Molecular cloning of EDG-3 and N-Shc genes from the puffer fish, *Fugu rubripes*, and conservation of synteny with the human genome[1]

Fuminori Yamaguchi[a,*], Kumiko Yamaguchi[a], Masaaki Tokuda[b], Sydney Brenner[a]

[a] *The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, CA 94704, USA*
[b] *Department of 1st Physiology, Kagawa Medical University, 1750-1, Ikenobe, Miki-cho, Kita-gun, Kagawa 761-0793, Japan*

Received 26 August 1999

Abstract EDG-3 is a receptor for sphingosine-1-phosphate mapped on human chromosome 9q22.1–q22.2. We used the compact *Fugu* genome for its linkage analysis. The *Fugu* EDG-3 was composed of one intron and two exons, encoding a 384 amino acid protein that has 56.9% homology with the human EDG-3. Approximately 3 kb apart, a neuronal Shc (N-Shc) gene was identified. It spans 7 kb containing 12 coding exons, and has an overall 53.4% similarity with the human protein. We mapped the human N-Shc gene to chromosome 9q21.3–q22.2. This is the first report of the genomic structure and the linkage of these two genes conserved between *Fugu* and human.

Key words: Human; EDG-3; N-Shc; Chromosome mapping; *Fugu rubripes*

1. Introduction

G protein-coupled receptors (GPCRs) are large superfamilies which transduce extracellular signals into intracellular pathways by coupling to G proteins, and which share a common structure of seven transmembrane domains [1]. The sequences of these domains are well conserved for each isoform, allowing the isolation of homologues even from distantly related species. The receptor, endothelial differentiation gene 1 (EDG-1), was discovered as an immediate early gene induced by the tumor promoter phorbol myristic acetate [2]. Sphingosine-1-phosphate (S1P) [3,4] and lysophosphatidic acid (LPA) [5] have recently been identified as ligands for it. We previously reported the isolation of a related gene, EDG-3, from a human genomic library and mapped it to human chromosome 9q22.1–q22.2 [6]. We were interested to obtain further information on this receptor and to do this, we used the compact genome of *Fugu rubripes* in which a number of GPCR families have been studied [7–9]. In this paper, we report the isolation of the *Fugu* EDG-3 receptor and N-Shc genes, which we found to be closely linked. We show that this linkage is also conserved in the human genome.

*Corresponding author. Fax: (1) (510) 647-0699.
E-mail: bunchan@molsci.org

[1] Accession numbers: AF164114 for the *Fugu* EDG-3 gene, AF163839 for the *Fugu* N-Shc gene.

Abbreviations: EDG, endothelial differentiation protein; GPCR, G protein-coupled receptor; N-Shc, neuronal Shc; LPA, lysophosphatidic acid; FISH, fluorescence in situ hybridization; PHA, phytohemagglutinin; PTB, phosphotyrosine binding; SH2, Src homology 2; CH, collagen homology; S1P, sphingosine-1-phosphate; DAPI, 4′6′-diamidino-2-phenylindole dihydrochloride; UTR, untranslated region

2. Materials and methods

2.1. Isolation and sequencing of cosmid clone

A partial sequence of the *Fugu* EDG-3 gene was obtained by PCR with degenerate primers using *Fugu* genomic DNA as a template. Specific EDG-3 primers (FEDG-F: TGTAGATGATGGGATTCATGGCCG, FEDG-R: TCCTCCCGCTCTACACCAAGAAAT) were then designed and used for the PCR screening of serial dilutions of a *Fugu* cosmid library on the sets of 96 microtiter plates. The cycling time used was 0.5 min at 95°C, 0.5 min at 60°C, 1.0 min at 72°C, for 35 cycles. Positive cosmids were picked and grown in 3 ml cultures overnight and the DNA was purified. For sequence determination of the full-length gene, selected fragments of the X23 cosmid were subcloned into a plasmid vector and sequenced by a combination of random sequence of sonicated DNA fragments (shotgun sequence) and primer walking methods. The sequence was determined on both strands. The sequence data were assembled by the Lasergene software and contigs were searched by BLAST using the SwissProt database.

2.2. Determination of cDNA sequence and exon-intron boundaries

To determine the exact exon-intron boundaries, *Fugu* liver or brain RNA was prepared by Trizol reagent from Gibco-BRL, and 3 μg each of total RNA was used for the first strand cDNA synthesis, using the first strand cDNA synthesis kit from Gibco-BRL. RT-PCR reaction was carried out to amplify the *Fugu* EDG-3 fragment using the liver cDNA, and the N-Shc fragment using brain cDNA. The PCR fragment was subcloned into the plasmid vector and the sequence was determined on both strands. The cDNA sequence was compared with the corresponding part of the genomic sequence of each gene and exon-intron boundaries were determined.

2.3. Long distance PCR of Fugu genes

The specific primers for the *Fugu* EDG-3 and N-Shc gene were made at 5′ and 3′ of each gene (N-Shc_5′F: CGCCTATTAGCAATCACAGGCCACCTGA, N-Shc_3′R: CCAATTCTGTCCGCTGGAAGTGAACTGT, EDG-3_5′F: GTTGTAGTGCAGGTAGATCAGAGGGTTG, EDG-3_3′R: ACGGAGGGAGATGAGCTGCAGACGGAAA). The long distance PCR was performed using LA Taq polymerase from Takara, with various combinations of these primers. The X23 cosmid DNA was used as a template and PCR was carried out with cycling of 0.5 min at 95°C, 0.5 min at 62°C and 4 min at 68°C for 35 cycles. The PCR sample was run on agarose gel with molecular weight markers. The sequence of the PCR product was partially determined to confirm that the correct fragment had been amplified by PCR.

2.4. Fluorescence in situ hybridization (FISH) of the human N-Shc gene

Lymphocytes isolated from human blood were cultured in α-minimal essential medium (α-MEM) supplemented with 10% fetal calf serum and phytohemagglutinin (PHA) at 37°C for 68–72 h. The lymphocyte cultures were treated with bromodeoxyuridine (0.18 mg/ml) to synchronize the cell population. The synchronized cells were washed three times with serum-free medium to release the block and re-cultured at 37°C for 6 h in α-MEM with thymidine (2.5 μg/ml). Cells were harvested and slides were made using standard procedures including hypotonic treatment, followed by fixing and air-drying. The entire coding sequence of the human N-Shc gene was amplified from human cortex cDNA with specific primers, and this cDNA probe was biotinylated with dATP using the BRL BioNick labeling kit (15°C, 1 h). The procedure for FISH detection was preformed as previously described [20,21]. Briefly, slides were baked at 55°C for 1 h.

PII: S0014-5793(99)01210-7

A

Introns of the Fugu N-Shc gene							
Intron	Donor	Acceptor	Size(bp)	Intron	Donor	Acceptor	Size(bp)
1	gttgg	tccag	1810	7	gtttg	ttcag	203
2	gtatg	tccag	983	8	gtagt	attag	78
3	gtaca	ttcag	731	9	gtaag	tgtag	80
4	gtgag	gacag	72	10	gtacg	accag	315
5	gtgag	tttag	83	11	gtaag	caaag	80
6	gtggg	ctcag	67				

B

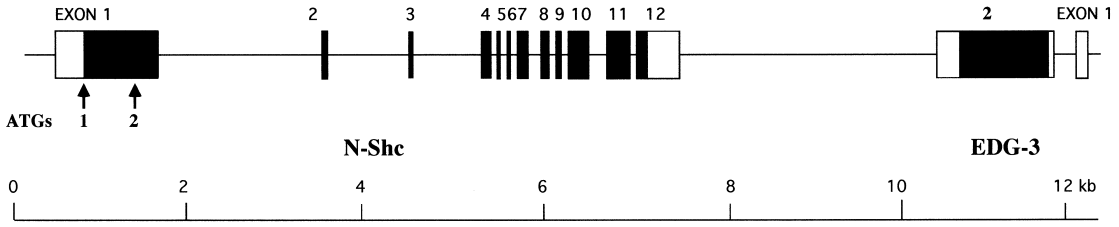

Fig. 1. A: Intron sizes and the donor and acceptor splice sites of the *Fugu* N-shc gene. B: Genomic organization of the *Fugu* N-Shc and EDG-3 genes. The sequences of the two genes were obtained by a cosmid clone isolated from a *Fugu* genomic library. The exon-intron boundaries were determined by comparing genomic and cDNA sequences of each gene. The exon numbers of both genes are shown. The arrows show the positions of the potential initiation codons in the N-Shc gene. The white box indicates untranslated regions and the black box indicates the coding regions.

After RNase treatment, the slides were denatured in 70% formamide in 2×SSC for 2 min at 70°C followed by dehydration with ethanol. Probes were denatured at 75°C for 5 min in a hybridization mix consisting of 50% formamide and 10% dextran sulfate. Probes were loaded on the denatured chromosomal slides. After overnight hybridization, slides were washed and detected as well as amplified. FISH signals and the 4′6′-diamidino-2-phenylindole dihydrochloride (DAPI) banding pattern were recorded separately by taking photographs, and the assignment of the FISH mapping data with chromosomal bands was achieved by superposing FISH signals with DAPI banded chromosomes.

3. Results

3.1. Isolation and characterization of Fugu EDG-3 gene

The *Fugu* EDG-3 gene was obtained by PCR screening of a cosmid library using *Fugu* EDG-3 specific primers. The DNA of one cosmid, X23, was digested with a number of restriction enzymes and a 3.5 kb *Cla*I fragment was used to determine the sequence of the full-length EDG-3 gene. To confirm exon-intron boundaries, a *Fugu* EDG-3 cDNA was isolated from a *Fugu* liver cDNA library by PCR and the PCR fragment was directly sequenced. Comparison of the cDNA and genomic sequences showed that *Fugu* EDG-3 gene is composed of two exons (Fig. 1B). The first small exon is untranslated, and the second exon contains a 43 bp 5′ UTR, 1.2 kb ORF and 3′ UTR. The intron between two exons is 0.3 kb in length. The alignment of the deduced amino acid sequences of EDG-3 (*Fugu* and human) and EDG-1 (human and mouse) is presented in Fig. 2. The *Fugu* gene encodes a 384 amino acid protein similar in size to the human EDG-3 protein (378 amino acids). At the amino acid level, the *Fugu* EDG-3 showed 56.9% homology with human EDG-3 and 43.8% with human EDG-1. The homology between *Fugu* and human EDG-3 increased to 73.9% in the transmembrane regions. A protein motif search indicated potential *N*-glycosylation sites and phosphorylation sites in the *Fugu* EDG-3 gene, some of them are shared with the human gene (Fig. 2).

3.2. Isolation and analysis of Fugu N-Shc gene

The cosmid was scanned by shotgun sequencing to find other genes. Sonicated 400–800 bp fragments were subcloned into a plasmid vector and sequenced. The sequences were searched against the SwissProt database, and revealed the presence of the N-Shc gene. The full-length sequence of this gene was determined, and a cDNA was isolated from a *Fugu* brain cDNA library by PCR and the fragment was sequenced. The genomic and cDNA sequences were compared to determine exon-intron boundaries (Fig. 1A). The *Fugu* introns range from 67 to 1810 bp in size, with the largest intron after exon 1. The genomic organization of the *Fugu* N-Shc gene is presented in Fig. 1B. The *Fugu* N-Shc gene spans 7 kb of genomic DNA and is composed of 12 coding exons with 2130 bp of ORF that covers the initiation methionine to the stop codon. Two in-frame initiation codons, corresponding to those of the human N-Shc gene, were found (Fig. 1B). The ORF beginning with these start codons encodes 660 and 505 amino acid proteins, and the predicted molecular weight of these proteins is 70.6 and 55.3 kDa, respectively. The alignment of the *Fugu* amino acid sequence with the human N-Shc and Shc genes is shown in Fig. 3. Together with the positions of the introns, the phosphotyrosine binding (PTB) and the Src homology 2 (SH2) domains are conserved in the *Fugu* gene, and the sequence similarity of these domains between *Fugu* and human is 80.6% and 77.7%, respectively. In addition, the consensus Grb2 SH2 site ('YVNT' or 'YVNV') is well conserved, but the surrounding region, called the collagen homology (CH), was only 42.7% similar. The overall homology between *Fugu* and human N-Shc protein is 53.4%, and between *Fugu* N-Shc and human Shc is 43.7%.

3.3. The physical distance of Fugu EDG-3 and N-Shc genes

The physical distance between the *Fugu* EDG-3 and N-Shc genes was determined by long distance PCR. Using specific primers for the *Fugu* EDG-3 and N-Shc genes, the two genes

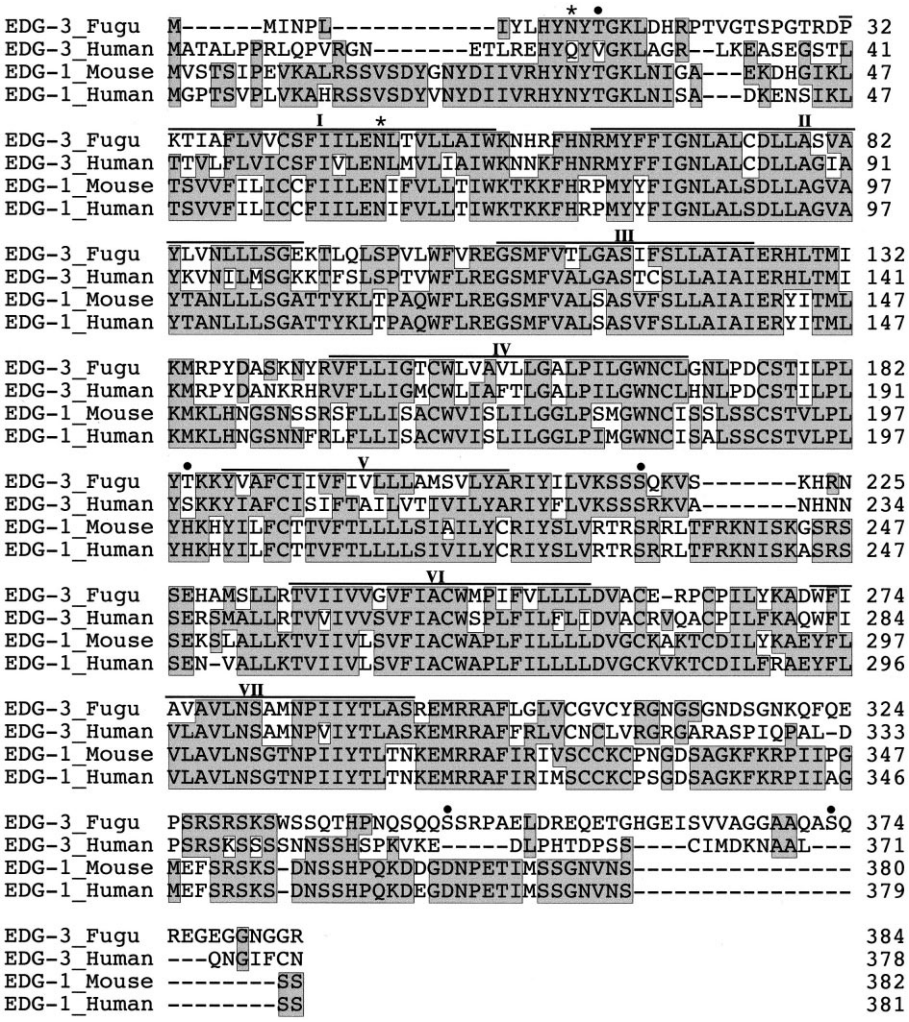

Fig. 2. Alignment of the deduced amino acid sequences of *Fugu* and human EDG-3 and EDG-1 genes. The sequence alignment was carried out using Clustal V software. The seven transmembrane domains are indicated by lines and Roman numerals. The conserved amino acid residues are shown in shaded boxes. The potential *N*-glycosylation sites (*) and phosphorylation sites (●) are marked. Amino acid numbers are indicated at the right of each row.

are approximately 3 kb apart in a tail to tail orientation (Fig. 4). The partial sequence of this fragment was determined to confirm that the intergenic region had been correctly amplified.

3.4. The chromosome mapping of the human N-Shc gene

Since the N-Shc gene has not yet been mapped in any species, human N-Shc cDNA was prepared by PCR from a human cerebral cortex cDNA library. The fragment was labeled and used for FISH mapping of the human N-Shc gene. Under the conditions used, hybridization efficiency was approximately 68% for this probe (68 out of 100 checked mitotic figures showed signals on one pair of the chromosomes). Since DAPI banding was used to identify the specific chromosome, the signal from the probe could be assigned to the long arm of chromosome 9, and a detailed analysis based on 10 mitoses showed that the human N-Shc gene maps to the chromosome 9q21.3–q22.2 region (Fig. 5A).

4. Discussion

The EDG receptors were originally classified as orphan receptors. EDG-1 was cloned from endothelial cells [2] induced by addition of PMA, which stops endothelial cell growth and induces its differentiation into a capillary-like tubule structure [10]. Recently S1P, a platelet-derived lipid mediator that was known to evoke adenylate cyclase inhibition, mitogen-activated protein (MAP) kinase activation and mitogenesis, was shown to bind to EDG-1 with high affinity [3,4]. In addition, LPA, a platelet-derived bioactive lipid structurally related to S1P, was reported to be a low affinity agonist of EDG-1 [5]. LPA binding to EDG-1 induces receptor phosphorylation, MAP kinase activation, as well as Rho-dependent morphogenesis and P-cadherin expression. It is thought that activation of the endothelial receptor EDG-1 by platelet-derived lipids LPA and SPP may be important for thrombosis and angiogenesis [5]. The human EDG-3 has been shown to activate serum response element-driven transcriptional reporter gene in response to S1P, dihydro-S1P and sphingosylphosphorylcholine, but not to LPA. This result indicated that EDG-3 is a functional receptor for S1P and perhaps other lysosphingolipids [11].

We cloned the EDG-3 gene from the *Fugu* genome, and found that it has two exons split by an intron. This exon-

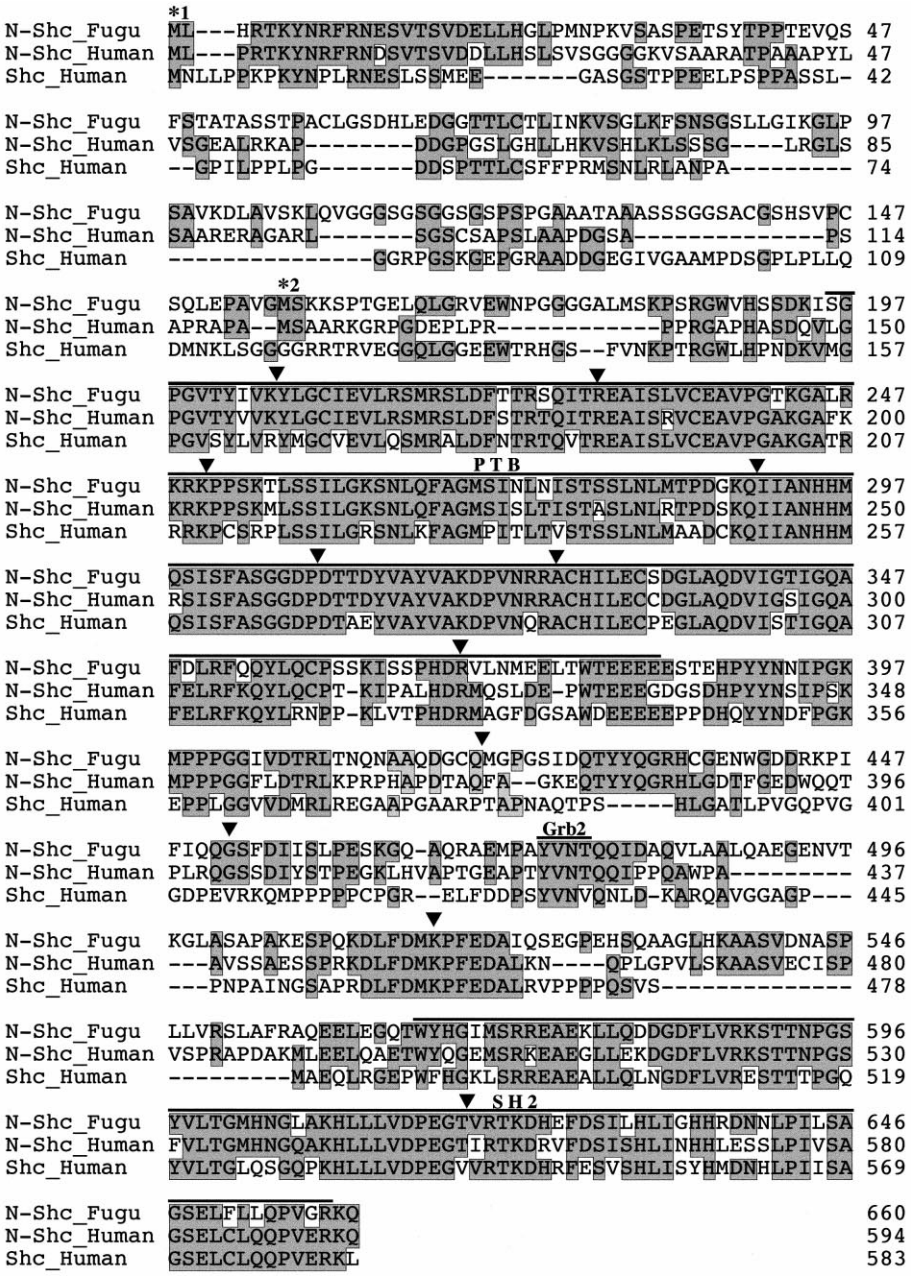

Fig. 3. Alignment of the deduced amino acid sequences of *Fugu* and human N-Shc and human Shc genes. The 70.6 kDa of *Fugu* N-shc, 64 kDa of human N-Shc and 62 kDa of human Shc protein were aligned. The conserved amino acid residues are shown in shaded boxes. The positions of *Fugu* N-Shc introns are marked (▼). The PTB, Grb2 binding and SH2 domains are shown with overlines. The alternative start methionine of the N-Shc gene is indicated (*1 or *2). Amino acid numbers are indicated at the right of each row.

intron structure is similar to that of the reported mouse EDG-1 gene, with a short 5′ untranslated exon, and a second exon containing a 1.2 kb ORF and 43 bp 5′ UTR (Fig. 1B). This gene structure, a small untranslated first exon, followed by a second exon encoding the whole protein sequence, is seen in other GPCRs such as cannabinoid type 1 and muscarinic receptors. In the case of the *Fugu* EDG-3 gene, the intron size is 0.3 kb which is smaller than that of the mouse EDG-1 gene (1.6 kb) (Fig. 1B). In the mouse EDG-1 gene, it was reported that the intron has a repressor site for gene expression [12]. Although the intron in the *Fugu* EDG-3 gene is smaller, the similarity of sequence and genomic structure of the human and *Fugu* genes suggests that the *Fugu* intron may have a similar function as seen in the mouse EDG-1 gene.

The *Fugu* N-Shc gene was located 3 kb away from the EDG-3 gene (Fig. 1B). The *Fugu* N-Shc gene consists of 12 coding exons, and the introns range in size from 67 to 1810 bp with the largest intron positioned between exon 1 and 2 (Fig. 1A). The genomic structure of N-Shc gene has not been determined in any species, and the human [13] and rat [14] genes have been sequenced only at the cDNA level. As the human Shc gene has high homology with the N-Shc gene and shares similar functional domains, its genomic structure was compared with that of the *Fugu* N-Shc gene. The human Shc

gene consists of 13 exons [15], whereas the *Fugu* N-Shc gene is composed of 12 exons. This difference is due to the existence of a 5′ untranslated exon in the Shc gene. At the protein level, two isoforms of human N-Shc protein (64 and 52 kDa) were detected by Western blot analysis [13,14]. As the human N-Shc gene has two potential initiation methionines, the mechanism of choosing one of these initiation sites controls the generation of two different alternative splicing variants of the N-Shc gene. 5′ RT-PCR of *Fugu* cDNA also suggested the existence of these two isoforms. The calculated molecular weight of these isoforms is 70.6 and 55.3 kDa, respectively, quite similar to those of the human N-Shc isoforms. On the other hand, the Shc gene contains three potential initiation methionines, and three splicing variants (66, 52 and 46 kDa) have been detected by Western blot analysis [15]. These splicing variants are generated with or without using the first untranslated exon by a distinct mechanism.

The similarity in sequence and genomic organization of the N-Shc and Shc genes suggests that these genes derive from a common ancestor. Both are adapter proteins, and are involved in phosphotyrosine signalling following extracellular stimulation through GPCR and growth factor receptors. For example, Shc was phosphorylated by EGFR following GPCR stimulation by thrombin, endothelin-1 or LPA [16]. Both Shc and N-Shc can bind to activated TrkB receptor following stimulation with brain-derived neurotrophic factor [13]. However, the distribution of Shc and N-Shc proteins is different; the former is expressed everywhere except in brain [17], whereas N-Shc is abundantly expressed in the central nervous system (CNS) and is an important adapter molecule in the CNS.

The PTB, CH and SH2 domains are found in both Shc and N-Shc genes and mediate protein interactions. The homology of the PTB and SH2 domains between *Fugu* and human N-shc is very high, suggesting that the *Fugu* gene may act as an adapter protein in a similar manner. However, the homology of the CH region is lower (42.7%), and was low between human Shc and N-Shc protein (34.3%). This domain is known to be responsible for the association with SH3-containing proteins [10,18,19]. In the CH domain, the consensus sequence 'YVNT' was found in both the *Fugu* and human genes and

Fig. 4. Long distance PCR of *Fugu* EDG-3 and N-Shc genes. Two primer sets, N-Shc_5′F and N-Shc_3′R, and EDG-3_5′F and EDG-3_3′R, were used for the PCR with four different combinations, and the PCR products were separated on a 0.7% agarose gel. Mk: 1 kb DNA marker; lane 1: N-Shc_5′F+EDG-3_5′F; lane 2: N-Shc_5′F+EDG-3_3′R; lane 3: N-Shc_3′R+EDG-3_5′F; lane 4: N-Shc_3′R+EDG-3_3′R.

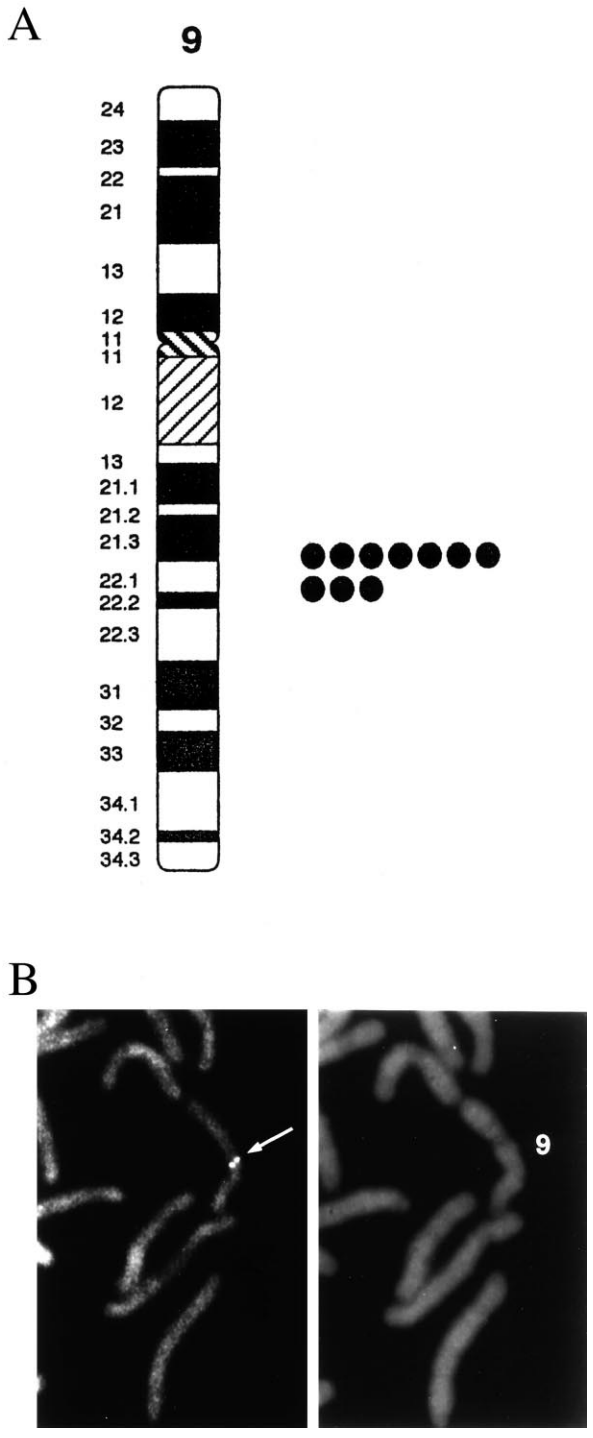

Fig. 5. FISH mapping of the human N-Shc gene. A: Diagram of FISH mapping results for probe N-Shc. Each dot represents the double FISH signals detected on human chromosome 9. B: Example of FISH mapping. The left panel shows the FISH signals on chromosome; the right panel shows the same mitotic figure stained with DAPI to identify chromosome 9.

matches the motif recognized by the Grb2 SH2 domain. The tyrosine 474 residue in the 'YVNT' motif of human N-Shc protein is a putative tyrosine-phosphorylation site and is considered to be important for the signalling [13]. The *Fugu* gene contains the corresponding tyrosine residue at position 424, which might be essential for N-Shc signalling. However, the divergence of the CH domain between *Fugu* and human N-shc protein may reflect different functions in the two species.

Acknowledgements: This research was financially supported by the Louis Jeantet Foundation, and by a Grant-in-Aid for Scientific Research (A) of the Ministry of Education, Science and Culture, Japan.

References

[1] Probst, W.C., Snyder, L.A., Schuster, D.I., Brosius, J. and Sealfon, S.C. (1992) DNA Cell Biol. 11, 1–20.
[2] Hla, T. and Maciag, T. (1990) J. Biol. Chem. 265, 9308–9313.
[3] Lee, M.J., Van Brocklyn, J.R., Thangada, S., Liu, C.H., Hand, A.R., Menzeleev, R., Spiegel, S. and Hla, T. (1998) Science 279, 1552–1555.
[4] Zondag, G.C., Postma, F.R., Etten, I.V., Verlaan, I. and Moolenaar, W.H. (1998) Biochem. J. 330, 605–609.
[5] Lee, M.J., Thangada, S., Liu, C.H., Thompson, B. and Hla, T. (1998) J. Biol. Chem. 21, 22105–22112.
[6] Yamaguchi, F., Tokuda, M., Hatase, O. and Brenner, S. (1996) Biochem. Biophys. Res. Commun. 227, 608–614.
[7] Yamaguchi, F. and Brenner, S. (1997) Gene 191, 219–223.
[8] Yamaguchi, F., Macrae, A.D. and Brenner, S. (1996) Genomics 35, 603–605.
[9] Macrae, A.D. and Brenner, S. (1995) Genomics 25, 436–446.
[10] Montesano, R. and Orci, L. Phorbol esters induce angiogenesis in vitro from large-vessel endothelial cells, (1987) J. Cell. Physiol. 130, 284–291.
[11] An, S., Bleu, T., Huang, W., Hallmark, O.G., Coughlin, S.R. and Goetzl, E.J. (1997) FEBS Lett. 417, 279–282.
[12] Liu, C.H. and Hla, T. (1997) Genomics 43, 15–24.
[13] Nakamura, T., Sanokawa, R., Sasaki, Y., Ayusawa, D., Oishi, M. and Mori, N. (1996) Oncogene 13, 1111–1121.
[14] Nakamura, T., Muraoka, S., Sanokawa, R. and Mori, N. (1998) J. Biol. Chem. 273, 6960–6967.
[15] Migliaccio, E., Mele, S., Salcini, A.E., Pelicci, G., Lai, K.M., Superti Furga, G., Pawson, T., Di Fiore, P.P., Lanfrancone, L. and Pelicci, P.G. (1997) EMBO J. 16, 706–716.
[16] Daub, H., Weiss, F.U., Wallasch, C. and Ullrich, A. (1996) Nature 379, 557–560.
[17] Pelicci, G., Lanfrancone, L., Grignani, F., McGlade, J., Cavallo, F., Forni, G., Nicoletti, I., Pawson, T. and Pelicci, P.G. (1992) Cell 70, 93–104.
[18] Harrison Findik, D., Susa, M. and Varticovski, L. (1995) Oncogene 10, 1385–1391.
[19] Pelicci, G., Lanfrancone, L., Salcini, A.E., Romano, A., Mele, S., Grazia Borrello, M., Segatto, O., Di Fiore, P.P. and Pelicci, P.G. (1995) Oncogene 11, 899–907.
[20] Heng, H., Squire, J. and Tsui, L.-C. (1992) Proc. Natl. Acad. Sci. USA 89, 9509–9513.
[21] Heng, H. and Tsui, L.-C. (1993) Chromosoma 102, 325–332.