GENOMICS **51,** 364–378 (1998)
ARTICLE NO. GE985400

# Complete cDNA Sequence, Genomic Structure, and Chromosomal Localization of the LPA Receptor Gene, *$lp_{A1}$/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 *$lp_{A1}$/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 (*$lp_{A2}$*) 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/$lp_{A1}$* subfamily of receptors compared to those of homologous orphan receptor genes.** © 1998 Academic Press

## INTRODUCTION

*Ventricular zone gene-1* (*vzg-1*/*$lp_{A1}$*[2]) encodes the first cloned and identified G-protein-coupled receptor (GPCR) for a lysophospholipid interacting specifically with lysophosphatidic acid (LPA, 1-acyl-2-*sn*-glycerol-3-phosphate) (Hecht *et al.,* 1996; Fukushima *et al.,* 1998; reviewed in Chun *et al.,* 1998). LPA is present in serum and produced during normal lipid metabolism in cells. The biological effects of LPA have been well documented and include mitogenesis, neurite retraction, stress fiber formation, transient increases in intracellular [$Ca^{2+}$], $Cl^-$ conductance changes, and smooth muscle contraction (Durieux, 1995; Moolenaar, 1995; Moolenaar *et al.,* 1997). Initial identification of the ligand for VZG-1 was made by overexpressing the gene in cell lines, resulting in increased responsiveness to LPA in cell rounding and adenylate cyclase inhibition assays. Heterologous expression of VZG-1 in mammalian cells has further demonstrated that it is necessary and sufficient for LPA signaling in different cell lineages and that it can mediate multiple LPA-dependent responses (Fukushima *et al.,* 1998). Independent confirmation that *vzg-1* encodes an LPA receptor came from studies of the human homologue of VZG-1 (EDG-2). Expression in yeast (which have no endogenous response to LPA) produced an LPA-dependent G-protein signaling cascade (Erickson *et al.,* 1998), and overexpression in lymphoid cells potentiated LPA-dependent serum-response-element activation (An *et al.,* 1997b).

Individual members of gene families often contain similar, if not identical, genomic structures. The published genes most closely related to *vzg-1,* termed *Edg1,* H218/AGR16, and edg-3 ($lp_{B1}$, $lp_{B2}$, and $lp_{B3}$, respectively)[2] (Hla and Maciag, 1990; MacLennan *et al.,* 1994; Okazaki *et al.,* 1993; Yamaguchi *et al.,* 1996), have recently been identified as specific sphingosine 1-phosphate (S1P) receptors (An *et al.,* 1997a; Lee *et al.,* 1998; Zondag *et al.,* 1998). Each of these shares 32–36% amino acid identity with *vzg-1* and 46–51% identity with one another (Chun *et al.,* 1998). The next

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@ucsd.edu.

[2] We now refer to *vzg-1* as "lysophospholipid receptor A1," *$lp_{A1}$*; to additional murine putative LPA receptors as *$lp_{A2}$* and *$lp_{A3}$*; and to the murine *Edg-1* subfamily of receptors as *$lp_{B1}$* (=*Edg1*), *$lp_{B2}$* (=H218/AGR16), and *$lp_{B3}$* (=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.

**TABLE 1**

**Primers Used in Polymerase Chain Reactions**

| | | | |
|---|---|---|---|
| mrec1 | TGCCCCTGCTTGGACTGACT | 513A | GCTATCGAGAGGCACATCAC |
| mrec2 | GAGGTACCTTAGCTGTGTCAGGA | 513B | CAATAAAGGCACCAAGCACAATGA |
| SK | CGCTCTAGAACTAGTGGATC | 513C | CTTCTGGGCCATTTTCAACC |
| SK′ | ATCCACTAGTTCTAGAGCGTT | 513F | GACTACAATCACCACCACCAC |
| T3 | AATTAACCCTCACTAAAGGG | 513G | ACTCCGGGATTGGTCTT |
| T7 | GTAATACGACTCACTATAGGGC | 513H′ | TCCTCCTAGCATGACCGAGATCTG |
| vzg.is2 | TATAGGAGTCTTGTGTTGCCTGTCC | 513K | AAGAGGCCCCCTGCACTGC |
| vzg.ix1 | AATCACTCAGGCATTTCATCTGCTG | 513M | GCGTGTTCATCATGTTGG |
| vzg.p2 | GGTTCTGCAAGTCACCCTTCT | 513N | TATTCAGTATGCCTTTCT |
| vzg.p6′ | CCAGGGTGTGTTGAGTCTCAGA | 513P | CAAGCTCACATTACTTTG |
| vzg.p8 | CAGTTCACRTGCTCACCAGAGA | 513Q | CAATCCAGCGAAGAAGTC |
| vzg.p10 | CCTTCTTCAGTGTATTCTTG | 513R | TGTGGTGGAATTGAGAAA |
| vzg.p13 | CTTGCATATGTGTGTACATG | 513T | ATACTTTTCCTCCATCAT |
| vzg.p16 | GCCTTCTTGAGGAATGTAGC | 513W | GTCCCTGCATAGTGAAGATA |
| vzg.z3 | AGTCCAAGTTCGGTTCACTCAT | 513Y | CCTACTTACTCTAGAAAACTTG |

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 clone 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 lysophospholipid 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 sequence 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 (Dieffenbach and Dveksler, 1995). SK oligonucleotide[3] (Table 1) (5 μg) from pBluescript was phosphorylated with polynucleotide kinase[4] and purified with NucTrap columns according to directions from the manufacturer (Stratagene). Dephosphorylated 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 μg) 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 $MgCl_2$, 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 μg/mL ethidium bromide (EtBr). As a control, in the secondary PCR only the SK′ oligo was used. PCR products were cloned into the *Xba*I site of pBS (the 513Y oligo has an A → C change that allows for creation of an *Xba*I site, and the SK′ oligo contains an *Xba*I site as well). These products were sequenced (as described below) using the T7 primer.

*RT-PCR detection of spliced exons.* Total RNA from E15 brain and cell line TR was used to generate cDNA template in the following way. A 40-μL reaction consisting of 1× Superscript first-strand synthesis buffer, 20 U RNasin, 0.5 mM each dNTP, 10 μg RNA, 2 μL (200 pmol) random hexamers, and 200 U Superscript was incubated at 23°C for 10 min, 42°C for 60 min, 95°C for 5 min; cooled on ice; diluted to 200 μL with $H_2O$; and stored at −20°C. PCR was done under reaction conditions similar to those of the 3′ RACE experiments, except that no *Pfu* was added and 5 μL of diluted cDNA was used as template. Primer combinations and the expected product sizes were 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/

[3] Oligonucleotide sequences are listed in Table 1. An Expedite Nucleic Acid Synthesis System (Millipore) was used to generate each oligo. Oligos either were used directly after resuspension in TE or were further purified with 15–20% polyacrylamide gels and SepPak reverse-phase column chromatography (Waters) (Ausubel *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 [$^{32}$P]dCTP (DuPont) and random hexamers (BM).

513T and 513W/vzg.z3 in both RNA samples, which are known to express the *vzg-1* transcript.

*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 $^{32}$P-labeled *Eco*RI/*Xho*I 2.2-kb *vzg-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 (Ausubel *et al.,* 1994). Restriction mapping was facilitated by the unique *Not*I and *Sal*I sites that release inserts from the λ phage. Initially, *Eco*RI, *Xho*I, *Hin*dIII, *Bam*HI, *Xba*I, *Bgl*II, and *Pst*I were used and later several other enzymes. Digests 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′ 180 bp), and 513N/P (near the 3′ end) and vector sequence was digested away and purified using the Qiaex 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 *Sph*I digestion (an *Sph*I 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, mrec, and 4, based on a published protocol (Israel, 1993) with several modifications. Approximately 5000 phage were grown by liquid lysate in each of the wells in a 96-well plate as follows: 1 mL of Xl1-blue (MRA) host cells at $OD_{600}$=5.0 in 10 mM $MgSO_4$ was mixed with 1 mL SM containing $10^6$ 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 513G/T primers) or exon mrec (mrec1/2), respectively. Each PCR tube contained 20 μL of 1× PCR buffer (15 mM $NH_4SO_4$, 60 mM Tris, pH 8.5), 2.0 mM $MgCl_2$, 0.5 μM each primer, 0.25 mM each dNTP, and 0.4 U *Taq.* The mix was heated to 90°C; then 1 μL of the amplified phage from an individual well was added and cycled 35× at 95°C for 30 s, 52–56°C for 30 s, 72°C for 2 min, with a final 72°C extension step for 10 min before cooling to 4°C. There were 10 wells positive for exon 4 initially and 2 for exon mrec. Wells were titered and positives plated out again in 96-well plates, this time using only 200 phage per well. Between 0 and 3 positives of 45 were identified in each of these secondary screens. To isolate single positive clones, approximately 500 phage from these secondary positive wells were plated out with NZY top agarose and screened by filter hybridization. Phage were grown by liquid lysate, and DNA was isolated with the Wizard λ Prep Kit (Promega). In this way 2 clones for exon 4 were isolated [1D(3E) and 6C(2A)] and 2 for exon mrec [4D(2D) and 5B(3D)]. One of each of these was characterized further by restriction mapping and Southern blotting. Distances and relative orientations were determined by long-distance PCR (Advantage Genomic Polymerase Kit; Clontech) using primers specific to the Lambda FIX II vector (T3 and T7) and primers specific to the exons (513G, 513T, mrec3, mrec4, 513K, and 513H′).

**TABLE 2**

**Subcloned Genomic Fragments**

| Donor DNA | Fragment subcloned | Name | Relative location of subclone |
|---|---|---|---|
| Clone 10a | 5.7-kb *Eco*RI/*Not*I | vzg 5.5 | Exon 3 |
| Clone 14a | 6.5-kb *Hin*dIII/*Not*I | HN6.5 | 5′ of exon 3 |
| Clone 3 | 7.2-kb *Eco*RI/*Eco*RI | RR9.0 | Exon 3 |
| Clone γ | 5.4-kb *Eco*RI/*Not*I | RN4.7 | Exon 1 |
| Clone γ | 10-kb *Eco*RI/*Not*I | RN10 | 5′ of exon 1 |
| Clone 4D(2D) | 7.0-kb *Eco*RI/*Not*I | RN7.0 | Exon 2 |
| Clone 4D(2D) | 4.8-kb *Eco*RI/*Not*I | RN4.8 | Exon mrec |
| Clone 1D(3E) | 3.0-kb *Xho*I/*Not*I | XN3.0 | Exon 4 |
| Clone 1D(3E) | 9.0-kb *Eco*RI/*Not*I | RN9.0 | Exon 4 |
| Clone 1D(3E) | 10-kb *Xho*I/*Xho*I | XX10 | 3′ of exon 4 |
| Clone 1D(3E) | 2.0-kb *Xho*I/*Not*I | XN2.0 | 3′ of exon 4 |
| vzg5.5 | 1.1-kb *Bgl*II/*Not*I | BN1.1 | 5′ of exon 3 |
| vzg5.5 | 1.2-kb *Xba*I/*Xba*I | BX1.2 | Exon 3 |
| vzg5.5 | 1.0-kb *Xba*I/*Xba*I | XX1.0 | 5′ of exon 3 |
| vzg5.5 | 0.6-kb *Xba*I/*Xba*I | XX0.2 | 5′ of exon 3 |
| vzg5.5 | 0.2-kb *Xba*I/*Xba*I | XX0.2 | 5′ of exon 3 |
| HN6.5 | 1.2-kb *Xba*I/*Xba*I | XX1.5 | 5′ of exon 3 |

*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 [on 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 (Ausubel *et al.,* 1994). Colonies were screened either by restriction digestion of minipreps (Ausubel *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 dideoxy method 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[5].

*Determination of allelic sequence differences in exon 3.* Single base changes from genomic DNA (gDNA) to cDNA and corresponding amino acid changes were A**A**C → A**G**C (Asn → ser), AT**G** → AT**A** (Met → Ile), and A**C**T → A**G**T (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/6J. PCR primers used were 513M/513B and 513M/513F, with amplification conditions the same as described for 513GT below. The first site was analyzed by digestion with the restriction enzyme *Hpa*I, which recognizes GTTA**A**C and would cut the PCR product if the sequence were the same as in 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 same as previously isolated cDNA subclones (which have GT-TA**G**C). An exon 3 subclone (BX1.2) was used as a positive control in the digestion. In addition, all three sites were analyzed by sequencing the Qiaquick-purified (Qiagen) PCR products using an automated sequencer and the 513M primer.

*RFLP detection and backcross panel mapping.* To find restriction fragment length polymorphisms (RFLPs) in the *vzg-1* gene, primers were used to amplify products from both *M. musculus* (C57BL/6J) and *M. spretus* genomic DNA (purchased from The Jackson Laboratory). Amplification targets (p13/p16, p10/p6, p8/p2, ix1/513Q, is3/513F, 513R/513P, and 513G/513T) were in known sequence in the region of exons 1, 3, or 4. Fifty-microliter polymerase chain reactions contained 1× PCR buffer (50 mM KCl, 10 mM Tris, pH 8.5), 1.5 mM $MgCl_2$, 0.5 $\mu$M each primer, 0.25 mM each dNTP, and 50 ng genomic DNA. The mix was heated to 95°C for 3 min and then cooled to 90°C; 1 U *Taq* + 0.1 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. Products were ethanol-precipitated after adding 1/10 volume 3 M NaAc, washed, dried, and resuspended in 30 $\mu$L of TE [however, the p13/p16 product could be digested with *Nde*I after twofold dilution to a final 0.5× buffer H (Boehringer Mannheim), 10 mM $MgCl_2$, and 1 mM DTT]. Each product was treated with enzymes (*Hae*III, *Hin*fI, *Mbo*I, *Nde*I, *Xba*I, *Bgl*II, *Ase*I, *Bst*XI, *Hin*dII, or *Taq*I) that would cut at one to six separate sites and analyzed by electrophoresis on 1.4% agarose gels containing EtBr. Five RFLPs of a total of 206 bp of restriction sites scanned were detectable. These were p13/p16 *Nde*I, p10/p6 *Hin*fI, 513R/513P *Hin*fI, 513G/513T *Taq*I, and 513G/513T *Hae*III. For both of the RFLPs used in typing, positive control sites within each of the PCR products ensured complete digestion was occurring. The 513GT *Hae*III RFLP is illustrated in Fig. 5A. The p13/p16 product was digested with *Nde*I into distinguishable 263- (C57BL/6J) or 240-bp (*M. spretus*) fragments, along with common 196- and 110-bp fragments. The 513GT *Hae*III RFLP was analyzed on 183 backcross samples, and the p13/p16 *Nde*I RFLP was analyzed on 186 backcross samples from The Jackson Laboratory. The formal names of the crosses are The Jackson Laboratory interspecific backcross panels (C57BL/6J × *M. spretus*) × C57BL/6J, called Jackson BSB, and (C57BL/6JEi × SPRET/Ei)$F_1$ × SPRET/Ei, known as Jackson BSS (Rowe *et al.,* 1994). Raw data were submitted to The Jackson Laboratory for comparison to known markers typed to the panel.

*Genomic Southern blot analysis.* Genomic DNA (10 $\mu$g) from either *M. musculus* (C57BL/6J) or *M. spretus* was 10-fold overdigested with the restriction enzyme *Eco*RI, *Pst*I, *Xba*I, or *Pvu*II; electrophoresed on a 0.8% agarose gel; acid-nicked by rocking in a 1:50 dilution of HCl for 20 min; reequilibrated by rocking 20 min in 0.4 M NaOH; and then blotted overnight using 0.4 M NaOH as transfer solution (Southern, 1975). The nylon filter was neutralized in 0.2 M Tris (pH 7.5)/2× SSC, blot-dried on Whatman filter paper, and UV-crosslinked. The blot was prehybridized for >15 min at 65°C using 0.1 mg/mL denatured sonicated salmon sperm DNA in Church buffer [0.5 M $Na_2HPO_4 \cdot 7H_2O$ (pH 7.2), 7% SDS, 1% BSA, 1 mM EDTA (pH 8.0)]. Probes purified with the Qiaex gel extraction kit (Qiagen) were labeled with [$^{32}$P]dCTP using random hexamers, heat-denatured with an equivalent aliquot of salmon sperm DNA, and added to the prehybridization mix with each blot. After incubation at 65°C overnight, the blots were washed for 15–30 min each time with 2× SSC/0.1% SDS, 1× SSC/0.1% SDS, 0.5× SSC/0.1% SDS, and 0.2× SSC/0.1% SDS at 65°C and finally with 2× SSC. After air drying, the blot was exposed to Kodak XAR film with two intensifying screens at −80°C for up to 7 days before being developed.

## RESULTS

### *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 cDNA 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).

**A)**

```
vzg-1 cDNA    1 GCACAGTGCT GCCCTCCGTA GGCTCCGGGT TGTGCTGGGG TGAGGCTTGG   50
mrec  cDNA

vzg-1 cDNA   51 GTTGGGTTGG CCCGGCGGCT GCGTGAACTG CGGAGCTGGA CCTAGCAGGC  100
mrec  cDNA

vzg-1 cDNA  101 TTACAGTTCC TCCTAGCATG ACCGAGATCT GATCAGCCAA CCCGCGCATT  150
mrec  cDNA

vzg-1 cDNA  151 GCTTTTTGTG CCTGGCACTG CAGTGCAGGG GGCCTCTTCA TCGCCCCAAA  200
mrec  cDNA    1                                  CTT GCCCCTGCTT GGACTGACTC   23

              1               M  A   A  A  S  T   S  S  P   V  I  S    12
vzg-1 cDNA  201 CTACAGCACT GTCATGGCAG CTGCCTCTAC TTCCAGCCCT GTAATTTCAC  250
mrec  cDNA   24 TTAAGATGCA GAATTGTTTT CTTGTGCTGG GCCAGCACCC AGTCCTGACA   73

             13 Q  P  Q  F   T  A  M   N  E  Q   Q  C  F  Y   N  E  S    29
vzg-1 cDNA  251 AGCCCC|AGTT CACAGCCATG AACGAACAAC AGTGCTTCTA CAATGAGTCT  300
mrec  cDNA   74 CAGCTA|AGTT CACAGCCATG AACGAACAAC AGTGCTTCTA CAATGAGTCT  123
```

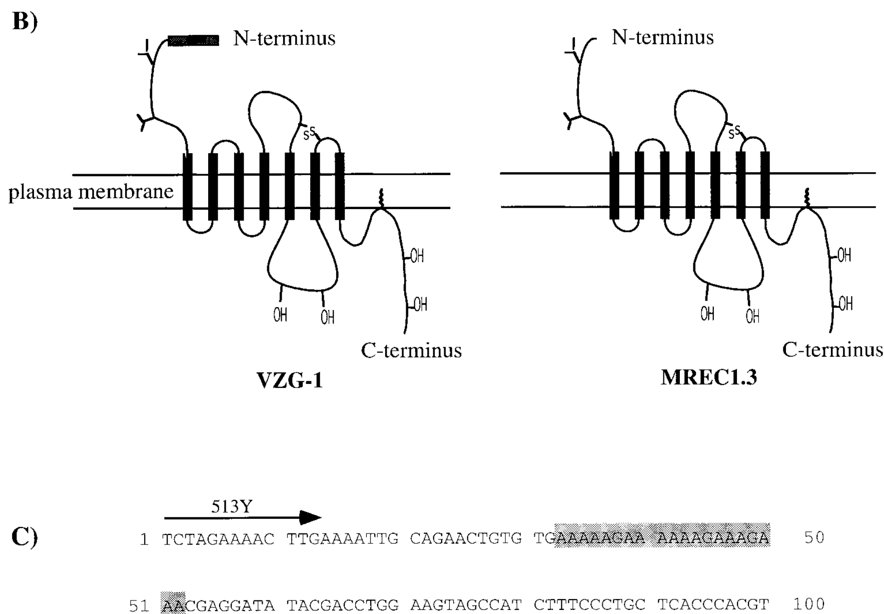


**FIG. 1.** Divergent 5′ and additional 3′ sequence in the *vzg-1* transcript. (**A**) The published cDNA sequences of *vzg-1* and mrec1.3 are shown aligned with one another. The sequences are identical beginning at bp 80 of mrec1.3 and bp 257 of *vzg-1* (after the solid vertical line). Note that sequence 5′ of the line is completely divergent. The start codon in each cDNA clone is indicated in boldface, and coding region is shaded, with predicted amino acid sequence shown above. (**B**) Schematic of the G-protein-coupled receptor structure, showing that the amino acid difference affects only the extracellular N-terminus of the receptor. Predicted carbohydrate, palmitoylation, disulfide bridge, and potential phosphorylation sites are indicated. (**C**) Partial sequence of the 3′ RACE product. The poly(A)-rich stretch at the end of the *vzg-1* cDNA clone (shaded) and part of the 513Y primer used in the secondary PCR are noted. Remaining sequence of the 3′ UTR was determined from the genomic clone (Fig. 3E).

### *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

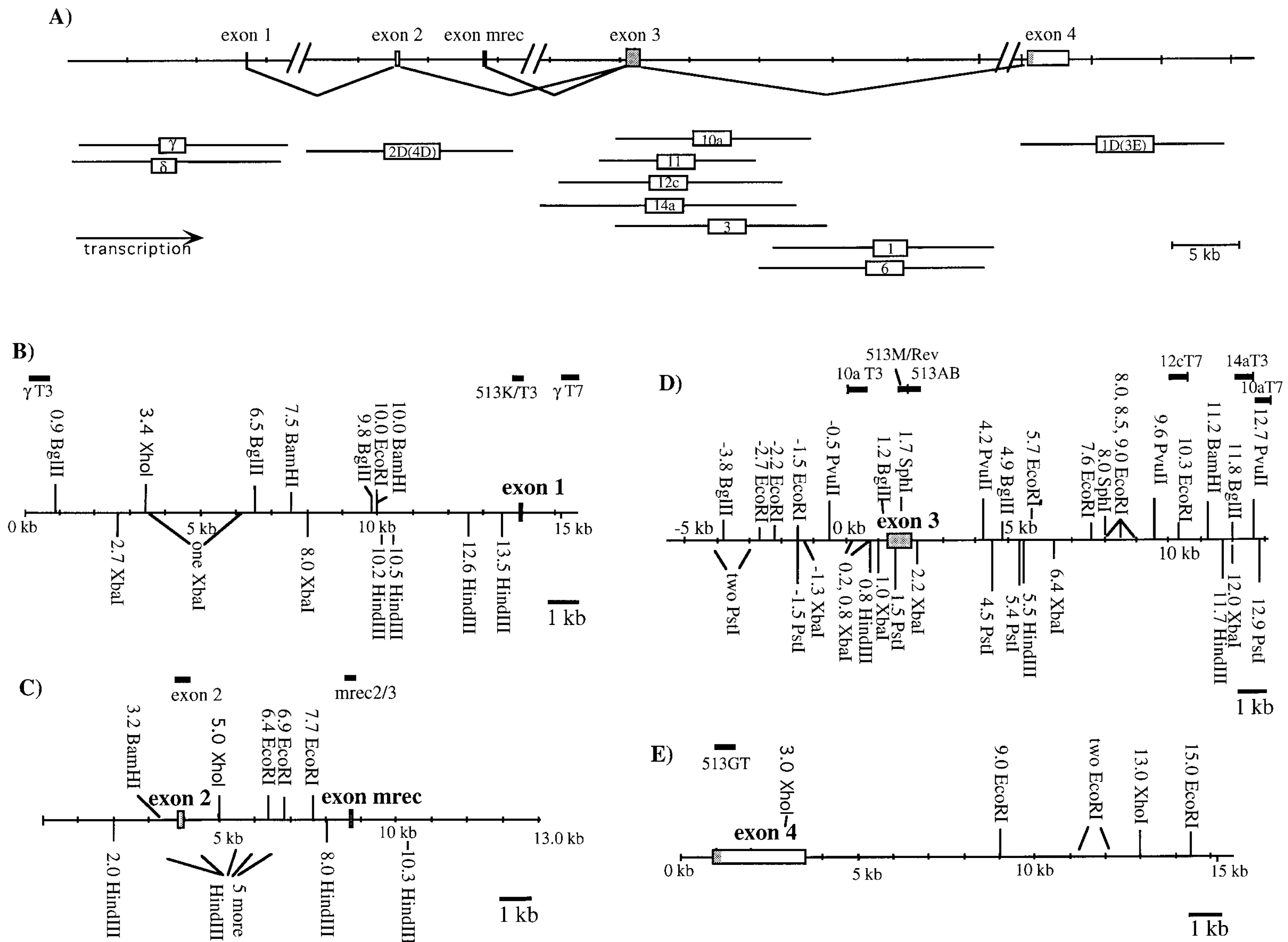


**FIG. 2.** Genomic structure of the *vzg-1* gene and restriction maps of genomic clones. (**A**) A genomic map for the relative placement of the five exons that encode the *vzg-1* and mrec1.3 genes. The distances between exons 1/2, mrec/3, and 3/4 have not been determined. Individual λ genomic clones that were isolated and characterized are indicated below the map. Clones 1, 3, and 6 were isolated using a 3′ probe from clone 10a in a chromosome-walking strategy with the intent of obtaining exon 4. (**B–E**) Restriction maps surrounding each of exons 1, 2/mrec, 3, and 4, respectively. Probes utilized are indicated above each map. Both sequencing and PCR allowed determination of orientations and distances of exons from genomic clone ends. In addition to the restriction enzyme sites shown, there are no *Sal*I or *Not*I sites in any of the inserts. For the exon 3 map, there also are no *Kpn*I or *Xho*I sites.

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 transcript 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).

### *Exon 4 Maps to Two Loci in M. spretus, But Only One in M. musculus*

To determine the location of the *vzg-1* gene in the mouse genome, RFLPs between *M. musculus* (C57BL/6J) and *M. spretus* were identified and used to screen an

**A) exon 1**

```
  1 gttgtgactc ttgtcctcca GCACAGTGCT GCCCTCCGTA GGCTCCGGGT TGTGCTGGGT GAGGCTTGGG TTGGGTTGGC CCGGCGGCTG CGTGAACTGC   100
101 GGAGCTGGAC CTAGCAGgta aagacacggc ggccggg                                                                       137
```

**B) exon 2**

```
  1 ggattccttg tgtttcgcag GCTTACAGTT CCTCCTAGCA TGACCGAGAT CTGATCAGCC AACCCGCGCA TTGCTTTTTG TGCCTGGCAC TGCAGTGCAG   100

  1                                          M  A   A  A  S   T  S  S   P  V  I  S   Q  P  Q                   15
101 GGGGCCTCTT CATCGCCCCA AACTACAGCA CTGTCATGGC AGCTGCCTCT ACTTCCAGCC CTGTAATTTC ACAGCCCCAG gtaagtgctt tgacctggtc   200
```

**C) exon mrec**

```
  1 agaccagtct ggttctcctc CTTGCCCCTG CTTGGACTGA CTCTTAAGAT GCAGAATTGT TTTCTTGTGC TGGGCCAGCA CCCAGTCCTG ACACAGCTAA   100
101 Ggtgagtccc agtatggttg g                                                                                        121
```

**D) exon 3**

```
 16                              F  T  A   M  N  E   Q  Q  C  F   Y  N  E   S  I  A   F  F  Y  N   R  S  G   K  Y  L    41
  1 ttcattcatc ttctttctta gTTCACAGCC ATGAACGAAC AACAGTGCTT CTACAATGAG TCTATCGCCT TCTTTTATAA CCGGAGTGGG AAATATCTAG   100
                                                    TMD I
 42 A  T  E  W   N  T  V   S  K  L   V  M  G  L   G  I  T   V  C  V   F  I  M  L   A  N  L   L  V  M   V  A  I  Y    75
101 CCACAGAATG GAACACAGTG AGCAAGCTGG TGATGGGACT GGGCATCACT GTTTGCGTGT TCATCATGTT GGCCAATCTC CTGGTCATGG TGGCAATCTA   200
                                                       TMD II
 76  V  N  R   R  F  H   F  P  I  Y   Y  L  M   A  N  L   A  A  A  D   F  F  A   G  L  A   Y  F  Y  L   M  F  N   108
201 CGTCAACCGC CGCTTCCATT TCCCTATTTA TTACTTGATG GCCAACCTGG CTGCTGCAGA CTTCTTCGCT GGATTGGCCT ACTTCTACCT GATGTTCAAT   300
                                                                          TMD III
109 T  G  P   N  T  R  R   L  T  V  [N] T  W   L  L  R  Q   G  L  I   D  T  S   L  T  A  S   V  A  N   L  L  A   141
301 ACAGGACCTA ATACCCGGAG ACTGACTGTT AACACGTGGC TCCTCCGGCA GGGCCTCATT GACACCAGCC TGACAGCTTC TGTGGCCAAC CTGCTGGCTA   400
                                                                                   TMD IV
142 I  A  I  E   R  H  I   T  V  F   R  M  Q  L   H  T  R   M  S  N   R  R  V  V   V  V  I   V  V  I   W  T  M  A   175
401 TTGCTATCGA GAGGCACATC ACGGTTTTCC GCATGCAGCT CCATACACGA ATGAGCAACC GGCGCGTGGT GGTGGTGATT GTAGTCATCT GGACTATGGC   500

176  I  V  M   G  A [M]  P [T] V  G   W  N  C   I  C  D   I  D  H  C   S  N  M   A  P  L   Y  S  D  S   Y  L  V   208
501 CATTGTGATG GGTGCTATGC CCACTGTGGG CTGGAACTGC ATCTGTGATA TCGATCACTG TTCCAACATG GCACCCCTCT ACAGTGACTC CTACTTAGTC   600
            TMD V
209  F  W  A   I  F  N  L   V  T  F   V  V  M   V  V  L  Y   A  H  I   F  G  Y   V  R  Q  R   T  M  R   M  S  R   241
601 TTCTGGGCCA TTTTCAACCT GGTGACCTTT GTGGTCATGG TGGTTCTCTA CGCTCACATC TTTGGCTATG TTCGCCAGAG GACTATGAGG ATGTCTCGGC   700
                                                          TMD VI
242 H  S  S  G   P  R  R   N  R  D   T  M  M  S   L  L  K   T  V  V   I  V  L                                     264
701 ATAGTTCTGG ACCCAGGAGG AATCGGGACA CCATGATGAG CCTTCTGAAG ACTGTGGTCA TTGTGCTTGg tgagttctgt attgactgt              771
```

**FIG. 3.** Sequences of individual exons and surrounding introns. (**A–E**) Sequences of exons 1, 2, mrec, 3, and 4, respectively, in capital letters with approximately 20 bp of surrounding intron in lowercase letters. Where coding sequence exists in the exons, the translated amino acid is indicated above the middle base of the codon. Predicted transmembrane domains (TMDs) are shaded. In (**D**), the three individual sequence differences between the genomic and the cDNA clones are indicated in boldface, and the corresponding codon and amino acid are boxed. In (**E**), the nine mRNA instability consensus sequences (ATTTA) are indicated in boldface, and the 3′ end of the *vzg-1* cDNA clone where priming by oligo(dT) occured is shaded. Four polyadenylation consensus sequences are boxed.

interspecific backcross panel for which the segregation pattern of over 1000 other loci was known. Two panels (BSB = $F_1$ × C57BL/6J and BSS = $F_1$ × *M. spretus*) of 94 progeny each were screened for the *Hae*III 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 the 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

**E) exon 4**

```
                               TMD VI
 265                   G  A  F  I  V  C  W  T  P  G  L  V  L  L  L  L  D  V  C  C  P  Q  C  D  V  L  A     291
   1 ccctccccctg tcgccatagG TGCCTTTATT GTCTGCTGGA CTCCGGGATT GGTCTTGTTA TTGCTGGATG TGTGCTGCCC GCAGTGCGAT GTCCTGGCCT   100
                 TMD VII
 292 Y  E  K  F  F  L  L  L  A  E  F  N  S  A  M  N  P  I  I  Y  S  Y  R  D  K  E  M  S  A  T  F  R  Q  I   325
 101 ATGAGAAGTT CTTCCTCCTC CTGGCCGAGT TCAACTCTGC TATGAACCCC ATCATCTACT CCTACCGCGA CAAAGAGATG AGCGCCACCT TCAGGCAGAT   200

 326  L  C  C  Q  R  N  E  N  P  N  G  P  T  E  G  S  D  R  S  A  S  S  L  N  H  T  I  L  A  G  V  H  S   358
 201 CCTGTGTTGC CAGCGCAACG AGAACCCTAA TGGCCCCACG GAAGGCTCTG ACCGCTCTGC CTCCTCCCTC AACCACACCA TTCTGGCTGG AGTTCACAGC   300

 359  N  D  H  S  V  V  *                                                                                         364
 301 AACGACCACT CTGTGGTTTA GAAGGAAGCC AGCCGGCCTC TGTGGATCTG TGAACCCCAC CCTACCCCCC ATTGCCAGGG CAAGGTGGGG AGCCAGAGGA   400
 401 GATGAGGACA CTCCTGTACT TAACACTAAC CAATGGCAGT ATTTGTCCCT AGACCCAAGA GACTTGAGGA TGAATTTATT TGGCAGGCCC CATCTTCTCC   500
 501 TTTGGAAAAC AGAAGGGGAC CGTCTTGTGG TGGAATTGAG AAATGGACTC TGGGGTGACC GTGTAGCATT CACTAACTAG ACTTAAAAGA TTTTATGTGG   600
 601 TTTGGCTTAA GCCAGGAAAA AAAAATCTGC TGAATTGAGT ATACAATCGA GTATACACAG GCTTCCCCTT TAAAGAACAA ACAATACATT GCATTTATTA   700
 701 ATGAGTATGT TTATGCCTGA CAGCATGTTT GTGATCGAAA AGACTGCTAA ACTGACATAG ATGAGTTGTT TTTTTTTTTT TGTTTTTTGT TTTTTTTTAC   800
 801 ATGATGGAGG AAAAGTATAA ATTAGAATGA TTTTTGTGTT TGTTTAGAAA GCAAGCATGT GGTGTGTGTA TTCAGTATGC CTTTCTTTAA AGATAAAAGG   900
 901 CCACTATTTT AAATCTTCTA GGGAATAGAA GAATCTAGTA AAAACCAGTA TTCATTTAGG CTACAGGAAA AACCATATCC TAATCAATTA CCTTTTAATT  1000
1001 AAAGTAATGA AATATACATG AAAGGCAAAG TAATGTGAGC TTGTCACCCA AAGAGTGTGT GCTCTCCAAA CGCTGGAGGA GATGAAGCTG TAGCGTTGTC  1100
1101 CCTGCATAGT GAAGATACCC ACGTGCGTTC TCAGTGCCAG ACCCTCAGTG GGACTTGTTT TAAAGCCTGT GGTTTTCCAA GTTAGAAAAT AATACCTACT  1200
1201 TACTATAGAA AACTTGAAAA TTGCAGAACT GTGTGAAAAA AGAAAAAAGA AAGAAAGATT GCTGATATAC GACCTGGAAG TAGCCATCTT TCCCTGCTCA  1300
1301 CCCACGTATG CCTATAGACA TATATTATAT ACCTTTTTTT TTTTTTACAT AATCGGGTTG ATATTGTAAA ATGTTTTGTA ACTTTTTCTT TCAGAATTGC  1400
1401 CAGGTTTTTC TATGGCATTT TTTTTCAAAA ACGTTAGCGG CTATATAATA TTCCATTTAA TGGATGCAGC TCAGTTTATT TAGTACCATT CCTGTATTGT  1500
1501 TGACTATGCT TTTGGCTTTT TTTTCATAAT ATTGGAAGAA TCTTGTGTAT ATAAAACTTT GCCAGTATCA TTGATGATTT ACTTGGGCTC CATTCCCATG  1600
1601 AAGGACCTAC TTTTAAAATG TGAAATGTCC TCATGTGTTG TCACTTTATA AAAAGGGATA TTCTACCCTC AAACTGCAGG GGTGACCATG CTGACAAAAA  1700
1701 GCTCAGAAAT TACCTTTTTA CAAAAGAAAC ACAGTGGCTA CTTTAGTTGC GAATGGGTTC TTGACAAGAT GTTTCCCAAT AACCCAGACC CTTAACATAA  1800
1801 CTAGCAAATT TACTTTAGTA GGAACTGGTG ATCCCTTTCT GGACACAAGA AAACACAAAG ACATCAAGCA TCAGAGTGAA TCCAGCAATG CAGTACCAAG  1900
1901 CAGCCTGGGT GGGGTGTGCT AGCAACTGGC TGCTGTGTGG TCTCCAGCAA TCACCCAATG GACACCCTGT GTTCAGAGAT GCCCAGTACA GTGTCCTGGA  2000
2001 GACTAAGGTC TCTAATGTCA TGCTATGAGG AACAAGGAGG AAGAAGAATA GTCCTGAAGC AGGTGTGGAG AGGTCAGGAG GACGTCTGAG AGAAGGCAGT  2100
2101 CCTGCCAAGA GTGGCATTCT AGGGAGGAGG AAGGCAGGGT GTGGCTTGTG TCTGGACAGT CAGTGGTAGG CATGTGACAT TTGCCTGGTG GAAAAAAAAT  2200
2201 AAGTAGGCGG AGATCCAGTT AGCAGGACAA AGATTTTGCT CGAGGATTCC CCCAATCCAA GAAATTTAAA CTGGAAATGA GTGAACCGAA CTTGGACTTT  2300
2301 TATTGATTCC CCTTTTATGG GGGAGGTAAG GACTATTTGA ATTGAAAAGC ATACTAATTG AACCTTAATA AATCATTCTC AATCAGTGTT TGGCGATGTG  2400
2401 TGGTTCTGGT CTGTGCTTTT TTTCTATTTT CGAGAAGCCT CTGGCTGTGC TGTGATGGCT TCTCTCTCTG ACCTTCTTTA ACAGTCTGAG GCAGCCCAAC  2500
2501 AACTGTTCCT TTAGCTCAGA TACTGGTTCC TCATCCATGA GATTCATGAG AGACGTGTTA CCTCAATGGA ATGAGTACTA GAGCAAGGTA TTTAGAGAGA  2600
2601 TTTTTTTTAT TATTTATTTA TTTATTTTGA ATAAAATGTA TGTAATAAAT AAGATAAAAT AAAAAGTANT ATATGACAAT TGGCAAAAAC TANACT      2696
```

**FIG. 3**—*Continued*

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% BB + S = BBS and 50% SB + S = SBS). In the BSS panel, ~50% of the progeny inherited one Chromosome 6 allele (BS + S = BSS and SS + S = SSS) and the other ~50% inherited two (BS + SS = BSSS and SS + SS = SSSS; 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.

```
consensus:  5'-CCCCCCCCCCC NCAG|G   exon    AA|GTAAGT -3'
               TTTTTTTTTTT
exon 1:     ------------------|   98 bp    AG|GTAAAGACAC
exon 2:     CCTTGTGTTTCGCAG|G   160 bp    AG|GTAAGTGCTT
exon mrec:  ------------------|   80 bp    AG|GTAAAGACAC
exon 3:     CATCTTCTTTCTTAG|T   748 bp    TG|GTGAGTTCTG
exon 4:     CCCCTGTCGCCATAG|G  ~2700 bp     |-----------
```

**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).

### *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.

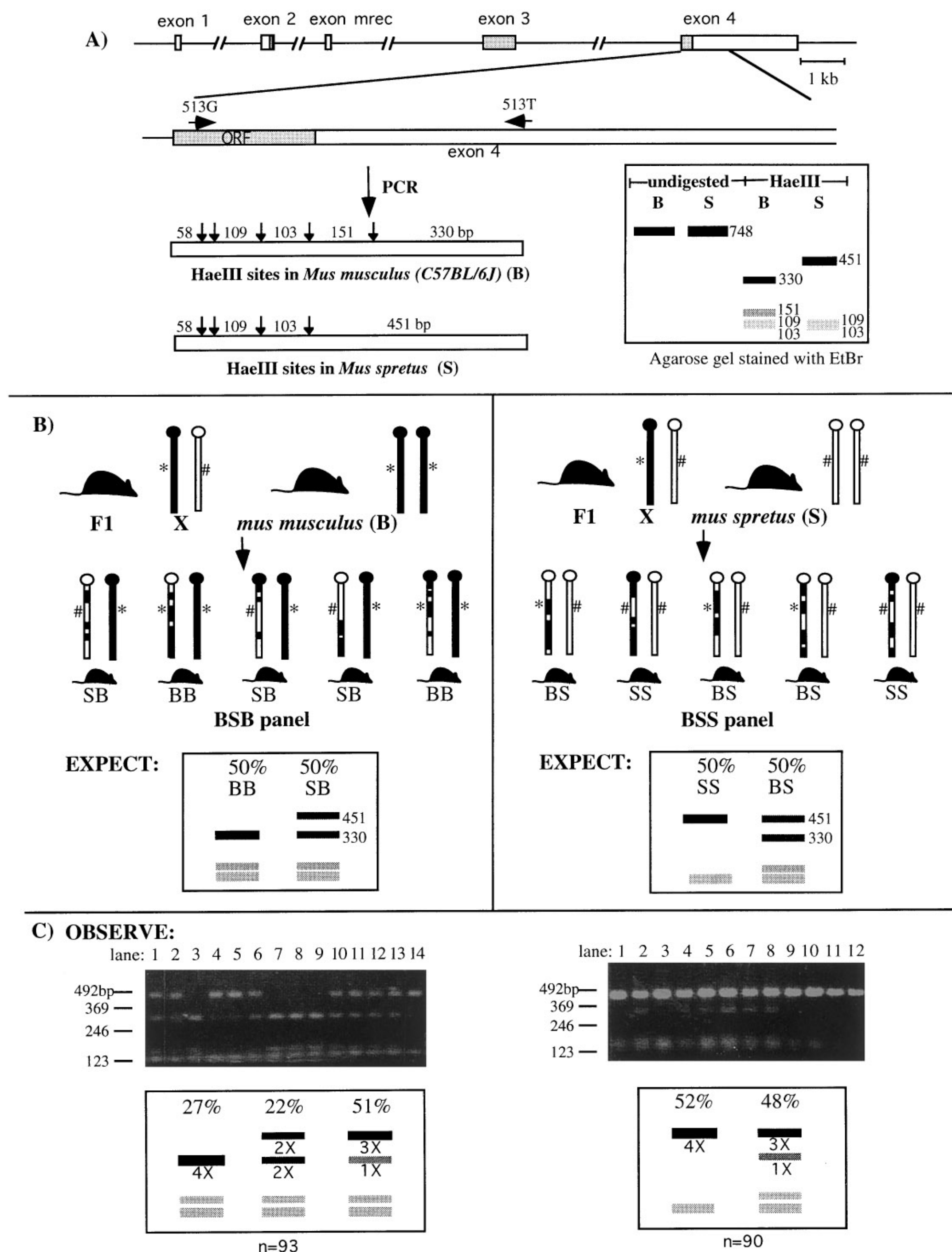


**FIG. 5.** Backcross panel screening with an *Hae*III RFLP in exon 4. (**A**) The location of 513GT PCR product amplification from exon 4 is schematically illustrated. Shaded regions in each of exons 2, 3, and 4 indicate coding region, while nonshaded regions are UTR. Five *Hae*III sites are present 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).

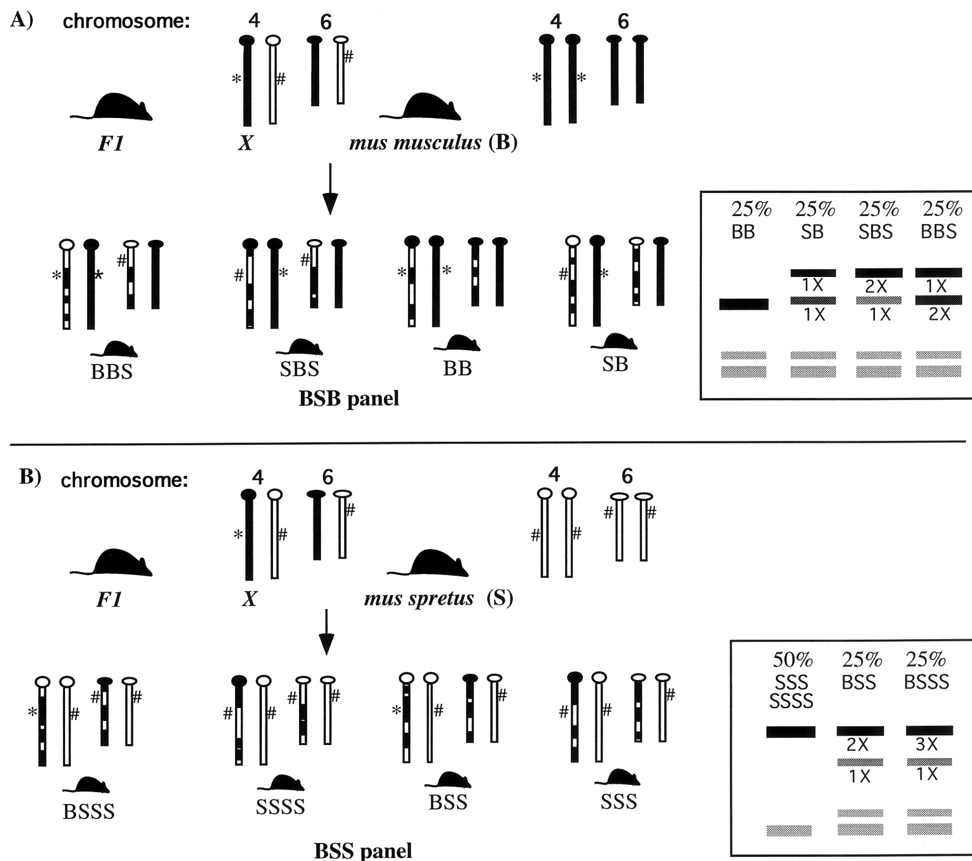


**FIG. 6.** Proposed explanation of the exon 4 backcross panel mapping results. (**A**) For the BSB panel, only one locus of exon 4 is present in *M. musculus,* while two unlinked loci of exon 4 are on chromosomes 4 and 6 in *M. spretus,* each equally capable of acting as template in the 513GT PCR. Symbols 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 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 *Nde*I RFLP from a PCR product (p13/p16) in the promoter. This time the expected Mendelian segregation ratio was observed, as well as the expected equimolar intensities for both restriction fragment products in the heterozygotes (data not shown). The segregation pattern was identical to the 513GT RFLP from the BSS panel and also identical to the markers *D4Mit44, D4Bir14,* and *D4Hun3* on Chromosome 4. These were 28.6–29.0 cM distal to the *Mos* gene, located at the centromere (Fig. 8A). Though Fig. 8A shows the mapped location using the BSS panel data alone, the BSB panel data confirmed the location. Figure 8B shows typing results for individual pairs of loci, while Fig. 8C is a higher resolution map of the genes and markers in the locale to which *vzg-1* mapped. Genes at this locus included *Nsk1* (neural fold/somite kinase-1), *Melk* (maternal embryonic leucine zipper kinase), *vc* (vacillans), *Gln3-1* (glutamine retrotransposon LTR), *Ambp* ($\alpha$-1 microglobulin/bikunin), and *Lv* ($\delta$-aminolevulinate dehydratase). Taking into account the Chromosome 4 mapping data from the BSS panel, the second *M. spretus* locus could be scored in 45 of the 94 BSB backcross animals. This showed linkage to markers on proximal Chromosome 6, including *D6Bir4* (lod score of 13.4) and *D6Mit1* (data not shown).

## DISCUSSION

The genomic characterization of *vzg-1*/*lp*$_{A1}$ 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-

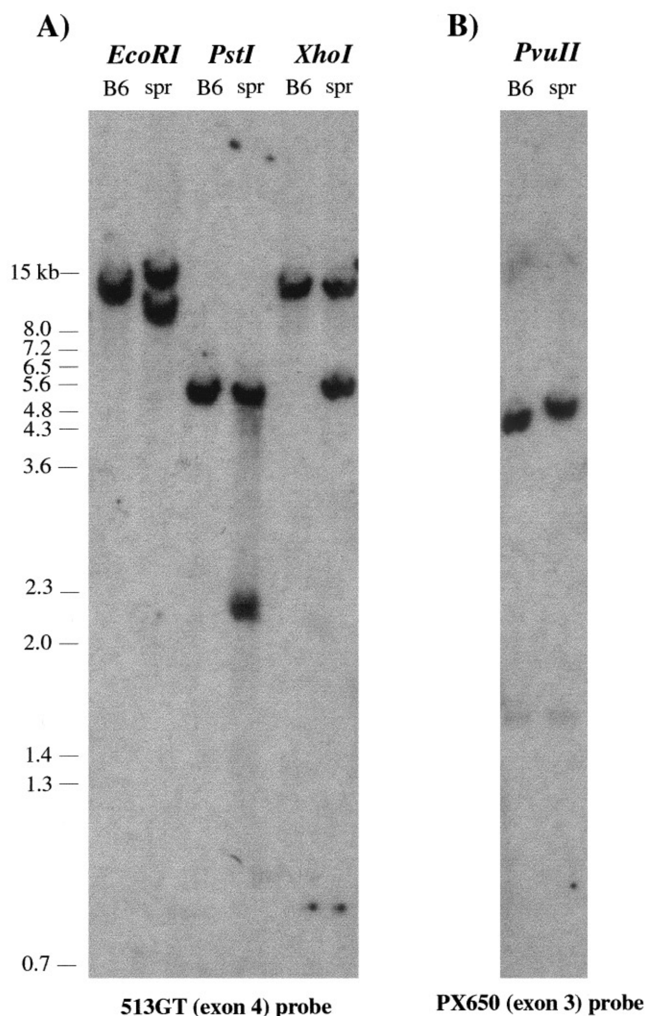


**FIG. 7.** Genomic Southern blots. Genomic DNA (10 $\mu$g) from either *M. musculus* (C57BL/6J; B6) or *M. spretus* (spr) was digested with *Eco*RI, *Pst*I, *Xba*I, or *Pvu*II, electrophoresed, and Southern blotted. Probes used were either (**A**) the 513GT PCR product (i.e., part of exon 4) amplified from the *vzg-1* cDNA plasmid or (**B**) the 650-bp *Pvu*II/*Xba*I fragment (PX650) from plasmid XX1.5 (whose insert is a 1.5-kb *Xba*I/*Xba*I fragment 1 kb upstream of exon 3). Using the 513GT probe, there is only one fragment that hybridizes from *M. musculus* with each restriction enzyme, while two fragments always hybridize from *M. spretus.* Only one fragment is observed hybridizing in each of the C57BL/6J and *M. spretus* samples using the PX650 probe. The C57BL/6J *Eco*RI band in the 513GT-probed blot is of slightly different size from both the *M. spretus* bands, likely due to an RFLP.

cus as vacillans, a gene whose mutation causes both neurological and systemic defects. These data thus provide a framework from which future experiments investigating the regulation and biological role of *vzg-1* can be designed.

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 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 $NH_2$-terminus, and no $NH_2$-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 I, 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 Richter, 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

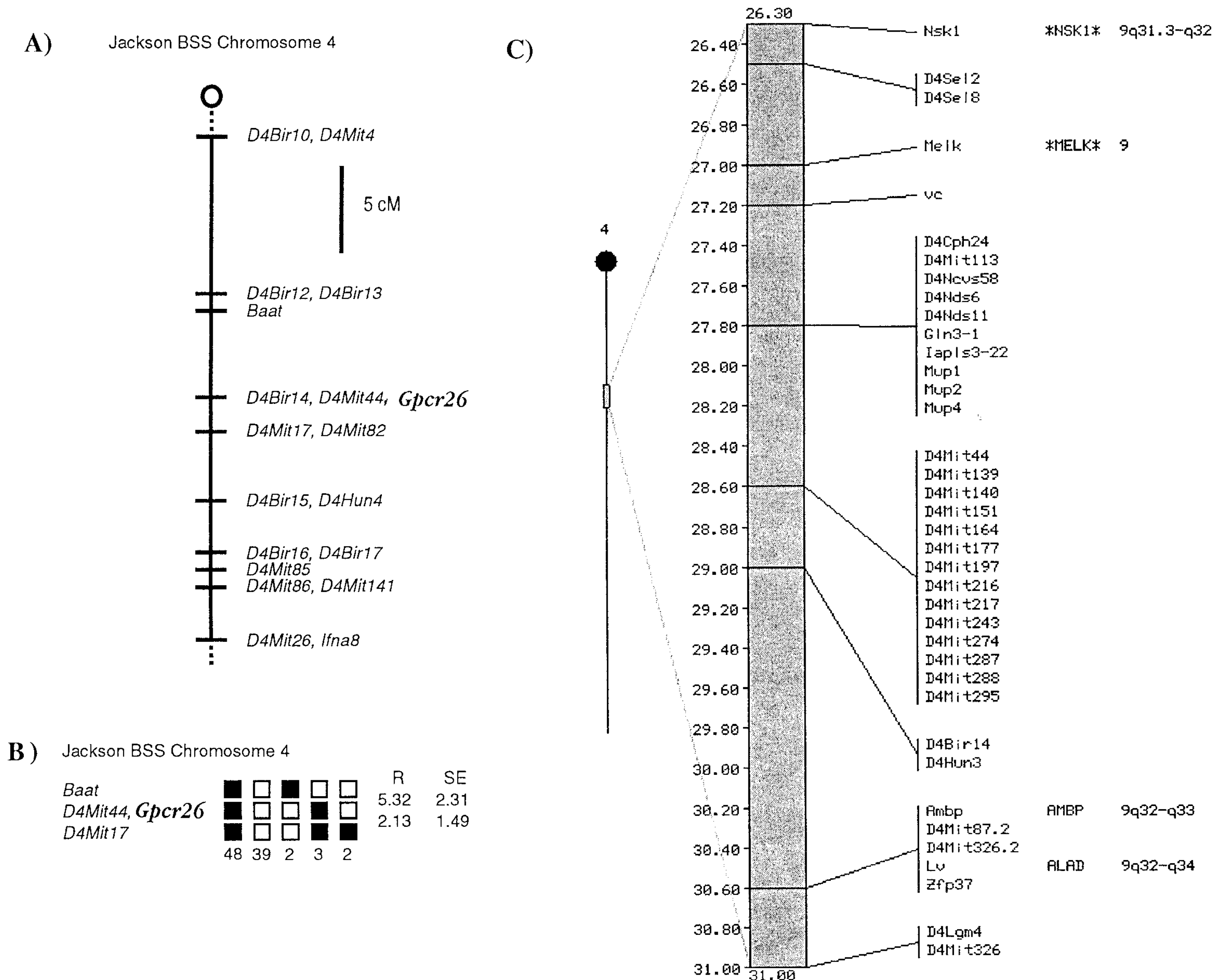


**FIG. 8.** Linkage map showing chromosomal placement of the *vzg-1* (*Gpcr26*) gene. (**A**) Map from The Jackson Laboratory BSS backcross showing part of Chromosome 4. The map is depicted with the centromere toward the top. A 5-cM scale bar is shown to the right. Loci mapping to the same position are listed in alphabetical order. Missing typings were inferred from surrounding data where assignment was unambiguous. Raw data from The Jackson Laboratory can be obtained from the World Wide Web address http://www.jax.org/resources/documents/cmdata. (**B**) Haplotype from The Jackson Laboratory BSS backcross showing part of Chromosome 4 with loci linked to *Gpcr26.* Loci are listed in order with the most proximal at the top. The black boxes represent the C57BL6/JEi allele and the white boxes the SPRET/Ei allele. The number of animals with each haplotype is given at the bottom of each column of boxes. The percentage recombination (R) between adjacent loci is given to the right, with the standard error (SE) for each R. (**C**) Closer detail linkage map of the Chromosome 4 region at which *vzg-1* has been placed. Syntenic genes from human chromosome 9q are shown to the right.

(Thompson *et al.,* 1995). 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 $lp_B$ 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 $lp_B$ members have 46–51% amino acid identity with one another, compared to only 32–36% identity with *vzg-1*/$lp_{A1}$. We hypothesized that members of the same subfamily as *vzg-1* would contain this same intron located in a con-

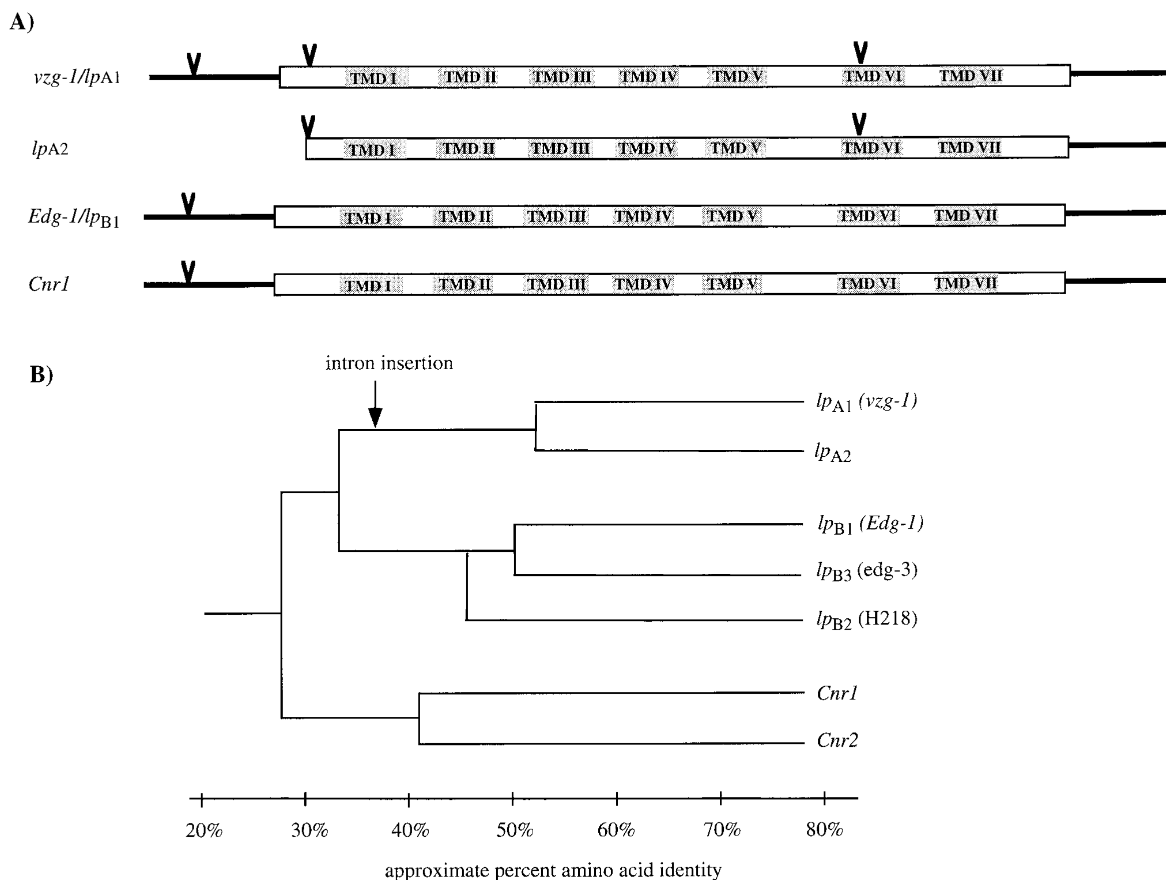


**FIG. 9.** Comparative genomic structures and proposed evolutionary divergence of lipid-type GPCR family members. (**A**) Genomic structures for *vzg-1*/$lp_{A1}$, $lp_{A2}$ (the genomic sequence from human chromosome 19), *Edg1,* and cannabinoid receptor *Cnr1* are shown. Schematics representing transcript UTR (solid horizontal line), coding region (boxed area) with transmembrane domains shaded, and relative intron locations (arrows) demonstrate that only the two closely related $lp_{A1}/lp_{A2}$ genes contain two conserved introns in their coding region (near the 5′ end and within TMD VI). (**B**) Dendrogram representing evolutionary divergence of the lysophospholipid 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: $lp_{A1}/lp_{A2}$ and $lp_{B1}/lp_{B2}/lp_{B3}$. The TMD VI intron (and perhaps the intron at the beginning of the coding region) apparently was inserted as the $lp_A$ ancestral gene diverged from the $lp_B$ ancestral gene. *Cnr1* is the central nervous system CB1 cannabanoid receptor and *Cnr2* is the periperal CB2 cannabinoid receptor.

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 $lp_{A2}$, being the second member in a putative subfamily of LPA receptors. As expected, the intron in $lp_{A2}$ is located at a strictly conserved site within the middle of TMD VI coding region (Fig. 9A). This result suggests that *vzg-1*/$lp_{A1}$ and $lp_{A2}$ 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, $lp_A$, and $lp_B$ 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 *Orm1,* 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-

rae *et al.,* 1996). 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 SSCPs.

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

Abood, M. E., Ditto, K. E., Noel, M. A., Showalter, V. M., and Tao, Q. (1997). Isolation and expression of a mouse CB1 cannabinoid receptor gene. Comparison of binding properties with those of native CB1 receptors in mouse brain and N18TG2 neuroblastoma cells. *Biochem. Pharmacol.* **53:** 207–214.

An, S., Bleu, T., Huang, W., Hallmark, O. G., Coughlin, S. R., and Goetzl, E. J. (1997a). Identification of cDNAs encoding two G-protein coupled receptors for lysosphingolipids. *FEBS Lett.* **417:** 279–282.

An, S., Dickens, M. A., Bleu, T., Hallmark, O. G., and Goetzl, E. J. (1997b). Molecular cloning of the human Edg2 protein and its identification as a functional cellular receptor for lysophosphatidic acid. *Biochem. Biophys. Res. Commun.* **231:** 619–622.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1994). "Current Protocols in Molecular Biology," Wiley, New York.

Ball, H. J., Shine, J., and Herzog, H. (1995). Multiple promoters regulate tissue-specific expression of the human NPY-Y1 receptor gene. *J. Biol. Chem.* **270:** 27272–27276.

Berget, S. M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270:** 2411–2414.

Birnstiel, M. L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3′ processing: The end is in site! *Cell* **41:** 349–359.

Chun, J., Contos, J. J. A., and Munroe, D. (1998) A growing family of receptor genes for lysophosphatidic acid (LPA) and other lysophospholipids (LPs). *Cell Biochem. Biophys.,* in press.

Clauser, E., Curnow, K. M., Davies, E., Conchon, S., Teutsch, B., Vianello, B., Monnot, C., and Corvol, P. (1996). Angiotensin II receptors: Protein and gene structures, expression and potential pathological involvements. *Eur. J. Endocrinol.* **134:** 403–411.

Dieffenbach, C. W., and Dveksler, G. S. (1995). "PCR Primer: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Durieux, M. E.(1995). "Lysophosphatidate Signaling: Cellular Effects and Molecular Mechanisms" Landes, Austin, TX.

Erickson, J. R., Wu, J. J., Goddard, J. G., Tigyi, G., Kawanishi, K., Tomei, L. D., and Kiefer, M. C. (1998). Edg-2/Vzg-1 couples to the yeast pheromone response pathway selectively in response to lysophosphatidic acid. *J. Biol. Chem.* **273:** 1506–1510.

Foletti, D., Guerini, D., and Carafoli, E. (1995). Subcellular targeting of the endoplasmic reticulum and plasma membrane $Ca^{2+}$ pumps: A study using recombinant chimeras. *FASEB J.* **9:** 670–680.

Fukushima, N., Kimura, Y., and Chun, J. (1998). A single receptor encoded by *vzg-1/lpa1/edg-2* couples to G-proteins and mediates multiple cellular responses to lysophosphatidic acid (LPA). *Proc. Natl. Acad. Sci. USA* **95:** 6151–6156.

Hecht, J. H., Weiner, J. A., Post, S. R., and Chun, J. (1996). Ventricular zone gene-1 (*vzg-1*) encodes a lysophosphatidic acid receptor expressed in neurogenic regions of the developing cerebral cortex. *J. Cell. Biol.* **135:** 1071–1083.

Hla, T., Jackson, A. Q., Appleby, S. B., and Maciag, T. (1995) Characterization of edg-2, a human homologue of the *Xenopus* maternal transcript G10 from endothelial cells. *Biochem. Biophys. Acta* **1260:** 227–229.

Hla, T., and Maciag, T. (1990). An abundant transcript induced in differentiating human endothelial cells encodes a polypeptide with structural similarities to G-protein-coupled receptors. *J. Biol. Chem.* **265:** 9308–9313.

Hla, T., Zimrin, A. B., Evans, M., Ballas, K., and Maciag, T. (1997) The immediate-early gene product MAD-3/EDG-3/IkappaB alpha is an endogenous modulator of fibroblast growth factor-1 (FGF-1) dependent human endothelial cell growth. *FEBS Lett.* **414:** 419–424.

Israel, D. I. (1993). A PCR-based method for high stringency screening of DNA libraries. *Nucleic Acids Res.* **21:** 2627–2631.

Kakar, S. S. (1997). Molecular structure of the human gonadotropin-releasing hormone receptor gene. *Eur. J. Endocrinol.* **137:** 183–192.

Kong, X. F., Schipani, E., Lanske, B., Joun, H., Karperien, M., Defize, L. H., Juppner, H., Potts, J. T., Jr., Segre, G. V., Kronenberg, H. M., *et al.* (1994). The rat, mouse, and human genes encoding the receptor for parathyroid hormone and parathyroid hormone-related peptide are highly homologous. *Biochem. Biophys. Res. Commun.* **201:** 1058.

Lee, M. J., Van Brocklyn, J. R., Thangada, S., Liu, C. H., Hand, A. R., Menzeleev, R., Spiegel, S., and Hla, T. (1998). Sphingosine-1-phosphate as a ligand for the G protein coupled receptor EDG-1. *Science* **279:** 1552–1555.

Liu, C. H., and Hla, T. (1997). The mouse gene for the inducible G-protein-coupled receptor edg-1. *Genomics* **43:** 15–24.

Long, M., de Souza, S. J., and Gilbert, W. (1995). Evolution of the intron–exon structure of eukaryotic genes. *Curr. Opin. Genet. Dev.* **5:** 774–778.

Lyon, M. F., Rastan, S., and Brown, S. D. M. (1996). "Genetic Variants and Strains of the Laboratory Mouse," Oxford Univ. Press, New York.

MacLennan, A. J., Browe, C. S., Gaskin, A. A., Lado, D. C., and Shaw, G. (1994). Cloning and characterization of a putative G-

protein coupled receptor potentially involved in development. *Mol. Cell. Neurosci.* **5:** 201–209.

Macrae, A. D., Premont, R. T., Jaber, M., Peterson, A. S., and Lefkowitz, R. J. (1996). Cloning, characterization, and chromosomal localization of rec1.3, a member of the G-protein-coupled receptor family highly expressed in brain. *Brain Res. Mol. Brain Res.* **42:** 245–254.

Maget, B., Tastenoy, M., and Svoboda, M. (1994). Sequencing of eleven introns in genomic DNA encoding rat glucagon receptor and multiple alternative splicing of its mRNA. *FEBS Lett.* **351:** 271–275.

Masana, M. I., Brown, R. C., Pu, H., Gurney, M. E., and Dubocovich, M. L. (1995). Cloning and characterization of a new member of the G-protein coupled receptor EDG family. *Recept. Channels* **3:** 255–262.

Matsuda, L. A., Lolait, S. J., Brownstein, M. J., Young, A. C., and Bonner, T. I. (1990) Structure of a cannabinoid receptor and functional expression of the cloned cDNA. *Nature* **346:** 561–564.

Moolenaar, W. H. (1995). Lysophosphatidic acid signalling. *Curr. Opin. Cell. Biol.* **7:** 203–210.

Moolenaar, W. H., Kranenburg, O., Postma, F. R., and Zondag, G. C. (1997). Lysophosphatidic acid: G-protein signalling and cellular responses. *Curr. Opin. Cell. Biol.* **9:** 168–173.

Munro, S., Thomal, K. L., and Abu-Shaar, M. (1993) Molecular characterization of a peripheral receptor for cannabinoids. *Nature* **365:** 61–65.

Murasawa, S., Matsubara, H., Kijima, K., Maruyama, K., Mori, Y., and Inada, M. (1995). Structure of the rat V1a vasopressin receptor gene and characterization of its promoter region and complete cDNA sequence of the 3′-end. *J. Biol. Chem.* **270:** 20042–20050.

Okazaki, H., Ishizaka, N., Sakurai, T., Kurokawa, K., Goto, K., Kumada, M., and Takuwa, Y. (1993). Molecular cloning of a novel putative G protein-coupled receptor expressed in the cardiovascular system. *Biochem. Biophys. Res. Commun.* **190:** 1104–1109.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. S., and Sharp, P. A. (1986) Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **55:** 1119–1150.

Peterfreund, R. A., MacCollin, M., Gusella, J., and Fink, J. S. (1996). Characterization and expression of the human A2a adenosine receptor gene. *J. Neurochem.* **66:** 362–368.

Proudfoot, N. J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.* **14:** 105–110.

Robakis, N. K., Mohamadi, M., Fu, D. Y., Sambamurti, K., and Refolo, L. M. (1990). Human retina D2 receptor cDNAs have multiple polyadenylation sites and differ from a pituitary clone at the 5′ non-coding region. *Nucleic Acids Res.* **18:** 1299.

Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J., and Birkenmeier, E. H. (1994). Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome* **5:** 253–274.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.

Schibler, U., and Sierra, F. (1987). Alternative promoters in developmental gene expression. *Annu. Rev. Genet.* **21:** 237–257.

Shaw, G., and Kamen, R. (1986). A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46:** 659–667.

Singer, S. J. (1990) The structure and insertion of integral proteins in membranes. *Annu. Rev. Cell Biol.* **6:** 247–296.

Sirlin, J. L. (1956) Vacillans, a neurological mutant in the house mouse linked to brown. *J. Genet.* **54:** 42–48.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503–517.

Stebbins-Boaz, B., and Richter, J. D. (1997). Translational control during early development. *Crit. Rev. Eukaryotic Dev.* **7:** 73–94.

Thompson, H., Burson, J., Lang, J., Gross, K., and Sigmund, C. (1995). Tissue-specific expression of novel messenger ribonucleic acids cloned from a renin-expressing kidney tumor cell line (As4.1). *Endocrinology* **136:** 3037–3045.

Tsai-Morris, C. H., Buczko, E., Geng, Y., Gamboa-Pinto, A., and Dufau, M. L. (1996). The genomic structure of the rat corticotropin releasing factor receptor. A member of the class II G protein-coupled receptors. *J. Biol. Chem.* **271:** 14519–14525.

Wahle, E. (1995) Poly(A) tail length control is caused by termination of processive synthesis. *J. Biol. Chem.* **270:** 2800–2808.

Walter, P., and Johnson, A. E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10:** 87–119.

Watson, S., and Arkinstall, S. (1994). "The G-Protein Linked Receptor Factsbook," Academic Press, London.

Wickens, M., Anderson, P., and Jackson, R. J. (1997). Life and death in the cytoplasm: Messages from the 3′ end. *Curr. Opin. Genet. Dev.* **7:** 220–232.

Yamaguchi, F., Tokuda, M., Hatase, O., and Brenner, S. (1996). Molecular cloning of the novel human G protein-coupled receptor (GPCR) gene mapped on chromosome 9. *Biochem. Biophys. Res. Commun.* **227:** 608–614.

Zondag, G. C. M., Postma, F. R., van Etten, I., Verlaan, I., and Moolenaar, W. H. (1998). Sphingosine 1-phosphate signalling through the G-protein coupled receptor Edg-1. *Biochem. J.* **330:** 605–609.