Glycoconjugate Journal 20, 493–500, 2004
© 2004 Kluwer Academic Publishers. Manufactured in The Netherlands.

Identification of seven new α2,3-sialyltransferase III, *ST3Gal III*, transcripts from human foetal brain

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

We have recently cloned and sequenced 19 human *ST3Gal III* gene isotranscripts from peripheral blood leukocytes and identified very complex patterns of isotranscripts of this gene in neuronal tissues. We have now cloned and sequenced additionally seven new isotranscripts from foetal brain. These novel isotranscripts showed losses of complete exons along the whole length of the coding sequence. None of the new isotranscripts coded for proteins with the two (L- and S-) sialylmotifs intact. One of the isotranscripts belonged to the isoform *ST3Gal III*-B, five to the *ST3Gal III*-C isoform and one to *ST3Gal III*-D isoform of isotranscripts, which lacks exon 3, exons 3 and 4 and exon 4 respectively. Two of the C series isotranscripts, *ST3Gal III* C4 and C11 had both lost exons 12 and 13 containing the S-motif but had otherwise the L- and the VS-motifs intact. Three isotranscripts, *ST3Gal III* C5, C12 and D5, were similar in the 3′-end coding for an identical amino acid sequence unrelated to the original enzyme. Isotranscripts *ST3Gal III* C9 and B10 were distinctly different from all other forms identified so far. The splice variants reported here are unlikely to express enzymatic activities but may other biological functions.
Published in 2004.

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

Abbreviations and accession numbers: ST3Gal III **(*SIAT6, ST3N*), Galβ1-3(4)GlcNAcα2,3-sialyltransferase III (EC 2.4.99.6, L23768, E08204, Q11203); *ST3Gal III* A8 (AF425853); *ST3Gal III* C4 (AY167993); *ST3Gal III* C5 (AY167995); *ST3Gal III* C9 (AY167996); *ST3Gal III* C12 (AY167998); *ST3Gal III* B10 (AY167992); *ST3Gal III* D5 (AY167994); Glyceraldehyde-3-phosphate dehydrogenase, GAPDH (J04038, NM_002046); IV^{3}NeuAcnLcOse$_4$Cer, Neu5Acα2,3Galβ1,4GlcNAcβ1,3Galβ1,4Glcβ-Cer; IV^{3}NeuAcLcOse$_4$Cer, Neu5Acα2,3Galβ1,3GlcNAcβ1,3Galβ1,4Glcβ-Cer; IV^{3}NeuAcGgOse$_4$Cer, Neu5Acα2,3Galβ1,3GalNAcβ1,4Galβ1,4Glcβ-Cer; nLcOse$_4$Cer, Galβ1,4GlcNAcβ1,3Galβ1,4Glcβ-Cer; LcOse$_4$Cer, Galβ1,3GlcNAcβ1,3Galβ1,4Glcβ-Cer; GgOse$_4$Cer, Galβ1,3GalNAcβ1,4Galβ1,4Glcβ-Cer; LIF-CE: laser induced fluorescent capillary electrophoresis; NCAM: neural cell adhesion molecule; PBL: peripheral blood leukocytes; UTR: untranslated region.**

Introduction

To date 19 different human sialyltransferase genes have been identified that catalyse the transfer of sialic acid from CMP-Neu5Ac into an α2,3-, α2,6- or α2,8-linkage 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 [1]. 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" which participates in the binding of the donor substrate CMP-Neu5Ac and a smaller sequence of about 23 amino acids towards the C-terminal is termed the "S (short)-sialylmotif" which participates in the binding of both the donor and the acceptor [2–5]. Also a very short sequence, the "VS-motif", further down towards the C-terminal and comprising only 3–5 amino acids has been identified as a conserved sequence of importance for the enzymatic activity [6,7]. The human α2,3 sialyltransferase III, *ST3Gal III* gene is located on chromosome 1 [8] and spans over more than 223 kilo bases of genomic DNA and is distributed over 15 exons [9]. So far there is no evidence for alternative splicing in 5′-UTR or 3′-UTR [10] but various transcripts are produced by alternative splicing within the coding region. We have thus recently cloned and sequenced 19 human *ST3Gal III*

To whom correspondence should be addressed: Göran Larson, Tel.: 46-31-342-1330; Fax: 46-31-828458; E-mail: Goran.Larson@clinchem.gu.se

gene isotranscripts from peripheral blood leukocytes of one single individual [9]. The transcripts revealed a systematic pattern; exons 3, 4 and 5 were either inserted or deleted. The persistent feature made us classify the transcripts into five major isoforms. The *ST3Gal III*-A contains exons 3, 4 and 5, *ST3Gal III*-B lacks exon 3, *ST3Gal III*-C lacks exons 3 and 4, *ST3Gal III*-D lacks exon 4 and finally *ST3Gal III*-E lacks exons 3, 4 and 5. The alternatively spliced *ST3Gal III* isotranscripts showed a highly tissue-specific expression. The most pronounced expression of isoforms, with the catalytic part of the enzyme intact, was found in muscle tissues, uterus and in adrenal glands whereas the most complex patterns of isotranscripts were found in foetal and adult brain and in spinal cord. In foetal and adult brain at least 9 isotranscripts were identified. We have now cloned and sequenced additionally seven new isotranscripts from foetal brain and found them to be the result of alternative splicing of the ST3Gal*III* gene.

Materials and methods

Cloning and sequencing

cDNA was synthesized from total RNA-samples from normal whole brains pooled from 21 spontaneously aborted male and female Caucasian foetuses, ages 26–40 weeks (Cat no K4005-Z, Clontech, Palo Alto, CA, USA). The cloning and sequencing was performed as previously described [9]. Briefly, primers (STIII-11s and STIII-11as) were designed to cover the coding region of the *ST3Gal III* gene. The amplified 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 ABI PRISM 310 instrument (P-E Applied Biosystems, Foster City, CA).

Laser induced fluorescent capillary electrophoresis

The sense primer, STIII-11s, was labelled in the 5′-end with the fluorescent dye 6-carboxyfluorescein (6-FAM) without any other change of the primer used for cloning. The antisense primer, STIII-11as, was similarly labelled with the hexachlorinated analogue to 6-FAM (HEX). The glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene (Acc.no. NM_002046) transcript was used as an external positive control, using a 6-FAM labelled GAPDH sense primer and an unlabeled antisense primer, giving rise to a fragment of 692 bp. About 0.1 μg RNA was used in the RT-step and 10 μl of the transcribed cDNA was further used as template. The RT-PCR and capillary electrophoresis technique was performed as described in reference [9].

Results

Cloning and sequencing

We have recently reported our findings of the complex splicing pattern of the *ST3Gal III* gene in peripheral blood leukocytes and in other human tissues (Figure 1). In the electrophoretograms of *ST3Gal III* isotranscripts from foetal and adult brain and from adrenal gland we discovered several additional and unknown peaks (Figure 2). In order to characterize the isoforms corresponding to these peaks we have cloned and sequenced the amplified fragments in this size range obtained from foetal brain. Within our clones we isolated seven new isotranscripts covering the coding region of the human *ST3Gal III* gene from exon 2 to exon 14. All the deletions followed the GT-AG rule for splicing of intron sequences. The isotranscripts showed losses of complete exons coding for both the potential stem region as well as for the catalytic part of the enzyme (Figures 1 and 3).

The protein structures of these alternative splices are illustrated in Figure 3, where the amino acid sequences of the putative proteins are related to the ST3Gal III exons identified before [9]. None of the cloned isotranscripts coded for proteins with the two sialyl motifs intact. Although the isotranscripts thus did not code for catalytically active proteins the resulting proteins displayed some similarities. Five of the isotranscripts belonged to the C-series; C4, C5, C9, C11 and C12, and showed differences only from exon 7 and downward. The ST3Gal III C4 and C11 isotranscripts both showed losses of exons 12 and 13, the latter coding for the S-motif of the sialyltransferase. These isotranscripts did however contain exons 10, 11 and 14 coding for the L- and VS-motifs and resulted in proteins with completely identical N-terminals and a C-terminal sequence homologous to the C-terminal of the full-length protein. The C5, C12 and D5 isotranscripts coded for truncated proteins, which shared a unique 16 amino acid sequence located at the most C-terminal part of the polypeptide (Figure 3). Within our clones we also identified the new B10 isoform adding to the long list of *ST3Gal III* gene isotranscript [9].

Laser induced fluorescent capillary electrophoresis

After establishing the isotranscript clones, the amplified fragments were checked by capillary electrophoresis and compared with the tissue expression of *ST3Gal III* gene (Figures 2 and 4). The fragment sizes, sizes estimated by the software and elution times for each transcript are given in Table 1. The *ST3 Gal III* C12, C5, D5, C4 and C11 fragments are easily recognized in the electrophoretograms. However, the *ST3Gal III* C9 and B10 fragments were not expressed over the detection limits in the electrophoretogram of cDNA from foetal brain. In the electrophoretogram of foetal brain in Figure 2 there are still some minor unidentified peaks, which may well be hitherto uncharacterised ST3Gal III isotranscripts. The 750 bp fragment in Figure 4 was always found together with the *ST3Gal III* C12 isotranscript and probably reflects a dimerization of the original fragment of 368 base pairs (estimated size 375 bp).

The relative expression, described as percentages of GAPDH expression, of all *ST3Gal III* isotranscripts in foetal and adult brain and in spinal cord are listed in Table 2. Spinal cord did not express any of these new isotranscripts.

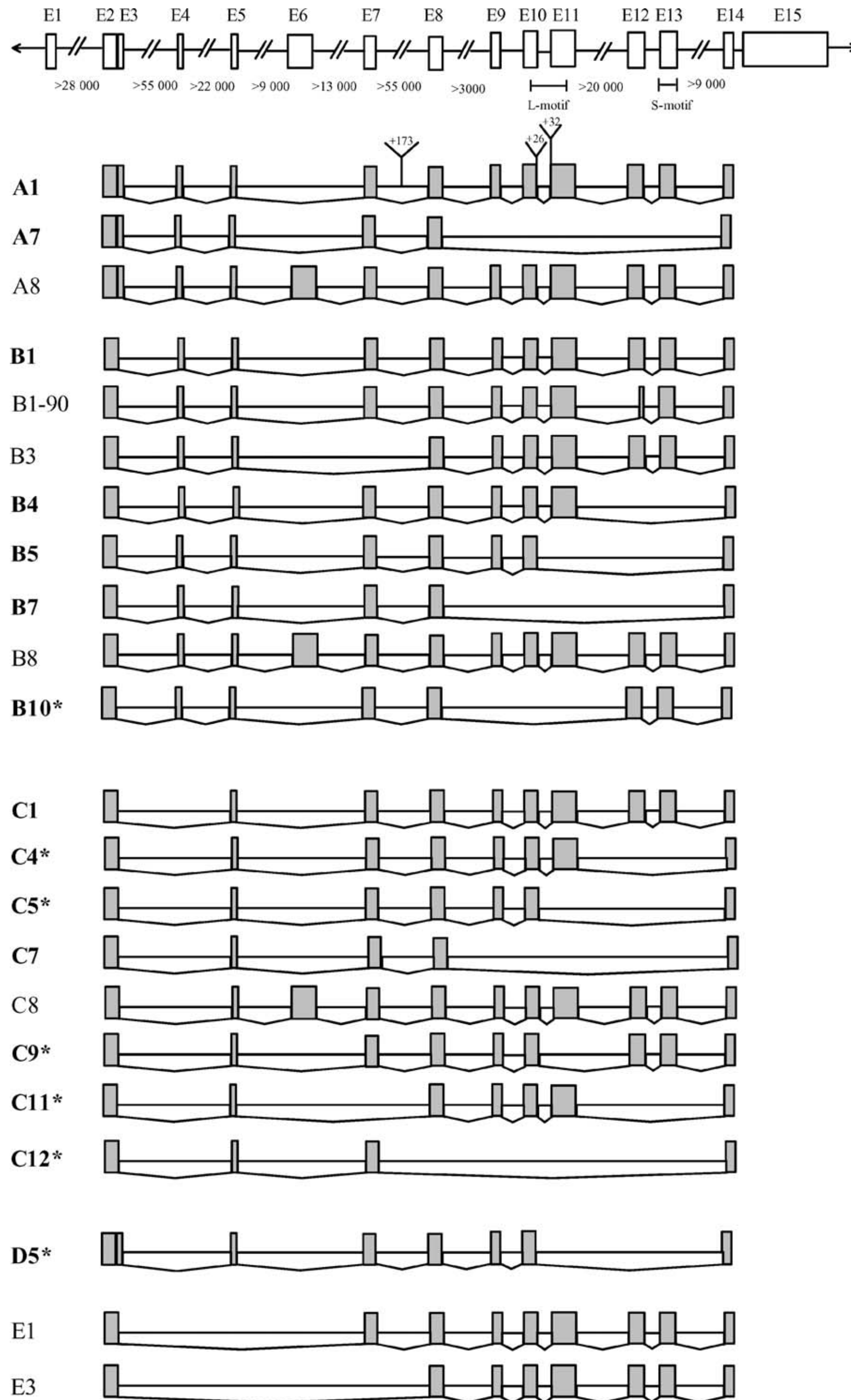

Figure 1. Illustration of cloned *ST3Gal III* transcripts and their relation to the *ST3Gal III* gene. Coding sequences are shown as grey boxes and L- and S-motifs are shown as bars below the gene sequence. Transcripts with insertions of intron nucleotides (26, 32 or 173 nt long) have been omitted but their positions are indicated in the illustration. Bold typing marks the isotranscripts identified in foetal brain and asterix marks the new isotranscripts defined in this paper.

Discussion

We have earlier identified and examined the tissue specific expression of 19 isotranscripts of the *ST3Gal III* gene. The most pronounced expressions were found in muscle tissues while the most complex patterns of isotranscripts were found in nervous tissues. The 7 new isotranscripts now cloned from foetal brain add to this complexity. In foetal and adult brain there were only

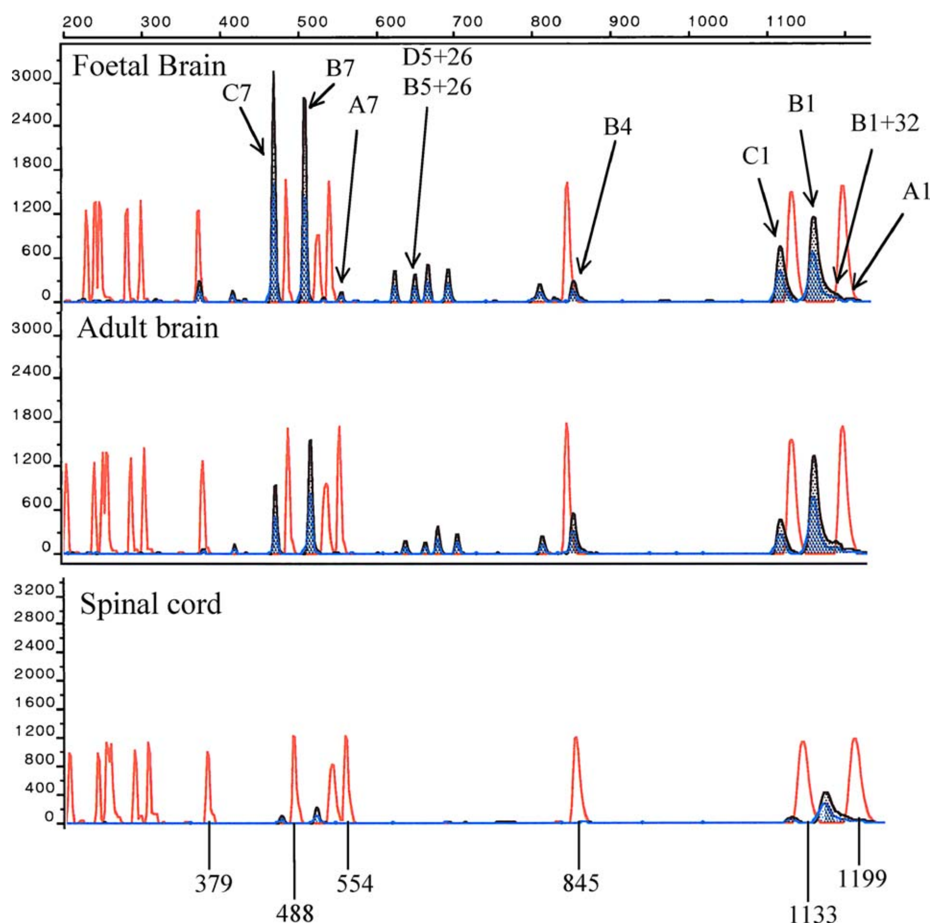

Figure 2. Electrophoretograms from LIF-CE analyses of amplified *ST3Gal III* transcripts obtained from foetal brain (top), adult brain (middle) and spinal cord (bottom). Identities of amplified fragments are given on top of each peak and size markers are given as bp lengths below reference peaks (Ref. 8).

two catalytically active isoforms expressed, *ST3Gal III* B1 and C1. These isotranscript showed a relatively lower expression, related to the GAPDH expression, than in other tissues [9]. All other isoforms expressed in foetal brain showed deletions in the sequence corresponding to complete or partial losses of the catalytic part of the enzyme. The biological roles of these isotranscripts, if any, are completely unclear and remain to be shown.

Table 1. Capillary electrophoretic data of fluorescence labelled amplified fragments from seven clones of *ST3Gal III* isotranscripts

Isotranscript[1]	*Fragment size (bp)*	*Estimated size (bp)*[2]	*Elution time (min)*[2]
C12	368	375	63.39
C5	623	638	73.16
D5	668	676	74.42
C11	717	721	75.80
B10	805	800	78.16
C4	810	813	78.52
C9	917	913	80.95

[1]Ordered after elution times.
[2]Refers to elution times of Figure 4.

When looking at the amino acid level there were, however, subgroups of polypeptides that exhibited identical new amino acid sequences in the C termini of the proteins (Figure 3). Thus, isoforms C4 and C11 included an intact exon 14 coding for the VS-motif of sialyltransferases. The C5, C12 and D5 isotranscripts all coded for a common stretch of 16 amino acids at the very end of the corresponding proteins.

The question is why foetal and adult brain and spinal cord display a more complex pattern of isotranscripts than other tissues. Gangliosides are especially enriched in nervous tissue and characterization of human foetal brain gangliosides has shown that, in addition to the dominating ganglioseries of glycosphingolipids, the $IV^3NeuAcnLcOse_4Cer$, $IV^3NeuAcLcOse_4Cer$ and $IV^3NeuAcGgOse_4Cer$ species are developmentally regulated and typically expressed between gestational week 10 and 15, and thereafter decreases to virtually zero [11]. The total RNA-samples used in the present study were obtained from whole

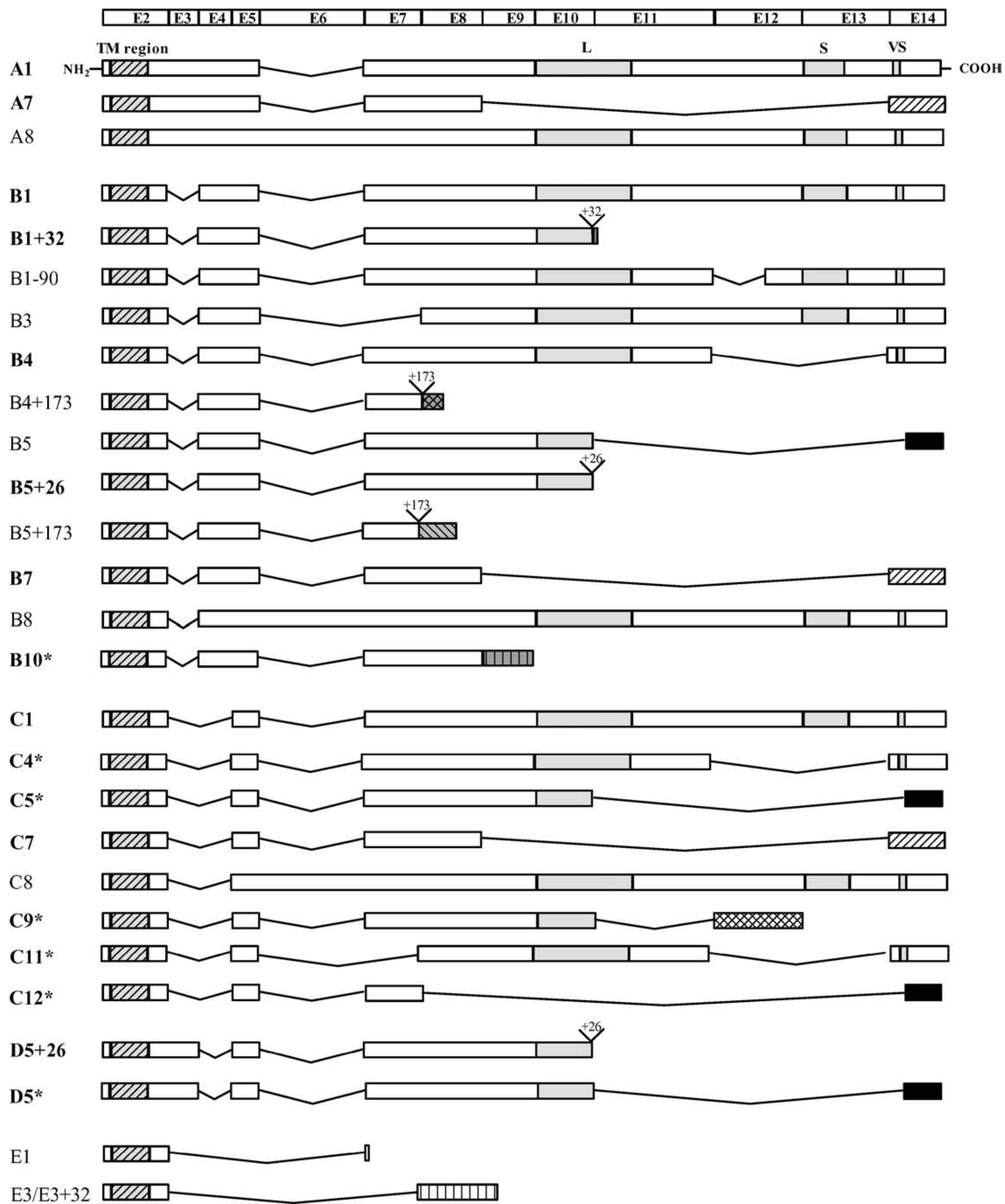

Figure 3. Schematic illustration of all potential proteins obtained from cloned *ST3Gal III* transcripts after ORF analysis. Isoproteins identified in foetal brain are marked by bold type and asterix indicates the isoforms presented in this paper. The N-terminal is marked NH_2-, the transmembrane region (TM region) is marked as a striped grey field, the L-, S- and VS-sialyl motifs are marked as grey fields and the C-terminal as –COOH. Black bars, squared or striped white or grey bars indicate altered amino acid sequences.

brains pooled from 21 spontaneously aborted foetuses, aged 26–40 weeks, and it would be interesting to analyse transcripts of brain tissues from earlier stages since these isoenzymes may play a role in the biosynthesis of lactoseries gangliosides. The isotranscript pattern seen in the total RNA sample from adult brain, which was purified from a normal whole brain of one single 28-year old male showed a similar pattern of complex expression. This indicates that the pattern is specific for brain tissue and not a reflection of individual variations. Analyses of mRNA corresponding to human α2,3-sialyltransferases have shown that all six enzymes are more or less expressed in adult brain and that all but *ST3Gal VI* are expressed in skeletal muscle [9,12–17]. One characteristic of nervous and muscle tissues is the presence of the neural cell adhesion molecule, NCAM,

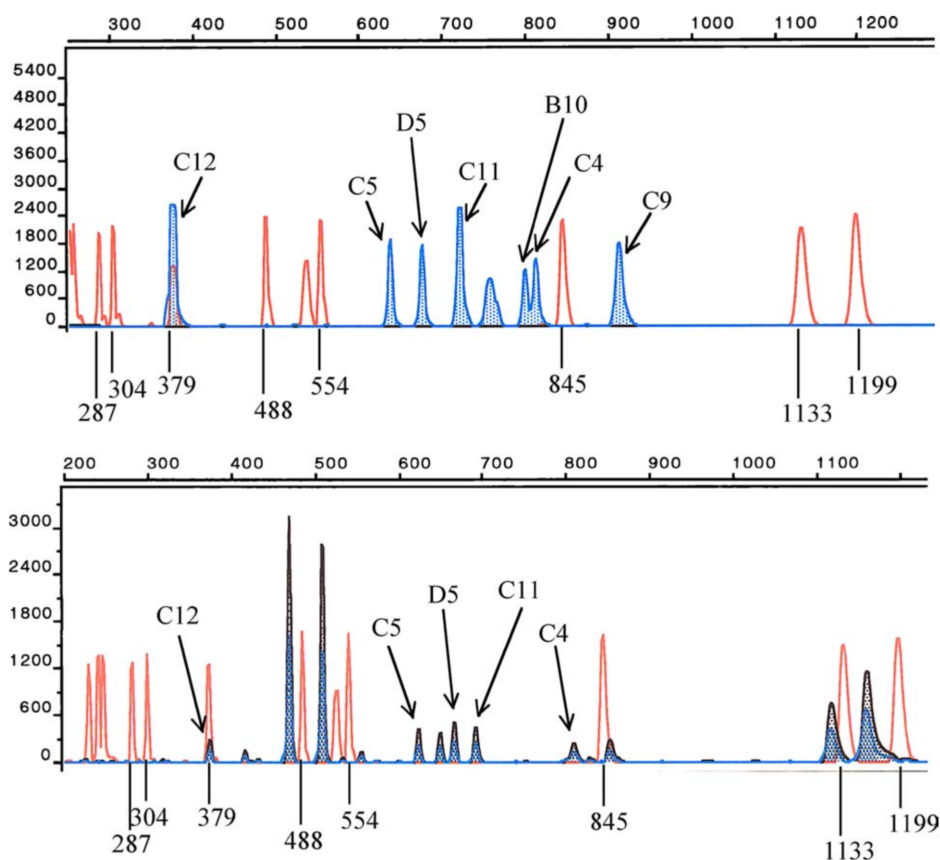

Figure 4. Electrophoretogram from LIF-CE analysis of a mixture of fluorescent-labelled fragments obtained from the seven *ST3Gal III* transcript clones (top) and amplified *ST3Gal III* transcripts obtained from foetal brain (bottom). Novel identities of amplified fragments are given on top of each peak and size markers are given as bp lengths below each peak. Sizes and elution times of labelled fragments are given in Table 1 and electrophoretic conditions are described in the methods section.

Table 2. The relative expression of the *ST3Gal III* isotranscripts (% of GAPDH) identified in foetal brain, adult brain and spinal cord

	Foetal brain	*Brain*	*Spinal cord*
A1	<1	<1	
A7	1		
B1	20	21	12
B1 + 32	3	2	5
B4	4	8	
B5 + 26/D5 + 26	6	3	
B7	33	17	9
B10			
C1	12	3	2
C4	1	1	
C5	4	1	
C7	38	10	4
C9			
C11	3	1	
C12	2		
D5	2		

which predominantly contains polysialic acids. The addition of poly-α2,8-linked sialic acid to NCAM plays a crucial role in neural development [18]. An addition of an α2,3-linked sialic acid, which could be catalysed by the action of either ST3Gal I, ST3Gal II or ST3Gal V [12,14,16,17,19] primes the polysialylation steps. However, ST3Gal III-B1 may also be responsible for this polysialylation priming since it has been shown that overexpression of ST3Gal III in neuroblastoma cells increased the polysialylation of NCAM [20].

The splicing pattern of *ST3Gal III* differs remarkably from that of other sialyltransferase genes. For example, The *ST3Gal IV* has been shown to produce at least six isoforms with alternatively spliced 5′-UTR and alternative promoters but show only minor structural variations within the coding region of the gene [8,13,21]. Two 5′-UTR alternatively spliced forms have been reported for *ST3Gal VI,* Type 1 and Type 2, which show a tissue specific expression. The transcriptional regulation depends on two different promoters P1 and P2 respectively [22]. In our laboratory, we have cloned and sequenced seven different isotranscripts of *ST3Gal VI* with variation within the coding region, but only one of these would probably exhibit any catalytic

activity (unpublished). Recently, Taniguchi *et al.* published the promoter site for the *ST3Gal III* gene and concluded that there were no alternative promoters and as a consequence there would be no alternative splicing in the 5′-UTR or 3′-UTR [10]. The exceptional complexity of splicing within the coding region of *ST3Gal III*, as we have seen it, might thus be a compensatory mechanism to alternative splicing in the 5′-UTR or 3′-UTR.

In RNA processing, as well as in all biological processes, errors are made and quality controls are needed. It sometimes happens that the splicing machinery makes mistakes that result in shifts in the open-reading frame. To avoid production of possibly truncated polypeptides two cellular quality control pathways ensure that only correctly processed mRNAs escape to the cytoplasm for subsequent translation. These processes are mRNA export and nonsense-mediated decay (NMD). During and after nuclear export, the mRNA is also subjected to a process called mRNA surveillance. If the open-reading frame has shifted to include a premature stop codon, the mRNA will trigger the NMD. A termination codon is recognized as premature if is located too far upstream from the original stop codon and positioned $>$50–55 nt upstream from an exon-exon junction [23–25]. Trigging of NMD results in increased degradation of the transcript and limits the expression of truncated polypeptides [23–25]. This suggests that the ST3Gal III B10 and C9 may not be translated to polypeptides since the splicing introduce new upstream stop codons and thereby probably trigger the NMD to target the mRNAs for rapid decay. However, the other ST3Gal III isotranscripts found in foetal brain and adult brain, which code for truncated proteins, C5, C12 and D5, also exhibit premature stop codons but these are located less than 50 nt upstream from the original stop codon and exon-exon border, and might thus escape the NMD. Taken together, the splicing pattern observed for *ST3Gal III* implies that the regulation of protein expression is controlled by alternative splicing rather than at the level of transcription or translation. Whether the *ST3Gal III* transcripts, lacking one or both sialylmotifs, and typically found in foetal brain really are translated to protein and thereby have biological functions remains to be proven.

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

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

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

4 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).

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

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

7 Jeanneau C, Chazalet V, Auge C, Soumpasis DM, Harduin-Lepers A, Delannoy P, Imberty A, Breton C, Structure-function analysis of the human sialyltransferase ST3Gal I: Role of N-glycosylation and a novel conserved sialylmotif, *J Biol Chem* **279**, 13461–8 (2004).

8 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).

9 Grahn A, Barkhordar GS, Larson G, Cloning and sequencing of nineteen transcript isoforms of the human alpha2,3-sialyltransferase gene, ST3Gal III; its genomic organisation and expression in human tissues, *Glycoconj J* **19**, 197–210 (2002).

10 Taniguchi A, Saito K, Kubota T, Matsumoto K, Characterization of the promoter region of the human Galbeta1,3(4)GlcNAc alpha2,3-sialyltransferase III (hST3Gal III) gene, *Biochim Biophys Acta* **1626**, 92–6 (2003).

11 Svennerholm L, Rynmark BM, Vilbergsson G, Fredman P, Gottfries J, Mansson JE, Percy A, Gangliosides in human fetal brain, *J Neurochem* **56**, 1763–8 (1991).

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

13 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).

14 Chang ML, Eddy RL, Shows TB, Lau JT, Three genes that encode human beta-galactoside alpha 2,3-sialyltransferases. Structural analysis and chromosomal mapping studies, *Glycobiology* **5**, 319–25 (1995).

15 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).

16 Kim YJ, Kim KS, Kim SH, Kim CH, Ko JH, Choe IS, Tsuji S, Lee YC, Molecular cloning and expression of human Gal beta 1,3GalNAc alpha 2,3-sialytransferase (hST3Gal II), *Biochem Biophys Res Commun* **228**, 324–7 (1996).

17 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).

18 Franceschini I, Angata K, Ong E, Hong A, Doherty P, Fukuda M, Polysialyltransferase ST8Sia II (STX) polysialylates all of the major isoforms of NCAM and facilitates neurite outgrowth, *Glycobiology* **11**, 231–9 (2001).
19 Harduin-Lepers A, Vallejo-Ruiz V, Krzewinski-Recchi MA, Samyn-Petit B, Julien S, Delannoy P, The human sialyltransferase family, *Biochimie* **83**, 727–37 (2001).
20 Georgopoulou N, Breen KC, Overexpression of alpha2,3 sialyltransferase in neuroblastoma cells results in an upset in the glycosylation process, *Glycoconj J* **16**, 649–57 (1999).
21 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).
22 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).
23 Lee CJ, Irizarry K, Alternative splicing in the nervous system: An emerging source of diversity and regulation, *Biol Psychiatry* **54**, 771–6 (2003).
24 Lykke-Andersen J, mRNA quality control: Marking the message for life or death, *Curr Biol* **11**, R88–91 (2001).
25 Singh G, Lykke-Andersen J, New insights into the formation of active nonsense-mediated decay complexes, *Trends Biochem Sci* **28**, 464–6 (2003).

Received 25 November 2003; revised 30 April 2004; accepted 26 May 2004