Tandem genomic arrangement of a G protein (*Gna15*) and G protein-coupled receptor (*s1p4/lpC1/Edg6*) gene

James J.A. Contos, Xiaoqin Ye, Valerie P. Sah, Jerold Chun*

Department of Pharmacology, School of Medicine, University of California at San Diego, La Jolla, CA 92093-0636, USA

Received 31 May 2002; revised 30 July 2002; accepted 6 September 2002

First published online 18 September 2002

Edited by Edward A. Dennis, Isabel Varela-Nieto and Alicia Alonso

Abstract **A genomic analysis of the *s1p4/lpC1/Edg6* mouse sphingosine-1-phosphate (S1P) G protein-coupled receptor gene revealed it to be located on central chromosome 10 and to consist of two exons with an intronless coding region. Surprisingly, we found the gene encoding the promiscuously coupling $G_{\alpha 15}$ protein (*Gna15*) located in tandem just upstream, an arrangement conserved in the human genome (on chromosome 19p13.3). Given that Northern blots demonstrated similar tissue distributions of the mouse *s1p4* and *Gna15* transcripts, we propose that transcription of the two genes may be under control of the same enhancer elements and that their protein products may couple in vivo.**

Key words: Lysophospholipid; Sphingosine; Sphingosine-1-phosphate; Signal transduction; Genomics; Mouse

1. Introduction

The *s1p4/lpC1/Edg6* gene is part of a family of eight related genes that encode G protein-coupled receptors (GPCRs) specifically activated by lysophospholipid molecules such as lysophosphatidic acid (LPA) and sphingosine-1-phosphate (S1P). Various cell types such as activated platelets and postmitotic neurons produce LPA and/or S1P [1,2]. These signaling molecules induce proliferative, morphological, and cell migratory changes on most cells and are believed to be involved in multiple biological processes, including neurogenesis, myelination, wound healing, angiogenesis, and immune system functions [3–5]. Of the eight related genes encoding receptors for LPA and S1P, three are specific for LPA (*lpa1–3*) and five for S1P (*s1p1–5*). Sequence relationships among the mouse genes clearly group the three LPA receptor genes (with 45–55% amino acid sequence identity), but only do so for four of the five S1P receptor genes (also with 43–55% amino acid sequence identity) [4]. The *s1p4* gene was difficult to place into either of these subclasses, with slightly more identity to S1P receptor genes (~38%) than to LPA receptor genes (~34%), suggesting it was an S1P receptor gene. Subsequent ligand activation studies confirmed that it encodes a receptor specifically activated by S1P [6,7].

Genomic structure analysis of *s1p4* could provide insight into the evolution of the eight lysophospholipid receptor genes. The coding regions for each of the *lpa* genes are divided between two exons, whereas for the *s1p1–3* genes, the coding region of each gene is within single exon, with only non-coding exon(s) upstream [8–12]. One would expect the genomic structure of *s1p4* to be similar to the other *s1p* genes. However, to date, genomic structure information has not been reported for *s1p4*.

Lysophospholipid receptors, as well as all GPCRs, couple to heterotrimeric G proteins, which consist of α, β and γ subunits. There are 15 types of mammalian G_α subunits, which are classified into four groups based on sequence similarity and general function [13]. Heterologous expression studies have demonstrated that most LP receptors can couple to multiple types of G_α proteins, including those in the $G_{i/o}$, $G_{12/13}$, and G_q classes, but not in the G_s class [4,5]. Although the $S1P_4$ receptor can couple to $G_{i/o}$ class proteins, its coupling to G_q class proteins has not been thoroughly examined [6,7].

The G_q class of G proteins (including G_q, G_{11}, G_{14}, and $G_{15/16}$) activates phospholipase C (PLC). PLC hydrolyzes membrane phospholipids, leading to production of inositol phosphates and diacylglycerol, which in turn induce protein kinase C activation and increases in cytosolic Ca^{2+}. G proteins typically show receptor specificity, in that they will couple only subsets of receptors to their effector proteins. Such receptor specificity is necessary, given the ubiquitous expression of most G_α proteins. The $G_{\alpha 15}$ protein is unique in that it couples promiscuously to nearly all GPCRs examined [14,15]. This promiscuous coupling of the $G_{\alpha 15}$ protein to GPCRs has been used to identify novel ligands for orphan GPCRs [16,17]. It is hypothesized that the receptor specificity of $G_{\alpha 15}$ lies in its highly restricted expression pattern, which is confined to hematopoietic tissues and lung [14].

We set out to characterize the mouse *s1p4* gene in order to understand its evolution, regulation, and function. In the process of analyzing the promoter, we discovered that the *Gna15* gene (encoding $G_{\alpha 15}$) was located in tandem immediately upstream. We then showed that the two genes were coexpressed in the same tissues, suggesting that the same local enhancer elements control transcription of both genes. If the dual expression is in the same cells, $S1P_4$ may functionally couple to $G_{\alpha 15}$ in vivo.

2. Materials and methods

2.1. Chromosomal mapping

An *Msp*I restriction fragment length polymorphism between *Mus*

*Corresponding author. Present address: Merck Research Labs MRLSDB1, 3535 General Atomics Court, San Diego, CA 92121, USA. Fax: (1)-858-202 5814.
E-mail address: jerold_chun@merck.com (J. Chun).

0014-5793/02/$22.00
PII: S0014-5793(02)03409-9

musculus (C57BL6/JEi) and *M. spretus* (SPRET/Ei) genomic DNA was identified 3′ of the poly(A) site in the *s1p4* gene. PCR, restriction digestion, and backcross panel analysis were done as described previously [9], except primers used to amplify genomic DNA were edg7h (5′-CGTGTTTAAGAATGAAAGGG-3′) and edg7n (5′-GGAGTTGTAGGCACACTTA-3′). Raw data can be viewed at http://www.jax.org/resources/documents/cmdata.

2.2. Genomic clone isolation and characterization

A PCR strategy [8] was used to isolate five mouse 129/SvJ genomic λ clones containing the *s1p4* gene. Primers were used that amplified a single 324 bp product from the *s1p4* gene coding region: edg7a (5′-CTGCTGCCCCTCTACTCCAA-3′) and edg7b (5′-ATTAATGGCTGAGTTGAACAC-3′). A 7.0 kb *Xho*I/*Not*I subclone containing the entire *s1p4* gene was manually sequenced entirely in both directions and was deposited in the EMBL database (accession number AJ489247).

2.3. cDNA clone analysis

5′ and 3′ rapid amplification of cDNA ends (RACE) was performed as described previously [9], with the exception of different gene-specific primers and adult spleen cDNA as template. The products were identical to a previously sequenced mouse cDNA clone (accession number AJ006074), although the 5′-RACE products terminated ~400 bp downstream. Expressed sequence tags were also used to align and determine gene transcript sequences of *s1p4* (accession numbers AA155468, AA155471, AA254425, AA451451, AI158066, AI158682, AI463732, AI481372, AI613663, AI645838, AI661326, AV079456, AV081387, AV081616, AV315591) and *Gna15* (AA571788, AA762974, AA959901, AI461852, AW492116).

2.4. Human sequence analysis

All sequences were downloaded from GenBank and analyzed using DNasis software. Accession numbers of human genomic clones containing the *Gna11*, *Gna15*, and *s1p4* (*EDG6*) genes were, respectively: AC005262, AC005264, and AC011547. Accession numbers of sequences used to determine the entire transcript sequences were *Gna11* (XM_009221, BF514534, BE873173, BE795320, BE395761, BE275993, AW375193, AI344423, AI097506, AA471045, BE885460), *Gna15* (NM_002068, AI660568, AI817049), and *s1p4* (NM_003775, BF663028, BF974516, AI869921, AI766542).

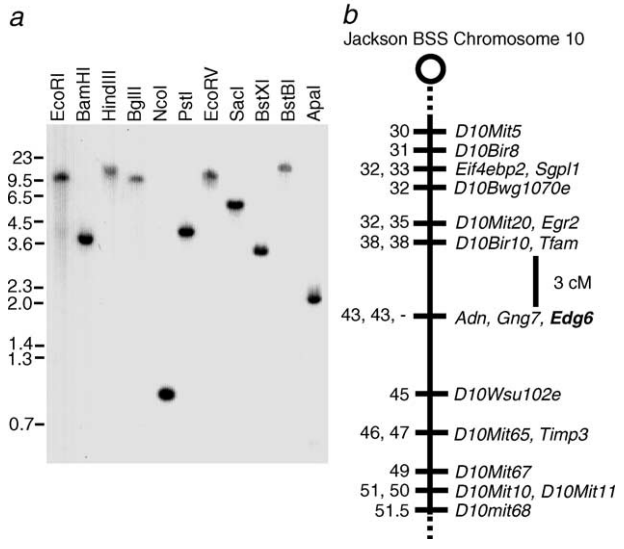

Fig. 1. Copy number and chromosomal mapping of the *s1p4*/*Edg6* gene. a: Southern blot of *M. musculus* genomic DNA (10 μg/lane) digested with the indicated restriction enzymes and probed with a fragment from the open reading frame of the *s1p4* gene. A single fragment hybridizing in each lane indicates the gene is single copy. b: Linkage map showing the *s1p4*/*Edg6* gene in the context of other genes mapped using the Jackson BSS panel (to the right) and their cM positions in the Mouse Genome Database (to the left). Genes also mapping to cM 43 include *gz*, *ji*, *mh*, *Gna11*, and *Gna15* (not shown).

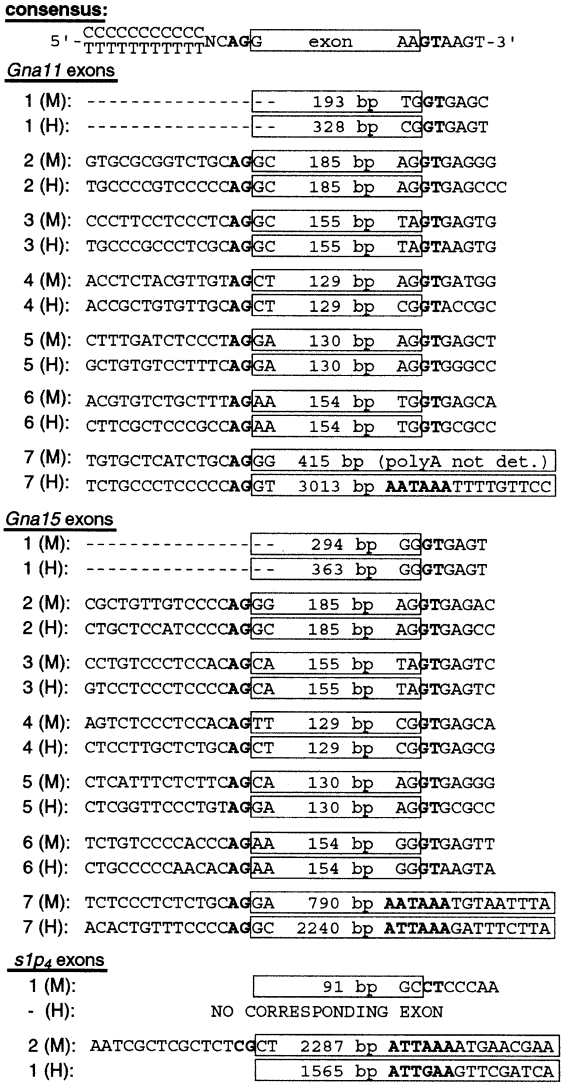

Fig. 2. Intron/exon boundaries and polyadenylation sites in the mouse (M) and human (H) *Gna11*, *Gna15*, and *s1p4* genes. Boxes indicate exons found in cDNA sequences with total bp in the exon noted. The nearly invariant AG and GT sequences that flank exons are shown in bold, as are the putative polyadenylation signal sequences of the terminal exons. The consensus polyadenylation signal sequence is AATAAA or ATTAAA.

2.5. Southern and Northern blot analysis

Preparation and probing of both the Southern and Northern blots were previously described [8,12]. To detect *s1p4* and *Gna15* genes, DNA fragments amplified from coding regions of the cDNAs were used.

3. Results and discussion

3.1. Chromosomal mapping of the mouse s1p4 gene

To further characterize the *s1p4* gene, we first determined that it was present as a single copy (Fig. 1a) and that it cosegregated with various markers and genes at cM 43.0 of chromosome 10 (Fig. 1b). Several genes that have not been identified are localized to this chromosomal region, including mocha (*mh*), grizzled (*gr*), and jittery (*ji*), which might be related to mutations in the *s1p4* gene. Interestingly, three oth-

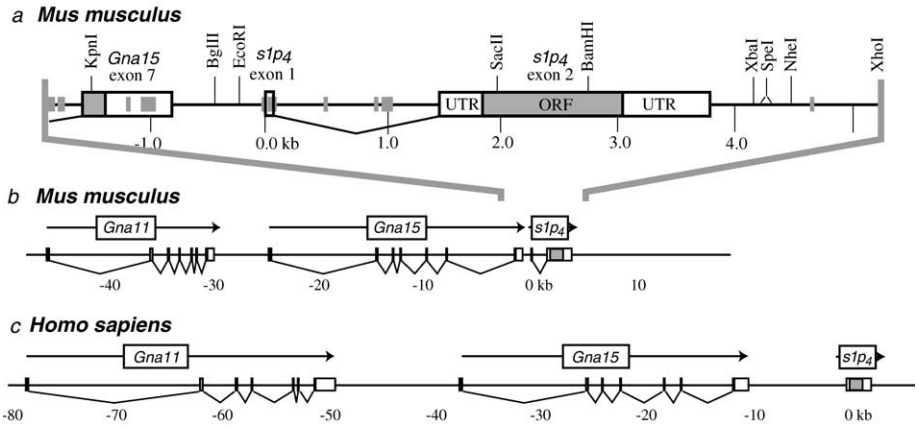

Fig. 3. Genomic maps. a: Genomic restriction map of the sequenced subclone containing the mouse *s1p4* gene, including part of the *Gna15* gene. Large rectangles represent exons and shading within them open reading frame (ORF). Smaller shaded rectangles represent repetitive elements. b,c: Genomic maps of the mouse (b) and human (c) regions containing the *Gna11*, *Gna15*, and *s1p4* genes. All distances are shown to scale.

er genes previously mapped to cM 43.0 encode G proteins to which $S1P_4$ might couple: *Gna11* and *Gna15*, which encode the G_q class $G_{\alpha 11}$ and $G_{\alpha 15}$ proteins, and *Gng7*, which encodes the $G_{\gamma 7}$ protein [18,19]. Each of the *s1p4*, *Gna11*, *Gna15*, and *Gng7* genes is also located at the same chromosomal locus on human chromosome 19p13.3 [20], suggesting conserved genomic arrangements in mammals.

3.2. cDNA sequence and genomic structure of the *s1p4* gene

To fully characterize the mouse *s1p4* gene, we compared cDNA with genomic sequences. The complete cDNA sequence was determined by aligning 5′- and 3′-RACE products, cDNA clones, and expressed sequence tags. All clones consistently terminated just downstream of the same polyadenylation signal sequence (Fig. 2). The 5′ end of the longest cDNA clone was 492 bp upstream of the start codon, although all RACE products terminated <100 bp upstream of the start codon. To determine the structure of the gene, genomic clones were isolated from a mouse 129/SvJ genomic DNA library. A 7.0 kb subclone, containing the complete *s1p4* gene, was sequenced entirely in both directions (Fig. 3a). The longest cDNA sequence was distributed between two exons of 91 and 2294 bp, with the intron located 401 bp upstream of the start codon (Fig. 3a). There was no TATA box in the vicinity of the putative transcription start site. Interestingly, the intron/exon boundary sequences did not conform to known consensus sequences, although the polyadenylation signal sequence did (Fig. 2). The coding region was uninterrupted, similar to other *s1p* genes, supporting the hypothesis that the *s1p4* gene diverged from an ancestral S1P receptor gene rather than from an ancestral LPA receptor gene.

We also compared human *s1p4* cDNA with genomic sequences. Like the mouse *s1p4* gene coding region, the human *s1p4* gene coding region is intronless. However, unlike mouse, the human gene does not have an initial exon encoding 5′-untranslated region (UTR) (Fig. 2). In addition, the polyadenylation signal sequence differed slightly from the consensus (Fig. 2).

3.3. Tandem genomic arrangement of the *Gna15* and *s1p4* genes in human and mouse

Surprisingly, BLAST searches of the *s1p4* putative promoter area revealed part of the deposited *Gna15* cDNA sequence to be located just upstream of *s1p4* exon 1 (Fig. 3a). By aligning expressed sequence tag sequences with the cDNA, we determined the remaining $\sim$600 bp of the *Gna15* cDNA sequence (i.e. 3′-UTR sequence through the polyadenylation site), which terminated $\sim$800 bp upstream of *s1p4* exon 1 (Fig. 2). The *Gna15* polyadenylation signal sequence conformed strongly to the consensus sequence (Fig. 2). It was previously determined that both the mouse *Gna15* and *Gna11* genes consisted of seven exons and were arranged in tandem over a total of 45 kb [21]. A genomic map of the region is shown in Fig. 3b.

We also characterized the human genomic region containing the *s1p4*, *Gna15*, and *Gna11* genes, using sequences deposited as part of the human genome project, expressed sequence tags, and cDNAs (Figs. 2 and 3c). Like the mouse genes, human *Gna11* and *Gna15* genes also had seven exons in the same relative positions of the cDNA, although the 3′-UTRs were significantly longer in both genes. All of the intron/exon boundaries conformed to consensus sequences (Fig. 2). As described above, the human *s1p4* gene consisted of only a single exon, without TATA box elements in the putative transcription initiation region. All three genes (*Gna11*, *Gna15*, and *s1p4*) were arranged in tandem, as in mouse (Fig. 3c). However, the spacing between the genes and many of the exons was larger, making the entire three-gene cluster occupy 80 kb rather than 50 kb.

3.4. Similar expression patterns of the *Gna15* and *s1p4* genes

Previous analyses demonstrated the mouse *Gna11* gene to be ubiquitously expressed [22] and the mouse *Gna15* and human *s1p4* genes expressed primarily in hematopoietic areas [22,23]. This suggested that the mouse *Gna15* and *s1p4* genes might have identical tissue distribution patterns. We compared the expression pattern of the two genes using Northern

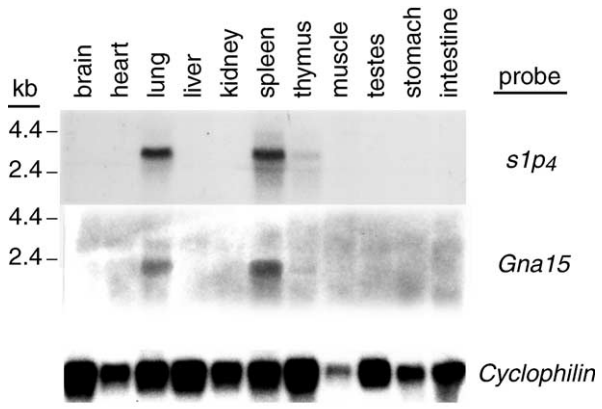

Fig. 4. Expression of the mouse *s1p4* and *Gna15* gene transcripts. A Northern blot of total adult mouse RNA (20 μg/lane) was first hybridized with a *s1p4* gene probe, then with a cyclophilin probe, and finally with a *Gna15* gene probe (the blot was exposed and stripped in between each hybridization). The differently sized *s1p4* and *Gna15* gene transcripts have identical tissue distribution levels, with highest levels found in spleen and lung and faint levels found in thymus.

blot, which demonstrated that the two transcripts were coexpressed at the same relative levels in all tissues examined, with the highest levels in adult spleen and lung (Fig. 4).

3.5. Discussion of coexpression of the Gna15 and s1p4 genes

Although our results suggest that the *Gna15* and *s1p4* transcripts are expressed in the same cells, definitive conclusions can only be drawn after further analyses using high resolution in situ hybridization or single cell transcript analysis. However, should it be true, it would implicate the same local enhancer elements in controlling transcription of both genes. This might occur through enhancers acting on two distinct promoters: one at the transcription start site of the *Gna15* gene and another at the transcription start site of the *s1p4* gene. Alternatively, both mature *Gna15* and *s1p4* gene transcripts might be produced from a single transcription unit, with enhancers acting on a single promoter at the start of the *Gna15* gene. In support of this mechanism is the report that transcription in mammalian cells often continues several kb past a polyadenylation site before RNA polymerase II disengages the DNA [24].

In addition, because $G_{\alpha 15}$ is known to couple promiscuously to GPCRs [14,15], coexpression of $S1P_4$ and $G_{\alpha 15}$ in the same cells would implicate $G_{\alpha 15}$ as a normal coupling partner to $S1P_4$ in vivo. Such coupling would link extracellular S1P signals to pertussis toxin-insensitive increases in cytosolic Ca^{2+}, which S1P is known to stimulate in many cell types [5]. This might explain the finding that erythroid cell differentiation is decreased when the human $G_{\alpha 15}$ protein (called $G_{\alpha 16}$) is inhibited or downregulated in the presence of serum [25], which is known to contain S1P [1].

Acknowledgements: We thank Dr. Joshua Weiner for help with the Northern blot and Carol Akita for expert technical assistance. This work was supported by a grant from the National Institute of Mental Health (Grant K02MH01723) and an unrestricted gift from Merck.

References

[1] Yatomi, Y. et al. (1997) J. Biochem. (Tokyo) 121, 969–973.
[2] Fukushima, N., Weiner, J. and Chun, J. (2000) Dev. Biol. 228, 6–18.
[3] Contos, J.J.A., Ishii, I. and Chun, J. (2000) Mol. Pharmacol. 58, 1188–1196.
[4] Fukushima, N., Ishii, I., Contos, J.J., Weiner, J.A. and Chun, J. (2001) Annu. Rev. Pharmacol. Toxicol. 41, 507–534.
[5] Hla, T., Lee, M.J., Ancellin, N., Paik, J.H. and Kluk, M.J. (2001) Science 294, 1875–1878.
[6] Van Brocklyn, J.R., Graler, M.H., Bernhardt, G., Hobson, J.P., Lipp, M. and Spiegel, S. (2000) Blood 95, 2624–2629.
[7] Yamazaki, Y. et al. (2000) Biochem. Biophys. Res. Commun. 268, 583–589.
[8] Contos, J.J. and Chun, J. (1998) Genomics 51, 364–378.
[9] Contos, J.J. and Chun, J. (2000) Genomics 64, 155–169.
[10] Contos, J.J. and Chun, J. (2001) Gene 267, 243–253.
[11] Liu, C.H. and Hla, T. (1997) Genomics 43, 15–24.
[12] Zhang, G., Contos, J.J.A., Weiner, J.A., Fukushima, N. and Chun, J. (1999) Gene 227, 89–99.
[13] Offermanns, S. and Simon, M.I. (1998) Oncogene 17, 1375–1381.
[14] Offermanns, S. and Simon, M.I. (1995) J. Biol. Chem. 270, 15175–15180.
[15] Zhu, X. and Birnbaumer, L. (1996) Proc. Natl. Acad. Sci. USA 93, 2827–2831.
[16] Chandrashekar, J., Mueller, K.L., Hoon, M.A., Adler, E., Feng, L., Guo, W., Zuker, C.S. and Ryba, N.J. (2000) Cell 100, 703–711.
[17] Chambers, J.K. et al. (2000) J. Biol. Chem. 275, 10767–10771.
[18] Wilkie, T.M. et al. (1992) Nature Genet. 1, 85–91.
[19] Danielson, P.E., Watson, J.B., Gerendasy, D.D., Erlander, M.G., Lovenberg, T.W., de Lecea, L., Sutcliffe, J.G. and Frankel, W.N. (1994) Genomics 19, 454–461.
[20] Lander, E.S. et al. (2001) Nature 409, 860–921.
[21] Davignon, I., Barnard, M., Gavrilova, O., Sweet, K. and Wilkie, T.M. (1996) Genomics 31, 359–366.
[22] Wilkie, T.M., Scherle, P.A., Strathmann, M.P., Slepak, V.Z. and Simon, M.I. (1991) Proc. Natl. Acad. Sci. USA 88, 10049–10053.
[23] Graler, M.H., Bernhardt, G. and Lipp, M. (1998) Genomics 53, 164–169.
[24] Hagenbuchle, O., Wellauer, P.K., Cribbs, D.L. and Schibler, U. (1984) Cell 38, 737–744.
[25] Ghose, S., Porzig, H. and Baltensperger, K. (1999) J. Biol. Chem. 274, 12848–12854.