Identification of Genetic and Epigenetic Similarities of \textit{SPHK1/Sphk1} in Mammals

Takuya IMAMURA\textsuperscript{1,\ast}, Nanami MIYAUCHI-SENDA\textsuperscript{1}, Satoshi TANAKA\textsuperscript{1} and Kunio SHIOTA\textsuperscript{1,\ast}\ast

\textsuperscript{1}Laboratory of Cellular Biochemistry, Animal Resource Sciences, Veterinary Medical Sciences, The University of Tokyo, 1–1–1 Yayoi, Bunkyo-ku, Tokyo 113–8657, Japan

(Received 25 May 2004/Accepted 5 July 2004)

\textbf{ABSTRACT.} In normal tissues, methylation of CpG islands is generally accepted to be limited to the inactive X-chromosome and imprinting clusters. Gene \textit{Sphk1} has shown complex organization, indicated by multiple alternative splicing and tissue-dependent DNA methylation within the limited area (T-DMR) of the CpG island in the rat. Comparisons among human, mouse and rat \textit{SPHK1/Sphk1} genomic DNA revealed five coding exons and association of a CpG island at the 5' end in common. We also found two novel subtypes, for a total of eight mRNA subtypes generated through selective usage of untranslated first exons. A 38-bp region at the 5'-end of T-DMR is highly conserved. This restricted area is specifically hypomethylated in the brain. Here, we examine the complex genetic/epigenetic features of the \textit{SPHK1/Sphk1} CpG island, and suggest that the T-DMR is the core target for tissue-dependent CpG island methylation.

\textbf{KEY WORDS:} alternative splicing, CpG island, DNA methylation, sphingosine kinase, tissue specific.

---

Sphingosine kinase (SPHK) is the enzyme catalyzing the production of sphingosine 1-phosphate, which is implicated in various biological events including cell growth, differentiation, and survival [6, 26]. Activity of SPHK has been found in several mammalian tissues [4, 14, 27, 28, 33]. Although SPHK is ubiquitously expressed, the level of its expression and subcellular localization is tissue dependent [27]. Recently, cDNAs for two isoforms of SPHK (SPHK1 and 2) have been cloned in humans and mice [11, 13, 18, 21, 27]. Recently, cDNAs for two isoforms of SPHK (SPHK1 and 2) have been cloned in humans and mice [11, 13, 18, 21, 28]. Furthermore, multiple subtypes are transcribed from rat \textit{Sphk1} through alternative usage of untranslated first exons [8]. Numerous isoforms suggest that multiple regulatory regions lie in the gene locus, all of which are suspected to pilot the exact expression level in respective cells.

\textit{Rat Sphk1} consists of GC-rich sequences, \textasciitilde60\% on average, over its entire coding region and its flanking sequences. In addition, the 5' end shows the characteristic feature of high CpG frequency, called a CpG island [8]. These CpG islands often lie in housekeeping gene loci and a subset of tissue-specific gene loci, providing evidence for their roles in gene expression [3, 7]. Using restriction landmark genomic scanning that enables us simultaneous detection of methylation status in thousands of genomic loci [25], we have recently demonstrated that there are cell type- and tissue-specific patterns of DNA methylation in the CpG islands [8, 22–24, 32], suggesting that genomic loci, in which methylation states are altered, are more common than had been realized. Thousands of CpG islands are under the regulation of tissue-dependent DNA methylation [32]. In the case of rat \textit{Sphk1}, this includes an \textasciitilde200-bp tissue-dependent, differentially methylated region (T-DMR) at the 5' edge of the CpG island, which seems to participate in RNA expression [8]. Studies on T-DMR of \textit{Sphk1} would clarify the features of cis-acting elements necessary for tissue-dependent methylation of CpG islands.

Here, we describe the genomic structures of \textit{SPHK1/Sphk1} in human, mouse and rat. We identified the multiple conserved regions in the CpG island, and demonstrated that the core target for epigenetic modification is localized within T-DMR.

\textbf{MATERIALS AND METHODS}

\textbf{Animal treatment and reagents:} C57/BL6J mice were from Oriental Yeast Co., Ltd. (Tokyo, Japan). They were kept under a lighting regime of 14 hr illumination and 10 hr darkness (lights on between 0500 hr and 1900 hr). To prepare genomic DNA and total RNA, brain and other tissues were removed following decapitation. They were dissected as appropriate, washed with 0.9\% NaCl, frozen in liquid nitrogen, and stored at \textasciitilde80\degreeCelsius until use.

All reagents, unless otherwise stated, were purchased from Wako Pure Chemicals (Osaka, Japan).

\textbf{Analysis by using databases:} We previously obtained genomic sequences of rat \textit{Sphk1} [7]. Genomic DNA sequences of human \textit{SPHK1} and mouse \textit{Sphk1} (accession nos. NT_010641 and NT_039521, respectively) were examined. We analyzed the GC content and CpG frequency of the genomic DNA fragments with a program (CpG view V1.5.2) provided by the National Institute of Infectious Disease (http://www.nih.go.jp/yoken/genebank/binaryFile/Mac/CpG-ViewV1.4.7.hqx). A homology search of \textit{SPHK1/Sphk1} among species was done using MacMolly Tetra V3.8 provided by Mologen Holding AG (http://www.mologen.com).

\textbf{RT-PCR:} Total RNA was prepared by homogenizing tissue directly in TRizol (Invitrogen; Carlsbad, CA). To
amplify SPHK1 subtypes, we designed the reverse primer (5'-GCGTGCAGTGTTGGTCAGGAGGTCTTCATTGG-3'; reverse complement to nt +1177 to +1207, in relation to the translation start site at +1) and specific forward primers based on the genomic DNA sequences as follows: for SPHK1a (5'-TGTTTCTCGGAGTTTTTCGACCCCT-3', nt -2484 to -2460), SPHK1c (5'-TTGCCAACTCTAGTTGGCAGCTCAAACC-3'; nt -3125 to -3101); SPHK1d (5'-ATCCTAACGCTGCTCTTCCCTCC-3', nt -1758 to -1734), SPHK1E (5'-AAGTTGAGCGAAAAGTTTGAGGCCG-3', nt -1004 to -980); SPHK1f (5'-AGTTCTGGGATTTTACGCAGCTGG-3', nt -528 to -504). A cDNA from human-derived cell lines (HeLa cells or HEK293 cells) was amplified in a cocktail consisting of GC buffer II for LA-Taq, enzyme and each primer set. Subjective PCR products were cloned and sequenced. We also amplified mouse Sphk1a using the following primers: 5'- ACCAAGGCATGTATTGCAGTGACGC-3', 5'-GACAGACTGAGCACAATGAGCC-3'.

DNA methylation analysis: In order to determine the DNA methylation status of individual CpG sites, genomic DNA was sequenced following bisulfite modification, as described previously [5, 8]. Briefly, 10 µg of Pvu II (TaKaRa; Kyoto, Japan) -digested genomic DNA was initially denatured with 330 mM NaOH. Then, sodium metabisulfite solution at pH 5.0 and hydroquinone (Nacarai; Osaka, Japan) were added at final concentrations of 2.0 M and 0.5 mM, respectively. Samples were incubated under mineral oil in the dark for 12 hr. The modified DNA was purified using Wizard DNA purification resin (Promega; Madison, WI). Modification was completed by treatment with NaOH (final concentration, 0.3 M) at 37°C for 15 min, followed by ethanol precipitation. Two µl of each sample were amplified using AmpliTaq Gold (Perkin Elmer; Norwalk, CT) and primer sets: forward, 5'-GAGAGGTGAGGAGGTTAGAGGATGA-3'; reverse, 5'-TAACTACCAAAATCTAATCCTACAC-3'.

RESULTS

Comparison of SPHK1/Sphk1 among human, mouse and rat: We examined genomic DNA sequences of human SPHK1 and mouse Sphk1, and located each sequence in the telomeric region of the chromosome, human Chr. 17 and mouse Chr. 11, respectively. Each genomic sequence contains exactly the entire protein coding sequence, consisting of five exons (Fig. 1). The polyadenylation signal, AATAAA, is in the last exon (see Fig. 2C). Each intron has GT... and ...AG nucleotides in its 5' and 3' ends, respectively. When genomic sequences were compared between human SPHK1 and rat Sphk1 [8] loci, several common structures were found (Fig. 1; see also Fig. 2C). The 5' ends of two cDNAs for human SPHK1 (GenBank acc. nos. AF238083 and AF266756), each of which had alternative untranslated first exons, showed high similarities with rat Sphk1e and -f, respectively (Table 1). Thus, we tentatively named these cDNAs human SPHK1e and -f (Fig. 1).

It is clear that SPHK1/Sphk1 contain CpG islands at their 5' ends, as documented in our previous study (Fig. 1; see also [8]). The size of rat Sphk1 CpG island is ~3.7 kb (~3295 to +406, in relation to the translation start site at +1), whereas human SPHK1 CpG island is ~4.1 kb (~3282 to +859). Each exon and intron is located similarly in the two species, except intron 2 is 330 bp for human SPHK1e and 561 bp for rat Sphk1e (Fig. 1; arrowheads). Since intron 2 is located at each 3' end of the CpG island, the 400-bp shorter sequence of the rat CpG island partly reflects the relatively smaller size of intron 2 in the rat.

Fig. 1. Schematic representation of SPHK1/Sphk1. The upper panel in each species shows the GC content and CpG frequency of the locus. The moving average of GC content is plotted on the graph as a jagged line, and CpG frequencies are shown as black bars. Below each graph, a short vertical line indicates the position of each CpG and CpG dinucleotide. The lower panel in each species shows the genomic structure. Closed black and gray boxes indicate the positions of protein-coding exons and multiple alternative first exons, respectively. The CpG island was formulated by high GC content of more than 50% and a CpG frequency score of more than 0.6 on average (Average GC content and CpG frequency in bulk DNA were reported to be about 40% and 0.25, respectively). Arrowheads indicate intron 2 in human and rat, the size of which is particularly different from each other.
To identify similar features of the CpG islands in SPHK1/Sphk1, we plotted the common sequences by 12-nt windows (Fig. 2). When mouse Sphk1 and rat Sphk1 were compared, the aligned region of over –7.5 kb, encompassing entire coding sequences and 5' flanking CpG island, showed ~80% similarity (Fig. 2A). The sequences adjacent to the first exons of the human SPHK1e and -f were aligned with those of rat Sphk1e and -f, thereby validating this analysis (Fig. 2C). In addition to these alternative first exons, five other regions appeared highly similar in the species examined. In the rat, these five regions corresponded to the 5' flanking region and first intron of Sphk1c, and the 5' flanking region, first exon and 5' end of the intron of Sphk1a (Fig. 2C), indicating that multiple conserved regions are embedded in the CpG island.

**Identification of SPHK1/Sphk1 subtypes:** To obtain insights into the alternative splicing of SPHK1/Sphk1, we searched the database of expressed sequence tags (EST) based on the sequence information of SPHK1/Sphk1 CpG island. We were successful in identifying cDNAs homologous to rat Sphk1a, -c, -d, -e, and -f, as shown in Table 1. Moreover, we found two novel subtypes (acc. nos. BB611745 and BI9114829) probably formed by alternative transcription of the first exons located in the CpG island (Table 1). We tentatively named these cDNAs SPHK1g/Sphk1g and -h.

We performed RT-PCR analysis using total RNA from human-derived cell lines, HeLa cells and HEK293 cells, and subtype-specific primers (Fig. 3). The specific bands of conventional subtypes, SPHK1e and -f were observed in each sample examined. In addition to these subtypes, we detected the specific products when aiming to amplify SPHK1a and -h, although SPHK1b, -c, -d, -g were not detected. Therefore, it would be necessary to determine if all of the subtypes are utilized in humans. Nonetheless, sequence analysis confirmed that all specific products were derived from mRNAs of SPHK1a, -e, -f, and -h. Therefore, the SPHK1/Sphk1 CpG island is clearly a region encompassing multiple untranslated first exons.

**DNA methylation analysis of T-DMR in Sphk1:** The T-DMR in the 5' edge of Sphk1 CpG island, which is the flanking region of subtype Sphk1a, serves to differentially methylate depending on rat tissue [8] (see also Fig. 1). To further characterize the SPHK1/Sphk1 T-DMR, we compared the rat T-DMR with corresponding human and mouse sequences. In the –200-bp region corresponding to T-DMR, these sequences were similar; the rat T-DMR showed 55% and 78% homologies with human and mouse sequences, respectively (Fig. 4A). In particular, a 38-bp region located at the 5' end of the rat T-DMR showed 89% and 92% homologies with those of mouse and human, respectively. We performed database search using UCSC genome browser (http://genome.ucsc.edu), and found that the 38-bp region is very unique since the database search indicated no similarities with any other genomic DNA sequences.

To determine if mouse Sphk1 is also under the regulation of DNA methylation, we investigated methylation status of the sequences between –3206 to –2533 using bisulfite-modified genomic DNA (Fig. 4B). Regional methylation was evident. For example, of 28 CpG dinucleotides, 20 of them were methylated in either of the heart-derived sequences,
indicating that the 5’ end of the CpG island in the mouse is capable of being methylated, similar to the rat. Of all the CpG sites that can be methylated, 8 sites are much higher methylated in heart and liver than in brain. On the contrary, there are no cases when CpG sites are more densely methylated in brain than in heart and liver. These observations were also similar to those in the rat, in which 8 sites have been shown to be methylated differentially depending on tissue [8]. In the mouse, there are 6 CpG sites (-3206, -3199, -3164, -3112, -3058 and -3028) that are aligned with T-DMR of rat Sphk1. It should be noted that all of these CpG sites are differentially methylated depending on tissue. Thus, T-DMR is also located around the 5’ edge of the mouse Sphk1 CpG island. In contrast, the region outside T-DMR is preferentially free of methylation in rat and mouse, which suggests that T-DMR includes important sequences to separate the areas of differential methylation from the unmethylated core of the CpG island.

The methylation pattern suggests that it is indicative of expression of Sphk1a, located downstream. Expression of mouse Sphk1a in the brain, heart and liver was analyzed by RT-PCR (Fig. 4C). Expression of Sphk1a was detected in brain, but not in heart and liver. In agreement with observations in the rat, expression of Sphk1a was inversely correlated with DNA methylation [8]. Based on these results, tissue-dependent methylation occurred in the boundary between the CpG island and CpG-depleted sequences. The methylation is related to the reduction of Sphk1a expression. The 5’ end of the CpG island is genetically and epigenetically conserved between mammalian species.

DISCUSSION

In the present study, we aimed to dissect the genetic and epigenetic features of SPHK1/Sphk1, focusing on CpG island methylation. Each gene consists of five protein-coding exons, flanked by the CpG island, which encompasses multiple alternative first exons and other conserved regions. The boundary between the CpG island and CpG-depleted sequences was defined as T-DMR, which appears in both mouse and rat.

Regions capable of being methylated in the CpG islands in mammals: The 38-bp highly-conserved region in T-DMR is located at the edge of the CpG island and near the densely-methylated CpG-depleted sequences (Fig. 4A). This suggests that the edge of the Sphk1 CpG island has been resistant to mutations triggered by aberrant DNA methylation and deamination during evolution. Methylation of CpG islands has been implicated in aging and cancer progression [36, 37]. Our finding clearly shows that methyl-

Table 1. Sequence comparison of alternative first exons in the SPHK1/Sphk1

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Species</th>
<th>3’-Boundary of alternative exon 1</th>
<th>Position of 3’-End</th>
<th>GenBank acc. nos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>rat</td>
<td>...GAAGCGATGCAAGGAGAGGag...</td>
<td>-2101</td>
<td>AB049571</td>
</tr>
<tr>
<td></td>
<td>mouse</td>
<td>...AAAGCGAGACCCTGATCCAGGag...</td>
<td>-2078</td>
<td>AF068748</td>
</tr>
<tr>
<td></td>
<td>human</td>
<td>...TTTAAAGGCGATGCAAGGag...</td>
<td>-2117</td>
<td>n. d.</td>
</tr>
<tr>
<td>b</td>
<td>rat</td>
<td>...CTTTAATCGGCTATCCAGGag...</td>
<td>+723</td>
<td>n. d.</td>
</tr>
<tr>
<td></td>
<td>mouse</td>
<td>...CTTTAATCGGCTATCCAGGag...</td>
<td>+838</td>
<td>AF068749</td>
</tr>
<tr>
<td></td>
<td>human</td>
<td>...CTTCAAGGCTATCGGCTATCCAGGag...</td>
<td>+493</td>
<td>n. d.</td>
</tr>
<tr>
<td>c</td>
<td>rat</td>
<td>...AAAGCGAGACCCTGATCCAGGag...</td>
<td>-2991</td>
<td>AB049572</td>
</tr>
<tr>
<td></td>
<td>mouse</td>
<td>...AAAGCGAGACCCTGATCCAGGag...</td>
<td>-2963</td>
<td>BG915874 (EST)</td>
</tr>
<tr>
<td></td>
<td>human</td>
<td>...AAAGCGAGACCCTGATCCAGGag...</td>
<td>-2990</td>
<td>n. d.</td>
</tr>
<tr>
<td>d</td>
<td>rat</td>
<td>...GCTGCATCTGGAAGCTAAAGag...</td>
<td>-1482</td>
<td>AB049573</td>
</tr>
<tr>
<td></td>
<td>mouse</td>
<td>...GCTGCATCTGGAAGCTAAAGag...</td>
<td>-1473</td>
<td>BE303629 (EST)</td>
</tr>
<tr>
<td></td>
<td>human</td>
<td>...CCCAGCTGCGAGGCGAGAAGCTAAAGag.</td>
<td>-1645</td>
<td>n. d.</td>
</tr>
<tr>
<td>e</td>
<td>rat</td>
<td>...CCGCTAGGAGGACAGCTCGGag...</td>
<td>-797</td>
<td>AB049574</td>
</tr>
<tr>
<td></td>
<td>mouse</td>
<td>...CCGCTAGGAGGACAGCTCGGag...</td>
<td>796</td>
<td>AA592274 (EST)</td>
</tr>
<tr>
<td></td>
<td>human</td>
<td>...CCGCTAGGAGGACAGCTCGGag...</td>
<td>-839</td>
<td>AF238083</td>
</tr>
<tr>
<td>f</td>
<td>rat</td>
<td>...GGACCTCGTATGCAACGAGATACGag...</td>
<td>+10</td>
<td>AB049575</td>
</tr>
<tr>
<td></td>
<td>mouse</td>
<td>...GGACCTCGTATGCAACGAGATACGag...</td>
<td>+10 n. d.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>human</td>
<td>...GGACCTCGTATGCAACGAGATACGag...</td>
<td>+10 n. d.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AF266756</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>rat</td>
<td>...CTTTGGAGGTTCATGCAACGAGATACGag.</td>
<td>-1593</td>
<td>n. d.</td>
</tr>
<tr>
<td></td>
<td>mouse</td>
<td>...CTTTGGAGGTTCATGCAACGAGATACGag.</td>
<td>-1587</td>
<td>BB611745 (EST)</td>
</tr>
<tr>
<td></td>
<td>human</td>
<td>no homologous regions</td>
<td></td>
<td>n. d.</td>
</tr>
<tr>
<td>h</td>
<td>rat</td>
<td>...CTTCCACCCACCCCTTCAAGAGGag...</td>
<td>-449</td>
<td>n. d.</td>
</tr>
<tr>
<td></td>
<td>mouse</td>
<td>...CTTCCACCCACCCCTTCAAGAGGag...</td>
<td>-434</td>
<td>n. d.</td>
</tr>
<tr>
<td></td>
<td>human</td>
<td>...CTTCCACCCACCCCTTCAAGAGGag...</td>
<td>-476</td>
<td>BJ111482 (EST)</td>
</tr>
</tbody>
</table>

a) Sequences shown in uppercase are those at the 3’ end of each exon 1 and those shown in lowercase are the donor splice sites, respectively.
b) Numbering is with regard to the translation start site of Sphk1a, at +1.
c) n. d., not determined.
Epigenetic similarities of Spk1 in mammals

Localization occurs at the Spk1 CpG island in a subset of normal cells. The DNA methylation patterns were conserved between species, being hypomethylated in brain and hypermethylated in heart and liver of both human and rat. This indicates that methylation of T-DMR is not clearly the consequence of pathological processes, but has a role in regulation of gene expression. Since T-DMR methylation is involved in silencing Spk1a subtype expression, aberrant methylation of T-DMR might result in abnormal gene expression. It is interesting to note that the locus containing rat Spk1, at telomeric region of Chr. 10, was identified as a susceptible region related to the autoimmune inflammatory disease, rheumatoid arthritis [30]. In humans, Spk1 has been mapped to Chr. 17q25 [18], which contains loci related to sclerosis, psoriasis, and epidermodysplasia [2, 12, 20, 29, 35]. The T-DMRs are dispersed in the genome [32], and thus, accumulation of T-DMR information would be very important in dissecting the epigenetic component of these diseases as well as in understanding normal development.

**T-DMR and conserved regions as multiple cis-acting elements:** An increasing number of CpG islands have been reported to be under the regulation of tissue-dependent methylation [8, 22–24, 32]. In the Spk1 CpG island, conservation among species of the 38-bp region in T-DMR raises the possibility that its function in mRNA expression is indispensable (Fig. 4). The DNA methylation patterns adjacent to the conserved region are also quite similar in the mouse.
mouse and rat. There is a growing body of evidence indicating that histone deacetylase, methyl-CpG binding protein and methyltransferase acts cooperatively for the inheritance of chromatin structure and gene expression pattern [10, 31]. Deacetylated histones constitute tightly-packed chromatin structure, repressing gene expression. We have previously shown that a methyl-CpG binding protein MeCP2 recruits methyltransferase 1 (Dnmt1) at methylcytosines, through which DNA methylation patterns and transcriptional repressor components could be maintained [10]. Until now, it remains unknown which modification precedes the other. It would be interesting to see whether the T-DMR methylation is established primarily or secondary to the histone modification for the regulation of SPHK1 expression.

Beyond the T-DMR, there are few CpG sites. These sites are constantly hypermethylated in the rat SPHK1 [8]. The T-DMR is the spacer element between hypermethylation and hypomethylation regions. We have not identified consensus sequences around T-DMR other than Sp1 recognition sites, which may be essential for protecting CpG islands from methylation [16, 19]. Furthermore, potential Sp1-binding sites are not conserved among species. Recently, we found that antisense RNA to SPHK1, Khps1, is expressed, and functions as a component for establishing and/or maintaining the DNA methylation pattern in the T-DMR [9]. One of the subtypes, Khps1α, starts from the 5’ end of intron 1 of SPHK1a. Note that the region of its transcription start site is mapped as a conserved region (Fig. 2C). Thus, the multiple conserved regions found in this study indicate a potential role in the regulation of tissue-dependent DNA methylation.

There are many untranslated first exons in the CpG island (Table 1). It has been postulated that the CpG island functions regardless of characteristic sequences such as TATA-box [7, 34]. Since all SPHK1/SPHK1 subtypes identified so far do not seem to share the same transcription start sites, multiple regulatory regions, in addition to T-DMR, may be embedded in the CpG island to express the gene in appropriate tissues or cells [8]. Other genes seem to share similar systems for their expression. For example, the mouse major histocompatibility complex class II I-Aβ gene contains a CpG island in its intron 2, where there are multiple transcription start sites for probable untranslated mRNA specifically expressed in testis [15]. Glucocorticoid receptor and cAMP-dependent protein kinaseRα genes also show the characteristic feature of multiple alternative splicing in the CpG islands [1, 17]. Identification of conserved regions, through comparing sequences among species (see Fig. 2), also would be effective for determining the regulatory sequences within CpG islands.

In conclusion, SPHK1/SPHK1 CpG island showed several similarities among mammalian species. In particular, T-DMR is a core target for tissue-dependent DNA methylation. The discovery of shared sequences and their epigenetic modifications near the CpG island suggests that combination of the locus-specific DNA methylation pattern represents the cellular quality.

ACKNOWLEDGEMENTS. This work was supported in part by a grant from the Program for Promotion of Basic Research Activities for Innovative Biosciences (K. S.), and Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists (T. I.).

REFERENCES


