Very Large G Protein-coupled Receptor-1, the Largest Known Cell Surface Protein, Is Highly Expressed in the Developing Central Nervous System*

We previously identified a member of the G protein-coupled receptor family, very large G protein-coupled receptor-1 (VLGR1). VLGR1 has a large ectodomain containing multiple calcium exchanger β repeats that resemble regulatory domains of sodium-calcium exchanger proteins. Similar repeats are found in the extracellular aggregation factor of marine sponges, which mediates species-specific cell aggregation. We now report that the protein encoded by the originally described human cDNA (now termed VLGR1a) is, in fact, at 1967 amino acids, the smallest of three expressed human isoforms. It is encoded by an alternative transcript that begins within intron 64 of the VLGR1 gene. The longest gene product, VLGR1b, is 6307 amino acids (6298 amino acids in mice) due to a much larger ectodomain containing 35 calcium exchanger β repeats and a pentraxin homology domain. VLGR1b is apparently the largest known cell surface protein. The VLGR1 gene comprises 90 exons and is >600 kb long. In situ hybridization studies with mouse embryo sections show that high level expression of VLGR1 is restricted to the developing central nervous system and eye. Strong expression in the ventricular zone, home of neural progenitor cells during embryonal neurogenesis, suggests a fundamental role for VLGR1 in the development of the central nervous system.

G protein-coupled receptors (GPCRs) are the largest family of related proteins presently known. There are more than 1000 known genes for such receptors, encompassing perhaps as much as 3% of all human genes. Arguably, GPCRs have an even larger impact on cells than their genetic representation implies; by recognizing a wide range of chemical ligands and other stimuli, from organic molecules (e.g., proteins, peptides, and odorants) to proteolytic activity, ions, and light, GPCRs provide cells with much of their ability to sample and respond to their environment.

Although all GPCRs have seven-transmembrane (7-TM) segments, these receptors fall into several families that have significant amino acid similarity within but not between families. Family 1 (or A) includes the rhodopsins, adrenergic and dopaminergic receptors, and receptors for other small organic ligands, whereas Family 2 (or B) consists mostly of receptors for peptide hormones such as secretin, and Family 3 (or C) consists of metabotropic glutamate receptors as well as the extracellular calcium sensor.

A recently defined subdivision within family 2, termed LNB-TM7, consists of proteins with large ectodomains composed of tandemly repeated units such as cadherin, thrombospondin, or epidermal growth factor repeats. Many of these putative GPCRs are expressed in the brain, and several have apparent functions in development. For example, Flamingo specifies ventral-dorsal polarity in Drosophila and is involved in dendrite formation, whereas latrophilin may play a role in synaptogenesis.

We previously identified a putative member of this subfamily termed very large G protein-coupled receptor-1 (VLGR1) (7). Although the 7-TM domain of VLGR1 was related to those of Flamingo and latrophilin, its putative ectodomain was unique, consisting mostly of seven imperfect calcium exchanger β (Calx-β) repeats resembling the calcium-binding regulatory domains of sodium-calcium exchangers. Calx-β repeats are also found in the core protein of the proteoglycan aggregation factor that mediates species-specific aggregation in marine sponges, suggesting that they can play a role in cell-cell adhesion (9).

Although the Calx-β repeats of VLGR1 indeed bound calcium (7), there were few additional clues as to the function of VLGR1. Its pattern of expression was uninformative, because VLGR1 was expressed at low levels in almost all adult human tissues examined. Therefore, we isolated cDNA encoding mouse VLGR1 with the immediate goal of analyzing gene expression by in situ hybridization of mouse embryos and with the ultimate aim of disrupting the gene.

To our surprise, we found that the originally described form of VLGR1 was actually encoded by an alternative transcript unique to humans that originated within an intron of a much larger gene. The longest form of VLGR1 is ~6300 amino acids in length and is encoded by a 90-exon gene spanning more than 600 kb. Thus, VLGR1 is by far the largest known protein that has a putative extracellular domain.

VLGR1 is highly expressed in the embryonic central nervous system, beginning at the time of development of the neural groove. In midgestation, VLGR1 expression is prominent in the ventricular zone (the region in which neurons forming the cerebral cortex arise) and in the eye. By late gestation, as...
neurogenesis is largely completed, VLGR1 expression declines. These findings suggest a role for VLGR1 in the development of the central nervous system.

MATERIALS AND METHODS

In Situ Hybridization—A 1.5-kb EcoRI fragment of our previously reported human VLGR1 cDNA, nucleotides (nt) 3282–4621, was radiolabeled and used to probe a commercially prepared B6/DBAF1J mouse genomic library in the bacteriophage A FIX II vector (Stratagene, La Jolla, CA). The inserts of hybridizing clones were subcloned into pbLue-script SK (Stratagene) and sequenced. Oligonucleotide primers were synthesized and used in a standard RT-PCR reaction with mouse kidney total RNA to produce a 1.6-kb cDNA fragment, which was cloned into pCR II (Invitrogen, Carlsbad, CA) and sequenced.

A 364-bp region of the 1.6-kb mouse clone, corresponding to nt 16310–16673 of full-length mouse VLGR1 cDNA (see below), was used as the template for RNA synthesis. Riboprobes were generated by in vitro transcription using the MAXIscript Kit (Ambion, Austin, TX) in the presence of [35S]UTP (1000 Ci/mmol) (Amersham Biosciences, Inc.). Tissue preparation and in situ hybridization was performed as previously described (10).

RNA Preparation and Rapid Amplification of cDNA Ends (RACE)—ICR mouse embryos were harvested at 14 days of gestation and eviscerated, and limbs were removed. The remains, consisting largely of head and spinal regions, were flash frozen in liquid nitrogen and stored at −80 °C until needed. Total RNA was extracted by the acid-guanidinium-phenol method according to manufacturers' protocols using either of two kits, the Ultraspec-II RNA Isolation System (BIOTECX, Houston, TX) or RNA STAT60 (Tel-Test B, Friendswood, TX).

Human embryonic RNA (a gift from Dr. W. Rainey, UT Southwestern) was isolated by the guanidinium thiocyanate-cesium chloride method (11) from fetal tissue at 15–19 weeks of gestation. Both 5′- and 3′-RACE were performed with the SMART Race cDNA Amplification Kit (CLONTECH, Palo Alto, CA). All PCRs were carried out in a PE 9600 thermocycler (PerkinElmer Life Sciences). Amplified DNA fragments were cloned into either of two vectors, pT-Adv (CLONTECH) or pCR 2.1-TOPO (Invitrogen). Automated sequence analysis was performed by the McDermott Center Core facility (UT Southwestern).

RT-PCR—Random hexamer-primed cDNA was produced with the RT-for-PCR Kit (CLONTECH). Equivalent amounts of cDNA, normalized to glyceraldehyde-3-phosphate dehydrogenase, were amplified with Advantage 2 PCR polymerase (CLONTECH) in a reaction containing 1× Advantage 2 buffer, 0.2 mM each dNTP, and 0.2 μM each oligonucleotide primer. Oligonucleotides used to discriminate between VLGR1a and VLGR1b (see “Results” for terminology) were as follows: VLGR1a (sense), 5′-GTATATGTATGTATCAGGCGGT-3′; VLGR1b (sense), 5′-TTCTATGATTGTTCGTACACTGC-3′; common antisense, 5′-GATCAGCCAACTATTTCTTC-3′. Cycling parameters included an initial denaturation at 95 °C for 2 min and then 10 cycles at 95 °C for 30 s, 69 °C for 30 s, 72 °C for 30 s with a 1 °C annealing temperature stepped down per cycle and then 32 additional cycles at 59 °C annealing temperature.

Oligonucleotides used to distinguish VLGR1b and VLGR1c were as follows: sense, 5′-CAATTGTCGAGTGGCGAGAAATC-3′; antisense, 5′-CGCAGAGCCATCTTCCAGTTCT-3′. Cycling parameters included an initial denaturation at 95 °C for 2 min followed by 36 cycles at 95 °C for 30 s, 68 °C for 1 min, and a final extension at 68 °C for 5 min. Aliquots were withdrawn from reactions at four-cycle intervals. Amplified products were fractionated on either 2% (w/v) Metaphor or 2% agarose. Relative abundance of the amplified fragments was determined by UV fluorescence after ethidium bromide staining. Image analysis was performed with an Eagle Eye II system (Stratagene).

Gene Analysis—Introns 65–82 and 88–89 were amplified with the Expand Long Template PCR System (Roche Molecular Biochemicals) (12) using 26–28-nt primers designed according to the manufacturer’s instructions. Splice junctions of introns 83–87 were amplified with a Genome Walker Kit (CLONTECH) (13).

RESULTS

In Situ Hybridization of VLGR1 to Mouse Embryos—A 1.6-kb mouse VLGR1 cDNA fragment was amplified from kidney total RNA using primers derived from mouse genomic DNA sequences; the cDNA was 80% identical to nucleotides 1957–3561 of the previously reported human VLGR1 open reading frame (7). A subclone was used as a probe for in situ hybridization with a panel of mouse embryo sections (Fig. 1). Very early embryos at day 6.75 showed no expression of VLGR1 in either the embryo or placenta (data not shown). However, by embryonic day 8.0 (Fig. 1A) a strong signal was seen in the ventral aspect of the neural groove, the rostral extremity of the notochord, and the neuroepithelium surrounding the invagination of the developing optic pit. By embryonic day 8.5 (Fig. 1B) VLGR1 was highly expressed in the ventral aspect of the developing floor of the brain. A transverse section at embryonic day 9.5
VLGR1 in the Developing Brain

Fig. 2. Structures of VLGR1 gene and cDNA. A, organization of the human VLGR1 gene. Exons are denoted by vertical lines, and every tenth exon is numbered. Gaps of indeterminate length are marked by doubled vertical dotted lines. Transcriptional start sites are indicated by arrows. B, VLGR1 cDNA structure. This is depicted by the long rectangle; the narrow vertical lines within the rectangle represent intron locations. The cDNA is also numbered every tenth exon. The filled circles represent segments encoding Calx-β repeats; the chevron denotes a segment encoding a PTX domain, and the cluster of seven triangles represents the region encoding the seven putative membrane-spanning segments. Gray bars below the rectangle denote the portions of the cDNA encoding the isoforms, VLGR1α, VLGR1β, and VLGR1γ. C, a diagram of the fragments used to assemble the complete mouse and human cDNA sequences. The thick lines represent the first mouse cDNA clone and the original 6.3-kb human VLGR1α clone, respectively. Fragments recovered by RACE are denoted by arrows, and the fragments recovered by RT-PCR are indicated by thinner lines.

(Fig. 1C) showed strong expression in the ventricular zone of the neuroepithelium in both the fourth ventricle and the telencephalic ventricle. Expression also extended out into the optic stalk. By embryonic day 10.5 (Fig. 1D), VLGR1 was expressed in both the inner layer of the optic cup (the future neural layer of the retina) and the outer layer (the future pigmented layer of the retina).

The same basic pattern of expression in the neuroepithelium, particularly the ventricular zone, continues through embryonic days 11.5–15 (Fig. 1, E–I). VLGR1 also continues to be expressed in almost all layers of the developing eye including the lens but not the cornea. Serial coronal sections at embryonic day 15.5 (Fig. 1, G–I) show VLGR1 expression extending through the brain and into the proximal region of spinal cord, but as neurogenesis slows during late gestation and the ventricular zone diminishes, the region of VLGR1 expression narrows. By adulthood, VLGR1 expression in the brain, as determined by in situ hybridization, is restricted to a subpopulation of cells in the mammillary nuclei of the hypothalamus (data not shown).

Isolation of Long Mouse and Human VLGR1 cDNAs — Because the pattern of expression of VLGR1 suggested that it might play a role in development of the brain, we decided to target the gene in mice, and we tried to determine the sequence of the 5′-end of the mouse cDNA as a first step in this project. We harvested day 14 embryos and dissected away peripheral tissue, leaving mostly the head and spinal region, and prepared total RNA. With the previously isolated 1.6-kb mouse cDNA sequence as a starting point, we performed two cycles of 5′-RACE to isolate additional 5′ sequences. When translated as an amino acid sequence, the new mouse sequence matched well with the human cDNA until a point corresponding to codon 21 of the latter sequence, just 3′ of the region encoding the putative signal peptide. The longest mouse cDNA clone had an open reading frame that extended to the 5′-end of the insert and contained part of another Calx-β repeat.

At the point at which the human and mouse sequences diverged, there was a potential splice acceptor in the human sequence, suggesting that the existing human and mouse sequences represented alternatively spliced transcripts. This interpretation was confirmed by searching the human genome database with the newly isolated mouse sequence. The corresponding human sequence was located 2.0 kb 5′ to the original presumed 5′-end of the gene (Figs. 2 and 3), and it ended with a splice donor sequence. Thus, the original human cDNA represented a transcript originating within an intron of a longer gene.

Four additional rounds of 5′-RACE were performed using mouse cDNA to extend the mVLGR1 sequence for ~13 kb. Each new cDNA fragment overlapped previously isolated sequences by at least several hundred nucleotides, thus confirming recovery of contiguous sequence. A final 5′-RACE reaction failed to extend the sequence, suggesting that the true 5′-end of the mRNA had been reached. The open reading frame in the longest cDNA began with an ATG codon in an adequate context for initiation of translation, and there were multiple stop codons upstream of this ATG. Finally, the 3′-end of the mouse cDNA was isolated in one round of 3′-RACE.
As the mouse cDNA sequence was being assembled, it was compared with the draft human genome database to ensure that aberrant clones were not isolated. Primers based on the corresponding human genomic sequences were then used in RT-PCR on human embryonic brain RNA to isolate the corresponding human VLGR1b cDNA in five segments. The 5' end of the human VLGR1b transcript was confirmed with a single RACE. The full mouse cDNA spanned 19,329 bp and encoded a protein of 6298 amino acids (Figs. 2 and 4), whereas the corresponding human cDNA was 19,320 bp and encoded a protein of 6298 amino acids (Figs. 2 and 4), whereas the corresponding human cDNA was 19,320 bp and encoded a protein of 6298 amino acids (Figs. 2 and 4). The first 29 amino acid residues of VLGR1b are mostly of seven Calx-

sections thereof. Of these, 23 are readily identified by the NCBI DART or CD-Blast utilities (14, 15), whereas the remainder are found using BLAST (11) to compare the ectodomain with itself. These repeats are spaced an average of 124 residues apart. The C-terminal 40 residues of each repeat comprise the most highly conserved portion of the 100-residue Calx-β motif (Fig. 4).

Based on distances between identifiable repeats within the amino acid sequence, there are potential spaces for nine additional repeats in which, nonetheless, no Calx-β homology can be identified. However, residues 1338–1498 (human VLGR1b numbering) have significant similarity to a pentraxin (PTX) domain. Such domains have a three-dimensional structure closely resembling legume lectins (16). Proteins with pentraxin domains include serum proteins such as serum amyloid P- and C-reactive proteins as well as several neuronal proteins (17). Serum amyloid P forms a pentameric complex and binds a wide variety of protein and nonprotein substances in a calcium-dependent manner. It is unclear whether the PTX domain in VLGR1b would bind calcium, since it lacks conserved residues known to mediate binding (16). Moreover, this domain lacks a highly conserved pentraxin signature sequence (HXCS/TWXK). Thus, the functional significance of the putative PTX domain in VLGR1b is uncertain.

As previously noted, the 7-TM domain of VLGR1 is similar to members of the secretin family of G protein-coupled receptors and particularly similar to members of a subfamily of receptors with large ectodomains. In all such receptors, the ectodomain is separated from the 7-TM domain by a G protein-coupled receptor proteolytic sequence at which the ectodomain can be specifically cleaved (18). Indeed, the ectodomain of VLGR1a can be cleaved when expressed in cultured cells, although the exact site of cleavage has not been determined (7).
VLGR1 gene corresponding to the original VLGR1a transcript was elucidated mainly by long range PCR techniques (12, 13). The remainder of the gene structure was deduced by queries to the NCBI draft human genome data base (19). The VLGR1 gene consists of 90 exons (Figs. 2 and 4); VLGR1a transcripts originate in intron 64 (Figs. 2 and 3). Exons 1 and 2 encode the putative signal peptide, and exons 3–74 encode Calx-β repeats, with the exception of exon 20, which encodes most of the putative PTX domain. Although the putative ectodomain of VLGR1 has a highly repetitive structure, there is no obvious correspondence between the occurrence of Calx-β repeats and exons. Each repeat is encoded by an average of ~2 exons, but introns are located neither at conserved positions within repeats nor at consistent positions with respect to the reading frame of translation. The overall intron distribution with respect to the reading frame is typical for the human genome (20), with 46 introns in “frame 0” (between codons), 27 in frame 1, and 16 in frame 2.

Exons 82–88, which encode the putative G protein-coupled receptor proteolytic sequence and seven-transmembrane segments, are very widely spaced over ~300 kb, not including two gaps of indeterminable length between exons 83 and 84 and between 87 and 88.

Although the total size of the VLGR1 gene cannot yet be precisely determined due to the gaps, it is ~600 kb based on the available human genome sequence (19). The 3′-end of VLGR1 and DSS618, a linkage marker flanking the 5′-end, are both on the same yeast artifical chromosome “Mega-YAC” clone 944B4 from the CEPH library (21) (not shown). Since this clone has an insert of 1050 kb, this sets an upper boundary to the size of the gene.

Abundance of Alternative Human VLGR1 Transcripts—Several human cDNA clones encoding VLGR1b had an 83-bp deletion corresponding to the use of a cryptic splice donor in exon 31. This organization shifted the reading frame and brought a stop codon into frame seven codons later (Fig. 3). These aberrantly processed transcripts potentially encode a 2296-amino acid protein with a signal peptide (but no other transmembrane segments), 15 Calx-β domains, and the PTX domain. We termed this putative truncated form of VLGR1 as VLGR1c.

To determine the relative abundance of VLGR1 isoforms, RNA from various human fetal tissues was subjected to RT-PCR; multiple aliquots were removed to ensure sampling during the exponential phase of each reaction (Fig. 5). Reactions to distinguish VLGR1a from VLGR1b used a common antisense primer and two transcript-specific sense primers. VLGR1b was ~4 times more abundant than VLGR1a in most tissues tested, despite wide variations in absolute levels of expression. This result suggests that the same transcriptional regulatory sequences may control VLGR1a and VLGR1b expression, despite the ~300-kb distance between the start of the VLGR1b mRNA and initiation of the VLGR1a mRNA within intron 64. The relative levels of VLGR1a and VLGR1b transcripts in human fetal brain were confirmed by an RNase protection assay (data not shown).

Reactions to distinguish VLGR1b and VLGR1c used a single primer pair. The results suggested that VLGR1c transcripts were expressed at ~1.5 times VLGR1b levels in most tissues examined (Fig. 5). In human fetal testis, VLGR1c is expressed almost exclusively.

Thus, a minority of mRNA transcripts encode full-length human VLGR1b. In contrast, mice predominantly express VLGR1b. We cloned and sequenced the 3′-end of mouse intron 64 (data not shown) and found that it does not encode a peptide with any similarity to the putative signal peptide specified by the corresponding sequence in human intron 64. In addition, there is no in-frame ATG initiation codon within the mouse intron. These data argue against the existence of the homologous mouse isoform of VLGR1a. Mouse VLGR1c transcripts are present at levels only ~0.5 times those of VLGR1b in fetal brain (data not shown).

DISCUSSION

Structure of VLGR1—The extremely large size of VLGR1b was unexpected. The long form of human VLGR1 was not detected in our previous studies because the primer used for a final confirmatory round of 5′-RACE was located 5′ of the splice acceptor of intron 64 and thus could only amplify transcripts originating in this intron. The 5′-RACE protocol used in the present study permitted isolation of much longer cDNA fragments at each step. Thus, it is unlikely that we missed even longer isoforms of VLGR1 due to failure of the 5′-RACE.

At ~6300 amino acid residues, VLGR1b is one of the largest proteins known, being comparable in size with giant muscle proteins such as obscurin and nebulin (both ~6600 residues) but smaller than titin (~27,000 residues) (22, 23). To the best of our knowledge, it is the largest protein putatively expressed on the cell surface. In particular, it is considerably larger than Flamingo, its closest rival among putative seven-transmembrane segment receptors, which is only ~3500 residues (5). It is not known whether the VLGR1 ectodomain exists as a complex with additional proteins that might further increase its apparent size.

At present, there is no known three-dimensional structure for a Calx-β repeat unit, so it is difficult to predict the size or shape of the VLGR1 ectodomain. If Calx-β repeats were globular, they would each be ~4 nm in diameter. If they were arranged in the VLGR1 ectodomain like beads on a string, the VLGR1b ectodomain would be ~180 nm long, comparable in length with a myosin filament.

Evolution of VLGR1—How did VLGR1 evolve? It has a highly repetitive ectodomain consisting mostly of Calx-β repeat units, followed by a putative 7-TM domain that significantly resembles other putative G protein-coupled receptors in the secretin family. Thus, VLGR1 presumably evolved by duplication of exons encoding one or a few Calx-β domains, with juxtaposition to a gene encoding a 7-TM receptor at some time during this process. There is no obvious correspondence, however, between repeat units and exon organization. Perhaps VLGR1 has existed for sufficient time that the original gene organization has been obscured by activation of cryptic splice sites or other mechanisms. The marked amino acid sequence divergence of the repeat units is consistent with this idea.

No homolog of VLGR1 exists in the Drosophila melanogaster or Caenorhabditis elegans genomes (24, 25). This absence suggests that VLGR1 arose after the common ancestor of arthropods and nematodes diverged from the ancestor of chordates during the Vendian (pre-Cambrian) epoch ~550 million years ago (26). On the other hand, the VLGR1 gene of the pufferfish Tetraodon nigroviridis is apparently organized similarly to the human gene; 29 exons with significant homology to human VLGR1 exons (the first corresponding to human exon 4, the last to human exon 80) can be identified in data bases, and all 42 introns that can be identified are identical to the corresponding human introns in position and phase. Thus, the current organization of VLGR1 was apparently established before the divergence of the ancestors of tetrapods and teleost fish. This conservation implies strong selective pressure early in the evolution of VLGR1 toward divergence of both intron position and amino acid sequence of the Calx-β repeats, followed by strong selection toward conservation of these properties. An obvious evolutionary impetus for repeat units to diverge would be stabilization of the number of such units; a gene with a large number of highly homologous exons would be prone to deletions.
due to unequal meiotic crossing over. The maintenance of a highly repeated structure encoded by divergent sequences implies that the extremely large size of the VLGR1 ectodomain is crucial to its function.

Speculations on the Function of VLGR1

Whereas the function of VLGR1b remains unclear, the presence of multiple Calx/H9252 repeats in the ectodomain suggests a role in protein-protein interaction that is perhaps calcium-mediated. A precedent has been set with the Calx/H9252 repeats in the highly polymorphic marine sponge aggregation factor. By both in vitro and in vivo analysis, marine sponge aggregation factor has been shown to mediate species-specific aggregation of sponge cells in Figure 4.

Amino acid alignment of human (top) and mouse (bottom) VLGR1. The intron locations and phases are represented by triangles (phase 0), open circles (phase 1), and closed circles (phase 2). The double underline at the N terminus depicts a putative signal sequence. The single underline denotes the 40-amino acid core Calx/H9252 repeats. The seven double underlines at the C terminus indicate the seven-transmembrane segments.

Fig. 4. Amino acid alignment of human (top) and mouse (bottom) VLGR1. The intron locations and phases are represented by triangles (phase 0), open circles (phase 1), and closed circles (phase 2). The double underline at the N terminus depicts a putative signal sequence. The single underline denotes the 40-amino acid core Calx/H9252 repeats. The seven double underlines at the C terminus indicate the seven-transmembrane segments.
a calcium-dependent manner (27).

As detected by in situ hybridization, VLGR1 is expressed predominantly in the central nervous system during development. Embryos at gastrulation, with all three germ layers evident and a clearly defined primitive streak (precursor to the neural groove), showed no VLGR1 expression. The earliest VLGR1 expression at embryonic day 8.0 is coincident with neurulation, which commences at about embryonic day 7.5 with the formation of the neural plate and then the neural groove. By mid-gestation, strong VLGR1 expression is seen in the ventricular zone, a specialized layer of neuroepithelium surrounding all of the ventricles in the brain. The ventricular zone is home to a proliferating population of neuronal precursor cells. As the cells differentiate and become postmitotic, they migrate out of the ventricular zone, moving radially along glial tracts to populate the cortex. During migration, age-dependent stratification occurs such that the younger cells bypass older cells and populate the more distal regions (28).

It is likely, therefore, that VLGR1 is a marker for a neuronal progenitor cell type. Indeed, VLGR1 expression in the narrowed ventricular zone at late gestation, when neurogenesis is essentially complete, is consistent with a recent report describing a potential pool of multipotent progenitor cells that remain into the early postnatal period (29). These expression data, considered together with data suggesting that VLGR1 exists only in chordates and perhaps only in vertebrates (i.e. animals with well developed central nervous systems) strongly imply an important role for VLGR1 in the development of the central nervous system. VLGR1 may have additional functions in the development of the eye, considering that it is expressed in the lens, which is not of neural origin.

But if VLGR1 is important for development of the central nervous system and if the strong conservation of the VLGR1b ectodomain in vertebrates reflects its importance, what functional roles can be ascribed to the alternative forms, VLGR1a and VLGR1c? VLGR1a apparently exists only in humans, and VLGR1c is expressed at considerably lower levels in mice than in humans relative to VLGR1b. Thus far, the mouse and human genomes appear to be very similar, raising the obvious question of what genetic differences...
VLGR1 in the Developing Brain

FIG. 5. Relative levels of expression of VLGR1 isoforms in human embryonic tissues. Ethidium bromide staining of RT-PCR fragments of RNA from the indicated tissues. The upper panel displays reactions that distinguish VLGR1b (921 bp) and VLGR1c (838 bp). Aliquots were taken at 24 cycles (lane 1), 28 cycles (lane 2), 32 cycles (lane 3), and 36 PCR cycles (lane 4). Measured at 32 cycles, the VLGR1b/VLGR1c ratio is as follows: lung, 1.0; testis, 0.0; kidney, 0.7; heart, 0.5; brain, 0.9. The reactions displayed in the lower panel distinguish VLGR1b (555 bp) and VLGR1a (515 bp). Aliquots were taken at 20 cycles (lane 1), 24 cycles (lane 2), 28 cycles (lane 3), and 32 PCR cycles (lane 4). Measured at 24 cycles, the VLGR1b/VLGR1a ratio is ~4 in all tissues except testis and liver.

ences between these species are responsible for development of the larger and more complex human cerebral cortex. Although changes might be postulated in amino acid sequence or expression patterns of transcriptional or morphogenetic factors, developing species-specific splice variants could provide a way to dramatically change the functional properties of such factors with very few nucleotide changes. For example, in humans, the N- and C-terminal portions of the ectodomain (residues 1–2296 and 4345–5892) are potentially present in molar excess over the remaining central portion due to the expression levels of VLGR1a and VLGR1c. Such excess might permit novel interactions of the central portion with other cell surface proteins, particularly if VLGR1 mediates homotypic interactions, as is the case for Flamingo (5), or if it functions as a homodimer, as has been shown for many other G protein-coupled receptors (30). However, we must know much more about the function and ligand(s) of VLGR1 to define its role in the development of the central nervous system.

Acknowledgments—We thank Heli Nikkila (who also isolated the first mouse cDNA clone), Leigh Pascoe, Kathleen Curnow, and Michael Si for preliminary studies on the structure of the VLGR1 gene; Bill Rainey for human embryonic RNA samples; and John Shelton and the Molecular Pathology Core for assistance with the preparation of mouse embryo sections and in situ hybridizations. We thank David Ruisell for helpful comments on the manuscript.

Note Added in Proof—A naturally occurring mutation of VLGR1 has been reported in the Frings mouse strain, which is prone to audiogenic seizures (31). This supports our speculation that VLGR1 is important for normal development of the central nervous system. The mutation is a deletion of 6835G of the cDNA (within exon 31), converting amino acid V2250 to a stop codon. It should prevent synthesis of VLGR1b but will truncate VLGR1c (the form encoded by transcripts containing the 83 nt deletion due to alternate splicing of intron 31) by only 63 amino acids. This mutation is reported as occurring in a gene termed MASS1, which has several transcripts, the longest of which actually consists of exons 6–39 of VLGR1 with initiation and termination of transcription in introns 5 and 39, respectively (31). We now have determined the frequency at which such transcripts occur in fetal and adult mouse brain mRNA. Transcripts originating in intron 5 are undetectable up to at 32 PCR cycles, whereas those terminating in intron 39 comprise ~20% of transcripts. Thus MASS1 transcripts are expressed, if at all, at low levels compared with VLGR1.

REFERENCES
1. Ji, T. H., Grossmann, M. & Ji, I. (1998) J. Biol. Chem. 273, 17299–17302
2. Beckert, J. & Pin, J. P. (1999) EMBO J. 18, 1723–1729
3. Stacey, M., Lin, H. H., Gordon, S. & McKnight, A. J. (2000) Trends Biochem. Sci. 25, 284–289
4. Gao, F. B., Kohwi, M., Brennan, J. E., Jan, L. Y. & Jan, Y. N. (2000) Neuron 27, 532–537
5. Usui, T., Shimn, Y., Shimada, Y., Hirano, S., Burgess, R. W., Schwartz, T. L., Takeichi, M. & Uemura, T. (1999) Cell 96, 585–595
6. Sudhof, T. C. (2001) Annu. Rev. Neurosci. 24, 933–962
7. Nikkila, H., McMillan, D. R., Nunez, B. S., Pascoe, L., Curnow, K. M. & White, P. C. (2000) Mol. Endocrinol. 14, 1351–1364
8. Schwartz, E. M. & Benzer, S. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 10249–10254
9. Fernandez-Busquets, X. & Burger, M. M. (1997) J. Biol. Chem. 272, 27839–27847
10. Shelton, J. M., Lee, M. H., Richardson, J. A. & Patel, S. B. (2000) J. Lipid Res. 41, 532–537
11. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) Biochemistry 18, 5294–5299
12. Cheng, S., Fockler, C., Barnes, W. M. & Higuchi, R. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 5695–5699
13. Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A. & Lukyanov, S. A. (1995) Nucleic Acids Res. 23, 1087–1088
14. Schultz, J., Copley, R. R., Doerk, T., Ponting, C. P. & Bork, P. (2000) Nucleic Acids Res. 28, 231–234
15. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000) Nucleic Acids Res. 28, 263–266
16. Emsley, J., White, H. E., O’Harra, B. P., Oliva, G., Srivivasan, N., Tickle, I. J., Bublendt, T. L., Pepys, M. B. & Wood, S. F. (1994) Nature 367, 338–345
17. Kirkpatrick, L. L., Matzuk, M. M., Dodds, D. C. & Perin, M. S. (2000) J. Biol. Chem. 275, 17796–17792
18. Sugita, S., Ichtchjenko, K., Khvotchev, M. & Sudhof, T. C. (1998) J. Biol. Chem. 273, 26715–26724
19. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Furke, R., Gage, D., Harris, K., Hesford, A., Howland, J., Kasas, L., Lehoucky, J., Rice, R., McKwan, P., McKernan, K., et al. & the International Human Genome Sequencing Consortium (2001) Nature 409, 860–921
20. Long, M., de Souza, J. S., Rosenberg, C. & Gilbert, W. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 219–220
21. Cohen, D., Chumakov, I. & Weissbach, J. (1993) Nature 366, 698–701
22. Keller, T. C., III (1995) Curr. Opin. Cell Biol. 7, 32–38
23. Young, P., Eker, E. & Gautel, M. (2001) J. Cell Biol. 154, 123–136
24. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoksins, R. A., Galie, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutten, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., et al. & the International Human Genome Sequencing Consortium (1999) Nature 401, 1087–1092
25. Hodgkin, J., Plasteck, R. H. & Waterston, R. H. (1996) Science 270, 410–414
26. Conway, M. S. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 4426–4429
27. Jarchow, J. & Burger, M. M. (1998) Cell Adhes. Commun. 6, 405–414
28. McConnell, S. K. (1989) Brain Res. 472, 1–23
29. Levers, T. K., Edgar, J. M. & Pracejus, D. J. (2001) J. Neurobiol. 48, 265–267
30. Bovier, M. (2001) Nat. Rev. Neurosci. 2, 274–286
31. Skradski, S. L., Clark, A. M., Jiang, H. M., White, H. S., Fu, Y. H., and Ptashe, L. J. (2001) Neuron 31, 537–544.
Very Large G Protein-coupled Receptor-1, the Largest Known Cell Surface Protein, Is Highly Expressed in the Developing Central Nervous System
D. Randy McMillan, Kathleen M. Kayes-Wandover, James A. Richardson and Perrin C. White

J. Biol. Chem. 2002, 277:785-792.
doi: 10.1074/jbc.M108929200 originally published online October 17, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M108929200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 31 references, 13 of which can be accessed free at http://www.jbc.org/content/277/1/785.full.html#ref-list-1