*Journal of Life Science* 2004, Vol. 14. No. 6. 1009~1017 ©JLS

# Genomic Structure Analyses of Five Kinds of Human Sialyltransferase Gene

**Nam-Young Kang, Sang-Wan Kim, Cheorl-Ho Kim[1] and Young-Choon Lee***

*Faculty of Biotechnology, College of Natural Resources and Life Science, Dong-A University, Busan 604-714, Korea, [1]Department of Biochemistry and Molecular Biology, College of Oriental Medicine, Dongguk University, Kyung-Pook 780-350, Korea*

Received November 29, 2004 / Accepted December 2, 2004

Sialyltransferases cloned so far show the remarkable tissue-specific expression, which is correlated with the existence of cell type-specific sialylated sugar structure in glycoconjugates. In the previous studies, we found various mRNA isoforms of human sialyltransferases generated by alternative splicing and alternative promoter utilization. To understand the regulatory mechanisms for specific expression of human sialyltransferase genes and for production of their mRNA isoforms, in this study, we have isolated and characterized five kinds of human sialyltransferase genes: hST3Gal II, hST8Sia II, hST8Sia III, hST8Sia IV, and hST8Sia V. The hST3Gal II gene is composed of six exons, which span over 17kb, with exons ranging in size from 46 to over 1017 bp. The hST8Sia III gene comprises over 10 kb, and consists of only four exons, which is much smaller and simpler than other human sialyltransferase genes. In contrast, three genes (hST8Sia II, hST8Sia IV and hST8Sia V) span more than 70 kb, and comprise five or more exons. All exon-intron boundaries follow the GT-AG rule. In particular, the sialylmotif L, which is a highly conserved region in all cloned sialyltransferases, was found in one exon of hST8Sia III, whereas this motif is encoded by discrete exons in the other human sialyltransferases. Exon structures of these sialyltransferase genes show the structural diversity, as found in other human sialyltransferase genes reported so far. We determined the transcription start site of hST3Gal II gene by the 5'-RACE and cap site hunting experiments.

**Key words** – Sialyltransferase, Genomic organization, Exon structure. Transcription start site

Sialic acid (NeuAc or Sia) residues occur at the terminal positions of the carbohydrate groups of three types of glycoconjugates (N-and O-glycosidically linked oligosaccharides of glycoproteins, and glycosphingolipids) and play important roles in a variety of biological processes, such as cell-cell communication, cell-matrix interaction, cell differentiation, invasiveness of a number of pathogenic organisms, clearance of asialo glycoproteins from circulation, protein targetting and adhesion of leukocytes to endothelial cells mediated by selectins [26]. In mammalian cells, sialic acids are usually found at the non-reducing terminal position of glycoconjugates sugar chains, α2,3- or α2,6-linked to a β-D-galactopyranosyl (gal) residue, or α2,6-linked to a β-D-N-acetylgalactosaminyl (GalNAc) residue or a β-D-N-acetyl-glucosaminyl (GlcNAc) residue [6,7]. Sialic acids are also found at glycoconjugates chain, α2,8-linked to sialic acid residues in gangliosides and in polysialic acid (PSA) which is a linear α2,8-homopolymer observed on several glycoproteins [6,7,19,22]. The biosynthesis of sialylated oligosaccharide sequences of glycoconjugates is catalyzed by a family of glycosyltransferases called sialyltransferases which reside at Golgi membrane [20]. To date, 18 different sialyltransferase cDNAs have been cloned from several animals and human by PCR and expression cloning methods [7,11,12-16,25]. Since sialyltransferases share the same sugar donors and recognize identical acceptor substrate, it was expected that they would exhibit similar protein sequences. Surprisingly, amino acid sequences of the cloned sialyltransferase cDNAs showed a little homology with the exception of the short consensus sequence called the sialylmotif [25]. However, they all share a common structural topology as in most other glycosyltransferases: a short $NH_2$-terminal cytoplasmic tail, a 13-20 amino acid signal anchor domain, an extended stem region followed by a large COOH-terminal active domain containing sialyl motifs L and S [20,25]. These enzymes so far cloned exhibit acceptor substrate specificities for glycoproteins and glycolipids and show remarkable tissue-specific expression, which is correlated with the existence of cell type-specific carbohydrate structure [20,25]. In general their expression appears to be regulated on the transcriptional level. Precise molecular knowledge of sialyltransferases is vital for understanding the regulatory mechanism for the sialylation of glycocon-

***Corresponding author**
Tel : +82-51-200-7591, Fax : +82-51-200-6993
E-mail : yclee@daunet.donga.ac.kr

jugates.

Previously we successfully have isolated and characterized various human sialyltransferase cDNAs, such as Galβ 1,3GalNAc α2,3-sialyltransferase (hST3Gal II) [11], NeuAcα 2,3Galβ1,3GalNAc α2,6-sialyltransferase (hST6GalNAc IV) [10], Siaα2,3Galβ1,4GlcNAc α2,8-sialyltransferase (hST8Sia III) [17], α2,8-sialyltransferase (hST8Sia V) [12] and GM3 synthase (hST3Gal V) [8]. We also have cloned the genomic DNAs encoding the hST6GalNAc IV and hST3Gal V, and analyzed their promoter functions as well as their genomic structures [8-10]. In addition, mRNA isoforms generated by alternative splicing and alternative promoter utilization were found in these genes [8,10].

As an essential step toward understanding the regulatory mechanisms for specific expression of human sialyltransferase genes and for production of their mRNA isoforms, it is necessary to establish the genomic organization of them and to analyze its promoter function. In this paper we report the genomic structure of five kinds of human sialyltransferase genes that have not analyzed yet. We also determined the transcription start site of hST3Gal II gene by the 5'-RACE and cap site hunting experiments.

## Materials and Methods

### Isolation of human sialyltransferase genomic clone and analysis of genomic structure

Human genomic DNA library (Clontech, Palo, CA) was screened by plaque hybridization method [21] with the [α -$^{32}$P]-radiolabeled cDNA probes. As probes of hST3Gal II, hST8ia II, and hST8Sia V, each fragment containing their full-length cDNAs, which has been previously isolated by our group [11,12,17] was used. Each fragment containing full-length cDNA of hST8Sia II and hST8Sia IV was amplified by polymerase chain reaction (PCR) using primers synthesized on the basis of cDNA sequences reported in the previous papers [19,22]. The sequences of the 5' and 3' primers for hST8Sia II amplifiaction were 5'-ATGCAG-CTGCAGTTCCGGAGCTGG-3' (nucleotides 1-24) and 5'-CTACGTGGCCCCATCGCACTGGCC-3' (complementary to the coding strand, nucleotides 1104-1128), respectively. The sequences of the 5' and 3' primers for hST8Sia IV amplifiaction were 5'-ATGCGCTCCATTAGGAAGAGGTGG-3' (nucleotides 1-24) and 5'-TTATTGCTTTACACACTTT- CCTGTTGTC-3' (complementary to the coding strand, nucleotides 1052-1080), respectively. The cDNA isolated from human fetal brain cDNA library (Clontech, Palo, CA) was used as a template for PCR. PCR amplification was carried out by 30 cycles consisting of 94℃ for 1 min, 55℃ for 1 min and 72 ℃ for 2 min. The PCR products (1.1 kb for hST8Sia II and 1.0 kb for hST8Sia IV) were subcloned into pGEM-T Easy cloning vector (Promega) and sequenced. Each fragment was labeled using Rediprime™ II DNA Labeling System (Amersham Bioscience) and used in the hybridization. Positive clones were isolated, and inserted DNA fragments were initially characterized by restriction digestion and Southern blot analysis [21]. Hybridization-positive fragments were subcloned into Bluescript plasmid vectors (Stratagene, La Jolla, CA), and their sequences were determined using Autocycle DNA sequencing kit and ALF express DNA sequencer (Amersham Bioscience). The size and location of introns were determined by long-distance PCR amplification with LA-Taq polymerase (Takara Biomedicals, Tokyo, Japan). Long-distance PCR was performed with exon-specific primers, and the isolated clones and human genomic DNA (Clontech, Palo, CA) were used as template. The amplified PCR products were ligated into apCR-XL-TOPO Cloning vector (Invitrogen, Groningen, The Netherland) and sequenced. The exon-intron boundaries of each gene were determined by aligning sequences of the subcloned DNA and PCR fragments with the sequence of the cDNA. The analyses of sequence data and the sequence homology were performed using the BLAST (N) program of the National Center for Biotechnology Information (NCBI). In addition, data analyses were performed by extensive search of GeneBank/EMBL human genome sequence database (HTG). Primer design used for genomic structure analysis was performed using the Primer 3 program of Whitehead Institute/MIT Center for Genome Research.

### 5'-Rapid amplification of cDNA ends (RACE) PCR

Amplification of the 5'-end of hST3Gal II was performed with the 5'-RACE kit (Gibco BRL, Life Technologies) according to the manufacturer's instructions, using 5 μg of mRNAs from human adult brain (Clontech, PaloAlto, CA). The gene-specific primer RTI (5'-ATGTTACCGTCAAAGTG-3', complementary to nucleotides 942-960) was used for the initial reverse transcription. After synthesis of the first strand cDNA, an Abridged anchor primer provided by the company and the gene-specific primer (P1, 5'-CCTTGCTG-AGGCGCTGCAGGCCGGC-3', complementary to nucleotides

843-867) were used in the first PCR. The second PCR was performed with an Abridged universal amplification primer and the gene-specific primer (P2, 5'-GCCACGGAG-AGGAACCACACCCGCAG-3', complementary to nucleotides 693-720). The PCR products were subcloned into pGEM-T Easy cloning vector (Promega) and sequenced.

### Determination of the transcription start site

The transcription start site of the hST3Gal II gene was determined by the cap site hunting method [18] with the GeneRacer™kit (Invitrogene, CA) according to the manufacturer's instructions. Total RNA (2 μg) from human fetal brain and human fetal liver (Clontech, Palo, CA) was treated with calf intestinal phosphatase to remove the 5'-phosphates of truncated mRNA and non-mRNA. Dephosphorylated RNA was treated with tobacco acid pyrophosphatase to remove the 5'-terminal $m^7$GpppN cap structure from intact and full-length mRNA, and the resulting 5'-end of decapped mRNA was ligated to the GeneRacer™RNA Oligo using T4 RNA ligase. The first-strand cDNA was synthesized from the ligated mRNA with the above gene-specific primer RTI using Superscript™ II RT. The first round of PCR was performed using the GeneRacer™ 5'-Primer and the above gene-specific primer P1. The second round of PCR (nested PCR) was performed using the GeneRacer™ 5'-Nested Primer and the above gene-specific primer P2. PCR reactions were carried out by 30 cycles consisting of 94℃ for 1 min, 55℃ for 45 sec and 72℃ for 1 min. PCR products were subcloned into pGEM-T Easy cloning vector (Promega) and sequenced. The transcription start site was determined by identification of the boundary sequence between GeneRacer™RNA Oligo and hST3Gal II mRNA sequence.

## Results

### Isolation and genomic organization of five kinds of human sialyltransferase gene

We have previously cloned the cDNAs encoding hST3Gal II (GeneBank Accession No. U63090), hST8Sia III (GeneBank Accession No. AF004668) and hST8Sia V (GeneBank Accession No. U91641) from the human liver and brain cDNA libraries. The cDNA fragments encoding hST8Sia II and hST8Sia IV were prepared by PCR amplification based on cDNA sequences reported previously [19,22]. Using these cDNA fragments as probes, two or four positive clones were isolated from human genomic DNA library and characterized. Most of these clones contained the entire gene structure for a corresponding gene. By comparison of the nucleotide sequences of the genomic clones with the cloned cDNA sequences, which included the 5'-RACE and cap site hunting products in the case of hST3Gal II, most of the exon-intron organization was defined including the exons coding for the entire open reading frame, and 5'- and 3'-UTRs. It was finally confirmed by comparing our data with each sequence of GeneBank/EMBL human genome sequence database (HTG) with BLAST (N) program. In addition, this analysis allowed us to find the chromosomal location of these genes. Exon location maps of these genes are shown in Fig. 1. The hST3Gal II gene in chromosome 16 consisted of 6 exons, ranging in size from 46 bp to over 1017 bp, with intron sizes of about 0.2-6.4 kb and spanning over 17 kb of genomic DNA (Table 1). The coding sequence was divided into six exons (exons 1-6), and the 5'-untranslated region (UTR) of the cDNA was contained in exon 1 (Fig. 1). Exon 1 contains an in-frame start codon (ATG) which starts an open reading frame of hST3Gal II. The hST8Sia II gene in chromosome 15 consisted of 6 exons, ranging in size from 63 bp to 285 bp, with intron sizes of about 3.9-35 kb and spanning over 70 kb of genomic DNA (Table 2). The coding sequence was divided into six exons (exons 1-6), and the short 5'-UTR of the cDNA was contained in exon 1 (Fig. 1). Exon 1 contains an in-frame start codon (ATG) which starts an open reading frame of hST8Sia II. The hST8Sia III gene in chromosome 18 consisted of 4 exons, ranging in size from 125 bp to over 2023 bp, with intron sizes of about 1.3-2.3 kb and spanning over 10 kb of genomic DNA (Table 3). The coding sequence was divided into four exons (exons 1-4), and the 5'-UTR of the cDNA was contained in exon 1 (Fig. 1). Exon 1 contains an in-frame start codon (ATG) which starts an open reading frame of hST8Sia III. The hST8Sia IV gene in chromosome 5 consisted of 5 exons, ranging in size from 132 bp to over 4832 bp, with intron sizes of about 7-43 kb and spanning over 96 kb of genomic DNA (Table 4). The coding sequence was divided into five exons (exons 1-5), and the 5'-UTR of the cDNA was contained in exon 1 (Fig. 1). Exon 1 contains an in-frame start codon (ATG) which starts an open reading frame of hST8Sia IV. The hST8Sia V gene in chromosome 18 consisted of 7 exons, ranging in size from 83 bp to 1394 bp, with intron sizes of about 1.4-51 kb and

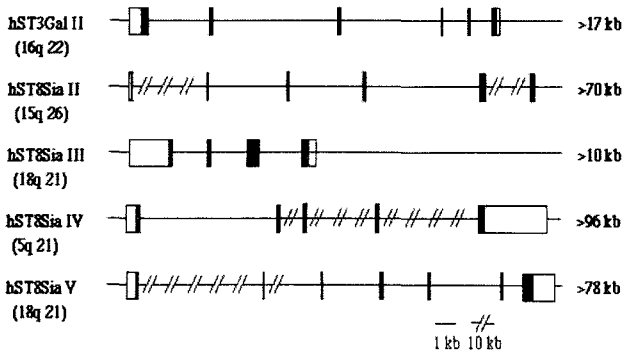


Fig. 1. Structures of the hST3Gal II, hST8Sia II, hST8Sia III, hST8Sia IV, and hST8Sia V genes. Exons are shown by boxes and introns are indicated byhorizontal lines. Black boxes represent coding regions and open boxes denote 5'-and 3'-UTRs.

spanning over 78 kb of genomic DNA (Table 5). The coding sequence was divided into seven exons (exons 1-7), and the 5'-UTR of the cDNA was contained in exon 1 (Fig. 1). Exon 1 contains an in-frame start codon (ATG) which starts an open reading frame of hST8Sia V. All exon-intron boundaries of these genes followed the GT-AG rule that introns begin with GT and end with AG (Tables 1-5) [1].

## 5'-UTR amplification of hST3Gal II by RACE-PCR

To obtain the extended sequence information of 5'-UTR of hST3Gal II mRNA, 5'RACE-PCR was performed using mRNAs isolated from human adult brain, and gene-specific primers P1 and P2 included in the exon 1. A single band of about 1 kb bp in human adult brain tissue was detected, respectively (Fig. 2, lane 2). After subcloning the PCR product and sequencing the clone (RA-P), we found that RA-P clone contained 956 bp of exon 1 (Fig. 3), which is 236 bp longer than the 5'-UTR length (720 bp) of hST3Gal II cDNA cloned previously (GeneBank Accession No. U63090). This results was used in the determination of transcription start site compared with those of cap site hunting described below.

## Identification of the transcription start site of hST3Gal II gene

As the first step toward characterizing the promoter of the hST3Gal II gene, the transcription start site was determined by the cap site hunting method with cap site cDNAs of human fetal liver and fetal brain. As shown in Fig. 2, cDNA with single band of about 1 kb was amplified by nested PCR using total RNA from human fetal brain, whereas three PCR products with different sizes (about 0.6 kb, 0.34 kb and 0.23 kb) were detected from human fetal liver. Sequence analyses of the amplified product revealed that CSH-1 clone with 1011 bp of human fetal brain cDNA contained 55 bp longer fragment than the above 5'-RACE product (Fig. 3). These results indicate that the G residue at 5'-terminus of CSH-1 clone might be the major initiation site of transcription of hST3Gal II gene.

## Discussion

Sialyltransferases are a family of more than eighteen members that catalyze the transfer of sialic acid from

Table 1. Exon/intron organization and boundary sequences of the hST3Gal II gene[a]

| Exon No. | Exon size (bp) | 5'-splice donor | Intron size (bp) | 3'-splice acceptor | Exon No. |
|---|---|---|---|---|---|
| 1 | >1017 | TGG TGG ATG gtaagagct ----<br>W W $M^{113}$ | 3015 | ---- cttccacag ATG CTG CAG<br>$M^{114}$ L Q | 2 |
| 2 | 194 | ATC ATG AG gtgagcccc ----<br>I M $R^{178}$ | 6435 | ---- tccctgcag G ATG AAT CAG<br>$M^{179}$ N Q | 3 |
| 3 | 180 | ATC CGA TT gtgagcttc ----<br>I R $F^{238}$ | 5131 | ---- tgcttacag C ACC TAC GCC<br>$T^{239}$ Y A | 4 |
| 4 | 46 | AAA GAA AAG gtaagctcc ----<br>K E $K^{253}$ | 265 | ---- tcctcccag GTC CAG ATC<br>$V^{254}$ Q I | 5 |
| 5 | 120 | TGT GAT GAG gtgggtagg ----<br>C D $E^{293}$ | 941 | ---- tccccgcag GTG AAC GTG<br>$V^{294}$ N V | 6 |
| 6 | 303 | CCAATCACG | | | |

[a] In exon/intron boundaries, exon sequences are shown in uppercase letters, and introns are shown in lowercase letters. The derived amino acid sequence is displayed below the nucleotide sequence. The amino acid numbering starts at the adenosine of the initiator methionine as +1.

Table 2. Exon/intron organization and boundary sequences of the hST8Sia II gene[a]

| Exon No. | Exon size (bp) | 5′-splice donor | Intron size (bp) | 3′-splice acceptor | Exon No. |
|---|---|---|---|---|---|
| 1 | >123 | GAA ATC GG gtaaatagc ----<br>E I $G^{33}$ | 35886 | ---- gtcttgcag G AAT TCG GGA<br>$N^{34}$ S G | 2 |
| 2 | 63 | TCT AAT AG gtttgtaaa ----<br>S N $R^{54}$ | 4135 | ---- cccttccag A GCT GAA GTT<br>$A^{55}$ E V | 3 |
| 3 | 129 | AGG ATC AG gtactggta ----<br>R I $R^{97}$ | 3976 | ---- ttttcacag G AAG CAG ATT<br>$K^{98}$ Q I | 4 |
| 4 | 258 | GTC ATC AG gtaacatgc ----<br>V I $R^{183}$ | 6025 | ---- ctccggcag G TGC AAC CTG<br>$C^{184}$ N L | 5 |
| 5 | 294 | GTT CGC GG gtgagcggc ----<br>V R $G^{291}$ | 19170 | ---- tttttccag A TAC TGG CTG<br>$Y^{292}$ W L | 6 |
| 6 | 285 | GCC ACG TAG<br>A T STOP | | | |

[a] In exon/intron boundaries, exon sequences are shown in uppercase letters, and introns are shown in lowercase letters. The derived amino acid sequence is displayed below the nucleotide sequence. The amino acid numbering starts at the adenosine of the initiator methionine as +1.

Table 3. Exon/intron organization and boundary sequences of the hST8Sia III[a]

| Exon No. | Exon size (bp) | 5′-splice donor | Intron size (bp) | 3′-splice acceptor | Exon No. |
|---|---|---|---|---|---|
| 1 | >2023 | GGA TTC CG gtgagtgcg ----<br>G F $R^{60}$ | 1373 | ---- cctctccag G TCA CAA TTT<br>$S^{61}$ Q F | 2 |
| 2 | 125 | CAT CAA AGG gtaggatag ----<br>H Q $R^{101}$ | 2386 | ---- ttattttag CAA GAA ATT<br>$Q^{102}$ E I | 3 |
| 3 | 559 | GTC AAC AGG gtgtatatt ----<br>V N $R^{287}$ | 2323 | ---- gttccacag TAC TGG AAA<br>$Y^{288}$ W K | 4 |
| 4 | 607 | AGAGCTATAGTT | | | |

[a] In exon/intron boundaries, exon sequences are shown in uppercase letters, and introns are shown in lowercase letters. The derived amino acid sequence is displayed below the nucleotide sequence. The amino acid numbering starts at the adenosine of the initiator methionine as +1.

Table 4. Exon/intron organization and boundary sequences of the hST8Sia IV gene[a]

| Exon No. | Exon size (bp) | 5′-splice donor | Intron size (bp) | 3′-splice acceptor | Exon No. |
|---|---|---|---|---|---|
| 1 | >423 | CTC ATC GG gtaaatgca ----<br>L I $G^{38}$ | 7056 | ---- ttgtttcag A GAT GGT GAA<br>$D^{39}$ G E | 2 |
| 2 | 132 | GAG ATA AG gtgagtttc ----<br>E I $R^{82}$ | 9052 | ---- caaatacag G AGG AAC ATA<br>$K^{83}$ N I | 3 |
| 3 | 258 | GTA ATA AG gtgagcttt ----<br>V I $R^{168}$ | 29946 | ---- cctcttcag G TGT AAT CTA<br>$C^{169}$ N L | 4 |
| 4 | 294 | GTC AGA GG gtaagtgac ----<br>V R $G^{266}$ | 43973 | ---- tttttccag T TAC TGG CTG<br>$Y^{267}$ W L | 5 |
| 5 | >4932 | CCTTTTTTTCACCTGT | | | |

[a] In exon/intron boundaries, exon sequences are shown in uppercase letters, and introns are shown in lowercase letters. The derived amino acid sequence is displayed below the nucleotide sequence. The amino acid numbering starts at the adenosine of the initiator methionine as +1.

CMP-NeuAc forming α2,3-, α2,6- or α2,8-linkages, depending on the acceptor carbohydrate chain. To date, 15 different human sialyltransferase cDNAs have been cloned and characterized [7]. The primary structures of these enzymes revealed a putative domain structure with a type II transmembrane topology, like other glycosyltransferases, consisting of a short N-terminal cytoplasmic domain, a transmembrane domain and a large C-terminal active domain. These enzymes have highly conserved sialylmotifs L and S which exist in all of the other sialyltransferases cloned so far, but not in other glycosyltransferases [7,25]. These also exhibit acceptor substrate specificities for glycoproteins and glycolipids and show remarkable tissue-specific expression. Among them, the genomic structures of five human sialyltransferase genes have been reported to date: hST3Gal I [2,23], hST3Gal IV [13], hST3Gal VI [24], hST6Gal I [27], hST8Sia I [3]. In addition, recently we have determined the genomic structures of hST3Gal V [4] and hST6GalNAc IV [10]. They have their coding sequences distributed over several exons and are dispersed in the human genome.

In this study, we determined the genomic structures of five human sialyltransferase genes (hST3Gal II, hST8Sia II, hST8Sia III, hST8Sia IV and hST8Sia V), which have not reported yet. Surprisingly, the genomic structure of the

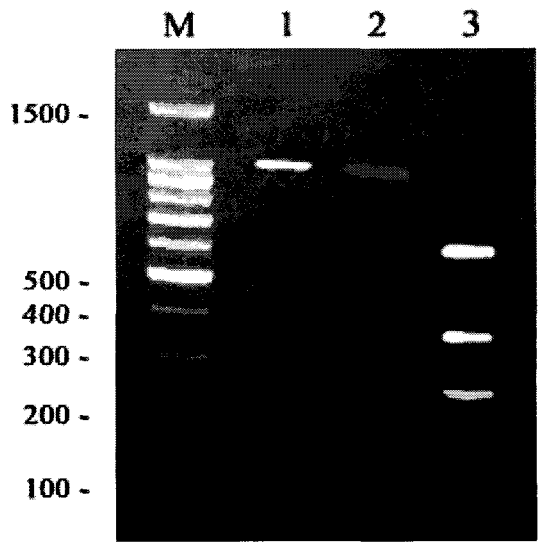


Fig. 2. Agarose gel (2%) analysis of PCR products amplified by the 5'-RACE and cap site hunting. Lanes 1, PCR product from human fetal brain by cap site hunting; lane 2, PCR product from human adult brain by 5'-PACE; lane 3PCR product from human fetal liver by cap site hunting; M, DNA size marker (100 bp ladder).

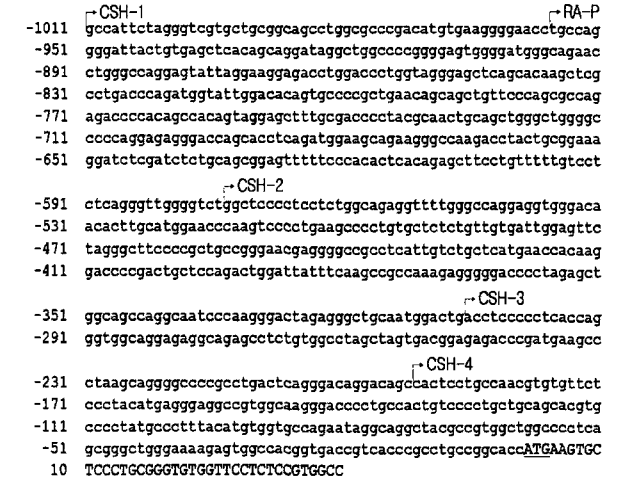


Fig. 3. Nucleotide sequences of PCR products amplified by the 5'-RACE and cap site hunting. CSH-1 shows the 1011 bp sequence amplified by cap site hunting from human fetal brain. CSH-2, CSH-3 and CSH-4 show the 624 bp, 346 bp and 232 bp sequences amplified by cap site hunting from human fetal liver, respectively. RA-P shows the 956 bp sequence amplified by 5'-RACE from human adult brain. The bold bases (ATG) indicates the initiation codon of hST3Gal II ORFA. The N-terminal amino acid sequence of hST3Gal II is shown as the capital letters.

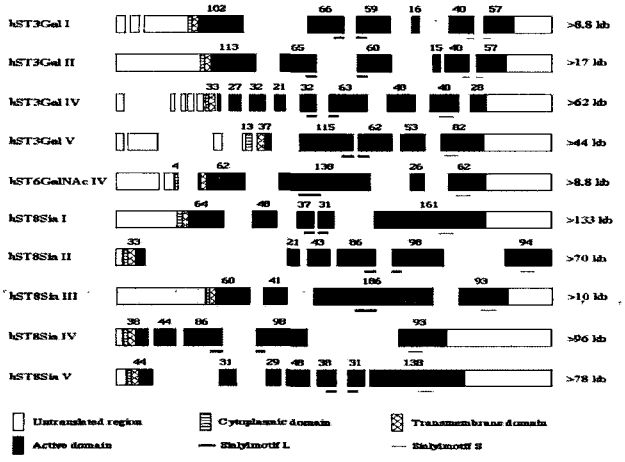


Fig. 4. Comparison of the exon structures of human sialyltransferase genes. The protein domain structure is represented schematically by a rectangle, which is subdivided to show the major structural elements of the protein. Thick and thin underlines show sialyl motifs L and S, respectively.

hST3Gal II coding region is almost identical to that of hST3Gal I, and moreover, the split patterns of coding sequence for sialyl motifs L and S are also identical (Fig. 4). Previously, our group and the other group have

demonstrated that mammalian ST3Gal I and II have similar substrate specificities and both can utilize Galβ 1,3GalNAc, *O*-Thr/Ser and glycolipid as substrate [4,5,11, 15,16]. Compared to ST3Gal I with the same acceptor substrate specificity, ST3Gal II exhibits much more acceptor substrate preference for glycolipids than for O-linked oligosaccharide of glycoproteins [14,15].

The hST8Sia II and IV also have a similar genomic structure including the split patterns of coding sequence for sialyl motif L and the same position of sialyl motif S on single exon (Fig. 4). These two enzyme can synthesize polysialic acid (PSA) on α2,3-linked sialic acid on neural cell adhesion molecule (N-CAM) without any initiators [19, 22,25], indicating that hST8Sia II has almost the same substrate specificity as hST8Sia IV. Based on these observations, we can hypothesize that the sialyltransferases that have similar substrate specificities arose from a common ancestral gene. On the other hand, the hST3Gal I gene is located on human chromosome 8 and this position is entirely distinct from the hST3Gal II gene located on human chromosome 16. Similarly, the hST8Sia II and IV genes are located on human chromosomes 15 and 5, respectively. These suggest that they have arisen from duplication and subsequent translocation events.

Our data presents that hST8Sia III is much smaller ant its organization much simpler than other human sialyltransferase genes reported so far, which span more than 17 kb, except hST3Gal I and hST6GalNAc IV, and comprise five or more exon (Fig. 4). In particular, the sialyl motif L of hST8Sia III, which is a highly conserved region in all cloned human sialyltransferases, is in one exon, as in the hST6GalNAc IV.

Sialyltransferases cloned to date exhibit remarkable tissue-specific expression, which is correlated with the existence of cell type-specific carbohydrate structure [6,7, 20,25]. In general their expression appears to be regulated on the transcriptional level. Therefore, precise molecular knowledge for transcriptional regulation of sialyltransferase genes is vital for understanding the regulatory mechanism for the sialylation of glycoconjugates. Because transcriptional regulations of sialyltransferase genes depends on their cell type-specific promoter functions, the determination of their transcription start sites is essential to isolate and characterize the promoter regions of their genes. In the previous papers [9,10], we determined transcription start sites of hST3Gal V and hST6GalNAc IV genes by cap site hunting and then isolated their promoter regions. 5'-RACE was also used to isolate the promoter regions of hST3Gal VI [24] and hST3Gal I [23].

In this study, we performed 5'-RACE and cap site hunting experiments to determine transcription start sites of hST3Gal II gene. Among the amplified PCR products, the longest fragment of 1011 bp from human fetal brain was considered to be a cDNA extended to the cap site of hST3Gal II mRNA. Therefore, we assigned the major initiation site of transcription of hST3Gal II gene to the G residue at a position -1011 (Fig. 4). Based on this finding, cloning of the upstream genomic region of this transcription start site is in progress to clarify promoter function for elucidation of the mechanism of regulation and tissue-specific expression of the hST3Gal II gene.

## Acknowledgments

This work was supported by the Korea Research Foundation Grant (KRF-2003-002-C00199).

## References

1. Breathnach, R. and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**, 349-383.
2. Chang, M. L., R. L. Eddy, T. B. Shows and J. T. Lau. 1995. Three genes that encode human β-galactoside α2,3-sialyltransferases. Structural analysis and chromosomal mapping studies. *Glycobiology* **5**, 319-325.
3. Furukawa K, M. Horie, K. Okutomi, S. Sugano and K. Furukawa. 2003. Isolation and functional analysis of the melanoma specific promoter region of human GD3 synthase gene. *Biochim. Biophys. Acta* **1627**, 71-78.
4. Gillespie, W., S. Kelm and J. C. Paulson. 1992. Cloning and expression of the Galβ1,3GalNAc α2,3-sialyltransferase. *J. Biol. Chem.* **267**, 21004-21010.
5. Giordanengo, V. S. Bannwarth, C. Laffont, V. Van Miegem, A. Harduin-Lepers, P. Delannoy and J.-C. Lefebvre. 1997. Cloning and expression of cDNA for a human Galβ 1-3GalNAc α2,3-sialyltransferase from the CEM T-cell line. *Eur. J. Biochem.* **247**, 558-566.
6. Harduin-Lepers, A., M. A. Recchi and P. Delannoy. 1995. The year of sialyltransferases. *Glycobiology* **5**, 741-758.
7. Harduin-Lepers, A., V. Vallejo-Ruiz, M. A. Krzewinski-Recchi, B. Samyn-Petit, S. Julien and P. Delannoy. 2001. The sialyltransferase family. *Biochimie* **83**, 727-737.
8. Kim, K.-W., S.-W. Kim, K.-S. Min, C.-H. Kim and Y.-C. Lee. 2001. Genomic structure of human GM3 synthase gene (hST3Gal V) and identification of mRNA isoforms in the 5'-untranslated region. *Gene* **273**, 163-171.

9. Kim, S.-W., S.-H. Lee, K.-S. Kim, C.-H. Kim, Y.-K. Choo and Y.-C. Lee. 2002. Isolation and characterization of the promoter region of the human GM3 synthase gene. *Biochim. Biophys. Acta* **1578**, 84-89.
10. Kim, S.-W., S.-H. Lee, K.-W. Kim, N.-Y. Kang, K.-S. Kim, C.-H. Kim and Y.-C. Lee. 2003. Genomic structure and promoter analysis of human NeuAcα2,3Galβ1,3GalNAc α 2,6-sialyltransferase (hST6GalNAc IV) gene. *Gene* **305**, 113-120.
11. Kim, Y.-J., K.-S. Kim, S.-H. Kim, C.-H. Kim, J. H. Ko, I.-S. Choe, S. Tsuji and Y.-C. Lee. 1996. Molecular cloning and expression of human Galβ1,3GalNAcα2,3-sialyltransferase (hST3Gal II). *Biochem. Biophys. Res. Commun.*, **228**, 324-327.
12. Kim, Y.-J., K.-S. Kim, S.-I. Do, C.-H. Kim, S.-K. Kim and Y.-C. Lee. 1997. Molecular cloning and expression of human α2,8-sialyltransferase (hST8Sia V). *Biochem. Biophys. Res. Commun.*, **235**, 327-330.
13. Kitagawa, H., M.-G. Mattei and J. C. Paulson. 1996. Genomic organization and chromosomal mapping of the Galβ1,3GalNAc/Galβ1,4GlcNAc α2,3-sialyltransferase. *J. Biol. Chem.* **271**, 931-938.
14. Kojima, N., Y.-C. Lee, T. Hamamoto, N. Kurosawa and S. Tsuji. 1994. Kinetic properties and acceptor substrate preferences of two kinds of Galβ1,3GalNAcα2,3- sialyltransferase from mouse brain. *Biochemistry.* **33**, 5772- 5776.
15. Lee, Y.-C., N. Kojima, E. Wada, N. Kurosawa, T. Nakaoka, T. Hamamoto and S. Tsuji. 1994. Cloning and expression of cDNA for a new type of Galβ1,3GalNAcα2,3- sialyltransferase. *J. Biol. Chem.*, **269**, 10028-10033.
16. Lee, Y.-C., N. Kurosawa, T. Hamamoto, T. Nakaoka and S. Tsuji. 1993. Molecular cloning and expression of Galβ 1,3GalNAcα2,3-sialyltransferase from mouse brain. *Eur. J. Biochem.*, **216**, 377-385.
17. Lee, Y.-C., Y.-J. Kim, K.-Y. Lee, K.-S. Kim, B.-U. Kim, H.-N. Kim, C.-H. Kim and D.-I. Do. 1998. Cloning and expression of cDNA for a human Siaα2,3Galβ1,4GlcNAc α2,8-sialyltransferase (hST8Sia III). *Arch. Biochem. Biophys.* **360**, 41-46.
18. Maruyama, K. and S. Sugano. 1994. Oligo-capping: A simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138**, 171-174.
19. Nakayama, J., M. N. Fukuda, B. Fredette, B. Ranscht and M. Fukuda. 1995. Expression cloning of a human polysialyltransferase that forms the polysialylated neural cell adhesion molecule present in embryonic brain. *Proc. Natl. Acad. Sci.* **92**, 7031-7035.
20. Paulson, J. C. and K. J. Colley. 1989. Glycosyltransferases: structure, localization and control of cell type-specific glycosylation. *J. Biol. Chem.*, **264**, 17617-17618.
21. Sambrook, J., E. F. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22. Scheidegger, E. P., L. R. Sternberg, J. Roth and J. B. Lowe. 1995. A human STX cDNA confers polysialic acid expression in mammalian cells. *J. Biol. Chem.* **270**, 22685-22688.
23. Taniguchi, A., I. Yoshikawa and K. Matsumoto. 2001. Genomic structure and transcriptional regulation of human Galβ1,3GalNAc α2,3-sialyltransferase (hST3Gal I) gene. *Glycobiology* **11**, 241-247.
24. Taniguchi, A., R. Kaneta, K. Morishita and K. Matsumoto. 2001. Gene structure and transcriptional regulation of human Galβ1,4(3)GlcNAc α2,3-sialyltransferase VI (hST3Gal VI) gene in prostate cancer cell line. *Biochem. Biophys. Res. Commun.* **287**, 1148-115.
25. Tsuji, S. 1996. Molecular cloning and functional analysis of sialyltransferases. *J. Biochem. (Tokyo)*, **120**, 1-13.
26. Varki, A. 1993. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology.* **3**, 97-130.
27. Wang, X., A. Vertino, R. L. Eddy, M. G. Byers, S. N. Jani-Sait, T. B. Shows and J. T. Lau. 1993. Chromosome mapping and organization of the human β-galactoside α 2,6-sialyltransferase gene. Differential and cell-type specific usage of upstream exon sequences in B-lymphoblastoid cells. *J. Biol. Chem.* **268**, 4355-4361.

## 초록 : 5종류의 인간유래 시알산전이효소 유전자들의 게놈구조 분석

강남영 · 김상완 · 김철호[1] · 이영춘*

(동아대학교 생명자원과학부, [1]동국대학교 한의과대학 생화학 분자생물학교실)

인간유래 시알산전이효소 유전자들의 특이적 발현과 그들의 mRNA isoform의 생성에 대한 조절기구를 이해하기 위하여 5종류의 human 시알산전이효소 유전자(hST3Gal II, hST8Sia II, hST8Sia III, hST8Sia IV, hST8Sia V)들의 게놈구조를 분석하였다. hST3Gal II 유전자는 17 kb이상의 게놈상에 46 bp에서 1017 bp의 길이를 가진 exon이 6개로 이루어져 있고, hST8Sia III 유전자는 10 kb이상의 게놈상에 125bp에서 2023bp의 길이를 가진 exon이 4개로 이루어져 있어 다른 human 시알산전이효소 유전자들보다 짧고 단순한 구조를 가지고 있었다. 반면에 다른 3종류의 유전자(hST8Sia II, hST8Sia IV, hST8Sia V)들은 70 kb이상의 게놈상에 5개이상의 exon으로 이루어져 있으며, 5종류 모두 exon-intron boundary는 GT-AG rule을 나타내고 있었다. 특히 모든 시알산전이효소에 고도로 보존되어 있는 sialylmotif L은 hST8Sia III 유전자에서는 하나의 exon에 존재하는 반면에, 다른 시알산전이효소 유전자에서는 분리된 exon에 존재하여 exon의 구조적 다양성을 나타내고 있다. 또한, 본 연구에서는 5'-RACE와 cap site hunting법에 의해 hST3Gal II 유전자의 전사개시점을 결정하였다.