Glycoconjugate Journal 19, 197–210, 2003

Cloning and sequencing of nineteen transcript isoforms of the human α2,3-sialyltransferase gene, *ST3Gal III*; its genomic organisation and expression in human tissues

Ammi Grahn, Giti Shah Barkhordar and Göran Larson*

Institute of Laboratory Medicine, Department of Clinical Chemistry and Transfusion Medicine, Sahlgrenska University Hospital, SE 413 45 Göteborg, Sweden

The recruitment of human peripheral blood leukocytes (PBL) to sites of infection and inflammation requires the surface expression of Sialyl Lewis x glycoconjugates (SLe^x) on white blood cells and their interaction with E- and P-selectins on activated endothelial cells. E-selectin has additionally been shown to interact with the sialyl Lewis a (SLe^a) epitope. Human *ST3Gal III* codes for an α2,3-sialyltransferase involved in the biosynthesis of both SLe^a and SLe^x epitopes, although the latter with a lower efficiency. We have cloned and sequenced human *ST3Gal III* gene transcripts from human peripheral blood leukocytes, covering the coding region of this gene. Within our clones we isolated 19 different transcripts with a wide variety of deletions from 45 to 896 nucleotides, and insertions of 26 to 173 nucleotides. Among the insertions we identified two new exons (E3, E6). In order to map and characterise the *ST3Gal III* gene we used the GenBank database and "computer-cloned" and characterised the genomic organisation of the *ST3Gal III* gene. The coding sequences of the *ST3Gal III* gene stretch over a gene sequence of approximately 223 Kb comprised of 15 exons. RT-PCR and laser-induced fluorescent capillary electrophoresis (LIF-CE) were used to examine the expression of this gene in twenty-one human tissues, which showed a highly specific tissue expression pattern. Neural and muscular tissues showed the most complex patterns and were distinctly different from all other tissues examined.
Published in 2003.

Keywords: **sialyltransferase, *ST3Gal III*, transcripts, alternative splicing, capillary electrophoresis**

Abbreviations: ***ST3Gal III*** **(*SIAT6, ST3N*), α2,3-sialyltransferase III (EC 2.4.99.6, L23768, E08204, Q11203); Human genome project chromosome 1 (Acc. No. AC005193, AC004950, AC005038); Glyceraldehyde-3-phosphate dehydrogenase, GAPDH (J04038, NM_002046); *ST3Gal III* isoform A1 (AF425851); *ST3Gal III* isoform A7 (AF425852); *ST3Gal III* isoform A8 (AF425853); *ST3Gal III* isoform B1 (AF425862); *ST3Gal III* isoform B1-90 (AF425854); *ST3Gal III* isoform B3 (AF425855); *ST3Gal III* isoform B4 (AF425856); *ST3Gal III* isoform B7 (AF425857); *ST3Gal III* isoform B8 (AF425858); *ST3Gal III* isoform C1 (AF425859); *ST3Gal III* isoform C7 (AF425860); *ST3Gal III* isoform C8 (AF425861); *ST3Gal III* type D2+26 (AF425863); *ST3Gal III* type B1+32 (AF425864); *ST3Gal III* type B4+173 (AF425865); *ST3Gal III* type B5+26 (AF425866); *ST3Gal III* type B5+173 (AF425867); *ST3Gal III* type E1 (AF425868); *ST3Gal III* type E3+32 (AF425869); LIF-CE, laser induced fluorescent capillary electrophoresis; PBL, peripheral blood leukocytes.**

Introduction

Cell surface glycoproteins have been shown to have key functions in intercellular adhesion mechanisms. The recruitment of human peripheral blood leukocytes (PBL) to sites of infection and inflammation requires the surface expression of Sialyl Lewis x glycoconjugates (SLe^x) on the white blood cells and of E- and P-selectins on the activated endothelial cells [1–3]. It is also known that E-selectin binds the Sialyl Lewis a (SLe^a) epitope [2,4]. Endothelial tumors especially, colon carcinomas and pancreatic carcinomas, often express SLe^a and the interaction between SLe^a and E-selectin is considered important for the metastatic properties of these carcinoma cells [5]. This

*To whom correspondence should be addressed: Göran Larson, Institute of Laboratory Medicine, Department of Clinical Chemistry and Transfusion Medicine, Sahlgrenska University Hospital, SE 413 45 Göteborg, Sweden. Tel.: +46-31-342-1330; Fax: 46-31-828458; E-mail: Goran.Larson@clinchem.gu.se

is not only of diagnostic and prognostic importance but may also be of therapeutic interest, as recently reported in a 10 year follow-up study from Japan [6] where colon carcinoma patients were treated with cimetidine in parallel to standard chemotherapeutics. In this study there was a highly significant prolonged survival time for patients additionally treated with cimetidine and the effects were most pronounced for patients carrying tumors with high expression of SLe^x and SLe^a epitopes. The pharmacological mechanism of cimetidine in cancer treatment, possibly mediated by a down regulation of E-selectin expression on endothelial cells, was commented on in an editorial [7].

Sialyltransferases catalyse the transfer of sialic acid from CMP-sialic acid to the terminal position of carbohydrates on glycoproteins and glycolipids. Analysis of the protein sequences of sialyltransferases reveals, as for most glycosyltransferases, a topological feature characteristic for type 2 transmembrane proteins with a short N-terminal cytoplasmic tail, a hydrophobic signal-anchor sequence that spans the membrane, a lumenal stem region and a large catalytic domain that resides in the lumen of the Golgi [8]. All sialyltransferases cloned to date contain two conserved sialylmotifs in the catalytic part of the enzyme. A large sialylmotif of approximately 48 amino acids is termed the "L(large)-sialylmotif" and a smaller sequence of about 23 amino acids towards the C-terminal is termed the "S(short)-sialylmotif". The L-motifs participate in the binding of the donor substrate, CMP-Neu5Ac, and the S-motifs participate in the binding of both the donor and the acceptor substrates [9–11]. Each of the two motifs contains an invariant cysteine residue present in all sialyltransferases cloned [10,12]. The conserved cysteine residues participate in a disulfide linkage between the two motifs and are critical for conformation and activity of the enzymes. Additionally a very short sequence, the "VS-motif", further down towards the C-terminal comprising only 3-5 amino acids has been identified as a conserved sequence although its importance for enzymatic activity has not yet been published [13].

The human α2,3-sialyltransferase genes *ST3Gal IV* and ST3Gal *VI*, located on chromosome 11 [14] and 3 [NCBI, GenBank]] respectively, code for enzymes that catalyse the transfer of sialic acid to the type 2 precursor (Galβ1, 4GlcNAc) [15,16] and are critical for the biosynthesis of SLe^x. Human *ST3Gal III* is located on chromosome 1 [14] and acts preferentially on the type 1 precursor (Galβ1,3GlcNAc) and is thus the major candidate responsible for the biosynthesis of SLe^a [17]. *In vitro* studies of *ST3Gal III* showed an additional catalytic activity towards type 2 precursors, but with lower efficiency [17]. It is well established that in the biosynthesis of SLe^a the type 1 precursor has to be sialylated prior to fucosylation by an α1,4-fucosyltransferase (Fuc-TIII) which thus transfers fucose from GDP-fucose to the sialylated type 1 precursor (Figure 1). Fuc-TIII also acts on non sialylated type 1 precursor forming Le^a [18,19]. Thus, there is a possible competition for the type 1 precursor between the α2,3-sialyltransferase and the α1,4-fucosyltransferase in the Golgi. In the biosynthesis of SLe^x an α1,3-fucosyltransferase, Fuc-TVII, transfers fucose to the sialylated type 2 precursor. In granulocytes Fuc-TVII can only transfer fucose to the sialylated and not to the non sialylated type 2 precursor and is thus in those cells responsible for the final biosynthesis of SLe^x [20,21]. In PBL the Fuc-TIV and in the brain the Fuc-TIX competes with the α2,3-sialyltransferase for the type 2 precursor to form Le^x instead of the sialylated compound. The localisation, activities and detailed specificities of those transferases are thus critical for the carbohydrate epitopes being synthesised in each tissue [22].

Northern blot analysis of *ST3Gal III* has shown that this gene is predominantly expressed in skeletal muscle but also

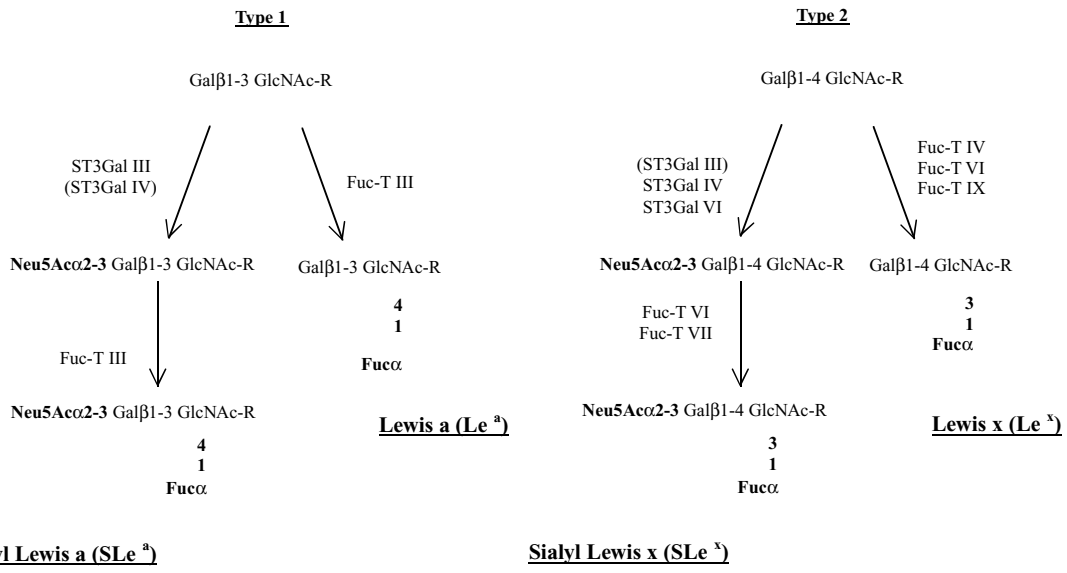

Figure 1. Biosynthesis of lacto series sialylated and non sialylated Le^a and Le^x antigens. Relevant sialyl- and fucosyltransferases are indicated for each step.

Table 1. *ST3Gal III* and GAPDH primers used for cloning and fragment analysis

Primers[a]	*Genes*	*Sequences*	*Annealing positions*[b]
ST III-11s	ST3Gal III	5′-TAGGAAATCGTAAATCATGTGAAG-3′	−1 to −24
ST III-11as	ST3Gal III	5′-TCAGATGCCACTGCTTAGATCAG-3′	1313 to 1335
FAM-GAPDH-s	GAPDH	5′-6-FAM-AGTCCACTGGCGTCTTCAC-3′	290 to 311
2-GAPDH-as	GAPDH	5′-GCCACAGTTTCCCGGAGG-3′	589 to 572
GAPDH-as	GAPDH	5′-GAGGTCCACCACCCTGTTG-3′	981 to 963

[a]For fragment analysis 6-FAM labelled sense and HEX labelled antisense primers were used.
[b]Numbers refer to the *ST3Gal III* A8 isotranscript and to the *GAPDH* gene (Acc. No NM_002046).

in several other human tissues [23]. In order to characterise all the candidate α2,3-sialyltransferases necessary for biosynthesis of SLe^x we have cloned and sequenced human *ST3Gal III, ST3Gal IV* and *ST3Gal VI* gene transcripts from PBL of one single individual. The *ST3Gal IV* isotranscripts have already been published [24]. Our clones have revealed a considerable heterogeneity in transcript isoforms corresponding to alternative splicing. Global database studies comparing genomes and Expressed Sequence Tags (ESTs) have recently shown that an unexpected high frequency of genes (38 to 42%) occur as alternatively spliced forms in these databases [25,26]. In a recent report Modrek et al. [25] revealed in a random sample of their database, that 74% of alternative splices really modified the protein product, whereas only 22% were alternatively spliced in the 5′-UTR and only 4% in the 3′-UTR. The most abundant category of alternatively spliced genes occurred in cell surface proteins/receptors (29%) including membrane anchored receptors, integral membrane proteins and proteins involved in cell surface adhesion. Twenty-nine per cent of the alternatively spliced genes encoded functions specific to the immune system cell surface receptors. Fourteen per cent of alternatively spliced genes showed tissue specificity for the minor isoforms. Within our clones from *ST3Gal III* gene, we have isolated 19 different transcripts with a wide variety of deletions from 45 to 896 nucleotides, and insertions of 26 to 173 nucleotides. Among the insertions we identified two exons not described before. In order to map and characterise the expression pattern of the various isotranscripts of *ST3Gal III* in human tissues, we have utilized the technique of RT-PCR and laser-induced fluorescent capillary electrophoresis (LIF-CE).

Materials and methods

"Computer-cloning" of human *ST3Gal III* gene

The human *ST3Gal III* mRNA sequences with Acc. no. L23768 [17] and Acc. no. E08204 were used to search the human genome database at the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) using BLAST human genome search (http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs). The BLAST results were analysed and aligned using the Lasergene software package (DNASTAR Inc., Madison, WI). ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was used for analysis of the open reading frames (ORFs).

Cloning and sequencing of *ST3Gal III* transcripts

Total RNA was prepared from blood samples (5 ml of EDTA anticoagulated blood) of one single individual using the Qiagen RNeasy midikit (Qiagen Ltd, UK). cDNA was synthesized using Promega Reverse transcription system with poly dT-primers (Promega Corp. Madison, USA). Primers were designed to cover the coding region of the *ST3Gal III* gene [17] using the human *ST3Gal III* sequence with Acc. no. E08204 (Sasaki K. et al. unpublished). All primers (Table 1) were purchased from Scandinavian Gene Synthesis AB (Köping, Sweden).

In the PCR reaction for *ST3Gal III* the final concentrations were 1.0 μM of STIII-11s and STIII-11as, 150 μM dNTP (Amersham Pharmacia Biotech AB, Uppsala, Sweden), 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Trizma base and 1 unit of Taq DNA polymerase (Roche Diagnostics GmbH, Mannheim, Germany). Ten μl of cDNA was used as template. The PCR-program was run at 85°C 10 min, 95°C 15 sec, 60°C 15 sec, 68°C 3 min for 30 cycles and finally 68°C for 10 min.

The amplified fragments were purified on Agarose gel (GIBCO BRL, Life Technologies, Paisley, Scotland) and recovered using GenElute Minus EtBr Spin Columns (Sigma, St Louis, MO, USA). Fragments were ligated into the pTAg cloning vector pCR2.1 and transformed into INVαF′ One Shot competent cells (Invitrogene BV, Groningen, the Netherlands). All selected clones were sequenced in both directions using Big Dye Terminator kit on Applied Biosystems 377 DNA sequencers and ABI PRISM 310 instrument (P-E Applied Biosystems, Foster City, CA). Sequence analyses and alignments were performed using the Lasergene software package (DNASTAR Inc., Madison, WI, USA).

PCR and RT-PCR amplification with fluorescent primers

Total RNA-samples from a panel of tissues were purchased from Clontech (Cat no K4005-Z, Clontech, Palo Alto, CA, USA). The panel contained human RNA samples from placenta, foetal brain, foetal liver, adrenal gland, lung, testis, skeletal

muscle, brain, spleen, liver, uterus, thymus, salivary gland, trachea, heart, kidney, bone marrow, spinal cord, prostate and colon. Total RNA from peripheral blood leukocytes was prepared as described above.

Sense primers for all the *ST3Gal III* transcripts were labelled in the 5′-end with the fluorescent dye 6-carboxyfluorescein (6-FAM) without any other change in the primers used for cloning. The antisense primers were similarly labelled with the hexachlorinated analogue to 6-FAM (HEX) (Scandinavian Gene Synthesis AB) (Table 1). The glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene (Acc. no. NM_002046) transcript was used as an external positive control, with a 6-FAM labelled GAPDH sense primer and an unlabelled antisense primer, giving rise to a fragment of 692bp. About 0.1 μg RNA was used in the RT-step and 10 μl of the transcribed cDNA was further used as template. In the PCR reaction the final concentrations were 600 nM of FAM-STIII-11s and HEX-STIII-11as, 1.0 mM $MgCl_2$, 50 mM KCl, 10 mM Trizma base and 1 unit of Taq DNA polymerase (Roche Diagnostics GmbH, Mannheim, Germany). In the PCR reaction for GAPDH the final concentrations were 40 nM of FAM-GAPDH-s and GAPDH-as, 1.0 mM $MgCl_2$, 50 mM KCl, 10 mM Trizma base and 1 unit of Taq DNA polymerase. The standard PCR-program was 85°C 10 min, 95°C 15 sec, 60°C 15 sec, and 68°C 3 min for 30 cycles.

Laser induced fluorescent capillary electrophoresis

Electrophoretic analyses were performed on an ABI PRISM 310 instrument equipped with ABI PRISM GeneScan Analysis Software package (P-E Applied Biosystems). One μl of the amplified products was mixed with 0.5 μl TAMRA-labelled GS-2500 size standard (P-E Applied Biosystems) and 12 μl of distilled water. The amplified GAPDH fragment was diluted 1:20 in distilled water and mixed with the size marker as above. A 1 ml Hamilton syringe and a 47 cm × 50 μm capillary with 3% GeneScan polymer and 1X Genetic Analyser Buffer with EDTA were used (P-E Applied Biosystems). The injection time was set at 5 sec at 15 kV and electrophoreses were run at 42°C at 2 kV with 90 min detection time using virtual filter set C, which records the fluorescent light intensities in four windows centred around 530, 542, 567 and 590 nm. Filter set C was used to provide maximal separation of the different dyes and to keep a low signal to noise ratio.

Results

"Computer-cloning" of the human *ST3Gal III* gene

The BLAST search revealed three human genomic clones containing *ST3Gal III* homologous sequences; AC005193, AC004950 and AC005038, all located on the (−)-strand of the human chromosome 1. These three clones overlapped with 200 bases. According to these clones, the *ST3Gal III* gene is a gene with a length of approximately 223 Kb. We identified 13 exons when comparing the two human cDNA sequences (Acc. no. L23768 and Acc. no. E08204) with the three genomic clones. The intron-exon junctions showed perfect agreement with the GT-AG consensus sequence [27,28]. The exon-intron splice junction sequences and the corresponding locations on the human clones are presented in Table 2. In Table 3 the exon lengths and corresponding cDNA nucleotides and protein amino acid numbers are shown. The cDNA sequences revealed two separate stop-codons localized at nt 1126–1128 and nt 1556–1558 for clone L23768 and E08204 respectively. In the sequences of E08204 the three nucleotides corresponding to the stop-codon of the sequence of L23768 was lost, possibly spliced off, and the sequence continued without disrupting the open reading frame. Since our anti sense primer STIII-11as included, at the 3′- end, the stop codon for the sequence L23768, we cannot exclude the presence of such an isoform in the tissues examined.

Table 2. *ST3Gal III* exon and intron splice junctions and corresponding position in the genome

Intron	*Intron length (bp)*	*Splice junction sequence*		*Position on human genomic clone (nt)*	*Human genomic clone Acc. no.*
		Exon : Intron	*Intron : Exon*		
E1/E2	28 537	CCGCTGTG: **gt**aagggc	ttttct **ag**:GTCATTTA	98 477 – 69 941	AC005193
E2/E3	No intron	GGACTCCA:	:GTAAGTAT	No intron	AC005193
E3/E4	55 679	TGTTGCAG: **gt**cagtta	ggttgc **ag**: ATTCAGTG	69 748 – 14 071	AC005193
E4/E5	22 754	AGGCTCAG: **g**taccaac	tttttc **ag**:AGTATGAT	14 022 – 84 684	AC005193/AC004950
E5/E6	9 823	TCTAAACT: **gt**gagtag	ctttgc **ag**:CTCACCGA	84 636 – 74 813	AC004950
E6/E7	13 337	CTGCCCAG: **gt**gagtct	tcccat **ag**:GCCTGCTG	74 651 – 61 315	AC004950
E7/E8	56 051	TTCCCCCG: **gt**aagtgc	tattcc **ag**:GTTCTCCA	61 219 – 5 169	AC004950
E8/E9	3 758	AGGTCAAG: **gt**atgttg	gtctgc **ag**:ACAATCTG	5 073 – 1 316	AC004950
E9/E10	869	TTGGACAG: **gt**gagcca	cccctc **ag**:CCTCCGCT	1 251 – 383	AC004950
E10/E11	278	GTGGTGAG: **gt**gagctc	ctctgt **ag**:ACTGAATT	286 – 190 554	AC004950/AC005038
E11/E12	20 129	AGAGAGTG: **gt**aagctc	gtccac **ag**:AGTGCATC	190 366 – 170 238	AC005038
E12/E13	230	GCCGGGGG: **gt**gagata	tcttct **ag**:AACATCCC	170 089 – 169 860	AC005038
E13/E14	9 229	TCAAAGAG: **gt**tcgggg	cctttc **ag**:TCCTGGAC	169 712 – 160 484	AC005038
E14/E15	50	GCATCTGA: **gt**gggccc	agcagc **ag**:CCAGCACC	160 392 – 160 343	AC005038

Table 3. *ST3Gal III* gene exons and corresponding cDNA nucleotides and protein amino acid numbers

Exons (no.)	*Exon lengths (bp)*	*cDNA nucleotides (position no.)*[a]	*Amino acids (position no.)*[a]
1	121	−151 – (−30)	Noncoding
2	148	−29 – 118	1 – 39
3	45	119 – 163	40 – 54
4	48	164 – 211	55 – 70
5	43	212 – 254	71 – 85
6	162	255 – 416	86 – 139
7	93	417 – 509	140 – 170
8	95	510 – 604	171 – 201
9	64	605 – 668	202 – 223
10	96	669 – 764	224 – 255
11	187	765 – 951	256 – 317
12	147	952 – 1098	318 – 366
13	147	1099 – 1245	367 – 415
14	90	1246 – 1335	416 – 444
15	884	1336 – 2219	Noncoding

[a]The *ST3Gal III* isotranscripts A8 (Acc. no. AF425853), being the longest transcript not showing any exon losses, is used throughout this manuscript as the basis for all numbering of nucleotides and potential amino acids.

Characterisation of 19 *ST3Gal III* transcripts from human peripheral blood leukocytes

Within our clones we isolated 19 different cDNA transcripts covering the coding region of the human *ST3Gal III* gene from exon 2 to exon 14. We found a wide variety of deletions from 45 to 896 nucleotides, and insertions of 26 to 173 nucleotides (Figure 2, Table 4). The deletions were identifiable as partial or complete losses of one or several exons. The structural consequences of these alternative splices are illustrated for the potential proteins in Figure 3. The amino acid sequence of the putative protein corresponding to the A8 isotranscript is illustrated in Figure 4. Seven of the 19 isotranscripts (A1, A8, B1, B3, B8, C1, C8) code for proteins with identical N-terminals, transmembrane regions and C-terminal catalytic regions with the L- and S-sialylmotifs intact but with differences only in the stem regions. One of the isotranscripts (B1−90) codes for a protein similar to B1 but with a loss of 30 amino acids of exon 12 in the catalytic region. Another isotranscripts (B4) codes for an intact protein having specifically lost both exons 12 and 13 and thus the S-sialylmotif. Three of the isotranscripts (A7, B7, C7) code for proteins with the N-terminals and the transmembrane regions unaffected and with relatively intact stem regions but with no catalytic domains at all. Two isotranscripts (E1, E3) code for proteins having the N-terminal and transmembrane regions intact but show losses of not only the catalytic domain but also the typical stem region of the original enzyme.

Among the insertions we identified two different kinds. One is exemplified by the two sequences of 45 nt and 162 nt, between nt 118–119 and nt 254–255 in the coding sequence of the published *ST3Gal III* gene [17]. These insertions did not disrupt the open reading-frame (ORF) of the transcripts, followed the GT-AG rule for splicing of intron sequences and thus we considered these insertions as novel exons (E3 and E6) of this gene. The other kind of insertions, containing 26, 32 or 173 nt, all disrupted the ORF and resulted in truncated membrane proteins (Figure 3). The insertions of 26 and 173 nt, but not the insertion of 32 nt, followed the GT-AG rule for splicing of intron sequences. Among these potential proteins there were two transcripts, in which the ORF analysis suggested an alternative ORF (orf 2); B4+173 and E1. The orf 2 of B4+173 suggested a soluble protein, lacking the transmembrane part as well as the S-motif of the sialyltransferase, and thus probably lacking enzymatic activity. The orf 2 of the E1-transcript, on the other hand, coded for a soluble protein with conserved L- and S-sialylmotifs. However, this sequence did not follow the Kozak's consensus sequence for translation [29].

Nomenclature for *ST3Gal III* transcripts

As seen in Figure 2 there are some persistent features for the transcripts. All transcripts contain exons 2, 8 and 14 but most other exons are inconsistently expressed. (Exons 2 and 14 contain the sequences of our cloning primers). The 45 nt of exon 3, the 48 nt of exon 4 and finally the 43 nt of exon 5 made us classify the transcripts into five major isoforms (*ST3Gal III*-A to -E). *ST3Gal III*-A includes both exon 3 and exon 4. *ST3Gal III*-B lacks exon 3 but contains exon 4 and corresponds to the previously published sequence of *ST3Gal III* [17]. *ST3Gal III*-C lacks both exons 3 and 4. *ST3Gal III*-D includes exon 3 but lacks exon 4. *ST3Gal III*-E finally lacks all exons 3, 4 and 5. Within each isoform we identified a variety of transcripts, which forced us to propose a more detailed nomenclature for the *ST3Gal III* transcripts (Figure 2, Table 4). Thus, apart from the noted differences in exons 3, 4 and 5 the A1, B1, C1, and E1 transcripts all lack exon 6 but contains exon 2, 7, 8, 9, 10, 11, 12, 13 and 14. Similarly, besides the differences corresponding to the isoform features the A7, B7 and C7 all lack exon 9 to 13 and the A8, B8 and C8 transcripts all contain exon 6. Remaining *ST3Gal III* isotranscripts have so far only been isolated as one or two of the five main isoform (A–E).

LIF-CE resolves 17 of 19 *ST3Gal III* isotranscripts

All our isolated and sequenced clones were analyzed with LIF-CE to determine the elution times and software estimated sizes for each fluorescence labelled isotranscript. The results are shown in Table 5. The fragments, all covering the coding sequences, varied in size from 439 bp to 1335 bp which resulted in elution times between 65 and 84 min under the conditions used. The software generally overestimated the fragment size by a mean of 4.1% (min 1.7%, max 7.7%, SD 1.8%) depending on amplified fragments, on electrophoretic intervals as well as different fluorochrome labeling of amplified fragments and reference size markers. The elution times and thus

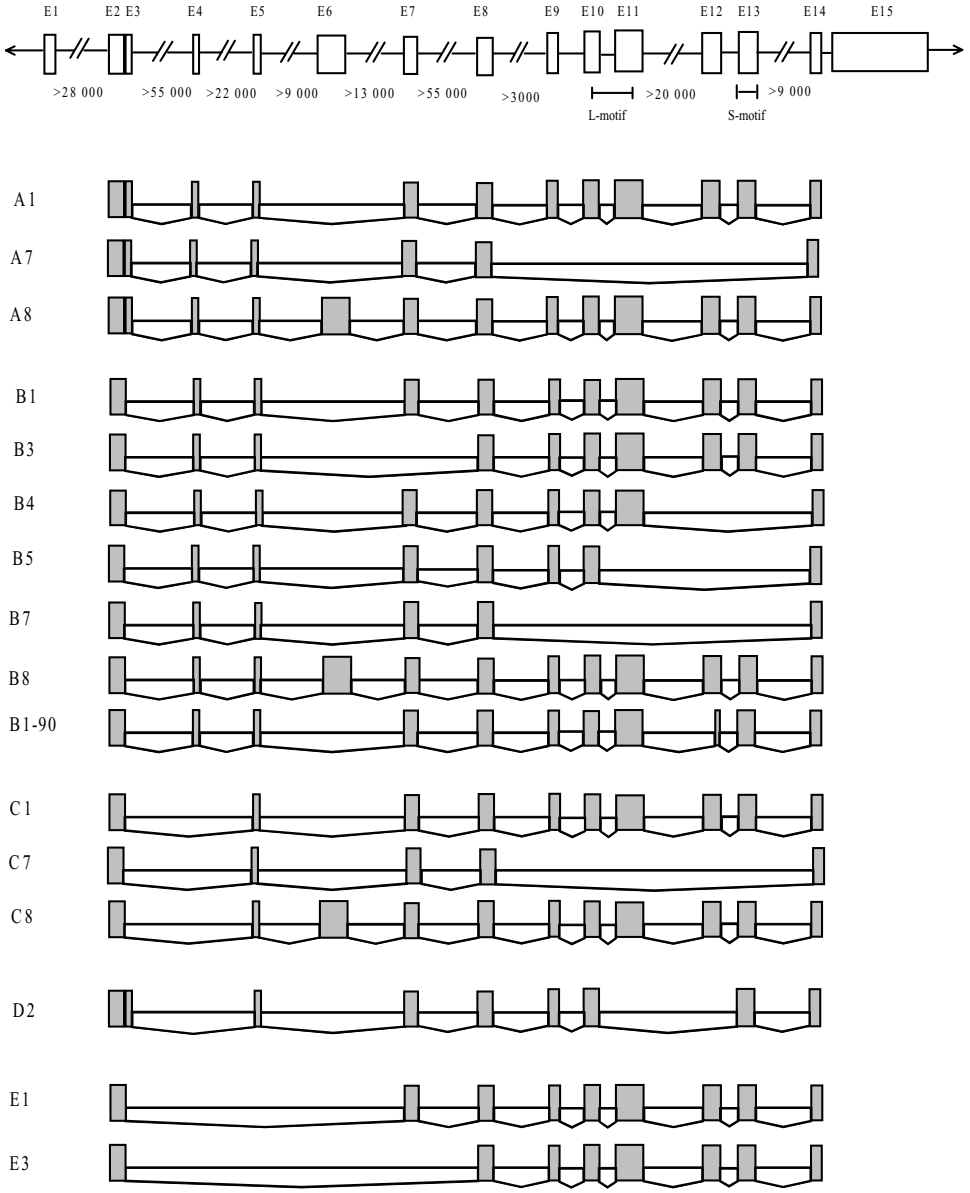

Figure 2. Illustration of 16 of 19 cloned *ST3Gal III* transcripts and their relation to the *ST3Gal III* gene. Coding sequences are shown as grey boxes and L- and S-motifs are shown as bars below the gene sequence. Transcripts with insertions of intron nucleotides (26, 32 or 173 nt long) have been omitted. Exact lengths of introns and exons are given in Tables 2 and 3.

estimated sizes were however very stable for each fragment with a total variation (CV%) of 0.02% [24]. Figure 5 illustrates the electrophoretic mobilities of fluorescence double-labelled fragments of all the 19 isotranscripts cloned and covers all fragments and reference markers in one run. Most isotranscripts are well resolved from each other but in some cases fragments of different isotranscripts are partially or completely overlapping. Thus, D2+26 and B5+26 and B1−90 and B3, which differ by only 3 bp respectively, do not separate under the conditions used. Other fragments, such as B1−90 and E1 which differ only by 1 bp, show a tendency to separate illustrating that not only size but also nucleotide composition and sequence and thus conformation is of importance for separation under these non-denaturating conditions.

ST3Gal III isotranscripts show a tissue specific expression

The different human tissues analysed by LIF-CE showed a noticeable difference in *ST3Gal III* expression. Not only did the pattern of isotranscripts vary between tissues but also the relative amounts described as percentages of GAPDH expression. Only 7 of the 19 splice variants encode for proteins with the catalytic domains completely intact and probably represent active enzymes. Only 2 of those (B1 and C1) were detected in

Table 4. *ST3Gal III* isotranscript names, splice losses, insertions and potential protein lengths

Isoform	*Isotranscript*	*Deletions*		*Insertions lengths (nt)*[c]	*cDNA fragment lengths (nt)*	*Potential protein lengths (aa)*
		Identities of deletions	*lengths (nt)*			
A	A1	E6	162		1173	390
	A7	E6, E9, E10, E11, E12, E13,	803		532	170
	A8[a]	None			1335	444
B	B1	E3, E6	207		1128	375
	B1−90	E3, E6 and 90 nt of E12	297		1038	345
	B1+32	E3, E6	207	32	1160	189
	B3	E3, E6, E7	300		1035	344
	B4	E3, E6, E12, E13	501		834	277
	B4+173	E3, E6, E12, E13	501	173	1007	115
	B5[b]	E3, E6, E11, E12, E13	688		647	202
	B5+26	E3, E6, E11, E12, E13	688	26	673	186
	B5+173	E3, E6, E11, E12, E13	688	173	821	121
	B7	E3, E6, E9, E10, E11, E12, E13	848		487	155
	B8	E3	45		1290	429
C	C1	E3, E4, E6	255		1080	359
	C7	E3, E4, E6, E9, E10, E11, E12,E13	896		439	139
	C8	E3, E4	93		1242	413
D	D2[b]	E4, E6, E11, E12	691		644	201
	D2+26	E4, E6, E11, E12	691	26	670	185
E	E1	E3, E4, E5, E6	298		1037	43
	E3[b]	E3, E4, E5, E6, E7	391		944	74
	E3+32	E3, E4, E5, E6, E7	391	32	976	74

[a]Transcript A8 is considered the full length cDNA against which all transcripts are compared.
[b]Transcripts B5, D2 and E3 were not cloned in this study but are included for comparison.
[c] Insertion of 26 nt corresponds to an intron sequence (nt 261-286) on the human genomic clone AC004950, insertion of 32 nt to an intron sequence (nt 190556-190587) on the human genomic clone AC005038, and insertion of 173 nt corresponds to an intron sequence (nt 20058-20230) on the human genomic clone AC004950.

most of the tissues analysed. The other 5 variants with the catalytic domain intact (A1, A8, B3, B8, C1 and C8) were poorly expressed or not detectable with the RT-PCR and LIF-CE technique. In fact, 8 of the 19 isotranscripts cloned from PBL were not detected at all in any of the other tissues. Two splice variants (B7 and C7), which encode for proteins completely lacking the catalytic domain but containing the N-terminal, the transmembrane region and the putative stem region of the intact enzyme, were strongly expressed in neural tissues. These truncated proteins may have some function but are highly unlikely to express enzymatic activities.

Thus, *ST3Gal III* B1 and C1 isoforms dominated quantitatively in most tissues. The highest relative expressions of B1 were found in uterus (80%), skeletal muscle (72%) and adrenal glands (69%) and the lowest in kidney and colon where no *ST3Gal III* isoforms were detected at all (Table 6). Tissues related to muscles showed a general pattern with a dominance of B1 isoform and expression of B4 and B1+32 isotranscripts but no C7 and B7 isotranscripts (Figure 6). The B4 isotranscript showed a very tissue specific expression with the highest amounts in skeletal muscle (11%), uterus (8%) and adult brain (8%) and no expression at all in 14 other tissues. B1+32, which codes for a truncated protein, was expressed in 13 tissues (Table 6). The C1 form had its highest expression in uterus (15%), skeletal muscle (13%) and foetal brain (12%). Foetal brain showed the most complex pattern of expression with 8 isotranscripts identified and a dominance of the C7 and B7 isoforms (38% and 33% respectively) (Table 6). A7 isotranscript was detected only in foetal brain. In fetal brain 5 additional transcripts were detected around 650 bp (Figure 6), which we now have cloned and found to be *ST3Gal III* isotranscripts (data not shown). Adult brain and spinal cord expressed a pattern similar to foetal brain with high amounts of B7 and C7 isotranscripts. The isotranscripts B4+173, B5+173, B1−90/B3, E1, C8, A8 or E3+32 were below detection level in all tissues under the standard conditions used. B8 isotranscript was detected only in skeletal muscle, comprising less than 1% of GAPDH expression.

Discussion

In order to characterise all the candidate α2,3-sialyltransferases involved in the biosynthesis of SLe^x, the minimal structural requirement for E-, P- and L-selectin ligands, we have now cloned and sequenced the transcripts covering the coding region of human *ST3Gal III* gene from PBL of one single individual. This

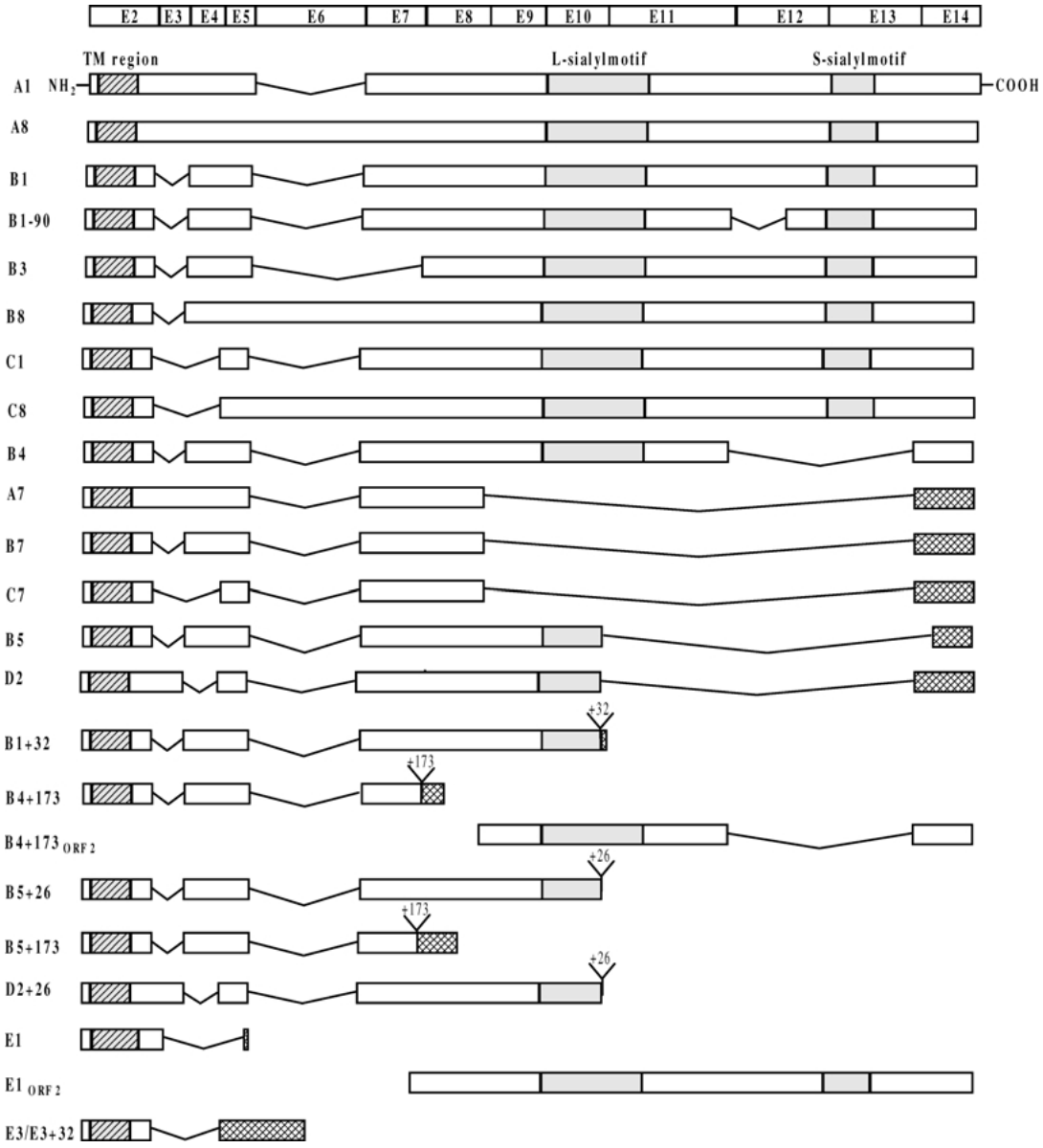

Figure 3. Schematic illustration of 21 potential proteins obtained from 19 cloned *ST3Gal III* transcripts after ORF analysis. The N-terminal is marked NH_2- , the transmembrane region (TM region) is marked as a striped grey bar, the L- and S- sialylmotifs are marked as grey bars and the C-terminal as ─COOH. Striped white bars indicate altered amino acid sequences. Insertions of intron sequences are indicated at Y signs at the relevant positions.

sialyltransferase is specifically responsible for the sialylation of the type 1 precursor, necessary for the final biosynthesis of SLe^a, an epitope known to interact with E-selectin and to be of importance in the diagnosis and possibly future treatment of gastrointestinal carcinomas in man [6,7]. We have earlier published data for *ST3Gal IV* transcripts in human PBL [24].

Within our *ST3Gal III* clones we identified 19 different transcripts with deletions as well as insertions. To understand the relationship between the different transcripts, we aligned the *ST3Gal III* transcripts with sequences found in the human genome database. As a result of this alignment, we identified the exon-intron junction sequences of published data [17]. From this we could conclude that all of our transcripts, except the two including the insertion of 32 nt, resulted from alternative splices according to the GT-AG consensus rule for splicing [28,30]. Among our clones we found two novel insertions of 45 nt and 162 nt, which did not disrupt the reading frame for translation. Judged from the positions of these sequences in the human genomic clones we concluded them to be exon 3 and exon 6, respectively. Additionally, we found transcripts with a deletion of 48 bp corresponding to exon 4 and to nucleotides 119 to 166 of the published sequence and nucleotides 164–211 in our

```
 M   G   L   L   V   F   V   R   N   L   L   L   A   L   C   L   F   L     18
ATG GGA CTC TTG GTA TTT GTG CGC AAT CTG CTG CTA GCC CTC TGC CTC TTT CTG    54

 V   L   G   F   L   Y   Y   S   A   W   K   L   H   L   L   Q   W   E     36
GTA CTG GGA TTT TTG TAT TAT TCT GCG TGG AAG CTA CAC TTA CTC CAG TGG GAG   108

 E   D   S   S   K   Y   S   H   S   S   S   P   Q   E   K   P   V   A     54
GAG GAC TCC AGT AAG TAT AGT CAC TCT AGC TCA CCC CAG GAG AAG CCT GTT GCA   162

 D   S   V   V   L   S   F   D   S   A   G   Q   T   L   G   S   E   Y     72
GAT TCA GTG GTT CTT TCC TTT GAC TCC GCT GGA CAA ACA CTA GGC TCA GAG TAT   216

 D   R   L   G   F   L   L   N   L   D   S   K   L   S   P   R   T   L     90
GAT CGG TTG GGC TTC CTC CTG AAT CTG GAC TCT AAA CTC TCA CCG AGG ACT CTC   270

 C   T   V   V   F   G   L   D   C   I   L   E   S   P   G   E   P   K    108
TGC ACG GTG GTT TTT GGC CTT GAC TGC ATA TTG GAA TCA CCT GGA GAG CCT AAA   324

 K   L   L   M   P   A   S   H   P   L   E   I   L   K   S   L   S   E    126
AAA TTA CTG ATG CCT GCA TCC CAC CCT CTA GAG ATT TTG AAG TCA CTG AGC GAG   378

 D   T   A   F   A   L   G   F   L   K   L   P   R   P   A   E   L   A    144
GAC ACA GCC TTT GCA TTA GGA TTT TTA AAG CTG CCC AGG CCT GCT GAA TTA GCC   432

 T   K   Y   A   N   F   S   E   G   A   C   K   P   G   Y   A   S   A    162
ACC AAG TAC GCA AAC TTT TCA GAG GGA GCT TGC AAG CCT GGC TAT GCT TCA GCC   486

 L   M   T   A   I   F   P   R   F   S   K   P   A   P   M   F   L   D    180
TTG ATG ACG GCC ATC TTC CCC CGG TTC TCC AAG CCA GCA CCC ATG TTC CTG GAT   540

 D   S   F   R   K   W   A   R   I   R   E   F   V   P   P   F   G   I    198
GAC TCC TTT CGC AAG TGG GCT AGA ATC CGG GAG TTC GTG CCG CCT TTT GGG ATC   594

 K   G   Q   D   N   L   I   K   A   I   L   S   V   T   K   E   Y   R    216
AAA GGT CAA GAC AAT CTG ATC AAA GCC ATC TTG TCA GTC ACC AAA GAG TAC CGC   648

 L   T   P   A   L   D   S   L   R   C   R   R   C   I   I   V   G   N    234
CTG ACC CCT GCC TTG GAC AGC CTC CGC TGC CGC CGC TGC ATC ATC GTG GGC AAT   702

 G   G   V   L   A   N   K   S   L   G   S   R   I   D   D   Y   D   I    252
GGA GGC GTT CTT GCC AAC AAG TCT CTG GGG TCA CGA ATT GAC GAC TAT GAC ATT   756

 V   V   R   L   N   S   A   P   V   K   G   F   E   K   D   V   G   S    270
GTG GTG AGA CTG AAT TCA GCA CCA GTG AAA GGC TTT GAG AAG GAC GTG GGC AGC   810

 K   T   T   L   R   I   T   Y   P   E   G   A   M   Q   R   P   E   Q    288
AAA ACG ACA CTG CGC ATC ACC TAC CCC GAG GGC GCC ATG CAG CGG CCT GAG CAG   864

 Y   E   R   D   S   L   F   V   L   A   G   F   K   W   Q   D   F   K    306
TAC GAG CGC GAT TCT CTC TTT GTC CTC GCC GGC TTC AAG TGG CAG GAC TTT AAG   918

 W   L   K   Y   I   V   Y   K   E   R   V   S   A   S   D   G   F   W    324
TGG TTG AAA TAC ATC GTC TAC AAG GAG AGA GTG AGT GCA TCG GAT GGC TTC TGG   972

 K   S   V   A   T   R   V   P   K   E   P   P   E   I   R   I   L   N    342
AAA TCT GTG GCC ACT CGA GTG CCC AAG GAG CCC CCT GAG ATT CGA ATC CTC AAC  1026

 P   Y   F   I   Q   E   A   A   F   T   L   I   G   L   P   F   N   N    360
CCA TAT TTC ATC CAG GAG GCC GCC TTC ACC CTC ATT GGC CTG CCC TTC AAC AAT  1080

 G   L   M   G   R   G   N   I   P   T   L   G   S   V   A   V   T   M    378
GGC CTC ATG GGC CGG GGG AAC ATC CCT ACC CTT GGC AGT GTG GCA GTG ACC ATG  1134

 A   L   H   G   C   D   E   V   A   V   A   G   F   G   Y   D   M   S    396
GCA CTA CAC GGC TGT GAC GAG GTG GCA GTC GCA GGA TTT GGC TAT GAC ATG AGC  1188

 T   P   N   A   P   L   H   Y   Y   E   T   V   R   M   A   A   I   K    414
ACA CCC AAC GCA CCC CTG CAC TAC TAT GAG ACC GTT CGC ATG GCA GCC ATC AAA  1242

 E   S   W   T   H   N   I   Q   R   E   K   E   F   L   R   K   L   V    432
GAG TCC TGG ACG CAC AAT ATC CAG CGA GAG AAA GAG TTT CTG CGG AAG CTG GTG  1296

 K   A   R   V   I   T   D   L   S   S   G   I   *                       445
AAA GCT CGC GTC ATC ACT GAT CTA AGC AGT GGC ATC TGA                     1335
```

Figure 4. Amino acid of the putative protein corresponding to the nucleotide sequence of the *ST3Gal III* A8 isotranscript.

revised sequence. Burger et al. used a transcript with a similar deletion obtained from Jurkat cells, when generating an antibody against the human *ST3Gal III* enzyme. Although these authors did not describe the 48 nt sequence in detail it was probably the same deletion as we have seen in human leukocytes, *ST3Gal III* C1 [31].

Interestingly, only 7 of the 19 splice variants, that we have identified, code for proteins, which probably represent active enzymes. These proteins have the catalytic domains completely intact including the L-, S- and VS-motifs and differ only in the putative stem region. Of those 7 variants only 2 (B1 and C1) were highly expressed in most tissues, the other 5 being

Table 5. Capillary electrophoretic data of fluorescence labelled amplified fragments from 19 clones of *ST3Gal III* isotranscripts

Isotranscript[a]	*Fragment size (bp)*	*Estimated size (bp)*[b]	*Elution time (min)*[b]
C7	439	472	64.54
B7	487	516	65.96
A7	532	573	67.62
D2+26	670	702	71.38
B5+26	673	702	71.38
B5+173	821	844	74.96
B4	834	853	75.15
E3+32	976	993	77.75
B4+173	1007	1040	78.52
B3	1035	1062	78.85
B1−90	1038	1063	78.87
E1	1037	1076	79.07
C1	1080	1119	79.70
B1	1128	1162	80.30
B1+32	1160	1187	80.64
A1	1173	1209	80.93
C8	1242	1316	82.24
B8	1290	1359	82.73
A8	1335	1413	83.30

[a]Ordered after elution times.
[b]Refers to elution times of Figure 4.

scarcely expressed or not detectable using the RT-PCR LIF-CE technique. The remaining splice variants are unlikely to express enzymatic activities but may have other, as yet unknown, biological functions.

The common differences in the *ST3Gal III* A, B, C and D isotranscripts, which we report here, correspond to 15 amino acids

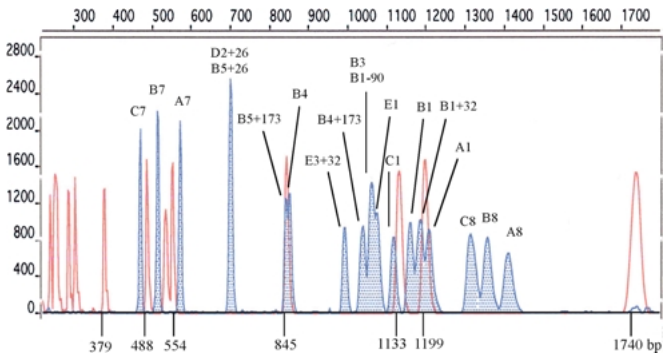

Figure 5. A typical electrophoregram of a mixture of labelled fragments obtained from the 19 *ST3Gal III* transcript clones together with appropriate size markers analyzed by LIF-CE. Identities of amplified fragments are given on top of each peak and size markers are given as bp lengths below reference peaks. Sizes and elution times of labelled fragments are given in Table 5 and electrophoretic conditions are described in the methods section.

(exon 3) and 16 amino acids (exon 4) which are found within the putative stem region of the α2,3-sialyltransferase. This part of the enzyme is considered of relevance for the localisation and retention within the Golgi system [8,32]. A recent report has confirmed that the cytoplasmic, transmembrane and stem regions of glycosyltransferases also specifies their functional localisation and stability within the Golgi [33]. De Vries et al. showed that recombinant Fuc-TIII α1,3/1,4-fucosyltransferase changes its acceptor specificity toward glycoproteins when truncated in the transmembrane and stem regions [34]. The differences in the stem region of enzymes coded for by *ST3Gal III* isoforms A, B and C reported here may contribute to the same phenomena.

Thus, the 7 enzymatic isoforms, which would result from translation of our cloned *ST3Gal III* isotranscripts (A1, A8, B1, B3, B8, C1 and C8), with identical and intact N-terminals, transmembrane regions and catalytic domains and which differ only in the stem regions might be located differently and have different functionality within the Golgi. The B1−90 isotranscript is unique in its loss of 90 nucleotides resulting in a potential protein with 30 amino acids less in the catalytic region. Whether this protein is enzymatically active remains to be shown.

Translated to protein isotranscript B4 would result in a peptide that specifically lacks the S-sialylmotif. It has been shown that L-motifs participates in the binding of the donor substrate CMP-NeuAc and the S-motif participates in binding of both the donor and acceptor substrates [9–11]. Datta et al. recently showed that the conserved cysteine residues of L- and S- motif participates in disulfide linkage between the both motifs and is important for proper conformation and activity of *ST6Gal I*. They also showed that the full-length mutant proteins, with substitutions of the conserved cysteines to alanine or serine in the L- and S-motifs, were retained in the endoplasmic reticulum, probably due to improper folding of the proteins [12]. Whether the *ST3Gal III* B4 protein, lacking the S-sialylmotif, is found in Golgi or is retained and degraded in the endoplasmic reticulum remains to be shown.

Three of our *ST3Gal III* isotranscripts (A7, B7, C7) would code for proteins that have lost their catalytic domains but are otherwise intact. The *ST3Gal III* E isotranscripts have lost exon 5, which would lead to an almost immediate truncation of the protein chains after the transmembrane part but prior to the stem region. Whether these proteins will be produced, correctly folded and leave the endoplasmic reticulum and finally localise to the Golgi and if so what functions they have there is a completely open question. SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptors) proteins have been identified as essential constituents of the Golgi system [35] and it would, at least in theory, be economical for the cell to use alternative splicing mechanisms to produce from one single gene proteins with quite different but maybe cooperative functions within this complex system of membrane proteins.

The five isotranscripts we identified with insertions of 26, 32 or 173 nt would, if translated, lead to production of truncated membrane proteins due to a shift in the ORF. We

Table 6. Expression of *ST3Gal III* isotranscript relative to GAPDH (100%) in twenty-one human tissues

Isotranscript	*Foetal brain*	*Brain*	*Spinal cord*	*Uterus*	*Skeletal muscle*	*Heart*	*Adrenal gland*	*PBL*	*Spleen*	*Testis*	*Placenta*
A1	–	–	–	3	–	–	1	–	–	–	–
B1	20	21	12	80	72	22	69	52	38	35	30
C1	12	3	2	15	13	5	10	8	3	4	6
E1	–	–	–	–	–	–	–	–	–	–	–
B1+32	–	2	5	9	3	1	10	5	3	1	2
B3/B1−90	–	–	–	–	–	–	–	–	–	–	–
B4	4	8	–	8	11	2	2	–	–	–	–
B4+173	–	–	–	–	–	–	–	–	–	–	–
B5+26/D2+26	6	3	–	4	1	–	–	–	–	–	–
B5+173	–	–	–	–	–	–	–	–	–	–	–
E3+32	–	–	–	–	–	–	–	–	–	–	–
A7	1	–	–	–	–	–	–	–	–	–	–
B7	33	17	9	10	–	–	1	–	–	–	–
C7	38	10	4	–	–	–	–	–	–	–	–
A8	–	–	–	–	–	–	–	–	–	–	–
B8	–	–	–	–	<1	–	–	–	–	–	–
C8	–	–	–	–	–	–	–	–	–	–	–

Isotranscript	*Foetal liver*	*Thymus*	*Salivary gland*	*Prostate*	*Trachea*	*Liver*	*Lung*	*Bone marrow*	*Kidney*	*Colon*
A1	–	–	–	–	–	–	–	–	–	–
B1	32	18	23	11	6	11	7	6	–	–
C1	5	2	–	–	1	–	–	–	–	–
E1	–	–	–	–	–	–	–	–	–	–
B1+32	2	4	3	–	–	–	–	–	–	–
B3/B1−90	–	–	–	–	–	–	–	–	–	–
B4	–	–	–	1	–	–	–	–	–	–
B4+173	–	–	–	–	–	–	–	–	–	–
B5+26/D2+26	–	–	–	–	–	–	–	–	–	–
B5+173	–	–	–	–	–	–	–	–	–	–
E3+32	–	–	–	–	–	–	–	–	–	–
A7	–	–	–	–	–	–	–	–	–	–
B7	–	–	–	–	–	–	–	–	–	–
C7	–	–	–	–	–	–	–	–	–	–
A8	–	–	–	–	–	–	–	–	–	–
B8	–	–	–	–	–	–	–	–	–	–
C8	–	–	–	–	–	–	–	–	–	–

believe these transcripts were obtained from pre mRNA since our preparative RNA technique would not exclude such structures and thus they would not become translated to proteins. Some of the isotranscripts with such insertions showed a very tissue specific expression, which might be a reflection of slower turnover of the gene products in those tissues.

The occurrence of alternative splicing of sialyltransferases has been reported before both for α2,3-sialyltransferases as well as for α2,6-sialyltransferases [14–16,36–38] although not specifically for the *ST3Gal III* gene. *ST3Gal IV* has been shown to have alternatively spliced forms, both in the 5′-UTR (A1, A2, B1, B2, B3 and BX) [14,15,36] and within the coding region [24]. In 1999, Kapitonov reported three alternative spliced forms of *ST3Gal V*; AFS 600, AFS 400 and AFS 800 [39]. AFS 600 was the only one that showed any enzymatic activity and was also homologous to the published *ST3Gal V* sequence [40]. When Okajima et al. cloned human *ST3Gal VI*, they reported a smear of bands on the Northern blot, which they presumed to be alternatively spliced forms of *ST3Gal VI* [16]. Recently a type 1 and type 2 isoform of *ST3Gal VI* has been reported, which differ in the 5′-UTR of the transcripts [38]. In fact, we have in our laboratory cloned and sequenced from human PBL 6 different isotranscripts from *ST3Gal VI* which differ within the coding region by deletions of 64, 325, 421 and 574 nucleotides. However, according to the ORF analysis, all the potential proteins seem to be non-functional enzymes, lacking both L- and S-sialylmotifs (data not shown). We also recently identified two completely homologous ESTs from the GenBank

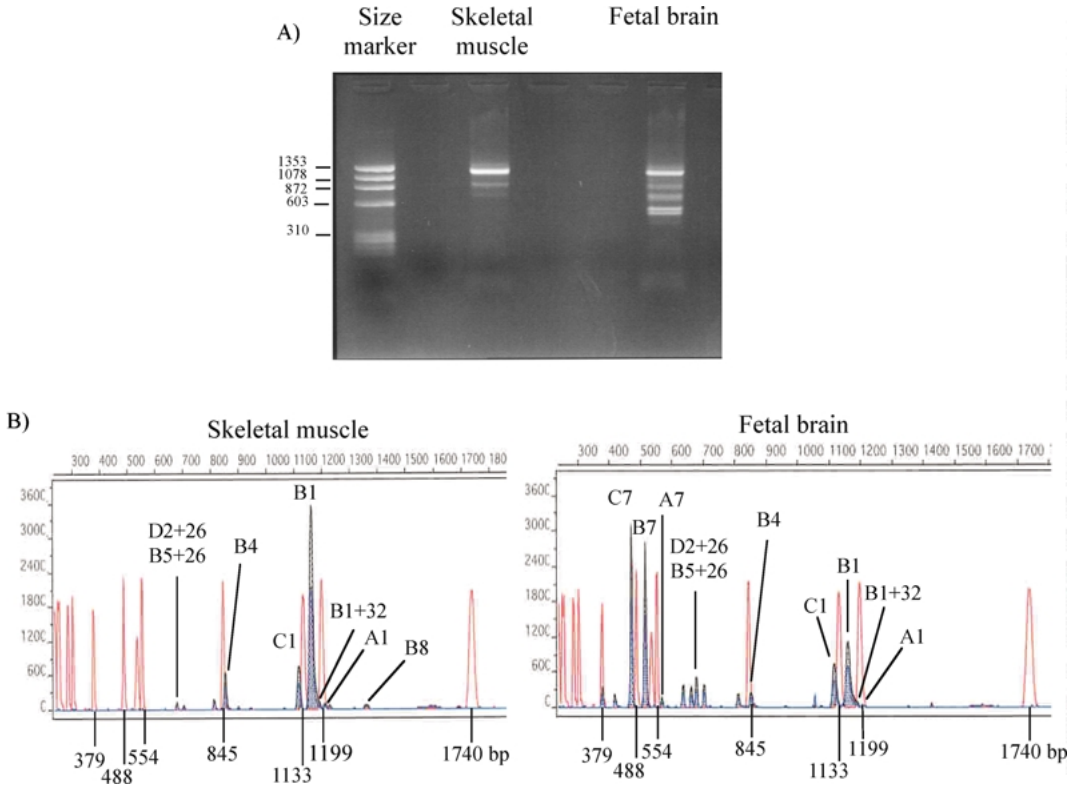

Figure 6. Comparisons of electrophoretic separations of amplified *ST3Gal III* transcripts obtained from skeletal muscle and foetal brain using commercial size markers on A). Agarose slab-gel electrophoresis and B) LIF-CE . Dominating transcripts are indicated on top of each peak in B. We recently identified the weak additional peaks between the markers at 554 and 845 bp as *ST3Gal III* isotranscripts.

database corresponding to yet another *ST3Gal VI* transcript (Acc. No. XM_002931, AF119391). The ORF analysis of those sequences revealed an alternative start codon 159 nucleotides up-streams to the published sequence, which would result in a 52 amino acids larger protein. The ORF analysis of the *ST3Gal III* transcripts described in this paper showed that translated to proteins, the isotranscripts may code for both membrane bound and soluble sialyltransferases as well as for membrane bound proteins lacking one or both of the two conserved L- and S-sialylmotifs (Figure 3).

In order to investigate if the 19 isotranscripts we had cloned were specific for human PBL, we used RT-PCR and LIF-CE and examined a panel of human tissues. Northern blot analyses have earlier been used to show that the human *ST3Gal III* gene has a tissue specific expression with a dominating expression in skeletal muscle but also expression in several other human tissues [23]. Northern blot gives rough qualitative information on mRNA sizes, gives relative abundance of transcript levels and gives information on alternatively spliced variants from one single gene. However, the technique exhibit relatively low resolution and sensitivity ($\sim 10^6$ copies of mRNAs). As illustrated in Figure 6, the relatively small size differences in our transcript sizes do not resolve well on agarose gel. The RT-PCR LIF-CE technique, as we used it, gives the possibility to separate and identify 17 of the 19 isotranscripts of the *ST3Gal III* gene (Figure 5) as well as to estimate the relative expression of these isotranscripts in many tissues in one run (Table 6). The method also makes it possible to detect even very low amounts of transcripts (in theory only one copy). However, using the Northern blot technique and commercial human tissues RNA blots Kitagawa and Paulson [23] detected expression of *ST3Gal III* in human colon not detected in our study.

The expression study of the 19 *ST3Gal III* isotranscripts revealed a highly tissue-specific pattern. *ST3Gal III* B1, the dominating isotranscript, was found in all tissues except kidney and colon where we didn't detect any *ST3Gal III* at all. The highest expression and most complex patterns of isotranscripts (Table 6) were found in muscle tissues, in tissues related to the nervous system and in adrenal glands. B7 and C7 isotranscripts were almost exclusively expressed in nervous tissues, uterus being the only exception. Notable differences in the expression patterns of *ST3Gal III* isotranscripts in foetal brain versus adult brain were the relative decreases of C1, C7 and B7 isotranscripts with B1 remaining essentially at the same level. However, these expression data has to be verified in future studies including tissues from many more selected individuals in order to draw more specific conclusions.

Our data has revealed an unknown complexity and expression profile of the *ST3Gal III* gene in human tissues. The complexity, which mainly results from alternative splicing mechanisms within the coding region of the gene, adds to the complexity that may involve also the 5′- and 3′-UTR of this gene. The biological significance and localization of *ST3Gal III* enzyme/protein isoforms suggested from these data will be the subject of our future studies.

Acknowledgments

This work was supported by grants from the Swedish Research Council (8266), from Governmental grants to Sahlgrenska University Hospital and from the programme "Glycoconjugates in Biological Systems" sponsored by the Swedish Foundation for Strategic Research. (A.G.).

References

1 Phillips ML, Nudelman E, Gaeta FC, Perez M, Singhal AK, Hakomori S, Paulson JC, ELAM-1 mediates cell adhesion by recognition of a carbohydrate ligand, sialyl-Lex, *Science* **250**, 1130–2 (1990).

2 Tyrrell D, James P, Rao N, Foxall C, Abbas S, Dasgupta F, Nashed M, Hasegawa A, Kiso M, Asa D, et al., Structural requirements for the carbohydrate ligand of E-selectin, *Proc Natl Acad Sci USA* **88**, 10372–6 (1991).

3 Polley MJ, Phillips ML, Wayner E, Nudelman E, Singhal AK, Hakomori S, Paulson JC, CD62 and endothelial cell-leukocyte adhesion molecule 1 (ELAM-1) recognize the same carbohydrate ligand, sialyl-Lewis x, *Proc Natl Acad Sci USA* **88**, 6224–8 (1991).

4 Berg EL, Robinson MK, Mansson O, Butcher EC, Magnani JL, A carbohydrate domain common to both sialyl Le(a) and sialyl Le(X) is recognized by the endothelial cell leukocyte adhesion molecule ELAM-1, *J Biol Chem* **266**, 14869–72 (1991).

5 Iwai K, Ishikura H, Kaji M, Sugiura H, Ishizu A, Takahashi C, Kato H, Tanabe T, Yoshiki T, Importance of E-selectin (ELAM-1) and sialyl Lewis(a) in the adhesion of pancreatic carcinoma cells to activated endothelium, *Int J Cancer* **54**, 972–7 (1993).

6 Matsumoto S, Imaeda Y, Umemoto S, Kobayashi K, Suzuki H, Okamoto T, Cimetidine increases survival of colorectal cancer patients with high levels of sialyl Lewis-X and sialyl Lewis-A epitope expression on tumour cells, *Br J Cancer* **86**, 161–7 (2002).

7 Eaton D, Hawkins RE, Cimetidine in colorectal cancer–are the effects immunological or adhesion-mediated? *Br J Cancer* **86**, 159–60 (2002).

8 Paulson JC, Colley KJ, Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation, *J Biol Chem* **264**, 17615–8 (1989).

9 Datta AK, Paulson JC, The sialyltransferase "sialylmotif" participates in binding the donor substrate CMP-NeuAc, *J Biol Chem* **270**, 1497–500 (1995).

10 Datta AK, Paulson JC, Sialylmotifs of sialyltransferases, *Indian J Biochem Biophys* **34**, 157–65 (1997).

11 Datta AK, Sinha A, Paulson JC, Mutation of the sialyltransferase S-sialylmotif alters the kinetics of the donor and acceptor substrates, *J Biol Chem* **273**, 9608–14 (1998).

12 Datta AK, Chammas R, Paulson JC, Conserved cysteines in the sialyltransferase sialylmotifs form an essential disulfide bond, *J Biol Chem* **276**, 15200–7 (2001).

13 Breton C, Imberty A, Structure/function studies of glycosyltransferases, *Curr Opin Struct Biol* **9**, 563–71 (1999).

14 Kitagawa H, Mattei MG, Paulson JC, Genomic organization and chromosomal mapping of the Gal beta 1,3GalNAc/Gal beta 1,4GlcNAc alpha 2,3-sialyltransferase, *J Biol Chem* **271**, 931–8 (1996).

15 Kitagawa H, Paulson JC, Cloning of a novel alpha 2,3-sialyltransferase that sialylates glycoprotein and glycolipid carbohydrate groups, *J Biol Chem* **269**, 1394–401 (1994a).

16 Okajima T, Fukumoto S, Miyazaki H, Ishida H, Kiso M, Furukawa K, Urano T, Molecular cloning of a novel alpha2,3-sialyltransferase (ST3Gal VI) that sialylates type II lactosamine structures on glycoproteins and glycolipids, *J Biol Chem* **274**, 11479–86 (1999).

17 Kitagawa H, Paulson JC, Cloning and expression of human Gal beta 1,3(4)GlcNAc alpha 2,3-sialyltransferase, *Biochem Biophys Res Commun* **194**, 375–82 (1993).

18 Hansson GC, Zopf D, Biosynthesis of the cancer-associated sialyl-Lea antigen, *J Biol Chem* **260**, 9388–92 (1985).

19 Zopf D, Hansson GC, The chemical basis for expression of the sialyl-Le(a) antigen, *Adv Exp Med Biol* **228**, 657–76 (1988).

20 Bengtson P, Larson C, Lundblad A, Larson G, Pahlsson P, Identification of a missense mutation (G329A;Arg(110) --> GLN) in the human FUT7 gene, *J Biol Chem* **276**, 31575–82 (2001).

21 Wagers AJ, Stoolman LM, Kannagi R, Craig R, Kansas GS, Expression of leukocyte fucosyltransferases regulates binding to E-selectin: Relationship to previously implicated carbohydrate epitopes, *J Immunol* **159**, 1917–29 (1997).

22 Toivonen S, Nishihara S, Narimatsu H, Renkonen O, Renkonen R, Fuc-TIX: A versatile alpha1,3-fucosyltransferase with a distinct acceptor- and site-specificity profile, *Glycobiology* **12**, 361–8 (2002).

23 Kitagawa H, Paulson JC, Differential expression of five sialyltransferase genes in human tissues, *J Biol Chem* **269**, 17872–8 (1994b).

24 Grahn A, Larson G, Identification of nine alternatively spliced alpha 2,3-sialyltransferase, ST3Gal IV, transcripts and analysis of their expression by RT-PCR and laser-induced fluorescent capillary electrophoresis (LIF-CE) in twenty-one human tissues, *Glycoconj J* **18**, 759–67 (2001).

25 Modrek B, Resch A, Grasso C, Lee C, Genome-wide detection of alternative splicing in expressed sequences of human genes, *Nucleic Acids Res* **29**, 2850–9 (2001).

26 Brett D, Hanke J, Lehmann G, Haase S, Delbruck S, Krueger S, Reich J, Bork P, EST comparison indicates 38% of human mRNAs contain possible alternative splice forms, *FEBS Lett* **474**, 83–6 (2000).

27 Breathnach R, Chambon P, Organization and expression of eucaryotic split genes coding for proteins, *Annu Rev Biochem* **50**, 349–83 (1981).

28 Reed R, Initial splice-site recognition and pairing during pre-mRNA splicing, *Curr Opin Genet Dev* **6**, 215–20 (1996).

29 Kozak M, At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells, *J Mol Biol* **196**, 947–50 (1987).

30 Smith CW, Valcarcel J, Alternative pre-mRNA splicing: The logic of combinatorial control, *Trends Biochem Sci* **25**, 381–8 (2000).
31 Burger PC, Lotscher M, Streiff M, Kleene R, Kaissling B, Berger EG, Immunocytochemical localization of alpha2,3(N)-sialyltransferase (*ST3Gal III*) in cell lines and rat kidney tissue sections: Evidence for golgi and post-golgi localization, *Glycobiology* **8**, 245–57 (1998).
32 Colley KJ, Lee EU, Adler B, Browne JK, Paulson JC, Conversion of a Golgi apparatus sialyltransferase to a secretory protein by replacement of the NH2-terminal signal anchor with a signal peptide, *J Biol Chem* **264**, 17619–22 (1989).
33 Grabenhorst E, Conradt HS, The cytoplasmic, transmembrane, and stem regions of glycosyltransferases specify their in vivo functional sublocalization and stability in the Golgi, *J Biol Chem* **274**, 36107–16 (1999).
34 de Vries T, Srnka CA, Palcic MM, Swiedler SJ, van den Eijnden DH, Macher BA, Acceptor specificity of different length constructs of human recombinant alpha 1,3/4-fucosyltransferases. Replacement of the stem region and the transmembrane domain of fucosyltransferase V by protein A results in an enzyme with GDP-fucose hydrolyzing activity, *J Biol Chem* **270**, 8712–22 (1995).
35 Nichols BJ, Pelham HR, SNAREs and membrane fusion in the Golgi apparatus, *Biochim Biophys Acta* **1404**, 9–31 (1998).
36 Taniguchi A, Matsumoto K, Down-regulation of human sialyltransferase gene expression during in vitro human keratinocyte cell line differentiation, *Biochem Biophys Res Commun* **243**, 177–83 (1998).
37 Wen DX, Svensson EC, Paulson JC, Tissue-specific alternative splicing of the beta-galactoside alpha 2,6-sialyltransferase gene, *J Biol Chem* **267**, 2512–8 (1992).
38 Taniguchi A, Kaneta R, Morishita K, Matsumoto K, Gene structure and transcriptional regulation of human Gal beta1,4(3) GlcNAc alpha2,3-sialyltransferase VI (hST3Gal VI) gene in prostate cancer cell line, *Biochem Biophys Res Commun* **287**, 1148–56 (2001).
39 Kapitonov D, Bieberich E, Yu RK, Combinatorial PCR approach to homology-based cloning: Cloning and expression of mouse and human GM3-synthase, *Glycoconj J* **16**, 337–50 (1999).
40 Ishii A, Ohta M, Watanabe Y, Matsuda K, Ishiyama K, Sakoe K, Nakamura M, Inokuchi J, Sanai Y, Saito M, Expression cloning and functional characterization of human cDNA for ganglioside GM3 synthase, *J Biol Chem* **273**, 31652–5 (1998).

Received 10 September 2002; revised 27 February 2003;
accepted 27 February 2003