATAAGG CGGCGGCGGG AGAAATGAGA AGCTGGGCGG 420
CGGAGGCCCC TCTAAGAAGG AGGCGGCAAAG GCAGGAGCGG ACTACCTTTA CGACCCAGCA 480
GCTCCACAGC TGAGGAGGCC CTCCCCCGAG GACCGCATAC CGAGACATGT CCAGCAAGGC 540
AGAAATCCTT GCTGAGACCA ACCCTACGGA AGGCCGAAGTC CGGTTTGGCC ACCACGAGCT 600
CGGGAGGCGT TTCTTGGCCT TGCCGCGTCC CGCCGCGAGG TCTGTGACAGA GATGCTAGG 660
GCCAATAGGA GAAAGAGGGA GGCGACACCG AGGGGCGGCC ATAGCAAGAA GGGCCCTGGG 720
CCCAGGTTCC ATGAGCCCTG GCAGGCGTACG GACCGATGCT ACCAAATGCT TTGGTCAAC 780
AACTGCGCGC CGAGGACGCT TACATCGGCG TCCGCAAGCA CGAGAGGTTT CCCCTTTGAC 840
AGCTCTACTG ATGAAGCCAC CCCGCTACAT CAGAGATGCT TTTCCCGACC CAAACCTGAT 900
TGCCCGATGA CAGTGCGCTG CAGCATGCTG CCCGCACAGG TGACAGGCCG CCGGGCTGCT 960
AGTCTCTACAA GCTCGCAAAC CTGGACACAG CAGATGCGCC CGCCGTCGTC TCCGGCGTGG 1020
CCGACGGCTG CCGTCTCTTA CGGCCCCCGG ACTCCCTCGT ATAGTTTATG AGCACAGCTG 1080
AACAGCGGCG CGGCTCCGCT GCAGCTGAGG AGGGGGTTGG GCGAGCTGCC CCGCGGCGCC 1140
AGCGTAGA AAGCGGCCCG CAACCGCTGT GCTGCGCACT ATGCACTGGA CGGGCGCGTG 1200
TGAGCCGCC CCAGCAAGCC GGGACTCATG GACCTTGCGG GATGGGCGAA CCCTGCGGCT 1260
GAAAGACTGG GAATTTAGCT AGAAGGCTCT GGGCGAAGAA AGAGAGGAGA AAGAGGAGAA 1320
CTATATAGAG AAAGAAGAAC CACGAGATCA AAGAGAGGAA TCCCTTGAGT CAAAAGGGAT 1380
GTGGTACAGG TGCTACATCG CATGCTAAAG ATGTTTCTAA AGTGTCCAGA GAGACATACA 1440
CAAAAGAAAT TTGTTACTCG ATATGACATT TTAGACATT TAAAGCTTG TTATAACTTA 1500
ATGTATACAT TTTAATTAG TAAATGATGC AAGAGATGTA TATATAGTA AAGGTCAAAT 1560
TTATTTTTA AAGGACCTG GATTATATAT CACACAGGAT TTATTTAAGG TTATCTTTTA 1620
AAATAAAGCA TTTTACTAGA AAGGACTATG GAAAGTTTCC GTGCGACAGA GGGAGTATAT 1680
ATATACATTG CCAAGCTCTG TTTTCCTATAC ATATTTATAT TTATTATAAA AATGCTGGAA 1740
TATCATTTTT AGAAATGTTT GCTGTTGAGA TGCAAATGAT TTTCGGAATC TGCATAACT 1800
AAACCTACCT GTCTATACAA TTCGGCCAAA TATATTTGT TTACTTTCGA CAAATAGAGA 1860
AAACAAATGCG TTTTATTTTT ATGGGACAT AAATATCATGCG ATACAAAGAA AAAAAA 1920
(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1929 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

CGAGAAAGCT GGGAGAGGAG AGAAAGAAAAA TGGCCAGTGCC GGGTGGATTT CGGAGCCG
60
AGTGGACCG GTCACTCTCG CGAGACCTGG CGCCCTCGAG AGCCCTGCAG TCCCTAGG
120
CCCCGTTCGT GCAGGTTCGG CAAGCCCACG TCAGCTGGAA TGGGCCCCGG
180
AAGCACTGAG CCGGCAGTAC AGACACTGCTC TGGCGGAGCC GCATAGGCTG AGGGCCCTCAT
240
TCATGGCAGT CTTGTGCCTG TTGGCGCCAA GGGCCGAAA TGGAGAACAA CTGGCGGAAAA
300
CTGTGTCGGA GTGGTGCACA ATAGGGTAA AGACCCCTTC TCTCGCGCGT GTACCGCGAC
360
GCGGTTTCAAT TTTTACAGGA AAAGTAAAGG CGGACAGGAA AAGATAGGAC ATGTTGGCGC
420
CGAGGACCCG TCCAGAGAGA AGCGCAGCAG CGGGCGAGGG ACTCAATTTCA CTAAGCCAGCA
480
GCTCGAGAC CTCGAGGACA CTTTGGGAGCA AAGCCCTAC CGACAGTATC CCATCCGCG
540
AGAAAATCGCT GGGTGACCCA AGCTTACGGA AGCGCGAGCC CGGCTAGGCT AGACCGAGAT
600
TGCGGGACCT GTGGCGCCCG TGCGGCCCCG CGCCGCGAGG TTGGGTCAGA GAAGCGCGG
660
GCCAATAAGA GACCGCGGAA AGCGGACCCG AGGGCGCCGA TGCGCAAGAA TGCGCTTTGGG
720
CCGGCACTCA AGCGGCTCAG GCAGCGTCATC AGGCAGATCG AGCCGTTAGCA AGCGGCG
780
AATGGGCTG CCGGGCGCT CAGGTCGAGC TTGCTGTCGG CAGAACGCTT CTGCTTTCTGC
840
AACCTCAGA GGCACAACG CCGTGCTCGT CAGGAGATGT TTTGCCCGCC CAAAGCTAC
900
TCACTCATGA GTAGTGCTCG CAGGAGGTGT GTCCTCGCGG TGCCCGGCGT CGGGGCGGCC
960
AGGCTCGAAA GGGCGAGTTA CTGGAGGCAA CGGACCCCG CTCGTTGCAAG TGGCGGCGG
1020
CCGGCGCCCG CCTGCCCTTA CGGCGCGGCG ACTCCGCGGT ACGTTTATAG GACACAGTGC
1080
AACCTGACGC TGCGACGCTT GGAGCTGAAG CCAAGCAAGC ACTCCGCTCCT CGGGCTACG
1140
AGGCTGAGA ACCGGGCCCTC CACCGCTGGT GCTGCGCAAT AGCAAGGCGGA CCGGGCGCGT
1200
TGGCGGCGCC CGGCGGACCGG GAGTTCTTGG GGCTGCGGAG GAGGGCGCC CCGGGCGG
1260
CAGAAGCTGG GATATTGGCT AGAAAGGCGT GGCGCATAAA CAGAGGGAGA CAGAGGGAGA
1320
CTATATAGA AAAAGAAGAC CCCGGAATCA AAGAGAAGAC CTGGTGTAGA CACAGCAGT
1380
GCTCTGACTG TGGCAGATCT TCGACTACAA ATGTTTCTAAAG CTTGGCGGTA GACAACATAAC
1440
CAAGAAAGTGG TTTTGCAGTG ATAGGACATT TTAATATAAT TAAATACCTG TTAATTTTTA
1500
AGTTTACATG TTTAATACAT TAAAATACGT AAGAGAGTA AATATACGGA AAATCCAAAT
1560
TTATTATAG AAGACACCTTG TTAAATATAT CACAAACTTG TTTTAAAGG TTAAGTTTTA
1620
AAATAAAGCG TTATATACAG AAGCCGTTAG GATTTTTGCC TCGCGAGCAA GGAGGTTAT
1680
ATATACCATG CGCACAGTTG TTTTACCTAA ATATATTAT TATTTAACAA ATGAGGCGA
1740
TATCGTTTTT AAGATAGTTT GGGTTGCGGA TGCAATGATG TTGCTGGAAC TCGTATATG
1800
AAGCCTGACC GTGGTAGAAAA TTTGGAGGTT TAACTTCTCA GCAAGAGACTA
1860
AAACAAATGAG TTTTTTTTTT TATGAGGTAAC AATATACCAG ACNACAAAAAA AAAAAAAA
1920
(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 515 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

ACCCGGCCCA GGGCCCGGGG ACTCCAGGAC TCTCCAGAGT GGCACACTCTG CCCGACAAAAG 60
ACTGAGAATT GTGCTAAGG TCTGCGCGA CTATGGGAAG GAAGAGGGGG GAAAAAGGAT 120
CAGAGAAAA GAAGCCACAG AATTCAGAAG GAGACGGCCT TGGATTCAAG AGGAAATGCCC 180
CCAGTCTTCT ACCTTTTTCG CTTAGATAT TCCCAACAGT TGGAGGACGC GTACCCCGAC 240
AAAGCTTTGCA CTTAGATATG CATTCAACAA TCTACTTAAG CGTCTTTATT TTGATAATTT 300
GCATCTYAA CATTTAGTG ACGAGAGGC TTTGATATAT CAGAAAAATGAA TATGACTTTT 360
TATAAAAGCA GTTCTTTCTA CTTATCTAC ACCTTTTATTA AAGCTTACCG TTAAAAATAT 420
AGCATTTAY ACAGAATCAG TTTAGATTTT TCCTTGGCGA GCAAAGGAAT GTATATACGT 480
AATCCACAC TCTTTGTTCTA AACATATTTA TTATT 515

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1775 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 234..1047

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

GGAGAAGACT GGAGAGGAG AGAAGAGAAC AGGCAAGTGG GGGGCAACGCG 60
AGTGGACCGG CTTAGCTCTT CCGACAGCTG AGCCGCTAGC TCTTCTTCAAG 120
GCCGCTTTGC GGGCCGGCGC GCGGTGGGCG TCCTGAAAA TCCCGGGGC 180
AACGGGTGAG AGGCGATCTC TCCGGTTAGC GCAATACGG AGGAAAGG 236
GAG ACC AAC TGC CCC AAA CTG GTG TCG GCG TGT CTT CAA TTA GAG AAA 284
Glu Thr Asn Cys Arg Lys Leu Val Ser Ala Cys Leu Gln Leu Glu Lys
5 10 15
GAT AAA AGC CAG CAG GGG AAG AAT GAG GAC GTC GGC GCC GAC CCG 332
Asp Lys Ser Gln Gln Gly Lys Asn Glu Asp Val Gly Ala Glu Asp Pro
20 25 30
TCT AAG AAG AGG CGG AGG CAG CCC ACT CAC TTT ACC AGC CAG 380
Ser Lys Lys Lys Arg Arg Arg Gln Arg Asp Thr His Phe Thr Ser Gln
35 40 45
CAG CTC CAG CAG GAG GCC ACT TTC CAC AGG AAC CCC TAC CCG GAC 428
Gln Leu Gln Gln Gln Ala Thr Phe Arg Asn Arg Tyr Pro Asp
50 55 60 65
ATG TCC ACA GCG GAA GAA ATC GCT GTG TGG ACC AAC CTT AGC GAA GCC 476
Met Ser Thr Arg Glu Glu Ile Ala Val Trp Thr Asn Leu Thr Glu Ala
CAG GTC CGG GCT GAC CGC CTT GCG ACC AAA TGG AGA AAG AGG
Arg Val Val Trp Phe Lys Arg Arg Arg Ala Lys Trp Arg Lys Arg
95 90 95

GAC GCC AAC CAG AGC GCC GAG CTG AAC AAT GGC TTC GGG CAG
Glu Arg Asn Gln Glu Ala Glu Leu Cys Asn Gly Phe Gly Pro Glu
100 105 110

TTC AAT GGG CTC ATG CAG CCC TAC GAC ATG TAC CCA GCC TAT TCC
Phe Aan Gly Leu Met Glu Pro Tyr Asp Met Tyr Pro Gly Tyr Ser
115 120 125

TAC AAC AAG TGG GCC AAG GCC CTG ATA TGG GCC CTG CTA TCC ACC
Tyr Asn Asn Trp Ala Ala Lys Gly Leu Thr Ser Ala Ser Leu Ser Thr
130 135 140 145

AAG AGC TTC TCC TTC TAC ATG AAC GTG ACC CCC GTG TCA TAC
Lys Ser Phe Pro Phe Aan Ser Met Aan Val Aan Pro Leu Ser Ser
150 155 160

CAG AGC ATG TTT TCC CCA CCC AAC TCT ATC TCG TCC ATG AGC ATG TCG
Gln Ser Met Phe Ser Ser Ile Ser Met Ser Ser Leu Ser Ser Ser
165 170 175

TCC AGC ATG GTG CCC TCA GCA GTG ACA GCC GTC CCG GGC TCC AGT CTC
Ser Ser Met Val Pro Ala Ala Val Thr Gly Val Pro Gly Ser Ser Leu
180 185 190

AAG AGC CTG ATT AAC TGG AAC AGG AGC TCG CTG AAT TCC
Aan Ser Leu Aan Leu Leu Aan Ser Ser Ser Ser Leu Aan Ser
195 200 205

GCG GTG CCG ACC CCT GCC TGT CCT TAC GCG CCC ACT CCT CCG TAT
Aa Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Pro Pro Tyr Pro
210 215 220 225

GTT TAT AGG GAC ACC TGG TAT AAC TCG AGG GCC AGC CTG AGA CTG AAA
Val Tyr Arg Asp Thr Cys Asn Ser Ser Leu Ala Ser Leu Arg Leu Lys
230 235 240

GCA AAG CAG GCC TCC ACC TTC GGC TAC GCC ACC GTG CAG AAA CCG GCC
Aa Lys Glu His Ser Ser Phe Gly Tyr Ala Ser Val Glu Lys Pro Ala
245 250 255

TCC AAC CTG AGT GCT TGC CAG TAT GCA GTG GAC CGG CCC GTG T
Ser Aan Leu Aan Ser Ala Cys Gly Tyr Ala Val Aan Aas Pro Val
260 265 270

GAC CGC ACC CAG CGGCCGG GATGCTTGG ACCGTGCCGG AGGGGCAAC TCCGCCCTTG
1107

AAGACTGCG AATATGCTA GAAATCTGG GGCACATAAG AAGGGGAGAG AAGAGAAG
1167

TATATAGCA AACGCCAAA ACTGAAATCAA AGAGAAAGCT CTTTGATTTC CAAAGGGATT
1227

TCTCAGTGT CTGACATCT TGACAGCAG TATTCAGAC AGTTGCAAGG ACAAATACAC
1287

AACAAAAGT TTTGACTCGA TATGACATCT TACGATCAGT ATACCCTGT GTATCTTTTAA
1347

GTTGGACATT GTTAACATTY AAATGACTGA AAGAGATGA AATATGCGA AAGTCAAAATT
1407

AATTATATAA AAACAGTTGTT TACTATATAT ACAACAGCTGT TTTTAAAGGT TATGCTTTTA
1467

ATAGAACATG TATATAGAGA AGGATAGCAT AATTTAGGTCT GCAGGCAAG GAGGTTTATA
1527

TACAAAAAC GAGCCCTTAC AGTTAGCTTAC TATTGACTTTAC TTTCAAATTTA
1587

ATGACCATCA GAAATGGCATG ATGACGATC GTTCGTTACG GCAGAGATG TAATGCAACT
1647

ACCTACCGTG CTTTATGCTC TCTCACATTTT TATATTGGTT TACTTTCATG CAATATGGA
1707

ACAATAGTGT TTTATTTTCA TGGGAGTAAA ATATACCTCA TACAAAAAG AAAAAAA
1767

AAAAAAAA
1775

(2) INFORMATION FOR SEQ ID NO: 121:
(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 816 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
  (A) NAME/KEY: CDS
  (B) LOCATION: 1..814

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

ATG GAG ACC AAC TGC CCG AAA CTC GTG GCG TGC GGG CAA TTA GAG
Met Glu Thr Aen Cys Arg Lys Leu Val Ser Ala Cye Leu Gln Leu Glu
### (2) INFORMATION FOR SEQ ID NO: 123:

(i) **SEQUENCE CHARACTERISTICS:**
- (A) LENGTH: 1929 base pairs
- (B) TTYPE: nucleic acid
- (C) STRANDNEDNESS: single
- (D) TOPOLOGY: linear

(ii) **MOLECULE TYPE:** DNA (genomic)
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

```
CGAGAAGACT GGAAGAGAAC AGAAGAAAC TCCAGTGCG GCCTAGGAAT CGAGGCCCC  60
AGTCAGCCCG TGATGCTCTT CGAGACTTGG CACCCCTAGG AGCCCCGAG TCTCTCTAGG 120
CCGAGCTTTC GGGGCGGCTG CUGCGACCCG GAGGCTCAGC TCAGTCGAAA TGCCCCGGG 180
AACAGTGAGG AGCGAGAAC AGACGCTTCT TCCGCGTAGC CCAGAAGTTG AGGCGCTTCT 240
TACAGGTCCT CTATTTCTGT TTGGGAGAT AAGGCGGAAA TGAGAACAACT CTCCGCGAAA 300
CTGCTGTCGG CGTGTGCGCA ATTAGGTAAG AACCCCCCTC TCTGGCGCCG GTCAAGGGAC 360
TGCTTGTCA AGAAGGATGG AAGAATGAGA CGTGGCGGGC  420
CGAGGACCGA TCTATGAGAA AGGCGCAAAG AGCGCGCGG TACATCTTGA CAGCGCAAGA  480
GCTCCACGAG CAGAGGCGCA CTTCGCGAGA GACGCCATAC CGGAGGAGT CACAGGCGGA  540
AGAAATCCCT GTCTGGAGCA AACTTACGGA AGCCGAGTCG CGGCTGGAGC AGCAAGGAGT  600
CTGGGAGGCT GTTGGGCTCC TTGGCCCAAC GCGCCCGAGG TTTGCTTCAA GAATCGTCGG  660
GCCGAAATGG GAAGAGGGA GCGAACGCG AGCGCCGAG TATGCAGAA TGCGCTGCGG  720
CGCAGCTTCA ATGGGCTCAT GCGCAGCTAC GAGCACATGT ACCAGACCTA TTCTCCACAC  780
AATCGGGCGC CCAAGGCGCT TACATCCGCG TCTCTATCCCA CAGAAGAAGT CCCCTCTCTC  840
AACCTCTAAG AGCPTCACCC CTCCTCATCA CAGAAGATGT TTCCCCCACC CAACCTCTAC  900
TCTCTCAGGA GCATGCTCCT CAGAGCTGTG TCTCCACAGG TCAGGGCTC CCCCCGGTCG  960
AGCTCTACAA GCTCTGAGAA CAGAACACCG CTAGCTAGCC CTCGCTCTCA TTCCGCGGCG 1020
CGAGGCTGCT CGCTCTCTTT CGCGGGCGCC ACTGCCTCCT ATGTCTTAGT GAACACGGGT 1080
AACCTGAGGC TGCGCCAGCT GAGACTGAAA GCCAAGCAGC AACTCCACCT CGCCTACGCC 1140
AGCTCTAGAG ACGCTGAGCC CACCTGGAGT CGTGGCAGAT ATCGAAGTGA CGGGGCGTGC 1200
TGAGCCGCGC CCACAGGCGC GGGATCTTAG GACCTGGCGC GTGGGCGCAA TCTGGCTCTT 1260
GAAAGACGTC GAAATATCTC AGAAGGCTGT GGGCTAAAG GAAGAGAGGA GAAGAGAAGG 1320
CTATATAGAG AAAAGAACAC CACTGAATCA AGAAGAGAC TCTTTATTG TCAAAGGGAT 1380
GTCCGACTGT TCTGACATCT TCTACATAA GATTTGTCTA CAGTGGCAAG GACACATCA 1440
CAACAAAGTG TTGGTGACTG ATATGACATT TAACCATACG TAAAGCTTG TTTATTATT 1500
AGTTGTATGT TCTTACTAGC TAATAAGACTG AAAAGATTGA TATATATGCA AATGGAACAT 1560
TAATTATATA AAACAGATTG TTATATATAT CAAACAAGTG TTTTAAAGG TTAAGCTTTA 1620
AAATAAAACA TTATATACAG AAGGGATTAG GATTTCGCGC TTTCTAGCGA GGGATGATAT 1680
ATACACTATA CCACACTGTA TTGTTTCTAC ATATATATAT TATATATAA AATTGCTGAA 1740
TATCATATTG AGATAAGTTT GTTGAGTGGA YGCAATTGTG TTCTCGAACG TGCATAGGCA 1800
AACCTACCCG GTATATACCA ATCCGACGAA TATATATGTT TAATCTTTCA GCAAATAGA 1860
AACAATAGTG TTATTTTTAC ATGGGATCAA AAATATATGC ATACAAAAAA AAAAAAAAAA 1920
AAAAAAAAA  1929
```

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1775 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
GAG ACC AAC TGC CGC AAA CGG TGC TCG CGG TGT CGT CAA TTA GAG AAA
Glu Thr Aan Cys Arg Lys Leu Val Ser Ala Cys Leu Glu Lys
5 10 15

GAT AAA AGC CAG CAG CGG AAG AAT GAG CAC G TG GCC CAC GAC CCG
Asp Lys Ser Gln Glu Gly Lys Aan Glu Asp Val Gly Ala Glu Asp Pro
20 25 30

TCT AAG AAG CCG CAA AGG CCG CAG ACT CAG TTT ACC GGC CAG
Ser Lys Lys Lys Arg Gln Arg Gln Arg Thr His Phe Thr Ser Gln
35 40 45

CAG CTC CAG CAG CTG GAG ACT TCC CAG AGG AAC CCC TAC CGG GAC
Gln Leu Gln Gln Leu Glu Ala Thr Phe Gln Arg Asn Arg Tyr Pro Asp
50 55 60 65

ATG TCC CCA CCG GAA GAA ATC GCT GTC TGG ACC AAG CT TCC ACG GAA GCC
Met Ser Pro Arg Gln Ile Ala Val Thr Asn Leu Thr Gln Ala
70 75 80

CGA GTC CGG GTT TGC TTC AAG AGG CAG CCG GCC AAG AAA TCG AAG AGG
Arg Val Arg Val Trp Phe Lys Asn Arg Ala Lys Trp Arg Lys Arg
85 90 95

GAG GCC AAC CAG CAG GCC GAG CTA TGC AAC AAT GCC TTC GGG CGC CAG
Glu Arg Asn Gln Gln Ala Leu Arg Cys Lys Gly Phe Gly Pro Gln
100 105 110

TTC AAT GGG CTC ATT CAG CCC TAC GAC GAC ATG TCC CCA GCC TAT TCC
Phe Asn Gly Leu Met Gln Pro Tyr Asp Met Tyr Pro Gly Tyr Ser
115 120 125

TAC AAC AAG TGG GCC CAC AAG GCC TCT ACA TCC GCC TCC GTA TCG ACC
Tyr Asn Asn Trp Ala Ala Lys Gly Thr Ser Ala Ser Leu Ser Thr
130 135 140 145

AAG AGC TTC CCC TTC TTC AAC TAC ATG AAC GTC AAC CCC CTG TCA TCA
Lys Ser Phe Pro Phe Asn Ser Met Asn Val Asn Pro Leu Ser Ser
150 155 160

CAG AGC ATG TTT TCC CCA CCC AAC TCT ATC TCG TCC ATG AGC ATG TCG
Gln Ser Met Phe Ser Pro Asn Ser Ser Met Ser Met Ser Ser
165 170 175

TCC AGC ATG GCG CCC TCA GCA GTC AAG GCC TCC CGG GCC TCC AGT TCC
Ser Met Val Pro Ser Ala Val Thr Gly Val Pro Gly Ser Ser Leu
180 185 190

AAC AGC CTG AAT AAC TGG AAC AAG CAG CGG TCG CTG AAT TCC
Aan Ser Leu Aan Aan Leu Asn Leu Ser Ser Leu Ser Leu Ser Aan
195 200 205

GCC GTG CGG AGC CTT GCC TGG GCC CGG AGC CCT CGG TAT
Ala Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Pro Thr Pro Pro Tyr
210 215 220 225

GT TAT AGG GAC AGC TGG TAC CGC CGC ACG CGC AGA CTG AAA
Val Tyr Arg Asp Thr Cys Asn Ser Leu Ala Leu Arg Leu Lys
230 235 240

GCA AAG CAG CAC TCC ACC TTC GCC ACC AGC GTC CAG AAA CGC GCC
Ala Lys Gln His Ser Ser Phe Gly Tyr Ala Ser Val Glu Pro Ala
245 250 255
(2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

Met Glu Thr Aen Cys Arg Lys Leu Val Ser Ala Cys Leu Gln Leu Glu
  1     5     10     15

Lys Asp Lys Ser Gln Gln Gly Lys Aen Gln Asp Val Gly Ala Glu Asp
  20    25    30

Pro Ser Lys Lys Arg Gln Arg Arg Gln Thr His Phe Thr Ser
  35    40    45

Gln Gln Leu Gln Gln Leu Ala Thr Phe Gln Arg Aen Arg Tyr Pro
  50    55    60

Asp Met Ser Pro Arg Gln Glu Ile Ala Val Trp Thr Aen Leu Thr Glu
  65    70    75    80

Arg Ala Arg Aen Gln Ala Ala Glu Leu Cys Lys Asn Gly Phe Gly Pro
  85    90    95

Arg Glu Arg Aen Gln Ala Ala Glu Leu Cys Lys Asn Gly Phe Gly Pro
 100   105   110

Gln Phe Asn Gly Leu Met Gln Pro Tyr Asp Aen Gln Pro Gly Tyr
 115   120   125

Ser Tyr Aen Aen Trp Ala Ala Lys Gly Leu Thr Ser Aen Ser Leu Ser
 130   135   140

Thr Lys Ser Phe Pro Phe Phe Aen Ser Met Aen Val Aen Pro Leu Ser
 145   150   155   160

Ser Gly Ser Met Phe Ser Pro Phe Ser Aen Ser Ile Ser Ser Met Ser Met
 165   170   175

Ser Ser Met Val Pro Ser Ala Val Thr Gly Val Pro Gly Ser Ser
 180   185   190
Leu Asn Ser Leu Asn Asn Leu Asn Asn Leu Ser Ser Pro Ser Leu Asn
195 200 205
Ser Ala Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Pro Thr Pro Pro
210 215 220
Tyr Val Tyr Arg Asp Thr Cys Asn Ser Ser Leu Ala Ser Leu Arg Leu
225 230 235 240
Lys Ala Lys Gln His Ser Ser Phe Gly Tyr Ala Ser Val Gln Lys Pro
245 250 255
Ala Ser Asn Leu Ser Ala Cys Gln Tyr Ala Val Asp Arg Pro Val
260 265 270

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 916 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..914

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

ATG GAG ACC AAC TGC CCG AAA CTG TCG GCC TCG TCT GTA TTA GAG
Met Glu Thr Asn Cys Arg Lys Leu Val Ser Ala Cys Leu Gln Leu Glu
1  5 10 15
AAA GAT AAA ACC CAG CAG GGG AAG AAT GAG GAC GTG GCG GCC GAG GAC
Lys Asp Lys Ser Glu Cys Gln Gly Lys Asn Gln Gly Ala Glu Asp
20 25 30
CGG TCT AAG AAG CGG CAA AGG CGG CAG ACT CAG TAT ACC AGC
Pro Ser Lys Lys Arg Cys Cys Arg Glu Arg Arg Thr His Thr Ser
35 40 45
CAG CAC CTC CAG CTC GAG GCC ACT TTC CAG AGG AAC GCC TAC CGG
Gln Glu Leu Glu Cys Leu Thr Phe Glu Arg Arg Arg Asp Thr
50 55 60
GAC ATG TCC CCA GCC GAA GAA ATC GCT GTG TCC ACC AAC CTT AGG GAA
Asp Met Ser Pro Arg Glu Ile Ala Val Trp Thr Asn Leu Thr Glu
65 70 75 80
GCC CGA GTG CGG GTT TCG TTC AAG AAT CGG GCC AAA TGG AGA AAG
Ala Arg Val Arg Val Trp Phe Asn Arg Ala Cys Gln Gly Asp Cys
85 90 95
AGG GAC GCC AAG CAG CAG GCC GAG CTA TCC AAG AAT GGC TCC GGG CCG
Arg Glu Arg Asn Cys Cys Cys Asp Pro Gly Arg Gly Asp Cys Gly
100 105 110
CAG TTC AAT GGG CTC ATG CAG CCC TAC GAC GAC AGC TAT CCA GCC TAT
Gln Phe Asn Gly Leu Met Glu Pro Tyr Asp Met Tyr Pro Gly Tyr
115 120 125
TCC TAC AAC TCG GCC AAG GCC CTT ACA TCC GCC TCC CTA TCC
Ser Tyr Asn Ser Trp Cys Cys Gln Ser Cys Cys Cys Cys Cys
130 135 140
ACC AAG GCC TCC TCC TTC AAC TCT TAC TCT AAC TCT ATG AAG CAC GCC TCT CTA
The Lys Ser Phe Pro Phe Asn Ser Glu Val Pro Aen Ser Leu Ser
145 150 155 160
TCA CAG AGC AGT TTT TCC CCA CCC TCC TCT ATC TCG TGC AGC ATG
Ser Glu Ser Met Phe Ser Pro Pro Asn Ser Ile Ser Ser Ser Ser Met
165 170 175
TGG GCC ATG GAT GGC TCA GCA GTG ACA GCC GTG GCC TGG GCC TAT
Ser Ser Ser Met Val Pro Ser Ala Val Gly Val Pro Gly Ser Ser
180 185 190
CTC AAC AGC CTG AAT AAC TGG AAC AAC CTG AGT AGC CGG TCG CTG AAT Leu Aan Ser Leu Aan Ser Leu Aan Ser Leu Aan Ser Pro Ser Leu Aan 195 200 205
TCC GGC GTG CGG AGC CCT GCC TGT CCT TAC GCC CGG ACT CCT CGG Ser Ala Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Thr Pro Thr Pro 210 215 220
TAT GTT TAT AGG GAC AGC TGT ACG GAC CTG GCC AGC CTG AGA CTG Tyr Val Tyr Arg Asp Thr Cys Asn Ser Ser Leu Ala Ser Leu Arg Leu 225 230 235 240
AAA GCA AAG CAG CAC TCC AGC TGC TAT GAC TGC CAG CAC TCC ACG Lys Ala Lys Gln His Ser Ser Ser Tyr Phe Gly Tyr Ala Ser Val Gin Lys Pro 245 250 255
GCC TCC AGT CGT AGT CTC TGC TAC TGC TAT CCA GTG GAC CCC CTG T Ala Ser Leu Ser Ser Cys Gin Tyr Val Asp Arg Pro Val 260 265 270

GA 816

(2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1259 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

CGGAGAAGCT GCGAGGAGG AAGARAGAA TGGCAGTGCC GCCTAGATT GGGAGGCCCG 60
AGTCGACCCG TGCAGCTTCG CCGAGCTTGG CACCCTCAGG AGGCTCTCA GAGCTCAGG 120
CCCAGGCTTC GCGGCTTCGC CTTGACCCGG GAGCTGCCTG TGCTAGGAGA TGCCGCCCGG 180
AGGCAGCTGG AGGCGGAGCG ACGAGCTCTC TCCGGTAAGCC CGGAGCTCTG 240
TACGAGGCTC TCTTTCGCTG GAGGTACAGA ACGCCCTTCG TGGCCTCGGG GTCAAGCGAC 360
TGCGTTCTGC TGTGTCGCGAA ATTAGCTAGA AACCCCTCTG TCTGCTGGCG GTACCGGAC 420
CGAGGAGGCG TCTAGAAAGA AGCGAGAGAAG CGAGGAGGCG AAGTATGCAG 480
GCTCCAGCAG CGGAGGAGCA CTGGTCCCCG GAGCCGCTAC CGGAGCAGCG CCCCCCGGG 540
AGGAAATCCT GTGAGGCCA ATCCTACGGA AGCGGCCGCTC CGGCGAGCGG AGCAGGGAT 600
CTGGAAGGCT GTGGCCCTCC TGCGGCCAAC CGCCGCCAGG TTTGGCTCGG GAACTCAGCG 660
GCCAAATGA GAAAGGAAGA GCAGACCGAC GAGCCGGAGC TATGCAAGA TGCGCTCGG 720
CGGACGTCTA ATGCGCTCAT GAGCACGACG ACGAGGCTGA TGGCTACACG 780
AACGACACG CGAAGGCCCT TACATTCCGC TGGCCTACCA CCAAGAGCTC CCCGCTTCTC 840
AACACATGCA AGCAGACCG GCCGCCCCATG GACGACGATG ACCCGCCGTT TGGCTACACG 900
TGGCCTACGA GCAGTCCGAT CAGCAGGCTG CCGCCCTGAG TGCGAGGCTG CGCCGCTCC 960
AGTCTCAACA GGCCTGAATT CTTGAGAACG CAGCTGCCG CTCCTGCTGG TCGGCCGCTG 1020
CGGAGCCCCG CTTGTTCTGG CGGCCCCGCG ACGCGCTCAGT ATGGTATAG AAGACCGGG 1080
AATCGGACG CGGCGACGCC GACGAGCTGG CGCGGCGGGT ACGCGGCTGG CGCGCGCGGG 1140
AGCGTGCAAG CGGCGGCTCTC CAGCGCTGAG CTTGGCCAGT ATGCGAGTGG CCGCGCGGAT 1200
TGAGAGCCAC CGCAGGGGC GAGTCTCCAG GACCTCGGCG GAGGGCGAC TCCGCCCCGT 1260
GAAGAGACGT GATATGCTC AGAGGTCGCG GGGCATCAA GAAAGGAGA GAAAGGAGA 1320
CTTATAGAG AAAAAAGAAC CACGTAACTCA AAGAGAAC GCCTCTGATT TCAAAGGGAT
1380
GTCTCTAGTG TCTGACACTT TTCATCACA GCTATTTCTA CAGGAGTAC AACACATACA 
1440
CAACAAAGTC TTTGCTCGT ATATGACATTT TTAACATTT TATAGCTTTG TTTATTTTTA
1500
AGTTTACGAT GCTTTACATT TAAATGACTG AAGAGATGA TATAATACGA AAATCACAAT
1560
TAATTTAATA AAGCGATGTTG TAGATTTATG CACAGACGTT GAAAAAGGG TTATGTCTTA
1620
AAAAAAGCA TGTATAAGCT AAGGGATAG GTTTTTCTGG TGCCAGACAA GGGAGGTGAT
1680
ATATCAAAATCCACACTCT TTTTTTCTAC ATATATATAT TTTATAAAAG AAGCTGCGAA
1740
TTACCTTTTTT ATGTTACATTT GCCGACACAA CTTATTTTTT ATCTTTTTCA GCAATATAG
1800
AAACAAATGTT TGTTCTTCTTT ATGGGAGTAAAA AATATACGCT ATACACAAAA AAAA
1860
AAA
1920
AAA
1929

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2295 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

CGAGAAGACT GGAGAAGAC AGAAGAAAAC TGGCAATGGC GGTAGATTT GGGAGCGCC
60
AGTCGACCCG TGGACTCTCT CGGAATCTTG CACCCCTCAG AGCCCTGCAG TCTCTCAGG
120
CCCGCTTCCG GCGCCCTTCC CGTCAAGCCG AGCCGCGTGCC TCCCGCTAAA CGGCCGCCG
180
AAAGCGCTGG AGGCGAGAAG AGGAACTCTCC TCCCAGTGCG CCCTAGTGGC AGGCGCTTCT
240
TACACCTCT TCCTTTCTCG TTTTTTCCAG ATCCGGGAAA TGCCACCCCG CTACCACAC
300
CTGCTTGCCTC GTGCTTTGCAG ATATTGTAAG AAGGGCTTCTC TCTGCCCCGG GTCACTCGAG
360
TGGTCTTGAC TTTTCTACGA AGAATATAGA AACCCCTTCTC TCTGCCCCGG GTCACTCGAG
420
CGAGGACCCG TCTAGAAGAG AGGGGCGAAG AGGCGCGCAG ATCTACTTTT CGACCAGAG
480
GCTCCAGAGC CGAGGACGCA CTTTTCCAGAG GAACCCTTAC CGGACATAGT CCAACACGA
540
AGAAATGCTT GTGGTGGCTA AGCCTACGGG AGCCGCGGTC AGCTGGTCAG AGCAACGGAT
600
CTGGAGGCTT GTTCTGCTCC TGGCCCAAC CGCCCCCGAG TTTGCTTCAA GAATCTCGAG
660
GCCAAGTGA AAGAAGGGA CCGCAGCGCAG AAGCGCGGAT TATCGAAGAA GCTCTCGCCG
720
GCCAGTCAGA AATGCTCAAG CGGACCTTAC GACAGACTGT ACCAGCGCTA TTCTCTTACA
780
AATCGGCGGG CCAAGGCGCA TACATCCGCC CTTCTATCCA CCAAGGGCTT CCCCCTCCTC
840
AACCTCAGA AGCGCCACCC CTGTCCATCA CACGAGCTGT TCCCGCCACG CAACTCTAC
900
TCGGCATGA GCATGCGTCT CAAGATGGAG TCCCTACGAG TGACGGCCGT CGCCGCTCCG
960
AGCTTCGCA CAGCTGGAAC CTGGAACGAC CGAGATCGCC TCTGCGCGGT TCTCGCGGCT
1020
CGGACGCCG CGTGGTCCTA CGGGCGCCCG ACTCTCCGCG ATGTGTTATAG GGAACAGCTG
1080
AACCTGAGCG TGCGCAAGCT GCGACTGAAA GGAACGCGC ACTCGAAGCTT CGGGTGCGCC
1140
AGCGGAGCA ACAGGGCGCT CAACCGTACT GCTTTGCAAT ATGCAAGTGA CGCGCCCGCTG
1200
TGACGCGCC CCAACGGCGG CGGATCTAGA ACCTCGCGCT GCCGCGCCCTT CCGCCGCCCTT
1260
GAAGACTCGG GAATTACCT AGAGATGCTG GGGACCTAAA GATAAGGAGA GAAGAGAGAG
1320
CTTTATAGG AAAAGGAAC CACCTGACTCA AAGAGAGAC GCCTTGTGATT CGAAGGGATG 1380
GTCTTCAGTG TCTGATACCT TCTACTACAA GATTTTCTCA AAGCTTGCAAG GACACTACA 1440
CAACAANAGT TTTGCTCTG ATATAGACATT TTAAACATAC CAGGAGCCTTT TTATTTTTTA 1500
AGTTrAGCATT TTTTTACATT TAAAATAGTGG AAGGATGTTA TATAATACGAA AAGGCTAAAT 1560
TATTTTTACT AAACAGCTGG TTTGATATCA CACAAAGTG TTTTTTTAGG TTAGGCTTGA 1620
AAATAAAGCAA TGTTTACAGG AAGGGATAG GTTTTTTCCG TGCGAGCCAA GGGGGTATG 1680
ATACTAAAGG CCAACACCTG TTTTCTCAG ATATATTATATTAAATAAA AAGGCTGGA 1740
TATCGATTTG AGAAAGTTTT GTGGTGTGGA YCGAATGGAG TTTCTGCAAC TGCAATCTAC 1800
AACCTACCTG GTCTCTACAA TCTGAGCAAA TTTTTTTTTT TACTTTTCTCCA GCAAATAGA 1860
AACAAAGTGG TTTTTTTTTT ATGGGAGTAA AATATACTGC ATACAAAAAA AAAAAAAAAA 1920
AAAAAAAAA 1929

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1929 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

CGAGAAACCT GGGAGAAGCC AGAGAGAAGA CAGCAAGTGG CGGCGAGTTT GGGAGCCCC 60
AGTCGACACCC TGGACACTCC CGCAACTCGG CACCGCTAGC AGCCTGCGAG TCTCTAGGGG 120
CCCGCATTCG GCAGCGCGCG CGCTGCAAGG GAGAAGGCCTG CGCTGACAAA TGGCCCCCGG 180
AACGAGCGG CGCAAGCGACG AACGAGCTCC TCCGGTGAGC CGGTAAGTG AGCGCGTCCT 240
TACCACCCT CGTTCTTCTCG CGGTACAGT ACGGGAGAAA TGGAGACCAA CGCGCGCACA 300
CTGTCGTCGCG TGTTTGTGCA CTGTTGCTCG CGAATGATAA AAGATATAGAG GCAAGGGCCG 360
TGGTTCTTCA GTTTCTTCTAG CGAATGAAAG AGATATAGAG AAGAAGATAGG AGCAGGGGC 420
CGAGACCGCC TCTAGAGGCA AAGGGCCAAAG GGGCGAGCGAG ACTCACATTCA CCAAGCAGCA 480
GCCGCAAGCC CGGCGAGGCC CTCTGACAGG GAAAGGTACT CACCGAGTTG CACCGGGCA 540
AGAATACTGC ATGGCAGCA AGCTACACCA AGCGCGAGTC GCAGTGCGAGG AAGAGGACGT 600
CTCGGATGGTC TGGCGCTCC CGGCGAGCGC GGGCGAGCGC TGGGTTTCCAG GAACTCGCAG 660
GCCAATAGGA GAGAAGAGGA GGCGAAGCCG AAGCGCGAGG TATCGCAAGAA GCGTCGCGGG 720
CCCGAGCCTC ATGGACTCGG CACGAGCTTC GAGCAAGTAG AAGAGCGGCTA TTGGTACACA 780
AAGGCGGCGG CCAAGGGGCCT ACATCTGCAGG CTTCTATAAG CCCAGGCTTCCCCGTTC 840
AATCTCATAG ACGCAGAACC CGCTGCACTAC CAGGAGATGG TTCCCGCCAC AACCTCTATT 900
TCTGCTATGC GCAGTGCAGC CGAGCTGGTG CCGTCAGGAG TGCCAGGCGGC CCGCGGCCGC 960
AACCTCAACA GCCTGAAATTT AATGACACAC TGGAGTACCC GCAGCGTGGG GTGCAGGCCT 1020
CCGCCGCGCT GTGGCGCTTT CGGGCGCCGCG ACTCTCCGCG AGTTTATAGA GAACAGCGTG 1080
AACGNCGCCG TGCCAGCCGT GAGACTGAAAG CAAAGAGCGC AACCAGCTCG GCAGTCAGGGC 1140
AGCAGTGCAGA AACCGGCCTC CAGACCTGATG CTGCAGCGAT ACGACAGTGG CCGCGCCCGG 1200
TGCGCCAGCC CGCAAGGCGC GGGATCGTAG AATCCCGCCG CTGGGCGCCCA CTCCCGCGCTT 1260
GAAGAGCTGG GAGTATGCT AGAGGCTGGG GGGACACTAA GAAAGGAGAA GAAGAGAAG 1320
-continued

CTATATAGG AAAAGGAAAC CACGUAATCA AAGAGAAGGC TCCCTTGATT TCAAGGGAGT

GTCTTCAGTG TCTGACATCT TTAGCAAAAG GTATTCTAA CAAGTGCAGA CACAGATAA

CACAACAAAT TTTGCACTGC ATAGACATT TTAACATTATTA AAAAACTTTGTA

GATTTGATCATG TAGTATGTTG TATATAGATG AAGAAAGGA TATATACGAA

TAATTTATAA AAGAAGTTG ATGACATGT TTTATTTAATA CACAGACTG TTTTAGAAA TACACCTTGA

AAATAAAGCA TGTATACAG AAGCCAGTAG GATTTTTGCG TGCCGAGCAAG GGGAGTAT

ATACAAAAGC CACACTGTA TGTATCTAC ATATATTATAT TATTATAAAA AAGGCTGGA

ATAAGTTTTC AGAAAGTTTGTGTTCTGAGA TCAATGGATG TTTCTGAAAC ATCCTATAC

AACCTTACCT GTGTATAACA TTTAGAACAA TGGTAATTTT TACTTITTA CGAAGATAG

AACAATGTTG TTTATTTTTATGGAGATCAA ATATATGUG ATACAAAAAAA AAAAAAAA

AAAAAAA

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1775 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 234..1047

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

CSGAGAAAGCT AGAGAGGACG AGGAAAGAAC TCCCAAGTGCG GCCTGATTTG CGGAAAGGCCC

AGTGCACCCCG TGAGCTCTTC CGCGACTGGGG CGCCCTTCAAGAG CACCCGCGAG TCCCTTACG

CCCCGCTTTCGGGCTCCATGGGC ACGCCGACGAG GCGCTCGTCC TCCCGTACGC CGGATACG

GAG ACC AAC TGC CGC AAA CTG GTC TCG GCC TG TGT TGT CAA TTA GAG AAG
Glu Thr Aaa Cys Arg Lys Leu Val Ser Ala Cys Leu Gln Leu Gln Leu Lys

5 10 15

GAT AAA AGC CAG CGG GGG AAG AAT GAG GAC GCT GCG GCC GAC GCC
Asp Lys Ser Glu Gln Gln Gly Lys Aaa Glu Arg Val Gly Ala Gln Aaa

10 15 20 25 30

TCT AAG AAG CGG CAA AGG CGG CAG ACT ACC ATG TAC AAC CAG
Ser Lys Lys Arg Ala Arg Arg Arg Arg Thr His Thr Ser Gln

35 40 45

CAG CTC GAG CAG CTG GAG GCC ACT TTC CAG AGG ACG GCC TAC CGG GAC
Gln Leu Gln Gln Glu Ala Thr Phe Gln Arg Aaa Arg Tyr Pro Aaa

50 55 60 65

ATG GTC ACA GCC GAA GAA AAT OCT GTC TGG ACC ACT CTT ACG GAA GCC
Met Ser Thr Arg Gln Glu Ile Ala Val Trp Thr Ser Thr Thr Glu Ala

70 75 80

CGA GTC CGG GGT TGG TTC AAG AAT CTT CGG GCC AAA TGG AGA AAG Arg Val Arg Val Thr Phe Lys Aaa Arg Pro Ala Lys Arg

85 90 95

GAG GCC ACC CAG CGG GAC GTA TGG AAG AAT GCC TTC GGS CGC CAG
Glu Arg Aaa Glu Gln Ala Leu Cys Lys Aaa Gly Phe Gly Pro Gln

100 105 110

TTC ATG GGC CTC ATG CAG CCC TAC GAC GAC ATG TAC CCA GCC TAT TCC

115 120 125 130 135
(2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

Met Glu Thr Asn Cys Arg Lys Leu Val Ser Ala Cys Leu Gin Leu Glu

Lys Asp Lys Ser Glu Glu Gly Lys Asn Glu Asp Val Gly Ala Glu Asp 20 22 25 30
Pro Ser Lys Lys Lys Arg Glu Arg Arg Glu Arg Thr His Phe Thr Ser 35 36 40 45
Gln Glu Leu Glu Glu Leu Glu Ala Thr Phe Glu Arg Asn Arg Tyr Pro 50 54 55 60
Asp Met Ser Thr Arg Glu Glu Ile Ala Val Trp Thr Asn Leu Thr Glu 65 70 75 80
Ala Arg Val Arg Val Trp Phe Lys Asn Arg Pro Ala Lys Trp Arg Lys 85 90 95
Arg Glu Arg Asn Glu Glu Ala Glu Leu Cys Lys Asn Gly Phe Gly Pro 100 105 110
Glu Phe Asn Gly Leu Met Glu Pro Tyr Asp Met Tyr Pro Gly Tyr 115 116 120 125
Ser Tyr Asn Asn Trp Ala Ala Lys Gly Leu Thr Ser Ala Ser Leu Ser 130 135 140
Thr Lys Ser Phe Pro Phe Asn Met Asn Val Asn Pro Leu Ser 145 146 150 155 160
Ser Gln Ser Met Phe Ser Pro Asn Ser Ile Ser Ser Met Ser Met 165 170 170 175
Ser Ser Ser Met Val Pro Ser Ala Val Thr Gly Val Pro Gly Ser Ser 180 185 190
Leu Asn Ser Leu Asn Leu Asn Leu Asn Ser Pro Ser Leu Asn 195 196 200 205
Ser Ala Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Pro Thr Pro Pro 210 215 220
Tyr Val Tyr Arg Asp Thr Cys Asn Ser Ser Leu Ala Ser Leu Arg Leu 225 230 235 240
Lys Ala Lys Glu His Ser Ser Ser Phe Gly Tyr Ala Ser Val Glu Lys Pro 245 250 255
Ala Ser Asn Leu Ser Ala Cys Glu Tyr Ala Val Asp Arg Pro Val 260 265 270

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 816 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..814

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

ATG GAG ACC AAC TOC CCG AAA CCG CTC TCG TCG OCG TOT CTC CAA TTA GAG
Met Glu Thr Asn Cys Arg Lys Leu Val Ser Ala Cys Leu Glu Leu Leu 1 5 10 15
AAA GAT AAA ACC CAG CAG CAG GAG AAT GAG GAC GTO GCG GCC GAC GAC
Lys Asp Lys Ser Glu Glu Gly Lys Asn Glu Asp Val Gly Ala Glu Asp 20 22 25 30 30
CCG TCT AAG AAG CCG CAA ACG CGG ACT CAC TTG ACC ACG
Pro Ser Lys Lys Arg Glu Arg Glu Arg Thr His Phe Thr Ser 35 40 45
CAG CAG CTC CAG CAG CAG CAG GAC ACT TTC CAG ACG AMC CTC CCG 192
Gln Gln Leu Gln Gln Leu Ala Thr Phe Gin Arg Aen Arg Tyr Pro
50 55 60
GAC ATT TCC ACA GCC GAA GAA ATC GCT GTG TGG ACC ACC CTT AGC GAA
Ase Met Ser Thr Arg Glu Glu Ile Ala Val Trp Thr Aen Leu Thr Glu
65 70 75 80
GCC GAA GTC CCG GTT TGG TTC ARG AGT CGT GCC GCC AAA TGG AGA AAG
Ala Arg Val Arg Val Trp Phe Aen Arg Pro Ala Lys Trp Arg Lys
85 90 95
AGG GAC GGG CAC CAG CAG GCC GGG CTA TGC RAG ART GCC TCC GGG CCG
Arg Glu Arg Aen Gin Gln Ala Ala Leu Cys Lys Aen Gly Phe Gly Pro
100 105 110
CAG TCC AAT GGG CTC ATT CAG CCC TAC GCC GAC AGT TCA CCC CCA GCC TAT
Gln Phe Aen Gly Leu Met Gin Pro Tyr Asp Met Tyr Pro Gly Tyr
115 120 125
tcc tac aac AAC TCG GCC GGC ACG CTT ACA TCC GCC TCC CTA TCC
Ser Tyr Aen Aen Trp Ala Lys Leu Thr Ser Ala Ser Leu Ser
130 135 140
acc aag ggc ttc TTC TCC TTC TTC AAC TCC ATT AAC GTC AAC CCG CTT TCA
Thr Lys Ser Phe Pro Phe Aen Ser Met Aen Val Aen Pro Leu Ser
145 150 155 160
tca cag agc agt tgg ttc CCA CCC CCC TCT ATC TCG ATG AGC ACG
Ser Gin Ser Met Phe Ser Pro Aen Ile Ser Ser Met Ser Met
165 170 175
tcg tcc aag ggg ctc tga cca gat gca gtc tcc ggc ggg gcc tcc agt
Ser Ser Ser Val Pro Ser Ala Val Thr Gly Val Pro Gly Ser Ser
180 185 190
tcc acc ctg atg aac tgg AAC TGG AAC AGT AGG CCG TCG CTG ATG
Leu Aen Ser Leu Aen Leu Leu Ser Ser Pro Ser Leu Aen
195 200 205
tcc gcg gtt cgg acc cct GCC TGG CCT TAC GCG CCG CAG CTT CCG Ser Ala Val Pro Thr Pro Ala Cys Pro Tyr Ala Pro Pro Thr Pro
210 215 220
tat gtt tat agc gac acc tgg AGC ACC ATC GCG GCC ACC AGT AGG
Tyr Val Tyr Arg Asp Thr Cys Aen Ser Ser Leu Ala Ser Leu Arg Leu
225 230 235 240
aaa gca aag cac cca TCC GGC GCC TAC GCC AGC GGG CAG AAC CCG Lys Ala Lys Gin His Ser Ser Phe Gly Tyr Ala Val Gin Lys Pro
245 250 255
tcc acc ctg agt CCT TCC CAG TAG gca gtg gac cgg ccc gtc T
Ala Ser Aen Leu Ser Ala Cys Gin Tyr Ala Val Aen Arg Pro Val
260 265 270
GA
816

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1929 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

CGAGAAAGCT GGAGGGAGCC AGAGGAAGAC TGGCAATGGCC GGTAGATTG CGGAGGCCC
60
AGTGACCGG TCCGACTCCG CCGACTGCGG CACGCTCCG AGCCGCCTC TCGCTCTAGG
120
CCCGCTCTTC CCTCGGCTGG CCGCGCTGGG CGCGCTCGG TGCGTCGAAA TGCGCGCCGGG
180
AACGCGCTCC AGCGCGCCTC ACGCGCTCCG TCCTCGCCGA CGCTCATGTG AGCGCGCTCC
240
TACGCGCTCC CTTCTCGCTAG TTTGCGACG ATACGGGCGA TGGAGAACCA CTGGCGCGA
300
CTGCTGTCG CGTCTGTCGA ATTAGGTAAG AACCCCCCTTC TCTCGCCCCGG GTCACCCGAC 360
TGCTTTTCCG TTTTCGAGAG AAGATTAAGAAC CAGGAGGAGG AGGATGGAGG AGGTGGGCGC
420
CGAGACCGC CTCTAGAAGAAC CAGCCGAAAG GCCCGACCCG ACTGACTTGA CCAGCCACAA
480
GCTCCAGCG CTGAGCCGCCT CAAGCCGCTAG ACAGGAGTG TCCCCCTGGAG 540
AGAATGCTCCTG GTGAGCGCCAA CTACATGGA CAGGCTGCTG CGGGAGGCTG 600
CTGGGAGCTC GTTGGCTCCCT TTTGCCCCAAC CGGGGCGCTG TTTGGTCACA GAATCGTCCG
660
GCCAAATAGGA AAAAGGAGGA GCGAAACCAG CAGGGCGACG TATGACGAGAA TGCTGTCGGG
720
CCCGAGTTCA ATGGGCTCAT GCAGCCCTAC GAGGACTAGT ACCAGGCTA TTTCTACAC
780
AACTGGGCG CGGGGGGGCT TACAATCCGC TTCCTATCCCA CCAAGATCTT CCCCTTCTCC
840
AACCTATAGA ACAGCACACC CTTGCATGCA CAGAGACAGT TTTCCCATCAC ACCACACTAT
900
TGCTCCAAGA GCATATGCGT CAGACGCTGT CCCCCAGCAG TGACAGGCTG CCGGGCCGCC
960
AGCTCATCAA GCTGGGAAATCG TAGGAGCAAC CTGAGATAGGC COTGGGCTGA TTTGGGCGAT
1020
CCGCGACGTC CTCTGCTCCTA CCGCGCCTCG ACTGCTGGCT ATGGTTATAG GAGAGCTGTG
1080
AACCAGGGCC TGGCAGGCTG GGATGCTGAA GCAAAGCGCC ACTGACGCTT GCCTGAGCC
1140
AGGCTGCAGA AAGCGCGGCG CAAACTGACT GCTGCTGAGT ATGCGAGGCA CGGGCCGGTG
1200
TGACCAGCC CAGAAGGGCC GGGATCGGTG GACCTGCGCG GATGGGGCGA CTGGGCCCCTT
1260
GAGAGCTGGT GAGATTGAGC GAGAGCTGTT GGGACTAAAG GAAAGGGAGA GAAAGAGAAG
1320
CTAATATAGA AAGAGGAACG CAGTCTGCTA AAGAGACAGG TCGCTCTGTG TCAAGGGATT
1380
GCTCCAGGCT GTGAGGCTCG AGATATCGAA TGTATGCAGA CAGTGGCAAG CAGACACATA
1440
CAACACAAAG TTTGGACTGG ATAGACATT ATACACTTAC TAAACATTAT CAAATGCTGCT
1500
AGTGTACCT TAGTACATC TAAATGCATG AAGGATGTA TATATACCGA AATGCCTAAAT
1560
TAATTTTATTAA AAGGATGCTG TAGATATATCAT CAAACAGAGC TTTTTAAAAA TTAGGGCTCA
1620
AANAAAGGAG TAGTCTATAGA CAGGGAGTAT TTTCCCCGCG TGGCGAGCCAG GGGAGTGATAT
1680
ATACAAAGCT CGACATGATG TTTTCTTAC ATATTATAT TATTAATAA AATGTGCGAA
1740
TATCATCTTT TTAGGAGTTT GTGTGGTGGG TGCAGTGATT TTCTGGCAGA ATCTAGTCTAC
1800
AACCACCTG GCTGAAACAAT TATATAGGTT TACTTCTCCCC GCAAATAATGC
1860
AACAATGTCG TTTTATTTTC ATGGGACTAA AATATCTGCG ATACAAAAAAA AAAAAAAAAA
1920
AAAAAAAAAAAAAAAAAAAAAAAA 1929

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1775 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 234..629

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

CGAGAAAGCT CGAGAGGAGC AGAAGAAAGAC TGCCAGGCGC GGGTAGATTG CGGAGGGCGC
60
AGTGCACCGG TGAGCTCTGT CCGACTTGGG CACCTGAGG CAGCCGGCCAG TCTGCTCAGG
120
CCGCGCTTCC GCGGGCTCGC COTGGACCCGC CAGGGCTCGC TGGCTGGAAA TCGCCCGCGG
180
-continued

AACGAGTGGG ACGCGGAGAC AGACGCCTTC TCCCGTGAAG CGATAACGGG GAA AGG
Met
1

GAG ACC AAC TGG CCC AAA CTT GTG TCG GGG TGT CTG CAA TTA GAG AAA
Glu Thr Aan Cys Arg Lye Leu Val Ser Ala Cys Leu Gln Leu Glu Lye
5 10 15

GAT AAA AGC CAT CGG GAG AAT GAG GAC GTG GCC GAC GAC CGG
Asp Lys Ser Gin Gin Lgy Lys Aen Glu Asp Val Gly Ala Glu Asp Pro
20 25 30

TCT AAG AAG AGG CGG CAA AGG CGG ACT CAC TTT ACC ACC CAG
Ser Lys Lys Lys Gin Aen Gin Gin Gin Thr Gin Thr Gin Thr Ser Gin
35 40 45

CAG CAG CAG CAG GAG GCC ACT TAC CAG AGG AAC GCC TAC CGG GAC
Gln Leu Gin Gin Leu Gin Ser Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin
50 55 60 65

ATG TCC ACA GCC GAA GAA ATC GGT TGG ACC AAG CCT ACG GAA GCC
Met Ser Thr Arg Glu Glu Ile Ala Val Thr Aen Leu Thr Gin Thr Gin
70 75 80

CGA GTC CGG GTT TGG TTC AAG GAT CGT CGG GCC AAA TGG AGA AAG AGG
Arg Val Arg Gin Val Trp Phe Lys Aen Arg Gin Gin Gin Gin Gin Gin
95 100 105 110

GAG GCC AAC CAG CAG GCC GAG CTA GTC AAG AAT GGC TTC GGG CCG GAC
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
125 130 135 140

TAC AAC AAG TGAGCGGCCA ACGGCTTAC ATCCGCCCTCC CTATCCACA
Tyr Aan Aan
135 140 145

AGAGCCTCCC CTCTCCAAC TCTATGAAA CTAACCCCTCT GTCATCAGAC AGCA GTTGG
729

CCCCCCTCAA CTCTATCCCG TCCATGAAAA CTCACTACGA CAGGATCCTG CCGGGA
709

CAGGCTCCCC GCGGCTCAGT CTCAGACGGA TGAATACCT GAGCAACTGT AGTAGGGCCG
849

CGCCTAGATC CAGGGCCTCC ACGCTCGTCC GGTGGGCTG CCTGGGCA CTGGCTTATG
909

TTTTATAGGGA CACCTGGTAC TCGAGGCGG CGCAGGGA AGCTAGGAA ACGAGCAC
969

CCATTCCGG TCTAGGGCC ACGGAGAAAA CTGGCTCAAC CAGGGTCCC CAGGGGCTA
1029

CAGGCGCG CCGGCTTGGA GGCGCACCAC CAGGGGCGG ATCCTTAGAG CTTGCGCATG
1089

GGGGGAATCC CGGGCGTGAA AGACTGGGAA TTAGCTAGA AGGTCTGGG CACTAAAGAA
1149

AGAGAGAAG AAGAGACTGA TTAGAGAAAA AGGAAGCAAC TGAATCAGAA AGAGAGC
1209

TTTATTCTCA AAGGATTGCA CTCCATCTTG GACATCTTC ACTGAGAA TTTCTACAG
1269

TTCAGGAC ACYACACAA ACAATUTTT TACTGGTA TACATTTTCA ACAATTTTTA
1329

AAGCTGTGA TTTTTATAGT TGGCATTGTT AATGATTAA AYCAAGAAA GGATATAT
1389

ATAGCAAAAT GCAAAATTTTTAGAAAA GCTGTTTGA GTAATCAA AACAAGTTTT
1449

TTAAGGTTA GCCGTTAAAA TAAAGAGTCA ATACAGAGAA GTAAATTGTT TTGCGCTCTG
1509

CGAGAGAAGG AGGTAGATTA CTAAACGCCA CACTGAGTGT TCTTACATT TAATATTAT
1569

TATATTTAAG AAGCTAGATC TCGTATTGCA ATAGATTTTG TATGAGCTATG AATGATT
1629

CTTCAACCTG TATAGCAAC CTAGGCTTTG TATACATTT GCTAGAATAT TATTGTATTT
1689

CTTTACACA AATAGGAAAA AATGTGTTTT TATTTTACAG GGAATAAAA ATACAGCAT
1749

CAAAAAAAAA AAAAAAAAA AAAAAA
(2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 132 amino acids
   (B) TYPE: amino acid
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

1  Met  Glu  Thr  Asn  Cys  Arg  Lys  Leu  Val  Ser  Ala  Cys  Leu  Gln  Leu  Glu
   5
20 Lys  Asp  Lys  Ser  Gln  Gln  Gly  Lys  Asn  Glu  Asp  Val  Gly  Ala  Glu  Asp
   25
35 Pro  Ser  Lys  Lys  Arg  Arg  Gln  Arg  Thr  His  Phe  Thr  Ser
   40
50 Gln  Gln  Leu  Gln  Gln  Leu  Gln  Ala  Thr  Phe  Gln  Arg  Arg  Arg  Tyr  Pro
   55
70 Asp  Met  Ser  Thr  Arg  Glu  Glu  Ile  Ala  Val  Trp  Thr  Asn  Leu  Thr
   75
85 Ala  Arg  Val  Arg  Val  Trp  Phe  Lys  Asn  Arg  Arg  Ala  Lys  Trp  Arg  Lys
   90
100 Arg  Glu  Arg  Asn  Gln  Gln  Ala  Glu  Leu  Cys  Lys  Asn  Gly  Phe  Gly  Pro
   105
110 Gln  Phe  Asn  Gly  Leu  Met  Gln  Pro  Tyr  Asp  Asp  Met  Tyr  Pro  Gly  Tyr
   115
130 Ser  Tyr  Asn  Asn

(2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 139 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
   (A) NAME/KEY: CDS
   (B) LOCATION: 1..396

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

ATG GAG ACC AAC TGC GCC AAA CTG GTG TCG GCC TCG CTG CAA TTA GAG
   1 5 10 15
AAA GAT AAA GGC CAG CAG GGG AAG AAT GAG GAC GTG GCC GCC GAG GAC
   20 25 30
Lys Asp Lys Ser Gln Gln Gly Lys Asn Glu Asp Val Gly Ala Glu Asp
   35
CGG TCT AAG AAG CCG CAA AGG CAG CCG ACT CAC TCT ACC ACC AGC
   40 45 50 55 60
Pro Ser Lys Lys Arg Gln Arg Arg Thr His Phe Thr Ser
   65
CAG CAC CTC CAG CAG CAG GCC ACT TTC CAG AGG AAC CCC TAC CCG
   70 75 80
Gln Gln Leu Gln Leu Gln Ala Thr Phe Gln Arg Arg Tyr Pro
   85
GAC ATG TCC ACA GCC GAA AAT ATC GCT GTC TGG ACC AAT CTT AGC GCA
   90 95 100
Asp Met Ser Thr Arg Glu Glu Ile Ala Val Trp Thr Asn Leu Thr Glu
   105
GCC CGA GTC CCG GTT TGG TTC AAG AAT CGT CCG GCC CCC TGG AGA AAG
   110 115
Ala Asp Val Arg Val Trp Phe Lys Asn Arg Arg Ala Lys Trp Arg Lys
   120
### DNA Sequence

```
AGG GAG GCC AAC CAG CAG GCC GAG CTA TGC AAG AAT GGC TTC GGG CCG
Arg Glu Arg Aan Gln Gln Ala Glu Cys Lys Aan Gly Phe Gly Pro
100 105

CAG TTC AAT GGG CTC ATG CAG CCC TAC GAC ATG TAC CCA GGC TAT
Gln Phe Aan Gly Leu Met Gln Pro Tyr Asp Asp Met Tyr Pro Gly Tyr
115 120 125

TCC TAC AAG AAC TGA
Ser Tyr Aan Aan
130
```

### Sequence Characteristics:
- **Length:** 1929 base pairs
- **Type:** nucleic acid
- **Strandedness:** single
- **Topology:** linear

### Molecule Type:
DNA (genomic)

### Sequence Description:
```
CGAGAAGGCT GGGAGGGAC AGAAGAAGAC TGCCAGTGGC GGGATGATTT CCGAGGCCCG
60

AGCTGACCCG TGGACTCCCT CGGACTCTGG CAGCCCTAGG AGGGCTGACG TCCGGTACGG
120

CCGGCTTTC GGGCGCTTCG CGTGGCGAGC GAGGCTCCGC TCGTCTGAAA TGGGGCCCCGG
180

AACGATGCG ACGGCGAGAC ACGACCTCTC TCGGCTGGAG CGTATGAGTC AGGGCTCTTC
240

TACGACCTTC CTGTTCTCTG TTTGCGATAG ATAGGGGAA AAGAGGACCA CCGTCGGAAA
300

CTGCTCTGG CCGTGCGGCA ATAGAGTATG ACGCCCTGTC TCAGCGCAGC TGCTGAGCG
360

TGCTTTACAG TTGCTCAGAA AAGTAAACCA CACGAGGAGG AAGAATGAGG ACGGCGGCG
420

CGAGGACGAC CTGAGGAGCA CTCTCCAGAG ACGCAGCTAC CCGAGCAGTGC CAGACGCGA
480

AGAATTCCCT GTGCTGGACCA ATCTTACGGA AGCCGGACTG CGGGCTGGAC AGGGCGAGCT
540

CTGGAGACGC TCTGGGTCCCG TTGGCCCCGAC CGGGCCCGAG TGGGGTGCAA GAATCTGGCG
600

GCCATAGAGA GAAAGAGGGA GGCAAGCCAG GAGCGGCGAT TATGGAAGGA TGCCGGTGGG
660

CCGGACTTGA ATGGCTGACT CAGCCGGTAC GAGCAAGATG ACCAGACGTT TCCCTAACAC
720

AACGAGCGCG CAGGGCGGCTC TACTACGGCC TCCCTATGCC CAGAGCGCTT CCCCGCTCTC
780

AACCTATGCA AGCTCAACCC CTTGCTATCA CAGACAGTGT TGGGGCCACC ACACCTATAC
900

TCCTCCAGAA GCTGTTCTGCA CGAGGAGGTT CCGCGAGGCTG AGGCTGGAGG GAGAGCTGAT
960

AGCTCTAACG GGTCCGATTA CTTGACCAAC CTGGATGACC CTTGACGGAAG TGGGGCGGCT
1020

CGGAGCGCTG CTTGTGCCTTA CGGCGGGCCCG ACCTGCGCTT AGTGGGATAG GAGAAGGTGC
1080

AACCTGAGGG CCGCGAAGCC CGAGCTGAGGA CGAGCTAGCT CAGCGTAG AACGAGCCGG
1140

AGCTGACGAGA ACGGCGGGTC CGACCTGATT GGTGGCAGT AGCGGCTGGC CGGAAGCCTG
1200

TGAGGGCGAC CGAACCGGCG GGATCTCAGG GACCTGCGCC TGCCCGCCCT
1260

GAGAAGCTGG GAAATTGCCT AGAAGAGCTG GCG CCTGAAA GGAAGAGAA GAAAGAGAGA
1320

CTATATAGAG AAAGAAGAAC CACTGAATCA AAGAAGAGGC TCTTCTGATT TCAAGGGAT
1380

GCTGTCAGGG CTGAGACATC TCTACTATCA GATATTTCAAA CAGTTGCGAG GAGCATAACA
1440

CAACAAAGGT TTTGACTGG ATATGACTT TTAAGTATT TATAAGCTTTG TTAATTGTAA
1500

AGTTTACAGT TTTTACAATC TAAATGACTG AAGAGTAGTA TATTATCGA ACGTCACAA
1560

TAATTATTAA AAGAGCGTTG TTTGATATAT CACAGACGTTG TTTTTAAGG TTAAGGCTTA
1620
```
AAATAAGCA TGTTTACAG AAGGCAATG GATTATAACG TTTGAGCAA GGAGATGAT 1680
ATACTAANG CGACACCTGA TGTTTCGAC ATATATTAT TATTAAAAA AAGTGGTCAA 1740
TACTAAGTGG GCTGTTGAG TGCTGATGAG TTTCTGAAAC TCTCAGTAC 1800
AAACCTACCT GTGTTAACA TTTCGTACAA TATTATGGT TTAATTTTCA GCAAATAGA 1860
AACAAATGCT TTATTTTTTATGGAATAA AATATACTAC ATACAAAAAA AAAAAAAA 1920
AARAAAAA 1929

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 484 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS  
(B) LOCATION: 281..484

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:
CACCAGCGCG CTCCCTAAAG AGCTCTCGAG CGTTTCCCTC CTCCCCACGT CCCACCTCT 60
GCTTGGGCTC CCTCTCTCTC GCGGCTGGAA ACTTTTTTGG GTGAGAGCT GAGCTAGAG 120
GAGACAGGCG TGAGATTGTGCG GCTTTCTGGC TCTTCCAGCT CTGGCTCAC 180
ACGGGACAC TTGGGACTCA GCTGCGCAC GAGAACCGG GGGCGCACTGG 240
CGTATGCCCT CGAATCAGAT GCTCCCGAC TCCCTGCTCC ATG AAC TGC ATG AAA 295
Met Asn Cys Met Lys 1 5

GCC CGG CTG CCC TCG GAG CAC GCA GCC GCG ACT AAG CTG TGC GCC 343
Gly Pro Leu Pro Leu Glu His Arg Ala Ala Gly Thr Lys Leu Ser Ala 10 15 20

GCC TCC TCA CCC TTC GCT CAC CAT CCC CAG GCG TTA GCC ATG GTT TGC 391
Ala Ser Ser Pro Phe Cys His His Pro Gln Ala Ala Met Ala Ser 25 30 35

GTC GCA GCT GCC CAT GCG CCC CCC CCC CCC TGG GAC TCC TCC AAA CAT AGA 439
Val Leu Ala Pro Gly Glu Pro Ser Leu Asp Ser Ser Lys His Arg 40 45 50

CTG GAG GTC GAT ACA TCC CCC CCC CCC TCG CAT ACT TCC ACC CTT GAA TGC GCA 484
Leu Glu Val Val Thr Thr Ser Asp Thr Ser Ser Pro Glu Val Ala 55 60 65

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:
Met Asn Cys Met Lys Gly Pro Leu Pro Leu Glu His Arg Ala Ala Gly 1 5 10 15

Thr Lys Leu Ser Ala Ala Ser Ser Pro Phe Cys His His Pro Gln Ala 20 25 30

Leu Ala Met Ala Ser Val Leu Ala Pro Gly Glu Pro Arg Ser Leu Asp 35 40 45
What is claimed is:

1. An isolated nucleic acid comprising the nucleotide sequence that is at least 95% identical to SEQ ID NO: 1 or 3 and encodes a protein that is a homebox regulator of gene expression.

2. The isolated nucleic acid of claim 1, comprising a nucleotide sequence that is at least about 98% identical to SEQ ID NO: 1 or 3.

3. The isolated nucleic acid of claim 2, wherein the nucleotide sequence is at least about 99% identical to SEQ ID NO: 1 or 3.

4. An isolated nucleic acid that hybridizes to the full complement to the nucleic acid consisting of SEQ ID NO: 1 under stringent hybridization conditions comprising 60x SSC at 45° C, followed by a wash step in 0.2xSSC at 50° C, wherein the nucleic acid encodes a protein that is a homebox regulator of gene expression.

5. The isolated nucleic acid of claim 4, which hybridizes to the full complement of the nucleic acid of SEQ ID NO: 1 or 3 under stringent conditions comprising 60xSSC at 45° C, followed by a wash step in 0.2xSSC at 65° C.

6. The isolated nucleic acid of claim 4, which encodes a polypeptide comprising the amino acid sequence that is set forth in SEQ ID NO: 7.

7. An isolated nucleic acid of claim 1, which is genomic DNA.

8. An isolated nucleic acid of claim 1, which is cDNA.

9. An isolated nucleic acid of claim 1, which is RNA.

10. An isolated nucleic acid of claim 1, further comprising a label.

11. A delivery complex comprising a nucleic acid of claim 1 and a targeting means.

12. A delivery complex comprising the nucleic acid of claim 1 and a targeting means.

13. An expression vector, comprising a nucleic acid of claim 1 operably linked to a transcriptional regulatory sequence.

14. An expression vector, comprising the nucleic acid of claim 1 operably linked to a transcriptional regulatory sequence.

15. A delivery complex of claim 14, wherein the targeting means is selected from the group consisting of: a sterol, lipid, virus or target cell specific binding agent.

16. An expression vector of claim 13, which is capable of replicating in a cell.

17. An isolated cell comprising an expression vector of claim 16.

18. A method of making a RIG polypeptide comprising the steps of:

a) culturing the cell of claim 17 in an appropriate culture medium to produce a polypeptide that is encoded by the nucleotide sequence that is at least 95% identical to SEQ ID NO: 1 or 3 and encodes a protein that is a regulator homebox of gene expression; and

b) isolating the polypeptide.

19. The isolated nucleic acid of claim 1, which encodes a polypeptide comprising the amino acid sequence that is set forth in SEQ ID NO: 7.

20. The isolated nucleic acid of claim 1, which encodes a polypeptide comprising the amino acid sequence that is set forth in SEQ ID NO: 7.

21. The isolated nucleic acid of claim 17, which encodes a polypeptide comprising the amino acid sequence that is set forth in SEQ ID NO: 27.

22. The isolated nucleic acid of claim 19, which encodes a polypeptide comprising the amino acid sequence that is set forth in SEQ ID NO: 27.

23. An isolated nucleic acid that encodes a polypeptide comprising the amino acid sequence that is at least about 95% identical to SEQ ID NO: 2 or 5, wherein the polypeptide is a homebox regulator of gene expression.

24. The isolated nucleic acid of claim 23, which encodes a polypeptide comprising the amino acid sequence that is at least about 98% identical to SEQ ID NO: 2 or 5.

25. The isolated nucleic acid of claim 23, which encodes a polypeptide comprising the amino acid sequence that is at least about 99% identical to SEQ ID NO: 2 or 5.

26. The isolated nucleic acid of claim 23, which encodes a polypeptide comprising the amino acid sequence that is set forth in SEQ ID NO: 7.

27. The isolated nucleic acid of claim 23, which encodes a polypeptide comprising the amino acid sequence that is set forth in SEQ ID NO: 27.

28. The isolated nucleic acid of claim 26, which encodes a polypeptide comprising the amino acid sequence that is set forth in SEQ ID NO: 27.

* * * * *