Complete cDNA Sequence, Genomic Structure, and Chromosomal Localization of the LPA Receptor Gene, \(I_{pA1}/vzg-1/Gpcr26\)

James J. A. Contos* and Jerold Chun†, 1

*Neurosciences Graduate Program and †Neurosciences and Biomedical Sciences Graduate Programs, Department of Pharmacology, School of Medicine, University of California at San Diego, La Jolla, California 92093

Received March 11, 1998; accepted May 19, 1998

The \(I_{pA1}/Gpcr26\) locus encodes the first cloned and identified G-protein-coupled receptor that specifically interacts with lysophosphatidic acid. A murine full-length cDNA of size consistent with that seen on Northern blots (3.7 kb) was determined using 3′ rapid amplification of cDNA ends. Analysis of genomic clones revealed that the gene is divided into five exons, with one intron inserted in the coding region for transmembrane domain VI and one exon encoding the divergent 5′ sequence in another published cDNA clone variant (orphan receptor mrec1.3). This structure differs from the intronless coding region for a homologous receptor, Edg1, but is identical to another more similar orphan receptor (\(I_{pA2}\)) that has been deposited with GenBank. Using backcross analysis, both exons 1 and 4 mapped to a proximal region of murine Chromosome 4 indistinguishable from the vacillans gene. Exon 4 also mapped to a second locus on proximal Chromosome 6 in Mus spretus, and this partial duplication was confirmed by Southern blot. The genomic structure indicates a distinct, divergent evolutionary lineage for the \(vzg-1/I_{pA1}\) subfamily of receptors compared to those of homologous orphan receptor genes. © 1998 Academic Press

INTRODUCTION

Ventricular zone gene-1 (vzg-1/\(I_{pA1}\)^2) encodes the first cloned and identified G-protein-coupled receptor (GPCR)

Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. AF075452, AF075453, AF075454, AF075455, AF075456.

1 To whom correspondence should be addressed at Department of Pharmacology, School of Medicine, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0636. Telephone: (619) 534-2659. Fax: (619) 822-0041. E-mail:jchun@usc.edu.

2 We now refer to vzg-1 as "lysophospholipid receptor A1," \(I_{pA1}\); to additional murine putative LPA receptors as \(I_{pA3}\) and \(I_{pA3}\); and to the murine Edg-1 subfamily of receptors as \(I_{pA1}\) (=Edg1), \(I_{pA2}\) (=H218/AGR16), and \(I_{pA3}\) (=edg-3) (Chun et al., 1998). This simple, consistent nomenclature system is being considered for widespread use at the Mouse Genome Database (Lois Maltais, Bar Harbor, ME, pers. comm., 1998). The names "edg-2" and "edg-3" are confusing because they have also been given to genes that encode a putative transcription factor and nuclear translocation regulator (respectively), identified in the same screen as Edg1 (Hla et al., 1995, 1997). We use uppercase to refer to the proteins VZG-1 and MREC1.3. The approved symbol for the vzg-1 gene is Gpcr26, which we use in the linkage maps.

Complete cDNA Sequence, Genomic Structure, and Chromosomal Localization of the LPA Receptor Gene, \(I_{pA1}/vzg-1/Gpcr26\)

James J. A. Contos* and Jerold Chun†, 1

*Neurosciences Graduate Program and †Neurosciences and Biomedical Sciences Graduate Programs, Department of Pharmacology, School of Medicine, University of California at San Diego, La Jolla, California 92093

Received March 11, 1998; accepted May 19, 1998

The \(I_{pA1}/Gpcr26\) locus encodes the first cloned and identified G-protein-coupled receptor that specifically interacts with lysophosphatidic acid. A murine full-length cDNA of size consistent with that seen on Northern blots (3.7 kb) was determined using 3′ rapid amplification of cDNA ends. Analysis of genomic clones revealed that the gene is divided into five exons, with one intron inserted in the coding region for transmembrane domain VI and one exon encoding the divergent 5′ sequence in another published cDNA clone variant (orphan receptor mrec1.3). This structure differs from the intronless coding region for a homologous receptor, Edg1, but is identical to another more similar orphan receptor (\(I_{pA2}\)) that has been deposited with GenBank. Using backcross analysis, both exons 1 and 4 mapped to a proximal region of murine Chromosome 4 indistinguishable from the vacillans gene. Exon 4 also mapped to a second locus on proximal Chromosome 6 in Mus spretus, and this partial duplication was confirmed by Southern blot. The genomic structure indicates a distinct, divergent evolutionary lineage for the \(vzg-1/I_{pA1}\) subfamily of receptors compared to those of homologous orphan receptor genes. © 1998 Academic Press

INTRODUCTION

Ventricular zone gene-1 (vzg-1/\(I_{pA1}\)^2) encodes the first cloned and identified G-protein-coupled receptor (GPCR)

Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. AF075452, AF075453, AF075454, AF075455, AF075456.

1 To whom correspondence should be addressed at Department of Pharmacology, School of Medicine, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0636. Telephone: (619) 534-2659. Fax: (619) 822-0041. E-mail:jchun@usc.edu.

2 We now refer to vzg-1 as "lysophospholipid receptor A1," \(I_{pA1}\); to additional murine putative LPA receptors as \(I_{pA3}\) and \(I_{pA3}\); and to the murine Edg-1 subfamily of receptors as \(I_{pA1}\) (=Edg1), \(I_{pA2}\) (=H218/AGR16), and \(I_{pA3}\) (=edg-3) (Chun et al., 1998). This simple, consistent nomenclature system is being considered for widespread use at the Mouse Genome Database (Lois Maltais, Bar Harbor, ME, pers. comm., 1998). The names "edg-2" and "edg-3" are confusing because they have also been given to genes that encode a putative transcription factor and nuclear translocation regulator (respectively), identified in the same screen as Edg1 (Hla et al., 1995, 1997). We use uppercase to refer to the proteins VZG-1 and MREC1.3. The approved symbol for the vzg-1 gene is Gpcr26, which we use in the linkage maps.

0888-7543/98 $25.00
Copyright © 1998 by Academic Press
All rights of reproduction in any form reserved.

364
closest related genes are the cannabinoid receptors (Cnr1 and Cnr2; Matsuda et al., 1990; Munro et al., 1993), which share approximately 29% identity with vzg-1. Genomic structures of both Edg1 and Cnr1 have recently been published (Abood et al., 1997; Liu and Hla, 1997). Because both are intronless in their coding region, but have introns in their 5’ untranslated regions (UTRs), one might expect that homologous receptor genes, such as vzg-1, would share a similar intron location pattern. However, a variant murine done termed mrec1.3 (Macrae et al., 1996) is identical to vzg-1 except for the initial 79 bp that diverge from the first 256 bp of vzg-1. This suggests alternate exon usage and a distinct genomic structure, since the putative intron occurs within the coding region of vzg-1.

As it is the first identified member of a growing family of lysosphospholipid receptor genes (Chun et al., 1998), we undertook an analysis of the complete cDNA sequence, genomic structure, and chromosomal localization of vzg-1. We show that additional 3’ UTR sequences accounts for the entire transcript size determined by Northern blot. Five exons account for published cDNA sequences, with a conserved intron located in transmembrane domain VI. The gene is located on mouse Chromosome 4 in Mus musculus (C57BL/6j) and Mus spretus, with a partially duplicated locus on Chromosome 6 in M. spretus.

MATERIALS AND METHODS

Generation and cloning of the 3’ RACE product. A modified 3’ RACE protocol in which a DNA oligonucleotide is ligated to dephosphorylated RNA was followed (Diefenbach and Dveksler, 1995). SK oligonucleotide (Table 1) (5 μg) from pBluescript was phosphorylated with polynucleotide kinase and purified with NucTrap columns according to directions from the manufacturer (Stratagene). Dethosphorylated total RNA from embryonic day 15 (E15) brain was used in the ligation reaction, and unligated oligo was removed using a Millipore100 spin cup (Millipore). The ligated RNA sample (10 μl) was reverse transcribed with 20 pmol of SK’ primer and eventually diluted to 80 μl with TE. The SK’ primer is complementary to SK, but includes two Ts at its 3’ end so only ligated RNAs that have AA at their end (i.e., a poly(A) tail) would act as template. Primary PCR was done using SK’ and 513W in a 50-μl reaction containing 1× PCR buffer (50 mM KCl, 10 mM Tris, pH 8.5), 2 mM MgCl2, 0.5 μM each primer, 0.25 mM each dNTP, and 2 μl of the diluted cDNA. The mix was heated to 98°C for 5 min and then cooled to 75°C for 5 min and 1 U Taq + 0.2 U Pfu was added and cycled 35× at 95°C for 30 s, 52°C for 30 s, 72°C for 3 min, with a final 72°C extension step for 10 min before cooling to 4°C. Secondary PCR was done in the same way, except SK’ and 513Y were the primers and 1 μl of a 1:20 dilution of the primary PCR product was used as template. Part of each reaction (10 μl) was electrophoresed on 1.4% Seakem LE agarose (FMC Bioproducts) gels containing 0.5 mM ethidium bromide (EtBr).

Specific products of the correct sizes were observed for both 513C/513T, 967 bp; 513W/vzg.z3, 1189 bp; and 513C/vzg.z3, 2437 bp. Products from reactions with 513C span exons 3 and 4. The vzg.z3 primer is located 80 bp 5’ of the first poly(A) consensus sequence (which is 1040 bp upstream of the end of the original vzg-1 cDNA clone), 513C is located in the coding region of exon 3, and 513W is located near the 3’ end of the original vzg-1 cDNA clone. Specific products of the correct sizes were observed for both 513C/365

### TABLE 1

<table>
<thead>
<tr>
<th>Primers Used in Polymerase Chain Reactions</th>
<th>513A</th>
<th>GCTATCGAGAGGCACACCATCAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>mrec1</td>
<td>TGCCCTCCTGGGACTGACT</td>
<td>513B</td>
</tr>
<tr>
<td>mrec2</td>
<td>GAGGTACCTTCTGGTCTCAGGA</td>
<td>513C</td>
</tr>
<tr>
<td>SK</td>
<td>CGCCCTGAACTATGGTACT</td>
<td>513D</td>
</tr>
<tr>
<td>SK’</td>
<td>ATCCACCTGCTTCTAGGTTT</td>
<td>513E</td>
</tr>
<tr>
<td>T3</td>
<td>AATATAACCCCTCACAAGGG</td>
<td>513F</td>
</tr>
<tr>
<td>T7</td>
<td>GTAATTCTGACTTATAGGCG</td>
<td>513G</td>
</tr>
<tr>
<td>vzg.is2</td>
<td>TATAGGAGTCTTGTGTTGCCT</td>
<td>513H</td>
</tr>
<tr>
<td>vzg.p10</td>
<td>CCTTCCTTCAGGTTATCCAG</td>
<td>513I</td>
</tr>
<tr>
<td>vzg.p6</td>
<td>CCAGGGTGTGTTGACTCAGA</td>
<td>513J</td>
</tr>
<tr>
<td>vzg.p8</td>
<td>CAGTTCACRTGCTCACCAGA</td>
<td>513K</td>
</tr>
<tr>
<td>vzg.p13</td>
<td>CTTGCATATGTGTATCATG</td>
<td>513L</td>
</tr>
<tr>
<td>vzg.p16</td>
<td>GCCTCTCGAGGAAATGACG</td>
<td>513M</td>
</tr>
<tr>
<td>vzg.z3</td>
<td>AGTCCAAGTCGGGTCCATCAT</td>
<td>513Y</td>
</tr>
</tbody>
</table>

3 Oligonucleotide sequences are listed in Table 1. An Expedite Nucleic Acid Synthesis System (Millipore) was used to generate each oligo. Oligos were either used directly after resuspension in TE or were further purified with 15–20% polyacrylamide gels and SepPak reverse-phase column chromatography (Waters) (Ausbel et al., 1994).

4 Enzymes used are from Boehringer Mannheim with the exceptions of T4 and RNA ligase (New England Biolabs), Taq and SuperScript ( Gibco BRL), restriction enzymes (New England Biolabs), Pfu polymerase (Stratagene), and Sequenase (Amersham). Chemical reagents were purchased from Sigma, with the exception of [32P]dCTP (DuPont) and random hexamers (BM).
Isolation of genomic clones by filter hybridization. Approximately 10^6 plaques from a mouse 129/SvJ genomic library in Lambda FIX II (Stratagene) were screened with a 32P-labeled EcoRI/XhoI 2.2-kb vsg-1 cDNA insert using the Pipes formamide hybridization protocol (Stratagene). Four clones were isolated after secondary and tertiary screens (10a, 11, 12c, and 14a), and their DNA was isolated by a plate lysate method (Auszubel et al., 1994). Restriction mapping was facilitated by the unique NotI and Sall sites that release inserts from the λ phage. Initially, EcoRI, XhoI, HindIII, BamHI, XbaI, BglII, and PstI were used and later several other enzymes. DIGs of each clone (0.2 μg) were electrophoresed on 0.8% agarose gels containing 0.5 μg/mL EtBr, blotted, and probed according to the protocol described below (Southern blots). Probes for various parts of the insert were amplified from the plasmid cDNA using primers 513A/B (ORF-specific), 513K/T3 (5' end), and 513NP (near the 3' end) and vector sequence was digested away and purified using the Qiagen gel-extraction kit (Qiagen). Each of the four genomic clones was positive for the ORF-specific probe (513AB), but not the other two probes from the UTRs. The orientation of exon 3 relative to restriction sites was determined using Splh digestion (an Splh site exists in the middle of exon 3) in combination with other enzymes and probes specific to 5' and 3' parts of exon 3. The library was screened two more times using the 513K/T3 and 513NP probes specific for upstream and downstream exons. While no clones were isolated with the 513NP probe, two were isolated (γ and δ) with the 513K/T3 probe, each containing exon 1. Another attempt at screening with a 3' UTR-specific probe was unsuccessful, so an alternate chromosomal walking strategy was used whereby the library was screened with a digoxigenin-labeled riboprobe transcribed from clone 10a using T7 polymerase (according to the Stratagene protocol for riboprobe hybridization and Boehringer Mannheim protocol for visualization). This screen yielded clones 1, 3, and 6, none of which contained exon 4.

Isolation of genomic clones by PCR. A PCR strategy was used to isolate genomic clones for exons 2, 3, and 4, based on a published protocol (Israel, 1993) with several modifications. Approximately 5000 ng were phage isolated by growing liquid in each of the wells in a 96-well plate as follows: 1 mL of X1-blue (MRA) host cells at Op_{o.d.} = 0.5 in 10 mM MgSO$_4$ was mixed with 1 mL SM containing 10^5 phage from the library, incubated at 37°C for 30 min and then diluted to 17 mL with NZY media, and 100 μL was aliquoted to each well in the plate. The plate was taped shut and placed in a 37°C shaker until the cells were visibly lysed (or a maximum of 8 h). Only polymerase chain reactions that gave a single specific fragment of the expected size from 5 ng of genomic DNA (in a 25-μL reaction) were used. A total of 96 or 43 of the wells were screened directly with reactions specific for exon 4 (using 513GT primers) or exon mrec (mrec4, mrec3, 513K, and 513H). Subcloning and sequencing of genomic clones. Lambda inserts were subcloned to allow larger preparations, further restriction mapping, and templates that could be used for sequencing. Table 2 lists the fragments subcloned and the names of the plasmids containing them. Subcloning was done using an in-gel ligation protocol (FMC Bioproducts, Inc.). DNA (λ or plasmid) was digested and electrophoresed in TAE buffer [0.8–1.0% SeaPlaque GTG low-melt agarose (FMC Bioproducts) gels containing EtBr]. Fragments were excised with a clean scalpel, melted at 70°C, and mixed at 37°C and then T4 ligase buffer, ATP, and T4 ligase were added. After incubation overnight at room temperature, the ligation mixes were diluted two- to threefold and melted at 70°C, and 10 μL was transformed into 100 μL RbCl-competent XL1-blue or DH5α cells using a heat-shock protocol (Auszubel et al., 1994). Colonies were screened either by restriction digestion of minipreps (Auszubel et al., 1994) or by direct PCR after growing overnight. 0.8 μL of grown bacteria was used as template in a 20-μL PCR and cycled 20× with the parameters described for 513GT below. Subcloned genomic fragments were sequenced using the dyeideox methodology with the T7, reverse, or T3 primers that flank the multiple cloning site of pBluescript or with oligonucleotides that were synthesized for the purpose (Sanger et al., 1977). Sequence was read into files using the DNAsis program, and contig maps were created with the same software program. All sequences were deposited with GenBank.

### TABLE 2

<table>
<thead>
<tr>
<th>Donor DNA</th>
<th>Fragment subcloned</th>
<th>Name</th>
<th>Relative location of subclone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clone 10a</td>
<td>5.7-kb EcoRI/NotI</td>
<td>vsg 5.5</td>
<td>Exon 3</td>
</tr>
<tr>
<td>Clone 14a</td>
<td>6.5-kb HindIII/NotI</td>
<td>HN6.5</td>
<td>5' of exon 3</td>
</tr>
<tr>
<td>Clone 3</td>
<td>7.2-kb EcoRI/EcoRI</td>
<td>RN9.0</td>
<td>Exon 3</td>
</tr>
<tr>
<td>Clone γ</td>
<td>5.4-kb EcoRI/NotI</td>
<td>RN4.7</td>
<td>Exon 1</td>
</tr>
<tr>
<td>Clone γ</td>
<td>10-kb EcoRI/NotI</td>
<td>RN10</td>
<td>5' of exon 1</td>
</tr>
<tr>
<td>Clone 4D(2D)</td>
<td>7.0-kb EcoRI/NotI</td>
<td>RN7.0</td>
<td>Exon 2</td>
</tr>
<tr>
<td>Clone 4D(2D)</td>
<td>4.8-kb EcoRI/NotI</td>
<td>RN4.8</td>
<td>Exon mrec</td>
</tr>
<tr>
<td>Clone 1D(3E)</td>
<td>3.0-kb XhoI/NotI</td>
<td>XN3.0</td>
<td>Exon 4</td>
</tr>
<tr>
<td>Clone 1D(3E)</td>
<td>9.0-kb EcoRI/NotI</td>
<td>RN9.0</td>
<td>Exon 4</td>
</tr>
<tr>
<td>Clone 1D(3E)</td>
<td>10-kb XhoI/XhoI</td>
<td>XX10</td>
<td>3' of exon 4</td>
</tr>
<tr>
<td>Clone 1D(3E)</td>
<td>2.0-kb XhoI/NotI</td>
<td>XN2.0</td>
<td>3' of exon 4</td>
</tr>
<tr>
<td>vzg5.5</td>
<td>1.1-kb BglII/NotI</td>
<td>BN1.1</td>
<td>5' of exon 3</td>
</tr>
<tr>
<td>vzg5.5</td>
<td>1.2-kb XbaI/XbaI</td>
<td>XX1.0</td>
<td>5' of exon 3</td>
</tr>
<tr>
<td>vzg5.5</td>
<td>1.0-kb XbaI/XbaI</td>
<td>XX1.0</td>
<td>5' of exon 3</td>
</tr>
<tr>
<td>vzg5.5</td>
<td>0.6-kb XbaI/XbaI</td>
<td>XX0.2</td>
<td>5' of exon 3</td>
</tr>
<tr>
<td>vzg5.5</td>
<td>0.2-kb XbaI/XbaI</td>
<td>XX0.2</td>
<td>5' of exon 3</td>
</tr>
<tr>
<td>HN6.5</td>
<td>1.2-kb XbaI/XbaI</td>
<td>XX1.5</td>
<td>5' of exon 3</td>
</tr>
</tbody>
</table>

Determination of allelic sequence differences in exon 3. Single base changes from genomic DNA (gDNA) to cDNA and corresponding amino acid changes were AAC → AGC (Asn → ser), ATG → ATA (Met → Ile), and ACT → AGT (Thr → Ser). To investigate whether these differences might represent RNA editing or allelic differences, a part of exon 3 was amplified from gDNA isolated from two separate mouse strains: BALB/c and C57BL/6. PCR primers used were 513M, 513B, and 513S, with amplification conditions the same as described for 513GT below. The first site was analyzed by digestion with the restriction enzyme Hpal, which recognizes GTTA and would cut the PCR product if the sequence were the same as the 129/SvJ gDNA subclone, but should not cut if the sequence were

---

5 Additional intron sequence (not shown in Fig. 3), including 1.2 kb upstream of exon 3 and approximately 200 bp in either direction from exon/intron boundaries, was deposited with GenBank.
The mrec1.3 Transcript Is a Variant of vzg-1

A published orphan murine receptor (mrec1.3) mostly identical to vzg-1 was remarkable in that its sequence differed only in the 5' end, including the first part of the coding region (Macrae et al., 1996). The DNA differences between the two genes are shown in Fig. 1A. Both vzg-1 and mrec1.3 contain ATG sequences upstream of the indicated start codons, but each of these is succeeded by in-frame stop codons. The divergent sequence in mrec1.3 continues until just after the start of the vzg-1 open reading frame. Because of this difference, the start codon in mrec1.3 is the second in-frame ATG of vzg-1, which leads to a predicted protein product that is truncated 18 amino acids relative to the vzg-1 translation product (Fig. 1B). This difference suggests that alternative splicing or use of multiple promoters occurs in the expression of the vzg-1 gene.

Additional 3' UTR Sequence Accounts for the vzg-1 Transcript Size

The vzg-1 mRNA size by Northern blot is approximately 3.8 kb (Hecht et al., 1996; Macrae et al., 1996), whereas the maximum size of the various cDNA clone variants is at most 2.2 kb, suggesting the presence of an additional 1.6 kb of 5' UTR and/ or 3' UTR sequence in the mRNA. Using 3' RACE, a product was amplified that went farther downstream from the 3' end of the vzg-1 cDNA clone (Fig. 1C). Sequence analysis of the vzg-1 3' RACE product using BLASTN identified a murine cDNA clone (clone 4.9, GenBank Accession No. U13370) with 99% identity that had been amplified in a differential screen for genes selectively expressed in a renin-expressing kidney tumor cell line but not in normal kidney (Thompson et al., 1995). Clone 4.9 extended 1.1 kb farther downstream from the original vzg-1 sequence and contained none of the coding region. Additional 3' UTR sequence through the poly(A) consensus sites was obtained from a genomic clone (see next section). The apparent reason for the original truncated vzg-1 cDNA clones is a string of As in part of the 3' UTR that led to priming by oligo(dT) (Fig. 1C).

Nine consensus sequences (ATTTA) known to cause mRNA instability were located in the 3' UTR (Shaw and Kamen, 1986). Accounting for a 250-bp poly(A) tail (the size on newly synthesized mRNA; Wahle, 1995) and the additional 3' UTR sequence determined here, transcript variant sizes range from 3.7 to 3.9 kb, in close agreement with the size estimated by Northern blot. To be certain that the additional putative 3' UTR sequence was actually present in the vzg-1 transcript from embryonic and postnatal brain, RT-PCR was used with primer combinations extending from vzg-1 exon 3 (or 3' UTR) and the 3' sequence of clone 4.9 (see Materials and Methods). Specific products of the predicted size were amplified, indicating a contiguous transcript containing our novel 3' UTR sequence and the ORF in exon 3 (data not shown).
The vzg-1 Gene Consists of Multiple Exons

The cDNA differences between mrec1.3 and vzg-1 suggested that at least three exons encoded the various vzg-1 transcripts. To examine this possibility, we isolated λ genomic clones by screening a mouse 129/SvJ genomic library, first with the entire 2.2-kb vzg-1 cDNA insert and later with probes or PCRs specific for 5' and 3' exons. Restriction mapping, Southern blot, and sequence analyses revealed that the vzg-1 cDNA was divided among four exons (termed exons 1, 2, 3, and 4), with a fifth exon encoding the 5' part of the mrec1.3 cDNA (termed exon mrec). The relative ordering and names of the genomic clones are shown in Fig. 2A, while individual restriction maps for genomic sequence surrounding each exon are shown in Figs. 2B–2E. The multiple nonoverlapping clones demonstrated that the vzg-1 gene spans at least 50 kb of genome. Only exon 2 and exon mrec were found linked on a common clone. Sequences of each exon and approximately 20 bp of surrounding intron sequence are shown in Figs. 3A–3E. Exons 1, 2, 3, and 4 encoded 98, 160, 748, and 2650 bp of vzg-1 cDNA, while exon mrec encoded the first 81 bp of the mrec1.3 cDNA that is divergent.
from vzg-1. The last 2 bases of exon mrec (AG) were identical to the last 2 bases of exon 2, thus the actual divergence point between the two cDNA clones is 2 bp away from the intron location. The vzg-1 ORF is divided among exons 2, 3, and 4, with exon 2 encoding the first 15 amino acids and exon 3 encoding transmembrane domains I through the middle of VI. Exon 4 contained the entire 3' UTR and at its 3' end has three polyadenylation consensus sequences (AATAAA) in close succession, indicating the probable termination area of the transcript (Fig. 3E). Another potential polyadenylation site was found 270 bp upstream of these three and may direct an alternative termination point. Sequencing of the vzg-1 genomic clones encoding exon 1, 2, 4, or mrec revealed no differences with the cDNA clones. However, three base differences were found in exon 3 that would be predicted to change 3 amino acids in the protein (shown in boldface and boxed, Fig. 3D). Analysis of PCR products generated from C57BL/6J and BALB/c genomic DNA demonstrated that the sequence differences were specific to 129/SvJ and are therefore simply allelic (data not shown). Intron boundary sequences conformed to eukaryotic donor and acceptor consensus sequences, including the strictly conserved GT and AG dimers at the 5' and 3' ends, respectively (Fig. 4).
interspecific backcross panel for which the segregation pattern of over 1000 other loci was known. Two panels (BSB = F₁ × C57BL/6J and BSS = F₁ × M. spretus) of 94 progeny each were screened for the HaeIII RFLP within the 513GT PCR product from exon 4 (Fig. 5A). Figure 5B illustrates a hypothetical recombination on the chromosome and the predicted segregation ratio of the RFLP. According to Mendelian segregation principles, approximately half of the progeny in each panel should be homozygous (BB or SS) and the other half heterozygous (BS). In addition, because an equimolar amount of C57BL/6J DNA and M. spretus DNA is present in heterozygous samples, equivalent amounts of each PCR product should be amplified. Thus, equivalent molar intensities of each restriction fragment should be observed for heterozygous samples. Interestingly, in our BSS panel screen, though the expected 52%:48% SS:BS Mendelian segregation ratio was observed (with a pattern localizing the gene to Chromosome 4), the C57BL/6J product in the heterozygotes was approximately one-third of the predicted intensity relative to the M. spretus product (Fig. 5C). For the BSB panel, the segregation ratio was 27%:73% BB:BS (i.e., non-Mendelian), considering only the presence or absence of the 451- and 330-bp restriction fragments (Fig. 5C). However, closer examination revealed that only one-third of the presumed BS heterozygotes had approximately equimolar intensities of the two restriction products. The other two-thirds had a lower intensity of the C57BL/6J product relative to the M. spretus product (similar to the heterozygotes in the BSS panel—Fig. 5C).

Our proposed explanation of these mapping results is presented in Fig. 6. There was one locus for exon 4 in C57BL/6J (on Chromosome 4), but two unlinked loci in M. spretus (on Chromosomes 4 and 6). In the BSB backcross panel, ~50% of the individuals did not inherit the
Chromosome 6 M. spretus allele and showed the expected segregation ratio (50% BB and 50% SB). Thus the predicted equimolar product intensities were observed here. The remaining 50% that inherited the Chromosome 6 allele had this extra M. spretus (S) template, which was observed in the final PCR product mixture (50% BB1S5BBS and 50% SB1S5SBS). In the BSS panel, 50% of the progeny inherited one Chromosome 6 allele (BS1S5BSS and SS1S5SSS; Fig. 6B). The net effect was that the segregation pattern was identical to a panel with only the Chromosome 4 locus, but the intensity of the M. spretus product was always greater than expected.

Southern Blots Confirm a Duplicated Exon 4 Locus in M. spretus

To confirm the hypothesis that there was one common locus of exon 4 in C57BL/6J and M. spretus, and that there was a duplicated locus present only in M. spretus, genomic DNA from both M. musculus (C57BL/6J) and M. spretus was digested with three separate restriction enzymes, and Southern blots were probed with the 513GT PCR product (Fig. 7A). With each restriction enzyme, the same fragment size present in C57BL/6J was also present in M. spretus (the Chromosome 4 locus), but for M. spretus DNA there was a second hybridizing fragment of different size with the same intensity as the Chromosome 4 fragment (this was the Chromosome 6 locus). A separate Southern blot was probed using a fragment located just upstream of exon 3, and only a single band was observed in both C57BL/6J and M. spretus DNA (Fig. 7B). This indicated that the gene was not duplicated upstream of exon 3.

**FIG. 4.** Alignment with consensus intron donor and acceptor sequences. The consensus sequence is shown at the top and exon/intron boundary sequences for each of the exons are shown below. The AG and GT (in boldface) are strictly conserved sites in all introns, while other conserved bases are found in approximately 50–91% of introns (Padgett et al., 1986).
**FIG. 5.** Backcross panel screening with an HaeIII RFLP in exon 4. (A) The location of 513GT PCR product amplification from exon 4 is schematically illustrated. Shaded regions in each of exons 2, 3, and 4 indicate coding region, while nonshaded regions are UTR. Five HaeIII sites are present in the PCR product from M. musculus (C57BL/6J; denoted B) and only four from a M. spretus (denoted S) template. An easily discernable restriction fragment length difference results: 451 bp from from M. spretus and 330 bp from C57BL/6J. (B) The mechanics of the predicted Mendelian segregation ratio are illustrated in both the BSB and the BSS backcross panels. Black represents chromosomal contribution from C57BL/6J, and white represents contribution from M. spretus. An asterisk (*) or B represents a C57BL/6J allele while a number symbol (#) or S represents a M. spretus allele for the vzg-1 gene on each chromosome. In each case, 50% of the progeny are expected to be heterozygous and 50% homozygous. For the heterozygotes, an equimolar ratio of the 330- and 451-bp fragments should be observed. (C) A sample of 12–14 individual typings is shown from each of the BSB and BSS panels. In the case of heterozygotes from the BSB panel, the 330-bp product appears to be either equimolar (e.g., lanes 1, 6, 10–13) or much less intense (e.g., lanes 2, 4, 5, 14) than expected relative to the 451-bp product. Percentages of the three types of observed restriction fragment ratios are indicated below the gel photo. For the BSS panel, although the ratio of homozygous:heterozygous typings is as expected (52%:48%), the intensity of the 330-bp fragment is always approximately 1/3 what is expected (e.g., lanes 2 and 5–8 in the photo).
The vzg-1 Promoter Maps Only to Chromosome 4

To obtain unambiguous backcross panel mapping data and determine further if 5' regions of the vzg-1 gene were duplicated on Chromosome 6, we analyzed the BSB and BSS backcross panels a second time using an NdeI RFLP from a PCR product (p13/p16) in the promoter. This time the expected Mendelian segregation ratio was observed, and this should happen in approximately 50% of backcross offspring (i.e., the BB and SB lanes). The remaining backcross individuals who inherited the M. spretus Chromosome 6 locus will show unexpected 451/330-bp relative product intensities (i.e., the SBS and BBS lanes). (B) For the BSS panel, the 451/330-bp relative product intensity ratios are always higher than expected, because all backcross progeny inherit one or two copies of the Chromosome 6 locus. However, this does not affect the overall segregation pattern, which is due to the Chromosome 4 locus.

**FIG. 6.** Proposed explanation of the exon 4 backcross panel mapping results. (A) For the BSB panel, only one locus of exon 4 is present in M. musculus, while two unlinked loci of exon 4 are on chromosomes 4 and 6 in M. spretus, each equally capable of acting as template in the 513GT PCR. Symbols are the same as in Fig. 5. If the M. spretus chromosome 6 locus is not inherited, then the expected 451/330-bp relative product intensities will be observed, and this should happen in approximately 50% of backcross offspring (i.e., the BB and SB lanes). The remaining backcross individuals who inherited the M. spretus Chromosome 6 locus will show unexpected 451/330-bp relative product intensities (i.e., the SBS and BBS lanes). (B) For the BSS panel, the 451/330-bp relative product intensity ratios are always higher than expected, because all backcross progeny inherit one or two copies of the Chromosome 6 locus. However, this does not affect the overall segregation pattern, which is due to the Chromosome 4 locus.

The genomic characterization of vzg-1/lpA1 reported here is necessary to obtain a complete understanding of factors controlling its expression, its evolutionary origins, and the relationship to other genes whose chromosomal location is known. A distinct, divergent evolutionary origin for vzg-1 and other putative LPA receptors is supported by the novel genomic structure, which includes two introns within the coding region. Chromosomal localization places vzg-1 at the same lo-
vestigating the regulation and biological role of a mechanism involved in the generation of VZG-1 isoform is predominant. We are currently investigating the distribution of the isoforms among various tissues using RT-PCR and Northern analysis.

Sequence the complete 3' UTR for the VZG-1 gene was important because elements within 3' UTRs are often critical in determining both the half-life of a transcript in the cytosol and the initiation of translation (Shaw and Kamen, 1986; Stebbins-Boaz and Rich-ter, 1997; Wickens et al., 1997). Termination of transcription most likely occurs just after one of the four poly(A) consensus sequences because each conforms exactly to known strong termination sites (AATAAA) (Birnstiel et al., 1985; Proudfoot, 1989). All are located within 300 bp of one another, and accounting for a poly(A) tail of approximately 250 bp, transcript sizes of 3.6–3.9 kb would result, consistent with Northern blots. Alternate transcription termination sites (e.g., exon 1) may be used in the generation of the 2.2-kb transcript observed in testes. (Birnstiel et al., 1985; Macrae et al., 1996) all contain the same initial 18 amino acids as VZG-1, with minor substitutions, suggesting that perhaps the VZG-1 isoform is predominant. We are currently investigat- ing the distribution of the isoforms among various tissues using RT-PCR and Northern analysis.

Sequencing the complete 3' UTR for the VZG-1 gene was important because elements within 3' UTRs are often critical in determining both the half-life of a transcript in the cytosol and the initiation of translation (Shaw and Kamen, 1986; Stebbins-Boaz and Rich-ter, 1997; Wickens et al., 1997). Termination of transcription most likely occurs just after one of the four poly(A) consensus sequences because each conforms exactly to known strong termination sites (AATAAA) (Birnstiel et al., 1985; Proudfoot, 1989). All are located within 300 bp of one another, and accounting for a poly(A) tail of approximately 250 bp, transcript sizes of 3.6–3.9 kb would result, consistent with Northern blots. Alternate transcription termination sites (e.g., an AATTAA located in exon 4) may be used in the generation of the 2.2-kb transcript observed in testes (J. A. Weiner, and J. Chun, unpublished observation; Macrae et al., 1996). A search through GenBank with the novel 3' UTR sequence from the 3' RACE product led to the identification of “clone 4.9”, a cDNA which appears to be solely 3' UTR of the VZG-1 transcript

Such alternate splicing mechanisms are well documented in several other GPCRs (Berget, 1995; Maget et al., 1994; Tsai-Morris et al., 1996). A second possibility is that the heterogeneous 5' ends are generated by use of alternative promoters, where they may direct gene expression in separate cells or tissues (Schibler and Sierra, 1987). These data thus provide a framework from which future experiments investigating these possibilities for the VZG-1 gene, including 5' RACE and primer extension analysis, will be presented elsewhere (J. J. A. Contos, and J. Chun, manuscript in preparation).

Different protein isoforms with potentially different functions result from translating the VZG-1 and mrec1.3 variant transcripts. Type IIIb plasma membrane proteins (i.e., with several transmembrane domains, an extracellular NH2-terminus, and no NH2-terminal signal sequence), including VZG-1 and other GPCRs, are usually directed to their target from the first and second transmembrane domains (Singer, 1990; Foletti et al., 1995). Because the initial 18 amino acids of VZG-1 (which are absent from MREC1.3) show no similarity to known consensus signal sequences (Walter and Johnson, 1994) and are 28 amino acids away from TMD 1, it is likely that they are not important for membrane targeting. Predicted translation products of the corresponding vzg-1 gene in other mammals (human edg-2, ovine edg-2, and bovine brec1.3; An et al., 1997a; Masana et al., 1995; Macrae et al., 1996) all contain the same initial 18 amino acids as VZG-1, with minor substitutions, suggesting that perhaps the VZG-1 isoform is predominant. We are currently investigating the distribution of the isoforms among various tissues using RT-PCR and Northern analysis.

Sequencing the complete 3' UTR for the VZG-1 gene was important because elements within 3' UTRs are often critical in determining both the half-life of a transcript in the cytosol and the initiation of translation (Shaw and Kamen, 1986; Stebbins-Boaz and Rich-ter, 1997; Wickens et al., 1997). Termination of transcription most likely occurs just after one of the four poly(A) consensus sequences because each conforms exactly to known strong termination sites (AATAAA) (Birnstiel et al., 1985; Proudfoot, 1989). All are located within 300 bp of one another, and accounting for a poly(A) tail of approximately 250 bp, transcript sizes of 3.6–3.9 kb would result, consistent with Northern blots. Alternate transcription termination sites (e.g., an AATTAA located in exon 4) may be used in the generation of the 2.2-kb transcript observed in testes (J. A. Weiner, and J. Chun, unpublished observation; Macrae et al., 1996). A search through GenBank with the novel 3' UTR sequence from the 3' RACE product led to the identification of “clone 4.9”, a cDNA which appears to be solely 3' UTR of the VZG-1 transcript

A mechanism involved in the generation of VZG-1 and mrec1.3 transcripts is either alternate splicing or use of multiple promoters. One promoter may lead to transcript initiation at a common point with subse-quent splicing together of exons 1/2/3/4 or 1/mrec/3/4. The isolated mrec1.3 cDNA would then be incomplete at its 5' end, lacking the sequence common

with vzg-1 (e.g., exon 1). Such alternate splicing mechanisms are well documented in several other GPCRs (Berget, 1995; Maget et al., 1994; Tsai-Morris et al., 1996). A second possibility is that the heterogeneous 5' ends are generated by use of alternative promoters, where they may direct gene expression in separate cells or tissues (Schibler and Sierra, 1987). At least two other GPCR genes have heterogeneous 5' end sequences, and for one of these (the NPY-Y1 receptor), multiple promoters (up to 12 kb apart) are utilized (Ball et al., 1995; Robakis et al., 1990). Experiments investigating these possibilities for the vzg-1 gene, including 5' RACE and primer extension analysis, will be presented elsewhere (J. J. A. Contos, and J. Chun, manuscript in preparation).

Different protein isoforms with potentially different functions result from translating the vzg-1 and mrec1.3 variant transcripts. Type IIIb plasma membrane proteins (i.e., with several transmembrane domains, an extracellular NH2-terminus, and no NH2-terminal signal sequence), including VZG-1 and other GPCRs, are usually directed to their target from the first and second transmembrane domains (Singer, 1990; Foletti et al., 1995). Because the initial 18 amino acids of VZG-1 (which are absent from MREC1.3) show no similarity to known consensus signal sequences (Walter and Johnson, 1994) and are 28 amino acids away from TMD 1, it is likely that they are not important for membrane targeting. Predicted translation products of the corresponding vzg-1 gene in other mammals (human edg-2, ovine edg-2, and bovine brec1.3; An et al., 1997a; Masana et al., 1995; Macrae et al., 1996) all contain the same initial 18 amino acids as VZG-1, with minor substitutions, suggesting that perhaps the VZG-1 isoform is predominant. We are currently investigating the distribution of the isoforms among various tissues using RT-PCR and Northern analysis.

Sequencing the complete 3' UTR for the vzg-1 gene was important because elements within 3' UTRs are often critical in determining both the half-life of a transcript in the cytosol and the initiation of translation (Shaw and Kamen, 1986; Stebbins-Boaz and Rich-ter, 1997; Wickens et al., 1997). Termination of transcription most likely occurs just after one of the four poly(A) consensus sequences because each conforms exactly to known strong termination sites (AATAAA) (Birnstiel et al., 1985; Proudfoot, 1989). All are located within 300 bp of one another, and accounting for a poly(A) tail of approximately 250 bp, transcript sizes of 3.6–3.9 kb would result, consistent with Northern blots. Alternate transcription termination sites (e.g., an AATTAA located in exon 4) may be used in the generation of the 2.2-kb transcript observed in testes (J. A. Weiner, and J. Chun, unpublished observation; Macrae et al., 1996). A search through GenBank with the novel 3' UTR sequence from the 3' RACE product led to the identification of “clone 4.9”, a cDNA which appears to be solely 3' UTR of the vzg-1 transcript

A mechanism involved in the generation of vzg-1 and mrec1.3 transcripts is either alternate splicing or use of multiple promoters. One promoter may lead to transcript initiation at a common point with subsequent splicing together of exons 1/2/3/4 or 1/mrec/3/4. The isolated mrec1.3 cDNA would then be incomplete at its 5' end, lacking the sequence common
While most data reported for clone 4.9 were consistent with those for vzg-1, Northern blots showed only expression in the cell line (As4.1) from which it was isolated and not in any endogenous tissue. One likely explanation for this is that exposure times adequate to observe the transcript in cell line As4.1 mRNA were inadequate to observe it in endogenous tissues.

The presence of two introns in the vzg-1 coding region is of note because most GPCR superfamily genes, including one for the homologous S1P receptor Edg1, have none (Fig. 9A). Of the GPCR genes that do contain introns (Clauser et al., 1996; Kakar, 1997; Kong et al., 1994; Maget et al., 1994; Murasawa et al., 1995; Peterfreund et al., 1996; Tsai-Morris et al., 1996; Watson and Arkinstall, 1994), none have one located in the middle of the TMD VI coding region, as we have determined for vzg-1. This result suggests that the intron was inserted after vzg-1 diverged from the lpB subfamily (i.e., Edg1, edg-3, and H218) (Long et al., 1995). The amino acid identities of these receptors lend further support for this hypothesis since lpB members have 46–51% amino acid identity with one another, compared to only 32–36% identity with vzg-1/lpaL. We hypothesized that members of the same subfamily as vzg-1 would contain this same intron located in a con-
served location. Indeed, a human genomic sequence from chromosome 19, containing two exons encoding a GPCR that is 57.4% identical to VZG-1, has recently been deposited with GenBank (cosmid 33799, GenBank Accession No. AC002306). We have tentatively termed this gene IpA2, being the second member in a putative subfamily of LPA receptors. As expected, the intron in IpA2 is located at a strictly conserved site within the middle of TMD VI coding region (Fig. 9A). This result suggests that vzg-1/IpA1 and IpA2 were derived from a common ancestral gene that contained the intron. A dendrogram based on amino acid identity showing the divergence of the known genes in the cannabinoid, IpA, and IpB receptor subfamilies is presented in Fig. 9B.

Unambiguous chromosomal mapping results are necessary to determine the relationship of vzg-1 to other previously mapped genes with known phenotypes. In the course of these studies, we found that exon 4 of vzg-1 is partially duplicated in M. spretus. The results indicate that the duplication does not include sequences upstream of either exon 3 or exon 1. Thus, it appears likely that only exon 4 and surrounding intronic sequences were duplicated on Chromosome 6 in M. spretus and not the entire gene. The additional mapping done with an RFLP in the promoter allowed definitive placing of the entire vzg-1 gene on proximal Chromosome 4 by the markers D4Mit44, D4Bir14, and D4Hun3, all of which had indistinguishable segregation patterns. The two closest markers for which there was a recombination event were D4Xrf422 and Orml1, located 3.2 and 2.1 cM proximal and distal to D4Mit44, respectively. The vzg-1 gene must therefore be located between 25.3 and 30.7 cM distal to the Mos gene, a region syntenic with human chromosome 9q31–q32. This result contrasts with the determination by Macrae et al. for the mrec1.3 gene, in which 86 backcross samples were screened with an single-strand conformation polymorphism (SSCP) difference within an exon 4 PCR product (Mac-

---

**FIG. 9.** Comparative genomic structures and proposed evolutionary divergence of lipid-type GPCR family members. (A) Genomic structures for vzg-1/IpA1, IpA2 (the genomic sequence from human chromosome 19), Edg1, and cannabinoid receptor Cnr1 are shown. Schematics representing transcript UTR (solid horizontal line), coding region (boxed area) with transmembrane domains shaded, and relative intron locations (arrows) demonstrate that only the two closely related IpA1/IpA2 genes contain two conserved introns in their coding region (near the 5′ end and within TMD VI). (B) Dendrogram representing evolutionary divergence of the lysospholipid receptor family and the related cannabinoid receptors. Branch points are based on the approximate percentage amino acid identities among the various receptors shown. Two subfamilies of vzg-1-type receptors are distinguished: IpA1/IpA2 and IpB1/IpB2/IpB3. The TMD VI intron (and perhaps the intron at the beginning of the coding region) apparently was inserted as the IpA ancestral gene diverged from the IpB ancestral gene. Cnr1 is the central nervous system CB1 cannabinoid receptor and Cnr2 is the peripheral CB2 cannabinoid receptor.
The placement of mrec1.3 “16.2 cM centromeric to Mos” was confusing because Mos is located at the centromere of Chromosome 4, and the proposition that mrec1.3 may be the gene responsible for the murine asp2 (audiogenic seizure prone) phenotype is not likely since asp2 is located 45.1 cM distal to the Mos gene. We believe the previously published determination for mrec1.3 was complicated by the fact that there are two copies of exon 4 in M. spretus, which could have led to misreading of several of the individual backcross SSCP.

The localization of vzg-1 to mouse Chromosome 4 allows comparison to previously mapped phenotypes that may be caused by mutations in vzg-1. The only such gene at this locus is vacillans (vc), a mutation in a mouse line characterized decades ago and now thought to be extinct (Lyon et al., 1996; Sirlin, 1956). Symptoms of vc homozygotes included a violent tremor when walking at P14, less aggressive behavior, a smaller overall size than littermates, muscular strength about half normal, a peak mortality rate at weaning, and sexual maturity in males occurring only after 5.5 months. Future experiments, including analysis of vzg-1 knockout mice, will allow more definitive conclusions regarding the association of vzg-1 with the vc gene.

ACKNOWLEDGMENTS

We thank Carol Akita for expert technical assistance; Lucy Rowe and Mary Barter for helpful discussions on the mapping results; and Joshua Weiner, Nobuyuki Fukushima, and Lois Maltais for critically reading the manuscript. This work was supported by the National Institute of Mental Health.

REFERENCES


