Glycobiology Insights

REVIEW

Comprehensive Analysis of Sialyltransferases in Vertebrate Genomes

Anne Harduin-Lepers

Unité de Glycobiologie Structurale et Fonctionnelle (UGSF), Université des Sciences et Technologies de Lille, UMR CNRS 8576, IFR 147, 59655 Villeneuve d'Ascq, France. Email: anne.harduin@univ-lille1.fr

Abstract: Sialyltransferases are a subset of glycosyltransferases involved in the biosynthesis of sialylated glycolipids or/and glycoproteins. The aim of this review is to provide a comprehensive review and analysis of vertebrate sialyltransferase genes. Primary structure/function relationships have been explored through the use of molecular phylogeny and phylogenomic approaches. Several animal sialyltransferase sequences have been identified in genomic databases on the basis of the presence of sialylmotifs. Depending on the glycosidic linkage formed and their monosaccharide acceptor, vertebrate sialyltransferases are arranged in four families of proteins (ST3Gal, ST6Gal, ST6GalNAc and ST8Sia), which are characterized by consensus peptides called family-motifs. Sialyltransferases families are further subdivided into 20 sub-families in mammals and more than 25 sub-families in lower vertebrates, each of them being characterized by conserved amino acid positions. From an evolutionary point of view, the genomic organization of the coding region of these sialyltransferase genes is highly conserved across vertebrate species suggesting that they evolved from common ancestral genes through multiple duplication events. Finally, comparative analysis of the sialyltransferase gene expression evolution in vertebrate adult tissues and during embryonic development have shown marked differences suggesting the influence of genetic and tissue specific factors.

Keywords: sialyltransferases, homolog, vertebrate genome, molecular phylogeny, phylogenomic, transcriptional regulation

Glycobiology Insights 2010:2 29–61

This article is available from http://www.la-press.com.

Introduction

In vertebrates, the transfer of sialic acid residues from activated sugar donors like CMP-β-Neu5Ac, CMP-β-Neu5Gc or CMP-β-KDN to the terminal non reducing position of galactose (Gal), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc) or sialic acid monosaccharide of glycoconjugate acceptors requires a subset of glycosyltransferases named sialyltransferases that catalyze the formation of α2-3, α2-6 or α2-8 glycosidic linkages. Up to date, a number of sialyltransferase genes have been identified in higher vertebrate genomes, mainly those of *Mus musculus* (Mmu), *Homo sapiens* (Hsa) and *Gallus gallus* (Gga). In the nineties, efforts have concentrated on the molecular cloning and biochemical characterization of each of the twenty genes encoding human sialyltransferases and their mouse counterparts (reviewed in),[1–3] which have been named according to the systematic nomenclature proposed by Tsuji et al.[4] Much less data are available from lower vertebrates like bony fish and invertebrate genomes. These recent years, bioinformatic strategies have been developed for comprehensive finding and analysis of glycosyltransferase genes.[1,3,5–7] Vertebrate sialyltransferase genes represent a multigene super-family of homologous sequences designated either orthologs or paralogs. Orthologs are related to common ancestor by speciation. They retain a same function in the course of evolution and belong to the same sub-family. Paralogs are related by duplication[8] within a genome and belong to different sub-families. They often evolve novel biochemical activities and show spatio-temporal expression. It is interesting to note that all the animal sialyltransferases identified up to date are grouped in the family 29 of the Carbohydrate-Active enZYmes (CAZy) database for glycosyltransferases[9] and this observation denotes their common modular organization and their common ancestral origin. Classically, vertebrate sialyltransferases have been divided into four families, namely the ST3Gal, ST6Gal, ST6GalNAc and ST8Sia depending on the glycosidic-linkage formed and the monosaccharide acceptor used.

The repertoire of sialylated structures encountered in tissues of animal species and in various pathological states such as cancer[10] relies on the molecular diversity of sialyltransferases and reflects their diverse expression pattern, which is reviewed here. New genes and novel biochemical processes arise mainly by gene duplication and divergence along animal evolution. However, other mechanisms such as transcriptional regulation, alternative splicing, post-translational and regulatory mutations among others, can serve to increase their functional diversity. In this context, the role of epigenetic regulation of glycosyltransferases has been suggested.[11] In spite of their multiplicity and crucial role in biology, still very little is known about sialyltransferase sequence/function relationships, tissue-specific expression and evolution. This comprehensive review of vertebrate sialyltransferase gene ontologies (i.e. molecular function, location and biological process) focus on molecular phylogeny and phylogenomics strategies as a way to get insights into their structure/function relationships.

Topology and Structural Features of Animal Sialyltransferases

It has long been known that sialyltransferase proteins share the same architecture within the *trans*-Golgi network[12] and this topology along the secretory pathway is conserved across vertebrate species from primitive chordate to mammals (recently reviewed in e-book sialobiology, chapter 5). Analysis of the deduced protein sequence indicated that animal sialyltransferases are type II transmembrane glycoproteins. The N-terminal cytoplasmic tail of sialyltransferase homologs is usually rather short, its amino acid composition being highly variable and not conserved across animal evolution. It is not essential for catalytic activity, but its role is still not clarified. It has been proposed that it could define subcellular localization and stability of sialyltransferases in intracellular dynamics.[13] A unique transmembrane domain (TMD) of 16–20 amino acids is located towards the N-terminus part of the protein. It is rich in Leu and Phe residues located mainly in the middle of the TMD with charged amino acid residues flanking both ends of the TMD[14] and appears to be more or less conserved across vertebrate species. It is known to function as a Golgi-membrane anchor and could play also a role in localization of the sialyltransferases. The stem domain can be defined as the region of the protein that lays upstream catalytic domain that is not necessary for catalysis. It is highly

variable in length among paralogs and orthologs and display high sequence variability. Although it is not essential for catalytic activity, it serves to extend the catalytic domain into the lumen of the Golgi acting as a flexible tether.[14] Proteolytic cleavage sites have been described in the stem region of sialyltransferases enabling secretion of soluble enzymes in body fluids.[12,15,16]

Vertebrate sialyltransferases show a large catalytic domain of about 250 amino acids oriented towards the lumen of the Golgi. Multiple sequence alignments of cloned sialyltransferases show very little overall sequence identity with the notable exception of four conserved peptide motifs named sialylmotifs identified in their catalytic domain[17] represented in Figure 1: Sialylmotif L or large, is located in the middle of the molecule and comprises 48–49 amino acids with 5 strictly conserved amino acids.[18] Sialylmotif S or small encompasses 23 amino acids with two invariant amino acids Gly and Cys, 11 amino acids apart from each other.[19] Sialylmotifs L and S contain Cys residues implicated in the formation of a disulfide bond.[20] Sialylmotif III shows 4 amino acids, where the Tyr residue is highly conserved.[21,22] Finally, sialylmotif VS or very short has 6 amino acids, where the highly conserved Glu is separated by 4 amino acids from the His also highly conserved.[23] These sialylmotifs were identified in all the sialyltransferases from vertebrate and invertebrate species and are typical of all the animal enzymes of the CAZy-family 29, irrespective of their linkage- and acceptor specificities. They serve as hallmarks in the identification of sialyltransferase homologs in the public nucleotide databases. Site-directed mutagenesis of the sialylmotifs L and S conserved amino acids residues in the rat ST6Gal I and in the human ST3Gal I have indicated that they are implicated in the donor (CMP-Neu5Ac) and acceptor binding, respectively.[17,24,25] Site-directed mutagenesis of the His residue in the sialylmotif VS of ST8Sia II, ST8Sia IV and ST3Gal I did not induce any conformational modification nor ability to bind substrates, but a loss of activity suggesting its implication in catalysis.[21,26] Similarly, mutations of His and Tyr residues in the sialylmotif III of the human ST3Gal I led to loss of activity and efficiency of catalysis.[21] Important roles of these motifs have been suggested in the folding and maintenance of 3D structure taking part in functional aspect common to all sialyltransferases, which was recently confirmed with the X-ray crystal structure of the porcine ST3Gal I.[27] More precisely, the His_{302}, Asn_{150} and Asn_{173} of the porcine ST3Gal I present in the sialylmotifs III and L of animal sialyltransferases and highlighted by a yellow star in Figure 1 are involved in stabilization of the phosphate group of the donor substrate CMP-Neu5Ac. These sialylmotifs represent residue conservation patterns at the super-family level suggesting structural and functional role conserved across evolution.

Family motifs of 4 to 20 amino acids that are specific of each sialyltransferase family i.e. ST3Gal, ST6Gal, ST8Sia and ST6GalNAc, have been reported.[28–30] Protein sequence alignments across vertebrate species revealed that each family share a common 4 amino acid sequence located 8 amino acids downstream the 3′-end of the sialylmotif L, named motif "a" in Figure 1[29,31] except for ST6GalNAc, since the 7 amino acids ST6GalNAc motif "a" is located 4 amino acids closer to the sialylmotif L (Fig. 1). Similarly, a second family motif, named motif "c" with 2 amino acids overlap at the 3′-end of the sialylmotif S, is shown in Figure 1.[2] More recently, Patel and Balaji used 47 cloned sialyltransferase sequences characterized mostly from mammalian species to delineate additional family motifs,[32] which are slightly refined in Figure 1. Motif "b" lies between sialylmotifs L and S, about 20 amino acids downstream sialylmotif L and varies highly in length among sialyltransferase families. Motif "d" is located downstream sialylmotif III in the ST6Gal family (Fig. 1) and motif "e" is found downstream sialylmotif VS in the ST8Sia[5] and ST6GalNAc families (Fig. 1). All these family motifs are described not only in vertebrate sialyltransferase sequences, but also in invertebrate sequences suggesting that they were also present in their last common ancestor (LCA). They represent a second level of amino acids residue conservation pattern of the protein in each sialyltransferase family and might be relevant for linkage specificity and for monosaccharide recognition although no site directed mutagenesis of conserved amino acids in family motifs has been reported yet. This hypothesis has been confirmed by Rao et al who demonstrated the interaction of Glu_{196} of the porcine ST3Gal I that lies in family motif "a" and Lys_{213} and Asp_{216} in family motif "b"

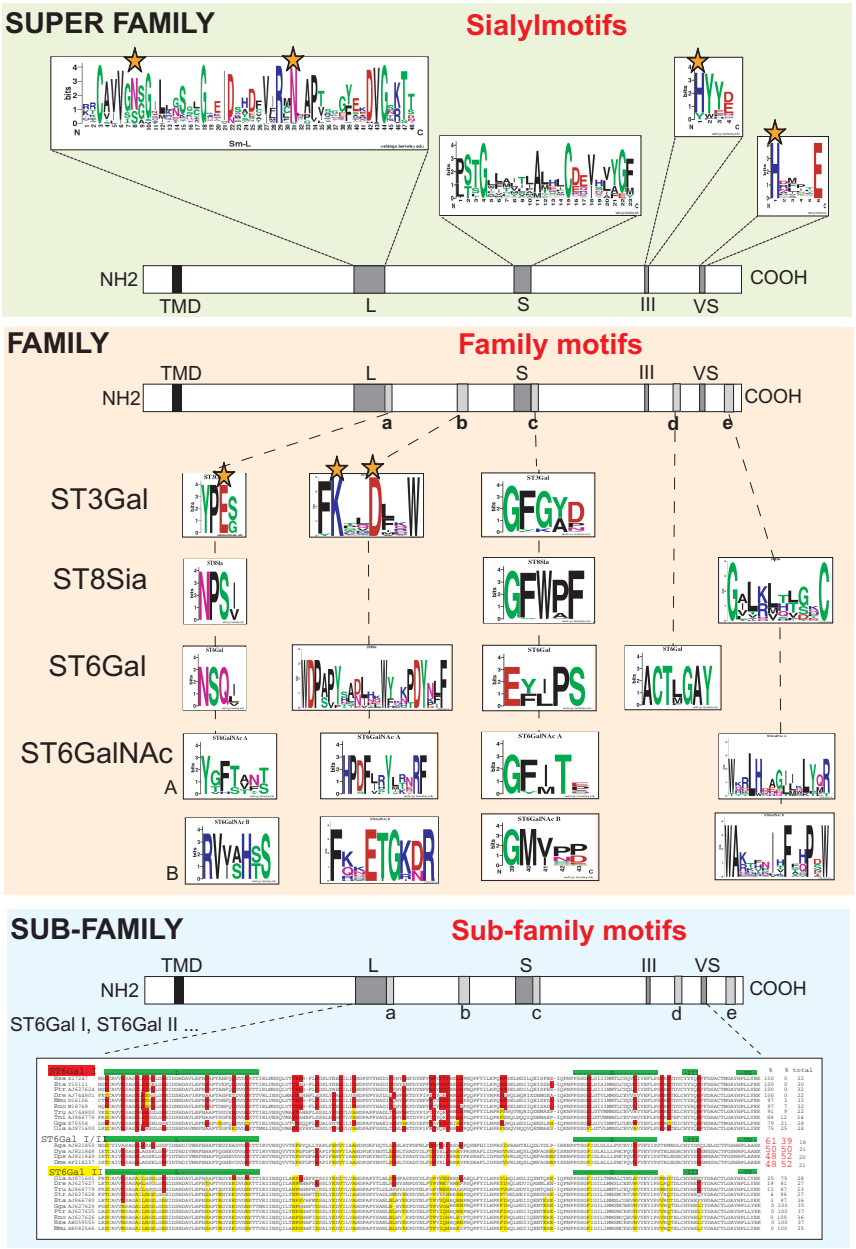

Figure 1. Sequence logo of super-family motifs, family motifs and sub-family motifs in the vertebrate sialyltransferase proteins. Graphical representation derived by the Berkeley weblogo tool[190,191] after multiple sequence alignments at the ClustalW site of the PBIL site (see http://pbil.univ-lyon1.fr/) illustrating the three levels of amino acid sequence conservation. Yellow stars indicate amino acid residues implicated in substrate binding as demonstrated by X-ray crystallography of porcine ST3Gal I.[27] **Upper panel** shows the first level of amino acid conservation represented by sequence logo of sialylmotif L, S, III and VS above a schematic representation of a sialyltransferase. These sialylmotifs characterize vertebrate sialyltransferases super-family. It shows frequency and occurrence of amino acids in four functional sites of 125 vertebrate sialyltransferase sequences. **Middle panel** shows sequence logo of linkage motifs with frequency and occurrence of amino acids characterizing each sialyltransferase family ST3Gal (42 vertebrate sequences), ST8Sia (37 vertebrate sequences), ST6Gal (12 vertebrate sequences), ST6GalNAc (11 ST6GalNAc I and ST6GalNAc II vertebrate sequences designed as ST6GalNAc (A) and 23 ST6GalNAc III, ST6GalNAc IV, ST6GalNAc V and ST6GalNAc VI vertebrate sequences, designated as ST6GalNAc (B). Family motifs “a”, “b”, “c”, “d” and “e” are positioned relative to the sialylmotifs as described previously by Patel and Balaji.[32] In the logos, one letter amino acid symbols are colored according to their chemical properties: polar amino acids (G, C, S, T, Y) are green, basic (K, R, H) are blue, acidic (D, E) are red, hydrophobic (A, V, L, I, P, W, F, M) are black and neutral polar amino acids (N, Q) are pink. The overall height of the stacks indicates the sequence conservation at a given position, while the height of symbols within the stack indicates the relative frequency of each amino acid at that position.[190,191] **Lower panel** illustrates sub-family motifs of the ST6Gal I and ST6Gal II sub-families as described previously.[1] Briefly, ClustalW alignment of the peptide sequences of the catalytic domain of the ST6Gal I and ST6Gal II sub-families are shown. Letters with red background represent conserved positions specific from ST6Gal I and letters on yellow background represent conserved amino acid positions specific for ST6Gal II. The total number of sub-family specific conserved amino acid positions is represented in the last column preceded by the relative proportions (%) of ST6Gal I and ST6Gal II specific amino acid positions, for each protein. The position of the sialylmotifs (L, S, III and VS) is indicated on a green background in the first line of each sub-family. The protostomia sequences flanked by horizontal thick lines cannot be classified in either of the two sub-families because they have similar proportions of ST6Gal I and ST6Gal II specific conserved amino acid positions (between 61 and 39%, bold red characters).

Table 1. Cloned human sialyltransferase cDNAs.

Sialyltransferase	Other name(s)	Acc num.	AA	Substrate	Structures formed	Ref.
ST3Gal I	ST3O, ST3GalA.1	L29555	340	O-GP/GL	**Neu5Acα2-3**Galβ1-3GalNAc- G_{D1a} G_{M1b} G_{T1b}	34
ST3Gal II	ST3GalA.2	X96667	350	O-GP/GL	**Neu5Acα2-3**Galβ1-3GalNAc- G_{D1a} G_{M1b} G_{T1b}	40
ST3Gal III	ST3(N)	L23768	375	GP	**Neu5Acα2-3**Galβ1-3/4GlcNAc-	43
ST3Gal IV	STZ	L23767	333	GP/GL	**Neu5Acα2-3**Galβ1-4GlcNAc- **Neu5Acα2-3**Galβ1-3GalNAc-	44
ST3Gal V	G_{M3} synthase	AB018356	362	GL	**Neu5Acα2-3**Galβ1-4Glcβ-Cer G_{M3} and possibly G_{M4}	47
ST3Gal VI		AF119391	331	GP/GL	**Neu5Acα2-3**Galβ1-4GlcNAc-	30
ST6Gal I	ST6(N)	X17247	406	N-GP	**Neu5Acα2-6**Galβ1-4GlcNAc-	55
ST6Gal II		AB059555	529	N-GP	**Neu5Acα2-6**Gal(NAc)β1-4GlcNAc-	61, 62
ST6GalNAc I		Y11339	600	O-GP	(Neu5Acα2-3)$_{0-1}$(Galβ1-3)$_{0-1}$GalNAc-Ser **Neu5Acα2-6**	66
ST6GalNAc II		AJ251053	374	O-GP	(Neu5Acα2-3)$_{0-1}$Galβ1-3GalNAc-Ser **Neu5Acα2-6**	70
ST6GalNAc III		AJ507291	305	O-GP/GL	Neu5Acα2-3Galβ1-3GalNAc-Ser **Neu5Acα2-6** and $G_{D1\alpha}$	1
ST6GalNAc IV		AJ271734	302	O-GP/GL	Neu5Acα2-3Galβ1-3GalNAc-Ser **Neu5Acα2-6** and $G_{D1\alpha}$	72
ST6GalNAc V	$G_{D1\alpha}$ synthase	AJ507292	336	GP/GL	Neu5Acα2-3Galβ1-4GlcNAc- **Neu5Acα2-6** and $G_{D1\alpha}$	1
ST6GalNAc VI	$G_{D1\alpha}$ synthase/$G_{T1a\alpha}$/ $G_{Q1b\alpha}$ synthase	AJ507293	299	GL	α-series gangliosides ($G_{D1\alpha}$ $G_{T1a\alpha}$ $G_{Q1b\alpha}$) and di-sialyl Lewis[a]	1
ST8Sia I	G_{D3} synthase-SAT-II	D26360	356 341	GL	**(Neu5Acα2-8)**$_{1-2}$Neu5Acα2-3Galβ1-4Glc-Cer G_{D3} and G_{T3}	86, 177 85, 175
ST8Sia II	STX	U33551	375	GP	**(Neu5Acα2-8)**$_n$Neu5Acα2-3Galβ1-4GlcNAc-	102
ST8Sia III		AF004668	393	GL/GP	**Neu5Acα2-8**Neu5Acα2-3Galβ1- **Neu5Acα2-8**Neu5Acα2-6GalNAc-	93
ST8Sia IV	PST	L41680	359	GP	**(Neu5Acα2-8)**$_n$Neu5Acα2-3Galβ1-4GlcNAc-	101
ST8Sia V	G_{T3} synthase-SAT-V	U91641	376	GL	G_{D1c}, G_{T1a}, G_{Q1b}, G_{T3}	88
ST8Sia VI		AJ621583	398	O-GP	**Neu5Acα2-8**Neu5Acα2-3Galβ1-3GalNAc-	91

Abbreviations: GL, glycolipids; GP, glycoproteins; *N*-GP, *N*-glycosylproteins; O-GP, *O*-glycosylproteins; –, this study.

illustrated in Figure 1 with the hydroxyl group 4 of the Gal residue found in the acceptor substrate Galβ1,3GalNAc.[27]

Multiple sequence alignments in the catalytic domain of each sialyltransferase ortholog across vertebrate species has led to the definition of a third level of amino acids residue conservation pattern called sub-family motifs.[1,5] Sialyltransferase sequences identified in invertebrates show intermediate values of sub-family-specific conserved amino acid positions in these pair wise alignments and thus appear to be intermediate sequences that do not belong to any of the vertebrate sub-families. These residues might be implicated in overall substrate specificity, in other words fine specificity, towards acceptor substrates.

Molecular Cloning and Molecular Phylogeny: From Sequence to Function

Biochemical characterization of each sialyltransferase activity has been achieved with a small number of cDNA cloned essentially from mammalian and avian sources to produce in mammalian cells recombinant proteins lacking the Golgi anchor and the cytoplasmic tail domain. Enzymatic assays were carried out *in vitro*, towards synthetic acceptors substrates, purified glycoproteins and glycolipids, shedding light on their substrate specificity (Table 1 and reviewed recently in).[33] Since it is not possible to determine directly and individually the enzymatic activity of all the newly identified sequences, these recent years, we have developed a combination of molecular phylogenetic and phylogenomic approaches to get further insights into the structure/function relationships of vertebrate sialyltransferases. As summarized in Table 2, our studies indicated that representatives of the twenty sialyltransferase sub-families were present in the most recent common ancestor of osteichthyans (bony fishes and tetrapods) and only a limited number of potential homologs of vertebrate sialyltransferase genes have been identified in the representative genomes of all major phylogenetic groups of invertebrate chordates.[1] Molecular phylogeny, with the construction of phylogenetic trees, multiple sequence alignments using predicted protein sequences, with the determination of sub-family-specific amino acid positions established the global evolutionary relationships between sialyltransferase sequences and enabled sequence-based prediction of functions of hundreds of sialyltransferase-related genes identified in various animal species.[1,3,5,6] Sialyltransferase genes are polyexonic and their coding region is divided into several coding exons. Thus, a third criterion of analysis is the conservation of the sialyltransferase intron/exon organization in their coding region.

1-The β-galactoside α2,3-sialyltransferase family (ST3Gal)

Six different β-galactoside 2,3-sialyltransferases cDNAs have been identified in various higher vertebrate genomes, which correspond to six ST3Gal gene sub-families.[1,3,6] These cDNAs have been cloned mainly from mammalian cells or tissues (Table 1) and recombinant enzymes were shown to catalyze the transfer of sialic acid in α2,3-linkage to the terminal Gal residues found on glycoproteins or glycolipids. In the phylogenetic tree of the ST3Gal family, two main branches clearly separate vertebrate ST3Gal I and ST3Gal II sub-families from the remaining four ST3Gal III, ST3Gal IV, ST3Gal V and ST3Gal VI sub-families (Fig. 2).

In the first branch, ST3Gal I and ST3Gal II use almost exclusively type III disaccharide Galβ1-3 GalNAc found onto glycoproteins (core1 of *O*-glycosylproteins) and glycolipids (asialo-G_{M1} and G_{M1a}).[34,35] ST3Gal I was also found to have little activity towards type I disaccharide (Galβ1,3GlcNAc).[36,37] The human, mouse and chicken ST3Gal I lead to the selective formation of Neu5Acα2-3Galβ1-3 GalNAc.[38] In addition, the mouse ST3Gal I has a preference for glycolipid acceptors leading to the biosynthesis of G_{M1b}, G_{D1a} and G_{T1b} at least in *in vitro* assays,[39] whereas the human ST3Gal II isolated from the CEM T-cell shows activity towards both glycolipids and glycoproteins.[40] It has been further suggested that ST3Gal II could have a recognition site for the ceramide moiety in addition to the one for the Galβ1-3 GalNAc moiety.[41] Four ST3Gal I-related and two ST3Gal II-related sequences have been identified in *D. rerio* and *T. rubripes* genomes (Table 2) and their fine biochemical characterization still awaits further studies. Recently, a fugu ST3Gal II recombinant protein (TruST3Gal II-r, AJ626817 in Table 2) was biochemically characterized *in vitro* and shown to have distinct acceptor specificity with no activity towards gangliosides (activity to G_{M1} oligosaccharide, but not

Table 2. Tissue pattern of expression of vertebrate sialyltransferases among vertebrate tissue samples.

	Acc num. GenBank	Gene symbol HUGO	Chrom. UNIGENE	Other gene symbols	Tissues	Ref.
ST3Gal I						
H. sapiens	L29555	*ST3GAL1*	8q24.22	SIAT4, SIAT4A	pl; lu; li; sm; fetal tissues	34
M. musculus	X73523	*st3gal1*	15D2			39
G. gallus	X80503		2			38
S. tropicalis	FN550106					–
D. rerio	AJ864512	*st3gal1-r1*	15	SIAT4-r1		1
	AJ864513	*st3gal1-r2*	15	SIAT4-r2		1
	AM287261	*st3gal1-r3*	Zv8_NA2151	SIAT4A-r3		–
	AM287262	*st3gal1-r4*	19	SIAT4A-r4		–
T. rubripes	AJ626816					1
ST3Gal II						
H. sapiens	X96667	*ST3GAL2*	16q22.1	SIAT5, SIAT4B	sm; he	40
M. musculus	X76989	*st3gal2*	8E1		sm; he	201
G. gallus	AJ585761		11			1
S. tropicalis	AJ585763					1
D. rerio	AJ783741	*st3gal2-r1*	11	siat5		1
II-r	AJ783740	*st3gal2-r2*	18	siat4		1
T. rubripes	AJ744805					1
II-r	AJ626817					42
ST3Gal III						
H. sapiens	L23768	*ST3GAL3*	1p34.1	SIAT6	he; br; sm	43
M. musculus	X84234	*st3gal3*	4D2.1		te; ov; fetal tissues	37
G. gallus	AJ865086		8			1
S. tropicalis	AJ626823					1
D. rerio	AJ626821	*st3gal3*	2	SIAT6		1
III-r	AJ626820	*st3gal3-r*	6	SIAT6-r		1
T. rubripes	AJ626818					1
ST3Gal IV						
H. sapiens	L23767	*ST3GAL4*	11q24.2	SIAT4C	he; pl; ki; sp; te; ov; fetal tissues	44
M. musculus	X95809	*st3gal4*	9A4			37
G. gallus	AJ866777		24			1
S. tropicalis	AJ622908					1
D. rerio	AJ744809		Zv8_NA1449	SIAT4C		1
T. rubripes	AJ865346					1
ST3Gal V- G_{M3} synthase						
H. sapiens	AB018356	*ST3GAL5*	2p11.2	SIAT9	br; pl; lu; sm; te; sp; fetal brain	47
M. musculus	Y15003	*st3gal5*	6C1			202
G. gallus	AY515255		4			203

(*Continued*)

Table 2. (*Continued*)

	Acc num. GenBank	Gene symbol HUGO	Chrom. UNIGENE	Other gene symbols	Tissues	Ref.
S. tropicalis	FN550108					
D. rerio	AJ619960	*st3gal5*	14	ST3Gal5		1
V-r	AJ783742	*st3gal5-r*	13	SIAT5-r		
T. rubripes	AJ865087					1
V-r	AJ865347					
ST3Gal VI						
H. sapiens	AF119391	*ST3GAL6*	3q12.1	SIAT10	he; pl; li	30
M. musculus	AK082566	*st3gal6*	16C1.2			204
G. gallus	AJ585767		1			1
S. tropicalis	AJ626744					1
D. rerio	Not identified					1
T. rubripes	Not identified					1
ST6Gal I						
H. sapiens	X17247	*ST6GAL1*	3q27–q28	SIAT1	pl; B-cells; li; ki; br	55
M. musculus	D16106	*st6gal1*	16B1		ly; hi; th; tr; sp; sk; li	57
G. gallus	X75558		9			59
S. tropicalis	Not identified					
D. rerio	AJ744801		21			1
T. rubripes	AJ744800					1
ST6Gal II						
H. sapiens	AB059555	*ST6GAL2*	2q11.2–q12.1	SIAT2	br; fetal br; te; th; si; co	61, 62
M. musculus	AK082506	*st6gal2*	17C			63
G. gallus	AJ627629		1			1
S. tropicalis	AJ627628					1
D. rerio	AJ627627	*st6gal2*	9			1
II-r	FN550105	*st6gal2-r*	6			–
T. rubripes	AJ866779					1
ST6GalNAc I						
H. sapiens	Y11339	*ST6GALNAC1*	17q25.1	SIAT7A	sg; sp; lm	66
M. musculus	Y11274	*st6galnac1*	11E2			166
G. gallus	X74946		18		te; he	68
S. tropicalis	Not identified					
D. rerio	AM287259		21			1
T. rubripes	Not identified					
ST6GalNAc II						
H. sapiens	AJ251053	*ST6GALNAC2*	17q25.1	SIAT7B	he; sm; li; si	70
M. musculus	X93999	*st6galnac2*	11E2		te; mg; ki; th; he; li; lu; sg	168
G. gallus	X77775		18		br; ki; li; lu; te	69
S. tropicalis	AJ620650					1
D. rerio	AJ634459		12			1

(*Continued*)

Table 2. (*Continued*)

	Acc num. GenBank	Gene symbol HUGO	Chrom. UNIGENE	Other gene symbols	Tissues	Ref.
T. rubripes	AJ634460					1
	AJ634461					1
ST6GalNAc III						
H. sapiens	AJ507291	*ST6GALNAC3*	1p31.1	SIAT7C	ki; br	1
M. musculus	Y11342	*st6galnac3*	3H3		br; he; lu; sp; ki; te; mg; em	73
G. gallus	AJ634455		8			1
S. tropicalis	Not identified					
D. rerio	AJ620947		2			1
T. rubripes	AJ634456					1
ST6GalNAc IV						
H. sapiens	AJ271734	*ST6GALNAC4*	9q34	SIAT7D	li; ki; pl; lu; in; sp; co; low expression else	72
M. musculus	Y19057	*st6galnac4*	2B		br; th; lu; he; sp; co	73
G. gallus	AJ620652		17			1
S. tropicalis	BC121364					205
D. rerio	AJ868430		21			1
T. rubripes	AJ646869					1
ST6GalNAc V						
H. sapiens	AJ507292	*ST6GALNAC5*	1p31.1	SIAT7E	br	1
M. musculus	BC055737	*st6galnac5*	3H3		br; sp	77
G. gallus	AJ646877		8			1
S. tropicalis	AJ646878					1
D. rerio	AJ646874		8			1
	AM287260		2			–
T. rubripes	AJ646873					1
ST6GalNAc VI						
H. sapiens	AJ507293	*ST6GALNAC6*	9q34.11	SIAT7F	br	1
M. musculus	AB035123	*st6galnac6*	2B		br; sp; co; li; te; he	77
G. gallus	Not identified					
S. tropicalis	FN550109					-
D. rerio	AJ646883		21			1
T. rubripes	AJ646880					1
ST8Sia I-G_{D3} synthase-SAT-II						
H. sapiens	D26360	*ST8SIA1*	12p12.1	SIAT8A	Fetal br; br;	85
M. musculus	X84235	*st8sia1*	6G3		em	94
G. gallus	U73176		1		re, br	173
S. tropicalis	AY652775				Fetal br; em	82
D. rerio	AJ715535		4		em; ov; fetal br	87
T. rubripes	AJ715534					1

(*Continued*)

Table 2. (*Continued*)

	Acc num. GenBank	Gene symbol HUGO	Chrom. UNIGENE	Other gene symbols	Tissues	Ref.
ST8Sia II-STX						
H. sapiens	U33551	*ST8SIA2*	15q26	SIAT8B	fetal br; fetal ki; he; th; br	102
M. musculus	X83562	*st8sia2*	7D2		em	100
G. gallus	AJ699419		10			1
S. tropicalis	BC121420					205
D. rerio	AY055462		18		br; ey; ov; li; mu	110
T. rubripes	AJ715538					1
ST8Sia III						
H. sapiens	AF004668	*ST8SIA3*	18q21.31	SIAT8C	fe br; fe li; br	93
M. musculus	X80502	*st8sia3*	18E1		br; te; em	94
G. gallus	AJ699420		Z			1
S. tropicalis	AJ715544					1
D. rerio	AJ715543		21		Em; in; br	1
T. rubripes	AJ715541	*st8sia3*				1
III-r	AJ715542	*st8sia3-r*				
ST8Sia IV-PST						
H. sapiens	L41680	*ST8SIA4*	5q21	SIAT8D	fe br	101
M. musculus	X86000	*st8sia4*	1D		sp; bm; ly; em	103
G. gallus	AF008194		Z			206
S. tropicalis	AM419014					5
D. rerio	AJ715545		10		br; mu; in	1
T. rubripes	Not identified					
ST8Sia V-G_{T3} synthase-SAT-V						
H. sapiens	U91641	*ST8SIA5*	18q12.3	SIAT8E	fetal br; br; sm; he	88
M. musculus	X98014	*st8sia5*	18E3		fe br; br; em	89
G. gallus	AJ704564		Z			1
S. tropicalis	AM422136					5
D. rerio	AJ715546		21		em; br	1
T. rubripes	AJ715547					1
ST8Sia VI						
H. sapiens	AJ621583	*ST8SIA6*	10p13	SIAT8F	low expression all tissues	91
M. musculus	AB059554	*st8sia6*	2A1		ki; he; sp; low else	92
G. gallus	AJ699424		2			1
S. tropicalis	AM422137					1
D. rerio	AJ715551		3		em; ov; in; br	1
T. rubripes	AJ715549	*st8sia6A*		SIAT8F		1
	AJ715550	*st8sia6B*		SIAT8F-r		1

Genes are named according to the HUGO gene nomenclature.
Abbreviations: bm, bone marrow; br, brain; co, colon, em, embryo; ey, eyes; he, heart; hi, hippocampus; in, intestine; li, liver; lu, lung; ly, lymphocytes; pl, placenta; ki, kidney; mg, lactating mammary gland; mu, muscle; ov, ovaries; pl, placenta; re, retina; sb, submaxillary gland; sg, salivary gland; si, small intestine; sk, skin; sm, skeletal muscle; sp, spleen; te, testis; th, thymus; tr, trachea.

to G_{M1}). As yet, there is no description of enzymatic activity of ST3Gal I and ST3Gal II for amphibians, cartilaginous fish or agnathans.

Multiple sequence alignments and determination of sub-family amino acid specific positions have demonstrated the existence of intermediate ST3Gal sequences in invertebrates, which cannot be ascribed to any of the two sub-families ST3Gal I or ST3Gal II. This observation further suggested that they might represent ancestral sequences orthologous to the common ancestor of ST3Gal I and ST3Gal II sub-families. Consequently, these sequences identified in the tunicate *C. intestinalis* were named ST3Gal I/II and were proposed to have similar enzymatic activity as ST3Gal I and ST3Gal II enzymes towards Galβ1-3GalNAc.[1] Complete biochemical characterization of the *C. intestinalis* ST3Gal I/II recombinant protein confirmed this hypothesis.[42]

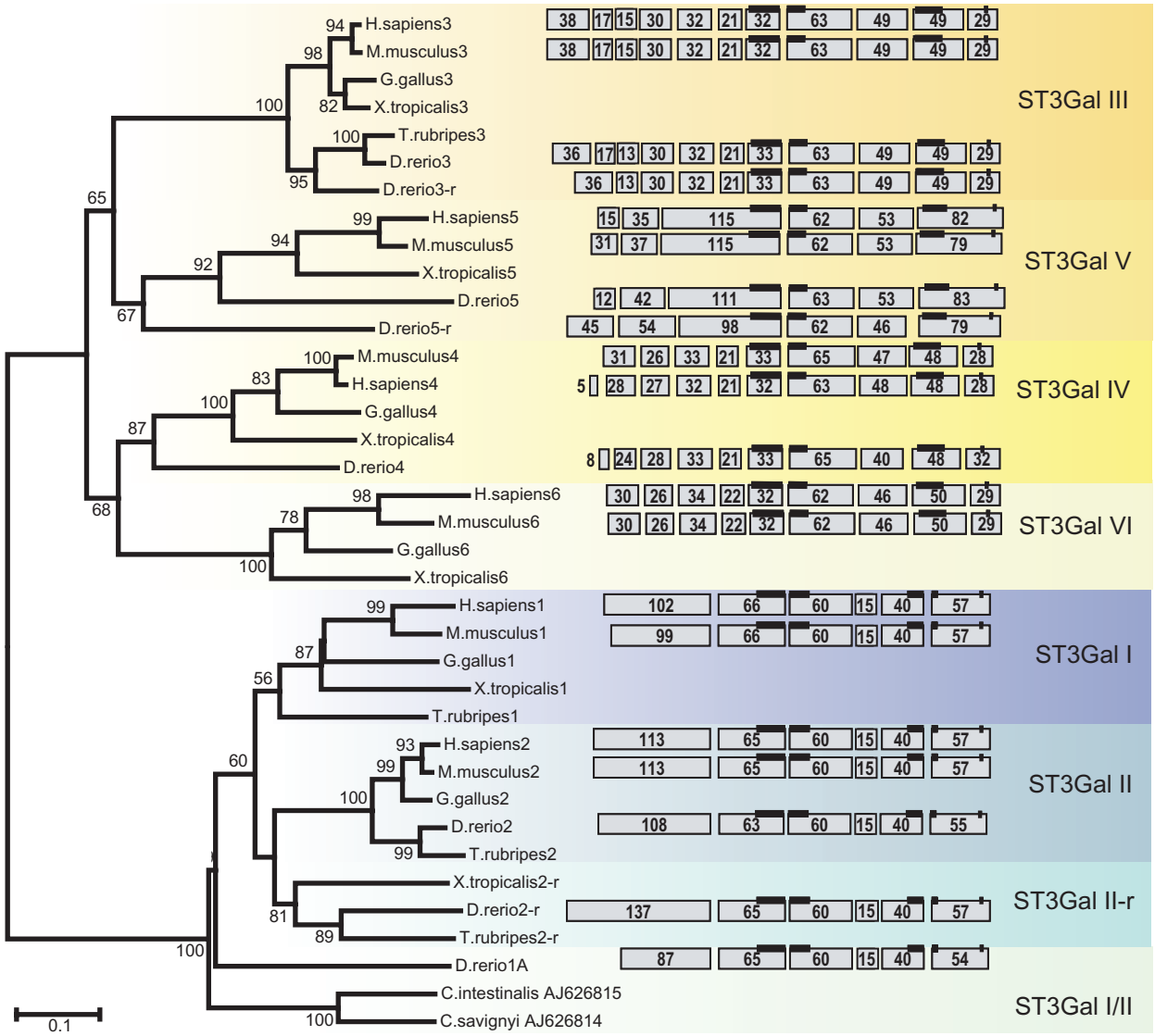

Figure 2. Neighbor-Joining phylogenetic tree of 36 sialyltransferase sequences of the ST3Gal family and juxtaposed, a schematic representation of the human, mouse and zebrafish gene organization. The evolutionary history was inferred using the Neighbor-Joining phylogenetic method.[192] The bootstrap consensus tree inferred from 500 replicates[193] is taken to represent the evolutionary history of the 36 taxa analyzed and the percentage of replicate trees are shown next to the branches. The evolutionary distances were computed using the Poisson correction method.[194] Phylogenetic studies were conducted in MEGA4.0.[195] 36 ST3Gal sequences, 162 informative amino acid positions out of the original 442 positions (36%) were selected with 6 GBLOCKS.[196] The scale bar represents the number of substitutions per site for a unit branch length. The ST3Gal I/ST3Gal II sub-tree was rooted with the tunicate *C. intestinalis* (AJ626815) and *C. savignyi* (AJ626814) sequences as outgroups. These two sequences represent orthologs to the common ancestor ST3Gal I/II present before the split of ST3Gal I and ST3Gal II sub-families. Coding exons are represented by grey rectangles and their size in amino acids is indicated within. Sialylmotif L, S and VS positions are indicated by a dark line above exons. Each sub-family name is indicated on the right side and the accession numbers in GenBank/EMBL of the invertebrate sequences that are not provided in Table 2 are indicated.

As illustrated in Figure 2, the exon/intron organization of the *st3gal* genes of these sub-families is highly conserved from fish to mammals. The members of these two gene sub-families show the same number of exons and intron positions and lengths are conserved. This observation emphasizes their common ancestral origin.

In the second branch of the ST3Gal phylogenetic tree, ST3Gal III, ST3Gal IV and ST3Gal VI show similar enzymatic specificity catalyzing the transfer of sialic acid on the Gal residue of the disaccharide Galβ1-3(4)GlcNAc of glycoprotein or glycolipids.

The human ST3Gal III uses preferentially type I disaccharide Galβ1-3GlcNAc and therefore represents the most probable candidate for the biosynthesis of sialyl-Lewis^a epitope, *in vivo*.[43] The mouse ST3Gal III exhibits high activity towards type I and type II (Galβ1-3/4GlcNAc) disaccharides and very low activity towards type III (Galβ1-3GalNAc) disaccharide.[37] Two ST3Gal III sequences identified in *D. rerio* genome (DreST3Gal III, AJ626821 and DreST3Gal III-r, AJ626820, in Table 2) still await further enzymatic characterization. The human ST3Gal IV and ST3Gal VI use preferentially Galβ1-4GlcNAc disaccharide as acceptor substrate[30,44,45] probably on different substrates in the glycoproteome of the cell.[46] This latter enzyme has been shown to be involved in the sialyl-3-paragloboside (Neu5Acα2-3Galβ1-4GlcNAcβ1-4Galβ1-4Glcβ1-Cer) biosynthesis, a precursor of the sialyl-Le^x on ceramide.[30] Interestingly, no ortholog of the human ST3Gal VI sequence could be identified in fish genomes suggesting that this ST3Gal VI subfamily might have been lost in teleosts. Finally, the recombinant human ST3Gal V[47] and the mouse ST3Gal V[29] were shown to use almost exclusively lactosyl-ceramide (Lac-Cer, Galβ1-4Glcβ1-Cer) as an acceptor substrate leading to the biosynthesis of G_{M3}. This enzyme is also known as the G_{M3}-synthase (Table 1) and it can be released in serum in a soluble form after cleavage of the N-terminus.[48] Noteworthy, the rat brain-purified enzyme shows a broader specificity utilizing both galactosyl-ceramide (Galβ-Cer) and asialoganglioside G_{A2} (GalNAcβ1-4 Galβ1-4Glcβ-Cer) as well as Lac-Cer.[49] Interestingly, Berselli et al reported on the expression of N-terminal 33 amino acids extended isoform of the human ST3Gal V in placenta, which uses Lac-Cer and also Gal-Cer, G_{A1} and G_{A2}.[50] Two related ST3Gal V sequences (DreST3Gal V, AJ619960 and DreST-3Gal V-r, AJ783742 in Table 2) were identified in the zebrafish genome.[1] The enzymatic properties of the corresponding recombinant proteins produced in hamster cultured cells were recently determined. In addition to a G_{M3}-synthase activity, DreST3Gal V-r (AJ783742) and the mouse ST3Gal V enzymes were found to have a G_{M4}-synthase activity leading to the formation of sialylated Gal-Cer, whereas DreST3Gal V (AJ619960) lacked this G_{M4}-synthase activity.[51] Interestingly, orthologs of the zebrafish ST3Gal V-r were identified in various fish genomes, but the gene seems to have disappeared from the tetrapod genomes (Fig. 2) during vertebrate evolution.

For the time being, no ancestral sequence for this second group of ST3Gal genes was reported in lower vertebrates nor in invertebrates, which raises the question of the divergence time of each of the subfamilies described.

St3gal4 and *st3gal6* genes show the same overall gene organization with 9 coding exons and similar exon/intron boundaries. However, *st3gal3* and *st3gal5* genes show 11 and 6 coding exons, respectively and slightly different exon/intron structures (Fig. 2). These additional intervening sequences found in the *st3gal3* genes are localized to the 5′ end of the genes, in the coding region corresponding to the stem region.

2-The β-galactoside α2,6-sialyltransferase family (ST6Gal)

This is the simplest of the sialyltransferase families since it comprises only two sub-families in higher vertebrates.[1] Members of both sub-families ST6Gal I and ST6Gal II were identified in all vertebrates from fish to mammals suggesting that the duplication leading to these two vertebrate sub-families event took place before fish radiation. In addition, their genomic organization appears to be highly conserved in vertebrate genomes (Fig. 3). β-galactoside α2,6-sialyltransferases ST6Gal I and ST6Gal II catalyze the transfer of sialic acid mainly to the terminal Gal residue of type II disaccharide through an α2,6-linkage leading to the formation of the Neu5Acα2-6Galβ1-4 GlcNAc- found on *N*-glycosylproteins and also to a

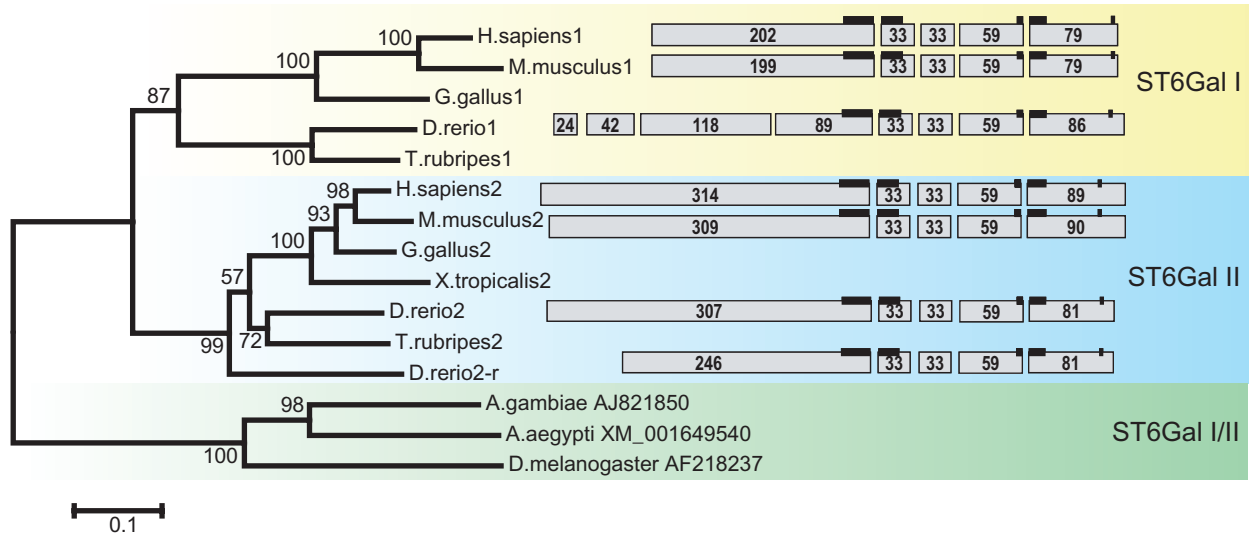

Figure 3. Neighbor-Joining phylogenetic tree of 15 sialyltransferase sequences of the ST6Gal family and, juxtaposed, a schematic representation of the human, mouse and zebrafish gene organization. The evolutionary history was inferred using the Neighbor-Joining phylogenetic method.[192] The bootstrap consensus tree inferred from 500 replicates[193] is taken to represent the evolutionary history of the 15 taxa analyzed and the percentage of replicate trees are shown next to the branches. The evolutionary distances were computed using the Poisson correction method.[194] Phylogenetic studies were conducted in MEGA4.0.[195] 15 ST6Gal sequences, 287 informative amino acid positions out of the original 575 positions (51%) were selected with 7 GBLOCKS.[196] The scale bar represents the number of substitutions per site for a unit branch length. The ST6Gal tree was rooted with the arthropods *D. melanogaster* (AF218237), the *A. aegypti* (XM_001649540) and the *A. gambiae* (AJ821850) sequences as outgroups. These three sequences represent orthologs to the common ancestor ST6Gal I/II present before the split of ST6Gal I and ST6Gal II vertebrate sub-families. Coding exons are represented by grey rectangles and their size in amino acids is indicated within. Sialylmotif L, S and VS positions are indicated by a dark line above exons Each sub-family name is indicated on the right side and the accession numbers in GenBank/EMBL of the insect sequences that are not in table 2 are indicated.

lesser extent, on *O*-glycosylproteins, glycolipids and free oligosaccharides.[52]

Bovine colostrum purified ST6Gal I was found to use also lactose (Galβ1-4Glc)[53] and LacdiNAc (GalNAcβ1-4GlcNAc)[54] as acceptor substrates. ST6Gal I is the first sialyltransferase, which has been cloned and cDNA has been obtained from various higher vertebrate species, including human,[55] rat,[56] mouse,[57] cattle[58] and chicken.[59] The recombinant ST6Gal I enzymes produced have shown a broad substrate specificity towards Gal(NAc)β1-4GlcNAc bearing substrates (Table 1).[58–60] Orthologs of the mammalian *st6gal1* gene could be identified in various fish genomes (Table 2),[1] but yet no biochemical characterization has been reported.

Mammalian recombinant ST6Gal II exhibits *in vitro*, more restricted substrate specificity towards a few Galβ1-4GlcNAc and GalNAcβ1-4GlcNAc bearing glycoconjugates that have not been identified yet.[60–64] Several ST6Gal II-related sequences were identified in lower vertebrate genomes,[1] but have not been yet biochemically characterized.

A unique sialyltransferase gene closely related to the ST6Gal family was identified in several insect genomes (*D. melanogaster, A. gambiae or A. aegyptii*). These sequences branched out from the ST6Gal phylogenetic tree before the separation of vertebrate ST6Gal I and ST6Gal II sub-families. In addition, multiple sequence alignments showed that these invertebrate ST6Gal sequences did not belong to any of the two vertebrate sub-families.[1] This further suggested that these protostomian sialyltransferases, named ST6Gal I/II, were orthologous to the common ancestor of the vertebrate ST6Gal I and ST6Gal II and were subsequently named ST6Gal I/II. ST6Gal I/II from *D. melanogaster* has been enzymatically characterized *in vitro* and demonstrated similar activity towards Gal(NAc)β1-4GlcNAc disaccharide,[65] as compared to vertebrates ST6Gal I and ST6Gal II.

3-The GalNAc α2,6-sialyltransferase family (ST6GalNAc)

Six different GalNAc α2,6-sialyltransferase cDNAs have been identified and cloned from mammalian and avian cells or tissues and enzymatically characterized. These enzymes expressed as recombinant proteins catalyze the transfer of sialic acid residues in α2,6-linkage to the proximal GalNAc residue of

O-glycosylproteins (ST6GalNAc I, ST6GalNAc II, ST6GalNAc IV) or on GalNAc residue of glycolipids like G_{M1b} (ST6GalNAc III, ST6GalNAc V, ST6GalNAc VI). As seen before for the ST3Gal family, the phylogenetic tree of the ST6GalNAc family shows two main branches clearly separating ST6GalNAc I and ST6GalNAc II sub-families on one hand from the remaining four sub-families ST6GalNAc III, ST6GalNAc IV, ST6GalNAc V and ST6GalNAc VI on the other hand (Fig. 4).

Recombinant ST6GalNAc I and ST6GalNAc II proteins have shown similar enzymatic activity *in vitro*, catalyzing the transfer of sialic acid onto GalNAc residues of the type III disaccharide Galβ1-3GalNAc-peptide (α2,3-sialylated or not) of mucin-type glycoproteins[35] (Table 1). Their activity is depending on the peptide moiety.[66–70] ST6GalNAc I is also known as the sialyl-Tn antigen synthase. No ST6GalNAc I sequence has been identified yet in the amphibian genomes nor in the fugu genome, but two ST6GalNAc II sequences were identified in the fugu genome (Table 2). However, none of these are enzymatically characterized yet. A rainbow trout (*Oncorhynchus mykiss*) ST6GalNAc II was characterized from ovaries and the recombinant protein expressed in COS-1 cells was shown to have a slightly different enzymatic specificity since it preferred Neu5Acα2-3Galβ1-3GalNAc-peptide rather than Galβ1-3GalNAc-peptide or GalNAc-peptide.[71]

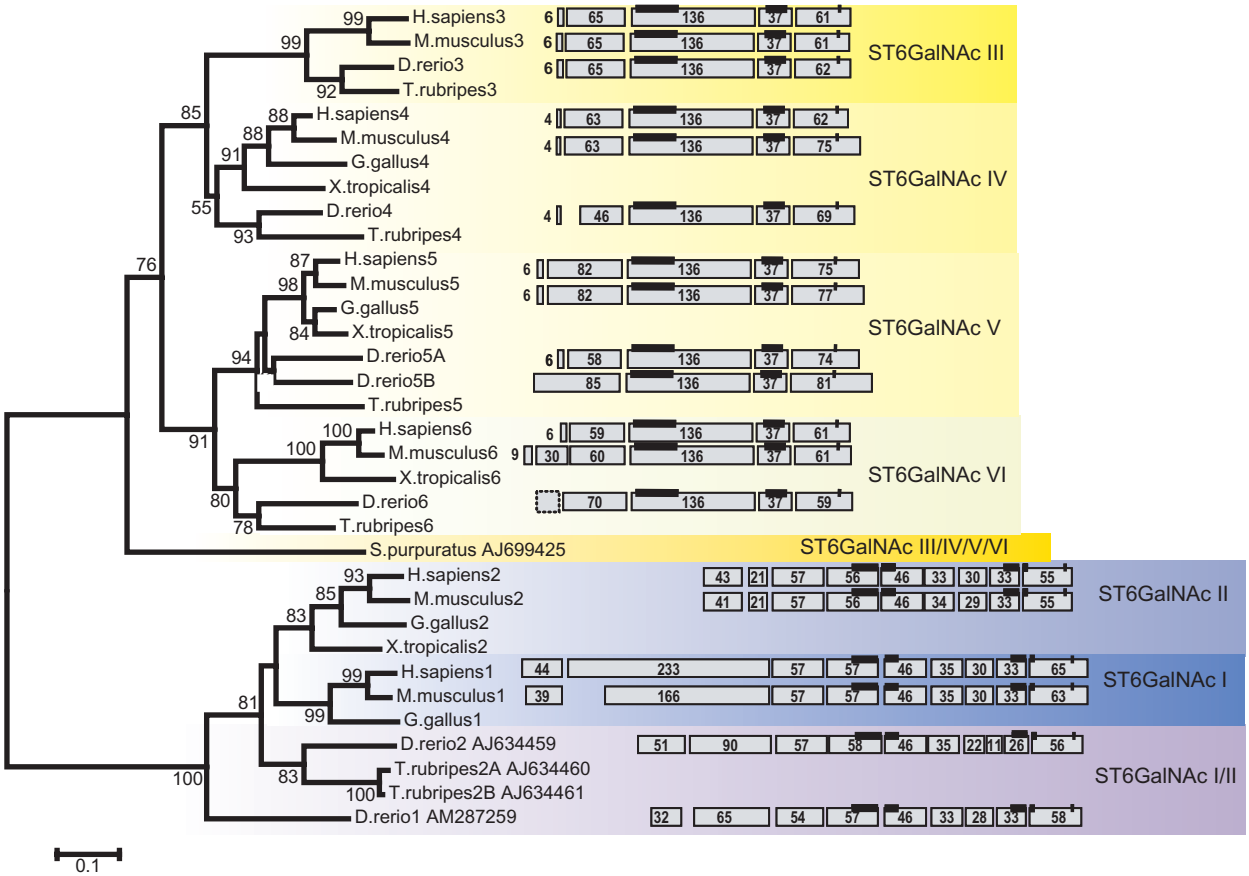

Figure 4. Neighbor-Joining phylogenetic tree of 34 sialyltransferase sequences of the ST6GalNAc family and, juxtaposed, a schematic representation of the human, mouse and zebrafish gene organization. The evolutionary history was inferred using the Neighbor-Joining phylogenetic method.[192] The bootstrap consensus tree inferred from 500 replicates[193] is taken to represent the evolutionary history of the 34 taxa analyzed and the percentage of replicate trees are shown next to the branches. The evolutionary distances were computed using the Poisson correction method.[194] Phylogenetic studies were conducted in MEGA4.0.[195] 34 ST6GalNAc sequences, 105 informative amino acid positions out of the original 698 positions (15%) were selected with 4 GBLOCKS.[196] The scale bar represents the number of substitutions per site for a unit branch length. Four genes from the bony fish *D. rerio* (AM287259 and AJ634459) and *T. rubripes* (AJ634460 and AJ634461) are orthologs to the common ancestor ST6GalNAc I/II present before the split of ST6GalNAc I and ST6GalNAc II sub-families. The gene from the sea urchin *S. purpuratus* AJ699425 is orthologous to the common ancestor ST6GalNAc III/IV/V/VI present before the separation of the four vertebrate sub-families. Coding exons are represented by grey rectangles and their size in amino acids is indicated within. Sialylmotif L, S and VS positions are indicated by a dark line above exons. Each sub-family name is indicated on the right side and the accession numbers in GenBank/EMBL of the sequences that are not in Table 2 are indicated.

On the first branch of the phylogenetic tree (Fig. 4), it was observed that the bony fish ST6GalNAc sequences branched out before the split of ST6GalNAc I and ST6GalNAc II sub-families locating the divergence time of these two sub-families before the emergence of amphibians and after the bony fish emergence. Multiple sequence alignments and determination of sub-family-specific amino acid positions further suggested that these fish sequences cannot be ascribed to any of the two sub-families ST6GalNAc I or ST6GalNAc II and thus might represent ancestral orthologs to the common ancestor.[1] We also speculated that the fish ST6GalNAc I/II could have similar narrow acceptor substrate specificity towards the core GalNAc residue of mucin-type *O*-glycosylproteins.

It is interesting to note that all the genes belonging to this first branch of the GalNAc phylogenetic tree have retained similar genomic organization along vertebrate evolution with 9 coding exons and conserved intron/exon boundaries (Fig. 4).

The second branch of the phylogenetic tree of the ST6GalNAc family contains the remaining four vertebrate sub-families ST6GalNAc III, ST6GalNAc IV, ST6GalNAc V and ST6GalNAc VI.

The mammalian ST6GalNAc III and the human ST6GalNAc IV show the most restricted substrate specificity using exclusively the Neu5Acα2-3Galβ1-3 GalNAc-trisaccharide found on either *O*-glycosylproteins or ganglioside G_{M1b}, which suggests that they do not discriminate between α- and β-linked GalNAc.[72–75] The mouse ST6GalNAc V and ST6GalNAc VI are specific for ganglioside acceptors leading to the biosynthesis of gangliosides of the α-series[76,77] (Table 1). Interestingly, Tsuchida et al have proposed indirect involvement of ST6GalNAc VI in synthesizing disialyl lactotetraosylceramide (Lc4), a precursor of disialyl Le^a. Indeed, the human ST6GalNAc VI and to a lesser extent ST6GalNAc V and ST6GalNAc III catalyze the transfer of a sialic acid residue onto the GlcNAc residue.[78,79] Orthologs of the human *ST6GALNAC* genes have been identified in the zebrafish genome,[1] but still await further biochemical characterization. No orthologous sequence could be identified for *ST6GALNAC* VI in the chicken genome.

ST6GalNAc III/IV/V/VI sequences were identified in the genome of echinoderms (*S. purpuratus*), which are branching out before the occurrence of the duplications at the origin of the four vertebrate sub-families[1,6] and in several invertebrate genomes. These sequences still await further biochemical characterization.

All these ST6GalNAc homologs retain similar genomic organization from fish to mammals (Fig. 4) with a highly conserved gene structure split into 5 coding exons.

4-The α2,8-sialyltransferase family (ST8Sia)

The six members of the mammalian ST8Sia family catalyze the transfer in α2,8-linkages of one to several sialic acids to another sialic acid of glycoproteins or glycolipids. As shown in Figure 5, detailed phylogenetic analysis of this ST8Sia gene family and dating the duplication events has pointed out the existence of four main branches in the phylogenetic tree.[1,5]

The first branch contains four vertebrate subfamilies, ST8Sia I, ST8Sia V, ST8Sia VI and ST8Sia VII that can be viewed as mono-α2,8-sialyltransferases, since they catalyze the transfer of a unique sialic acid residue in α2,8-linkage (Table 1). This branch comprises also several invertebrate sequences forming with the vertebrate sub-families a monophyletic clade.

ST8Sia I with gene representatives from agnathans to mammals was the first to emerge during animal evolution about 600 MYA. The tetrapod ST8Sia I also known as G_{D3}-synthase, has been cloned from various animal species and shows a strict specificity towards G_{M3} resulting in the formation of G_{D3}.[80–84] However, Nara et al reported the molecular cloning of a short isoform of the human ST8Sia I (341 amino acids) using other gangliosides as acceptor substrates yielding to the formation of G_{D3} and also G_{D1c}, G_{T1a} and G_{Q1b} *in vitro*,[85] whereas Nakayama et al reported the molecular cloning of a long isoform of the human ST8Sia I (356 amino acids) with an extended cytoplasmic domain capable of using G_{D3} to form G_{T3} *in vitro*.[86] The two amphibian ST8Sia I cDNAs cloned from *X. laevis* suggested also differential use of in frame start codons and alternative splicing of a unique *st8sia1* gene leading to the production of two protein isoforms with similar enzymatic activity towards G_{M3} yielding G_{D3}.[82] A zebrafish ST8Sia I cDNA has been identified,[1] cloned and the corresponding gene is found to be expressed mainly in the developing brain,[87] but it has not been enzymatically characterized yet.

St8sia5 gene, found from cartilaginous fish (*Callorynchus milii*) to mammals emerged about 563

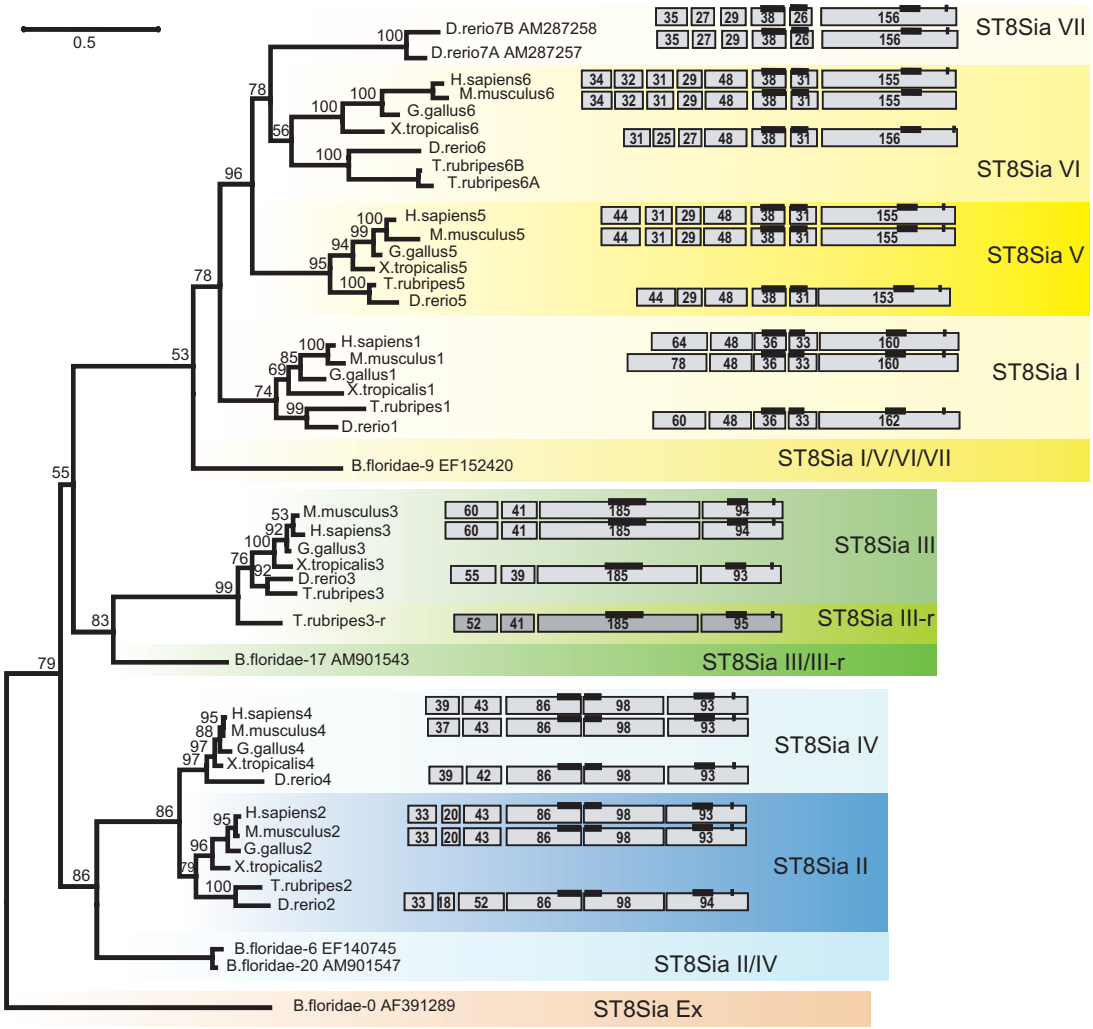

Figure 5. Maximum Likelihood phylogenetic tree of 34 sialyltransferase sequences of the ST8Sia family and, juxtaposed, a schematic representation of the human, mouse and zebrafish gene organization. This ML phylogenetic tree representation is simplified from ref.[1] Figure 3. Phylogenetic studies were conducted with Phyml, version 2.4.4.[197] Initially, 63 ST8Sia sequences, 201 informative amino acid positions out of the original 426 positions (47%) were selected with G-BLOCKS.[196] The bootstrap values were calculated from 500 replicates[193] and values >50% are shown next to the branches. The scale bar represents the number of substitutions per site for a unit branch length. The amphioxus *B. floridae* gene with the GenBank accession number EF140745 is orthologous to the common ancestor ST8Sia II/IV present before the split of the vertebrate ST8Sia II and ST8Sia IV sub-families. One gene from *B. floridae* with the GenBank accession number AM901543 is orthologous to the common ancestor ST8Sia III/III-r present before the separation of the vertebrate ST8Sia III and ST8Sia III-r sub-families and another one with the GenBank accession number EF152420 is orthologous to the common ancestor ST8Sia I/V/VI/VII present before the separation of the vertebrate ST8Sia I, ST8Sia V, ST8Sia VI and ST8Sia VII sub-families. Two zebrafish representatives of this latter new ST8Sia sub-family with the GenBank accession number AM287258 and AM287257 are represented. The tree was rooted with the invertebrate *B. floridae-0* sequence AF391289 as an outgroup. Coding exons are represented by grey rectangles and their size in amino acids is indicated within. Sialylmotif L, S and VS positions are indicated by a dark line above exons. Each sub-family name is indicated on the right side and the accession numbers in GenBank/EMBL of the sequences that are not in table are indicated.

MYA. ST8Sia V also known as the G_{T3}-synthase has been cloned from mammalian sources[88,89] and found to sialylate different gangliosides such as G_{D3}, but also G_{M1b}, G_{D1a} and G_{T1b} (Table 1).[2,90] Although not yet biochemically characterized in fish, DreST8Sia V was cloned from *D. rerio* sources and found to be highly expressed in the developing brain.[87]

Finally, ST8Sia VI found from bony fish to mammals and ST8Sia VII found essentially in Cyprinidae (*D. rerio*) and Salmonidae (*O. mykiss*) and in squamates (*A. carolinensis*) split out around 552 MYA.[5] This latter ST8Sia sub-family has disappeared in mammals and is not yet biochemically characterized. The human ST8Sia VI catalyzes the transfer of a single sialic acid residue mainly on α2,3-sialylated *O*-glycans of glycoproteins leading to the formation of diSia motifs and also to a lesser extent on α2,6-sialylated *O*-glycosylproteins.[91] The mouse ST8Sia VI showed a slightly broader acceptor specificity using more extensively α2,6-sialylated *O*-glycans of the bovine submaxil-

lary mucin, 3′-sialyllactose, 6'-sialyllactose and also gangliosides such as $G_{M3,}$ at least *in vitro* assays.[92]

Nowadays, several ancestral genes that are orthologous to the common ancestor of these four vertebrate sub-families are found in the genome of invertebrates like the amphioxus *B. floridae* or the sea urchin *S. purpuratus* and they are named consequently ST8Sia I/V/VI/VII. They form with the vertebrate sub-family a monophyletic clade and are all proposed to be mono-α2,8-sialyltransferases.

All these genes have kept a similar genomic organization and conserved intron/exon boundaries in their 3′-coding region (Fig. 5). However, each vertebrate sub-family shows a distinct number of coding exons and different intervening sequences are located in the 5′-region of the genes corresponding to the cytoplasmic, transmembrane domain and stem region of the sialyltransferases.

The second branch groups two vertebrate sub-families and invertebrate sequences: ST8Sia III sub-family with genes identified from bony fish to mammals (Table 2) and ST8Sia III-r sub-family with representatives found only in neognathi fish genomes[5] and which are not yet biochemically characterized. ST8Sia III has been cloned essentially from mammalian sources[93,94] and from the zebrafish.[87,95] ST8Sia III catalyzes the transfer of one to several sialic acid residues either on glycoproteins or glycolipids[96] and can be viewed as an oligo-α2,8-sialyltransferase.[97] ST8Sia III is thought to be implicated in the biosynthesis of G_{T3} and disialyl-motifs found on CD-166 (Table 1).[98]

Several ancestral genes orthologous to the common ancestor of the two vertebrate sub-families are found in the genome of *B. floridae* or *S. purpuratus* and consequently they are named ST8Sia III/III-r.

All the genes found in this second branch of the ST8Sia phylogenetic tree have retained a highly similar and simple gene organization with only 4 coding exons and identical intron/exon boundaries (Fig. 5).

The third branch comprises two vertebrate poly-α2, 8-sialyltransferase sub-families ST8Sia II and ST8Sia IV, both with genes identified from bony fish to mammals (Table 2) and also invertebrate sequences. ST8Sia II and ST8Sia IV have been cloned from various mammalian sources including human, mouse and hamster.[99–103] Mammalian ST8Sia II and ST8Sia IV catalyze the transfer several sialic acid residues on other sialylated glycoconjugates. Both enzymes are expressed in the nervous system of most vertebrates where they act mainly on the α2,3-sialylated *N*-glycans of N-CAM[104–106] resulting in an increased neuronal plasticity and migration in embryonic vertebrate embryos.[107] A ST8Sia II cDNA was cloned from *X. laevis* tailbud and assumed to be the major factor in N-CAM polysialylation.[108] Recently, two distinct ST8Sia II cDNAs have been identified in the whole embryo (rtSTXem) and ovary (rtSTXov) of rainbow trout and enzymatic activity of recombinant proteins assayed *in vitro*[109] showed very low level of activity towards the *N*-glycosylprotein N-CAM and the cortical alveolus *O*-glycosylprotein PSGP of fish oocytes for individual enzyme, but enhanced activity when assayed in conjunction with the rainbow trout ST8Sia IV (rtPST). ST8Sia II and ST8Sia IV were also cloned from the zebrafish, however N-CAM polysialylation was obtained *in vitro*, only with recombinant ST8Sia II protein.[110]

Ancestral genes that are orthologous to the common ancestor of these two vertebrate sub-families are found in the genome of the lamprey (*Petromyzon marinus*) and in the invertebrate genomes of amphioxus (*B. floridae*) or sea urchin (*S. purpuratus*) and they are named ST8Sia II/IV. None of these has been enzymatically characterized yet.[5]

ST8Sia II and ST8Sia IV genes found in this third group of ST8Sia have kept highly similar gene organization across vertebrate species showing 5 and 6 coding exons, respectively.

As illustrated in Figure 5, the last branch, which is the most external of the ST8Sia phylogenetic tree, comprises only invertebrate α2,8-sialyltransferase related genes that have been named ST8Sia EX for external. No ortholog could be identified in vertebrate genomes, which suggests the complete loss of these ancestral genes early in vertebrate evolution.

In summary, biochemical characterization of mammalian sialyltransferases has pointed out their similar substrate specificity and functional redundancy. Identification of sialyltransferase-related sequences, molecular biology and phylogenetic analysis suggested that the evolutionary origin of sialyltransferase genes predates deuterostome/protostome divergence. Orthologs to the twenty known mammalian sialyltransferase sub-families were identified in all the vertebrate genomes screened up to date. In addition, several new sub-families could be identified in particular in lower vertebrates,

although they remain to be enzymatically characterized. Four ancestral sialyltransferase gene families have been identified in various invertebrate genomes enabling to trace back the origin of these gene families.[1] These ancestral sialyltransferase families comprise sequences, which could not be ascribed to any of the vertebrate sialyltransferase sub-families suggesting that they are orthologous to the LCA of the different vertebrate sialyltransferase sub-families. The deuterostome and the protostome lineages share a unique ancestral ST6Gal family in their LCA. Four groups belonging to the ancestral ST8Sia family were identified in the LCA of echinoderm and chordate lineages.[5] Similarly, the first precursor gene of the ST3Gal family was identified in urochordates (*C. intestinalis* and *C. savignyi*) suggesting the occurrence of a unique ancestral ST3Gal family in the LCA of urochordates and vertebrates. Finally, a unique ancestral ST6GalNAc family was found in the LCA of echinoderms and vertebrates.[1] However, all these studies still appear insufficient to definitively resolve all the sialyltransferase relationships and further phylogenetic studies of each individual sialyltransferase family remain necessary to establish their evolutionary history.

Phylogenomics of ST8Sia as a Way to Gain Insights into their Evolutionary History

The use of the vast amount of genomic data now available in the public databases coupled to functional genetic experiments offers unprecedented possibilities to trace back the origins of the sialylation biosynthetic machinery in vertebrates. To further test historical orthologous relationships among sialyltransferase homologs of the ST8Sia family, genomic distribution and comparative synteny analysis extended to the whole chordate genomes available in public databases were recently undertaken.[111,112]

Essentially two genetic mechanisms may explain sialyltransferases evolution and the emergence of the twenty sialyltransferase sub-families described in mammals. One is gene duplication, the other is gene loss. The simplest case of gene duplication is tandem duplication of individual genes (TDT). Several examples have been described in fish genomes as for ST8Sia VIA and ST8Sia VIB in *T. rubripes*,[5] ST8Sia VIIA and ST8Sia VIIB in *D. rerio*[87] or ST8Sia IIA and ST8Sia IIB in *O. mykiss*.[109] Most of the time, the two copies of these duplicated genes are positioned immediately adjacent to one another on the same chromosome. Phylogenetic and phylogenomic studies illustrated that the initial expansion and subsequent divergence of ST8Sia I, ST8Sia III and ST8Sia V were the consequence of ancient gene duplications and translocations in the ancestral invertebrate genome, long before the emergence of vertebrates.[5] Other genetic processes, more global, should be emphasized. The first is whole genome duplication rounds (WGDR), suggested in the 1970s by Susumu Ohno, who first hypothesized that polyploidy has been an important factor in the evolution of vertebrates.[113] Two rounds of genome duplication (2R hypothesis) would have taken place early in the vertebrate lineage around 560 MYA and 520 MYA[114] as testified by the identification of paralogous regions, which are remnant of these regional chromosomal duplications in vertebrate genomes. These genetic events account for the emergence of several new sialyltransferase subfamilies early in vertebrate evolution as demonstrated for ST8Sia II and ST8Sia IV sub-families and ST8Sia VI and ST8Sia VII sub-families probably resulting from WGDR1 that took place around 550 MYA, before the divergence between jawed and jawless vertebrates.[5,114] The other process is a segmental duplication, in which only a portion of a given chromosome is duplicated, but no event involving a sialyltransferase gene was yet detected. Otherwise, some lineage-specific duplication events are difficult to interpret, in reason of an incomplete genome assembly. This is the case for example, of the four ST8Sia II/IV sequences described in the cephalochordate *B. floridae* or the eight ST8Sia I/V/VI/VII sequences identified in the urochordate *S. purpuratus*.[5] All these observations raises the question of the fate of duplicated vertebrate sialyltransferases non-, sub- or neofunctionalization[115] that is respectively, loss of one copy sharing of the ancestral function between the two duplicates or acquisition of a new function by one of the copies, and its link with the origin of evolutionary novelties.

Gene loss also contributed significantly to the emergence of divergent animal lineages.[116,117] Although uncertainties in the completion or assembly of sequenced chordate genomes remain, gene loss can be revealed by comparison of distant evolutionary organisms such as *C. intestinalis* in the urochordate phylum, which has experienced gene loss of almost all sialyltransferase ancestral gene families with the notable exception of the ST3Gal I/II family.[1]

Similarly, the ST8Sia EX sequences described in the non-vertebrate marine deutorostomia *B. floridae* and *S. purpuratus* have completely disappeared from vertebrate genomes.[5] Sialyltransferase sub-family loss was also described in various vertebrate genomes such as the ST8Sia IV sub-family, which has disappeared from neognathi fish like *T. nigroviridis*, *T. rubripes* or *O. latipes,*[5] ST8Sia III-r sub-family, which has disappeared from the Cyprinidae, Salmonidae and tetrapods or the ST8Sia VII sub-family, which was maintained only in the fishes *D. rerio* and *O. mykiss* and in the green lizard *A. carolinensis*. These observations further suggest an impact on the functional fate of surviving paralogs, which remains to be established.

Thus, combining data from our phylogenetic and phylogenomic approaches, we have reconstructed a birth and death model of divergent evolution of sialyltransferase genes, which helps to explain the persistence of redundant genes for hundred million years. This scenario not only highlights the ancestral lineage between vertebrate sialyltransferases, which has shaped the sialylated glycan repertoire known in the living organisms and also points to biologically significant function for sialoglycoconjugates conserved during evolution of vertebrates.

Control of Sialyltransferases Gene Expression and Pattern of Expression in Vertebrates

In Metazoans, sialyltransferase genes are polyexonic, their coding region is divided into several coding exons[4–10] and different length of introns spanning

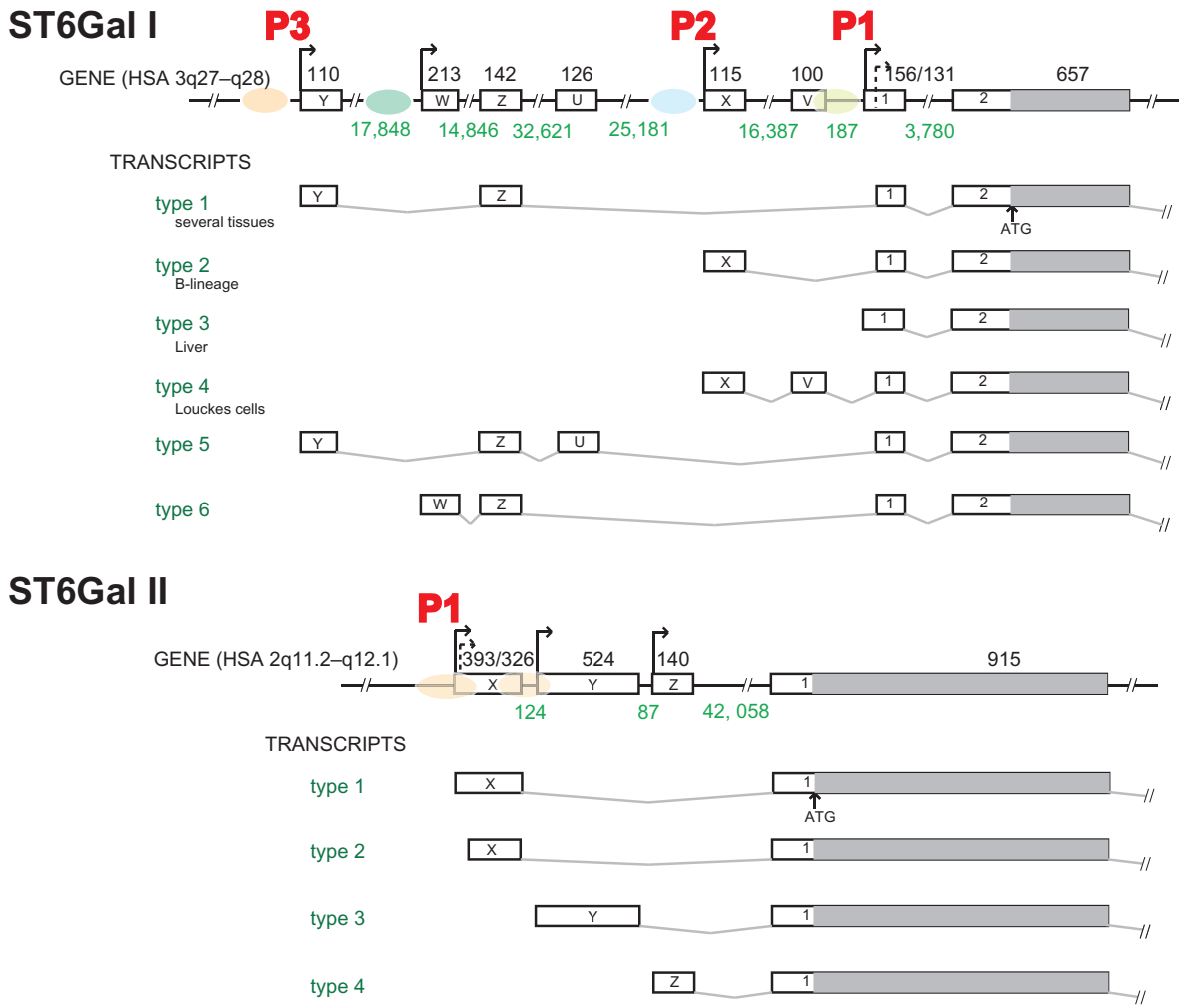

Figure 6. Schematic representation of the 5′-end of human ST6Gal genes and the major transcripts expressed. The boxes and lines connecting boxes indicate exons and introns respectively. Grey rectangles represent coding regions whereas open rectangles represent 5′-UT exon. Promoter regions are symbolized by colored ovals and TSS by bent arrows. Multiplicity of TSS is indicated by dashed bent arrows. Intron lengths are indicated in green below the gene schematic and the exon length in black above. The translation start codon AUG is indicated. Not drawn to scale.

over up to 100 kb. As reported above, their gene organization within the coding sequence is conserved during vertebrate evolution (Figs. 2–5). However, despite gene organization and good primary protein sequence conservation among vertebrate sialyltransferase orthologs, marked differences in tissue expression profiles have been noticed among vertebrate species as reported for bovine and human ST6Gal II[61,62,64] or for the rodent ST6Gal I and ST3Gal I.[118] In addition, spatio-temporal expression of sialyltransferase genes evidenced through Northern blot, RT-PCR and microarray analysis and whole mount *in situ* hybridization show a strong positive correlation between mRNA expression levels and enzyme activity levels, suggesting that cell surface sialic acid levels mainly depend on the mRNA level of sialyltransferase genes and their transcriptional regulation.[119,120] In order to elucidate the molecular basis of vertebrate sialyltransferase gene regulated expression, the structure and gene organization of human and mouse sialyltransferase genes has been widely analyzed and a few of their key *cis*-regulatory sequences delineated.[121–123] As reported before for other terminal glycosyltransferases like β1, 4-GalT1,[124,125] N-acetylglucosaminyltransferases III[126] and V[127] or FUT1,[128] sialyltransferase genes appear to be differentially expressed whereas glycosyltransferases involved in the synthesis of the core region of glycans are constitutively expressed.

Interestingly, analysis of the mammalian *ST6GAL1*, *ST6GAL2*, *ST3GAL2*, *ST3GAL4*, *ST3GAL5* and *ST3GAL6* genes revealed the existence of several 5′-untranslated (5′-UT) exons and heterogeneous transcriptional start sites (TSS) leading to several mRNA isoforms with potentially different translation rate or mRNA stability. Genomic sequence analysis of the 5′-flancking region of these human sialyltransferase genes led to the identification of various alternative promoter regions spread across hundred kb of linear genomic sequences, as illustrated in Figures 6–9. A few of these genomic promoter regions

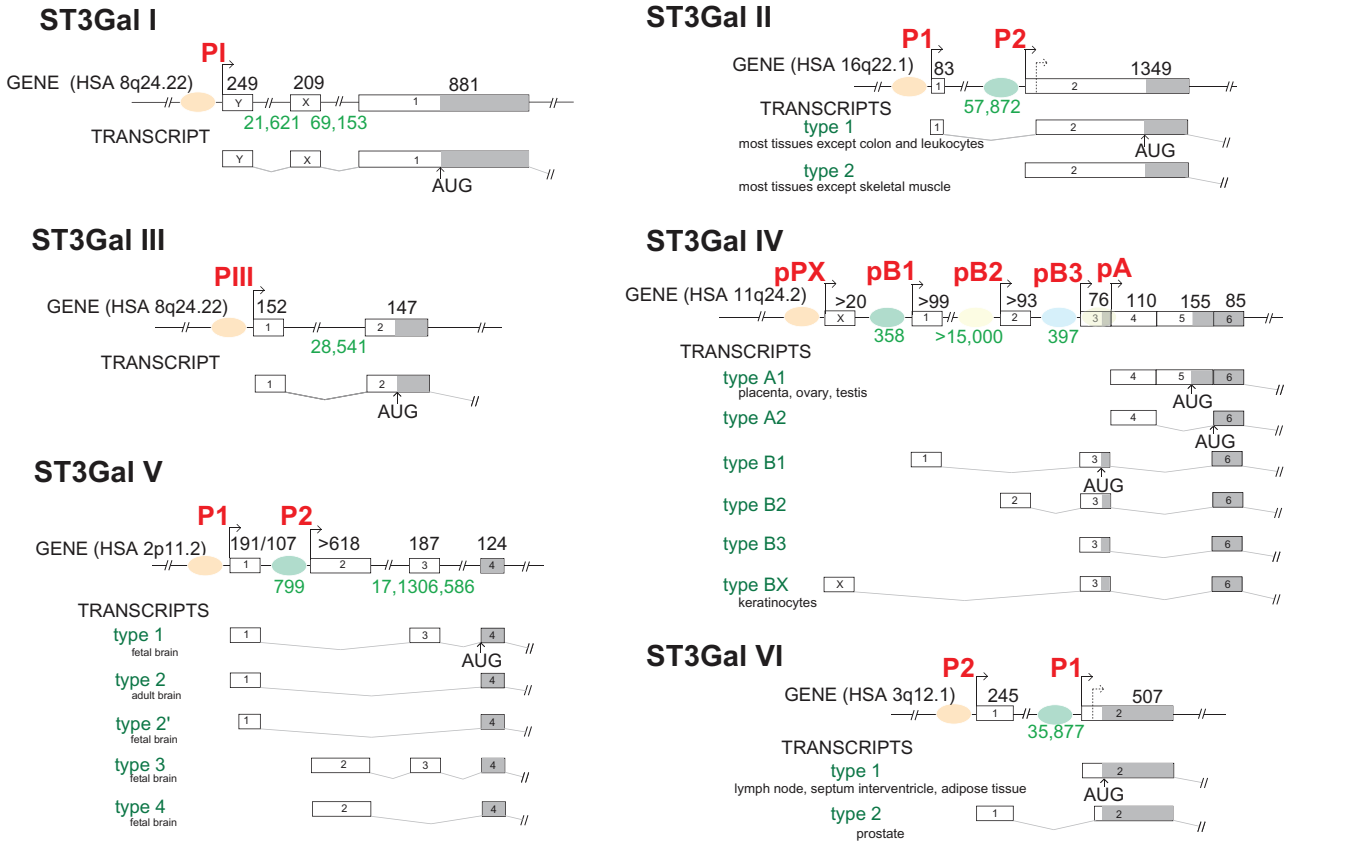

Figure 7. Schematic representation of the 5′-end of human ST3Gal genes and the major transcripts expressed. The boxes and lines connecting boxes indicate exons and introns respectively. Grey rectangles represent coding regions whereas open rectangles represent 5′-UT exon. Promoter regions are symbolized by colored ovals and TSS by bent arrows. Multiplicity of TSS is indicated by dashed bent arrows. Intron lengths are indicated in green below the gene schematic and the exon length in black above boxes. The translation start codon AUG is indicated. Not drawn to scale.

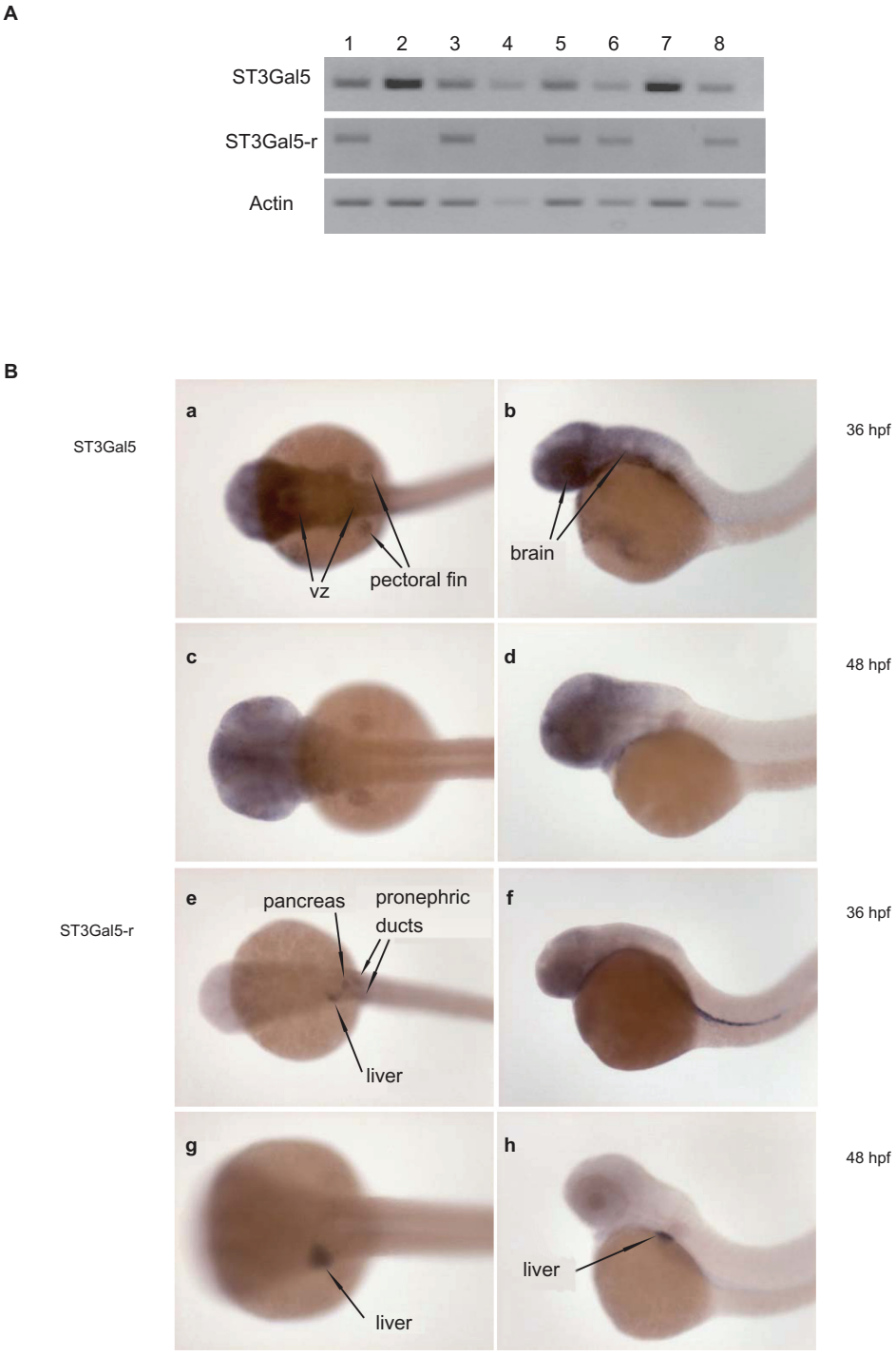

Figure 8. Expression pattern of zebrafish *st3gal5* and *st3gal5-r* genes **A**) in various adult tissues, by RT-PCR and **B**) during embryonic development, by whole mount *in situ* hybridization. **A**) Total RNA was isolated from 36 hpf embryos and various adult tissues of zebrafish and RT-PCR was performed, as previously described.[1] Lane 1: 36 hpf; lane 2: eggs; lane 3: liver; lane 4: brain; lane 5: intestine; lane 6: heart; lane 7: ovaries; lane 8: kidney. The zebrafish β-actin gene was used as a positive control of PCR amplification leading to the amplification of a 378 bp fragment, as described before.[1] Forward 5′-TGGCTGTGATATTCAAGTC and reverse 5′-TGTAGGGATGTTCTGGTC primers were used for the amplification of a zebrafish *st3gal5* cDNA (AJ619960) of 236 bp and forward 5′-TCCCACATACAGGAATACG and reverse 5′-CAGTCTCGGCACTAACATC primers were used for the amplification of a zebrafish *st3gal5-r* cDNA (AJ783742) of 424 bp. **B**) Expression of the *st3gal5* and *st3gal5-r* genes during embryonic development of the zebrafish was followed by whole mount *in situ* hybridization. Molecular cloning of *st3gal5* full-length open reading frame was achieved using the forward 5′-CGAGGACTTGGTGTTGAGG and the reverse 5′-GTGCTGCATTAGCAGTGTTTC primer pair for *st3gal5* and the forward 5′-CAGTAGGAGTATTT GGGAAGT and the reverse 5′-TGAGAATGGTCGAAACAGC primer pair for *st3gal5-r*. RNA probes were obtained as previously described[87] and ISH was carried out as described.[198–200] Panels a, c, e and g show dorsal views and panels b, d, f and h show lateral views (anterior is to the left). At 36 hpf (panels a, b, e and f) and at 48 hpf (panels c, d, g and h), *st3gal5* and *st3gal5-r* shows very distinct territories of expression: *st3gal5* gene is expressed mainly in the developing brain whereas *st3gal5-r* gene is expressed mainly in the developing digestive tract, in the liver, pancreas and anterior part of pronephric ducts.

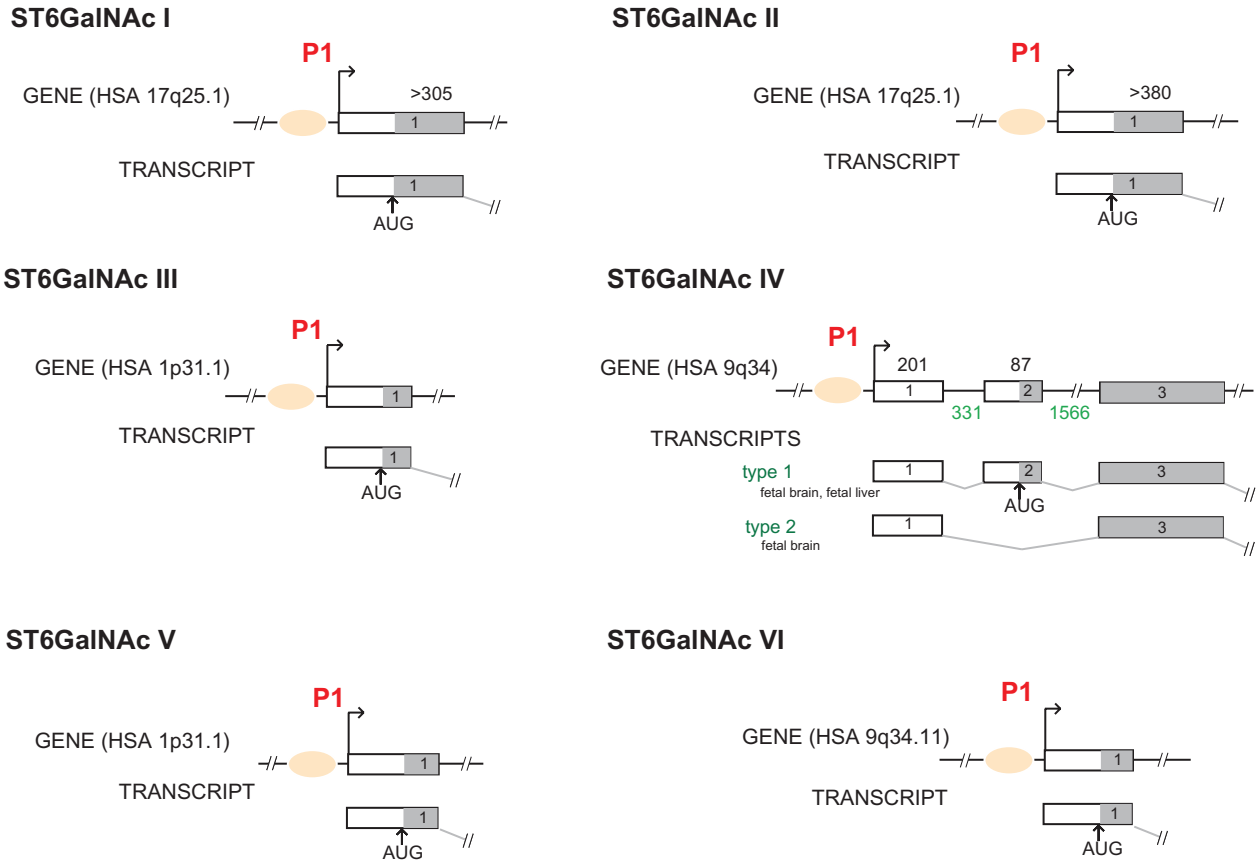

Figure 9. Schematic representation of the 5′-end of human ST6GalNAc genes and the major transcripts expressed. The boxes and lines connecting boxes indicate exons and introns respectively. Grey rectangles represent coding regions whereas open rectangles represent 5′-UT exon. Promoter regions are symbolized by colored ovals and TSS by bent arrows. Multiplicity of TSS is indicated by dashed bent arrows. Intron lengths are indicated in green below the gene schematic and the exon length in black above. The translation start codon AUG is indicated. Not drawn to scale.

were shown to govern differentially modulated sialyltransferase expression in specific cell types and/or at different time during embryogenesis.

On the other hand, human *ST3GAL1*, *ST3GAL3*, *ST6GALNAC1*, *ST6GALNAC2*, *ST6GALNAC3*, *ST6GALNAC4*, *ST6GALNAC5*, *ST6GALNAC6*, *ST8SIA1*, *ST8SIA2*, *ST8SIA3*, *ST8SIA4*, *ST8SIA5* and *ST8SIA6* genes described to date display a unique promoter region and ubiquitous transcription factors binding sites driving the synthesis of unique mRNA isoform or alternatively internally spliced RNA. They have a cell type specific and developmentally regulated expression pattern relying on the use of cell type specific transacting factors.

St6gal genes

The widely studied rat *st6gal 1* gene shows two upstream 5′-UT exons named −1 and 0 (reviewed in).[52] The use of two physically distinct promoters give rise to two major types of transcripts of 4.7 kb and of 4.3 kb expressed in many rat tissues and in liver,[129,130] respectively. In addition, three minor transcripts with coding exons IV–VI are found in kidney, but their functional significance remains obscure.[131,132] The mouse *st6gal1* gene is also widely expressed in almost all tissues screened[63] (Table 2). Its differential expression is also governed by the use of multiple alternative promoter regions[133] and Dalziel et al described the use of alternative 5′-UT exons L and Q specifically in lactating mammary gland driven by a lactogenic promoter P4.[134] The human *ST6GAL1* gene is highly expressed in many different human tissues (Table 2). The use of alternative promoters named P1, P2 and P3 illustrated in Figure 6[135–141] leads to the expression of several transcripts differing in their 5′-UT. One basal transcript (type 1) originally cloned from placenta,[55] is expressed in several different tissues and cells.[142] It displays two 5′-UT exons Y and Z, homologous to

the rat 5′-UT exons −1 and 0.[139] A second transcript (type 2) is described in mature B-cells. It lacks 5′-UT exons Y and Z but shows another 5′-UT exon X[139] with no rat counterpart described yet. A third transcript (type 3) described in the human hepatocarcinoma cells HepG2[135] shows no additional 5′-UT exon and an upstream extension of exon 1. It represents the major hepatic form of the human *ST6GAL1* gene. Three additional 5′-UT exons named U, V and W were described in the ST6Gal I transcripts (type 4–6) found in the lymphoblastoïd Louckes cells.[137] The bovine and chicken *st6gal1* genes show similar genomic structure in their 5′-UT region and transcript synthesis even though their tissue distribution differs markedly.[58,59] The fish *st6gal1* gene identified in *D. rerio*[1] shows a more restricted profile of expression in adult kidney, intestine, brain and ovary and in the developing embryonic brain (unpublished data).

The human *ST6GAL2* as well as the mouse *st6gal2* genes show a very restricted expression profile in adult and embryonic brain, although they are also expressed to a lower extent in thyroid gland, testis, small intestine and colon.[61–63] Four transcripts (type 1–4) arising from the brain specific expression of the human *ST6GAL2* gene were recently described and shown to differ in their 5′-UTR through the use of three alternative 5′-UT exons named X, Y and Z (Fig. 6).[143] These data further suggested the use of a large genomic promoter region for the specific expression of the *ST6Gal2* gene in brain. Bovine *st6gal2* gene encodes three transcripts primarily expressed in brain, although transcript a2 is also detected in other bovine tissues such as lung. They result from the alternative use of one 5′-UT exon named E0 and various TSS located in exon E0 and E1. These data further suggest the use of two distinct promoter regions P_A and P_B 27 kb apart from each other.[64] The zebrafish *st6gal2* genes identified show a less restricted tissue expression profile in intestine, muscle, brain and ovary and also during embryonic brain development (unpublished data). This observation raises the question of a differential gene expression pattern evolution in vertebrates.

The *D. melanogaster st6gal1/2* gene is expressed primarily during late embryogenesis and larval development of the central nervous system of the insect and also to a lesser extent in the adult head.[144] Interestingly, a splice variant of the *D. melanogaster* gene with an additional 69 nt sequence corresponding to the first intron (GeneBank accession number AF397532),[145] was found to be highly expressed in the embryonic salivary glands.[146] No further report was found concerning additional 5′-UT exon nor genomic regulatory region of the *DSiaT* gene.

St3gal genes

The human *ST3GAL1* gene shows a cell type specific and developmentally regulated expression in various adult and fetal tissues with high levels of mRNA expression in placenta, kidney and skeletal muscle.[34,147] As illustrated in Figure 7, it shows two upstream 5′-UT exons named X and Z located at huge genomic distance from the coding exons generating a unique transcript in colon cancer cells. Genomic regulatory sequences display ubiquitous transcriptional factors binding sites and no cell type specific transcriptional factors binding sites.[148] The mouse *st3gal1* gene is highly expressed in submaxillary gland and to a lesser extent in lung and heart (Table 2).[39] No promoter studies concerning the mouse *st3gal1* gene have been reported yet.

Similarly to the human *ST6GAL1* and *ST6GAL2* genes, the human *ST3GAL2* gene drives the expression of at least two transcripts in most human tissues, except colon and peripheral blood leukocytes for type 1 transcripts and except skeletal muscle for type 2 transcripts through alternative use of two distant promoter regions named P1 and P2 and two TSS, as shown in Figure 7.[149] Type 1 transcript shows an additional 5′-UT exon named exon 1.

The human *ST3GAL3* gene exhibits a cell type specific and developmentally regulated profile of expression and is abundantly expressed in skeletal muscle and to a lesser extent in various adult and fetal tissues.[34] Using RT-PCR followed by laser-induced fluorescent capillary electrophoresis, Grahn et al detected more than twenty-five distinct transcripts from peripheral blood leukocytes, fetal and adult brain produced by alternative splicing within the coding region.[150] The human *ST3GAL3* gene shows a unique 5′-UT exon, named exon 1 located more than 28 kb upstream exon 2 and probably a unique TSS (Fig. 7). So far, there is no evidence for alternative splicing in the 5′-UT region and a unique promoter region has been delineated.[151]

The human *ST3GAL4* gene, which is predominantly expressed in placenta, is probably the most

thoroughly studied for transcriptional regulation.[152–155] It comprises four 5′-UT exons and shows at least five TSS. Using a RT-PCR capillary electrophoresis method, Grahn et al have detected nine human mRNAs expressed in peripheral blood leukocytes.[156] The use of at least five alternative promoter regions (pBx, pB1, pB2, pB3 and pA) and the occurrence of alternative splicing in the 5′-UTR of the gene lead to the synthesis of six transcripts of type A1 expressed in placenta, testis and ovary, type A2, type B1, type B2 and type B3 expressed in several cell types and type Bx specifically found in germ cells.[149] It is interesting to note the existence of three start codons in three different and contiguous exons (exon 3, exon 5 and exon 6) generating proteins with different N-terminus cytoplasmic ends schematized in Figure 7.

The human *ST3GAL5* gene is expressed in numerous human tissues with higher expression levels detected in adult brain and also in testis, placenta, lung and spleen.[47,157] Analysis of the human *ST3GAL5* gene organization shows three 5′-UT exons named exon 1, exon 2 and exon 3 spread over 25 kb of linear genomic sequence.[158,159] It encodes multiple mRNAs alternatively spliced in their 5′-UTR.[13] Four transcripts that differ in their 5′-UTR (type 1, type 2, type 3 and type 4) are found in human fetal brain and an unique developmentally regulated mRNA isoform (type 2) is found in adult brain,[158–160] suggesting the use of at least two promoter regions P1 and P2.[123,158,160–162] An additional type 5 mRNA variant of the human *ST3GAL5* gene is expressed in placenta and in undifferentiated HL60 cells.[50] This transcript results from the use of an alternative TSS located in an additional 5′ exon containing another in frame start codon. These data further suggest the use of an alternative promoter named P3 in Figure 7. The mouse *st3gal5* gene is expressed in various adult tissues, to high level in brain, testis, heart, spleen and liver and to lower levels in kidney and thymus.[13,29] Widespread distribution of the mouse *st3gal5* gene was observed in several populations of neurons throughout the brain.[163,164] At least three mRNA variants containing different 5′-UT regions were identified in mouse tissues suggesting the use of alternative promoter regions. Two distinct genes *st3gal5* (AJ619960) and *st3gal5-r* (AJ783742) orthologous to the human *ST3GAL5* gene were identified in the zebrafish genome[1] and found to have a very distinct profile of expression in the zebrafish adult tissues and during zebrafish embryonic development as evidenced by RT-PCR and whole mount ISH (Fig. 8). *St3gal5* gene is mainly expressed in the developing and adult brain as well as in several adult tissues whereas *st3gal5-r* gene is expressed mainly in the developing and adult digestive tract, but not in the brain.

The human *ST3GAL6* gene has been found to be abundantly expressed in heart and liver. Analysis of the genomic organization of the human *ST3GAL6* gene shows the existence of a unique 5′-UT exon and transcription of this gene give rise to two distinct mRNAs (type 1 and type 2) in HepG2 cells and in prostate.[165] The transcriptional regulation of this gene depends on the use of two alternative promoters P1 and P2 (Fig. 7). Interestingly, it has been proposed recently that an epigenetic change by DNA hypermethylation of the human *ST3GAL6* gene *cis*-regulatory region could cause cancer-associated changes in the expression of carbohydrate determinants.[11]

St6galnac genes

The human *ST6GALNAC1* gene shows a tissue restricted expression pattern in submaxillary gland and lactating mammary gland (Table 2). It has no additional 5′-UT exon and exhibits a unique start codon in exon 1 (Fig. 9). A unique mRNA has been described in various human tissues suggesting the use of a single promoter region.[66] Interestingly, the mouse *st6galnac1* gene shows two identical ATG containing exons located more than 10 kb apart from each other and leading to the expression of the same protein. However, two distinct tissue-specific promoters govern the expression of a first transcript containing Exon 1b specifically in submaxillary glands and of a second transcript containing Exon 1a specifically in colon and mammary glands.[166]

The human *ST6GALNAC2* gene is more widely expressed in a tissue specific manner in heart, skeletal muscle and kidney.[70] Genomic sequence analysis suggested the existence of a unique promoter region (Fig. 9). Polymorphisms detected in the human *ST6GALNAC2* promoter were shown to influence transcriptional activity and contribute to genetic susceptibility to IgA nephropathy constituting the first example of regulatory polymorphism for a human sialyltransferase.[167] The mouse *st6galnac2* gene is highly and developmentally regulated in mammary

gland and testis like its human counterpart and shows no additional 5′-UT exon and a unique TSS.[168]

The expression pattern of the human *ST6GALNAC3* gene is restricted to a few human adult tissues mainly brain and kidney and differs from that observed in mouse and rat adult and embryonic tissues.[74,75,169] The mouse *st6galnac3* gene is highly expressed during brain development and shows a unique TSS located 71 bp upstream the start codon, suggesting a unique promoter region driving the expression the gene.[169]

The expression of the mouse *st6galnac4* gene is also developmentally regulated in brain[73,169] with one major TSS located 228 bp upstream the start codon and two extra TSS around 316 and 259 bp upstream ATG, giving rise to three differentially sized mRNAs. The human *ST6GALNAC4* gene shows constitutive expression in almost all human tissues and cancer cell lines examined and its genomic organization has been partially reported.[170] Alternatively spliced isoforms in the 5′-UTR were evidenced in fetal liver and fetal brain and its unique promoter region characterized.[171,172]

The mouse and the human *st6galnac5* and *st6galnac6* genes show a restricted expression pattern in adult brain, but their transcriptional regulation remain unknown.[76,77]

St8sia genes

As summarized in Table 2, the *st8sia1* gene shows a developmentally regulated expression in chick brain and retina,[173] in the amphibian embryo,[81,82] in the developing central nervous system of the zebrafish embryo[87] and in adult brain,[5] in the rat brain, spleen and testis[83] and in the mouse brain and retina.[174] Regulation of the expression of the human *ST8SIA1* gene has been studied in various human tumor cell lines, in the normal and malignant T-lymphocytes.[175,176] Functional analysis of the melanoma-, Fas induced Jurkat T cells- and glioblastoma-specific promoter region of the G_{D3}-synthase gene pointed towards similar conclusions: no additional 5′-UT exon was found despite multiple TSSs identified in human cells and the alternative use of two in-frame start codons leading to a 356 amino acids protein[86,177] or to a 341 amino acids protein.[85,175] As illustrated in Figure 10, a unique human *ST8SIA1* gene promoter region, not constitutively active was identified that drives the expression of multiple mRNAs during development and differentiation.[178–180] Interestingly, no similarity was found between rodent and human *st8sia1* promoter sequences and no common transcription factor binding sequences among these animal species could be identified.[90]

The mouse *st8sia2* gene is expressed mainly in the developing brain and also in fetal kidney and to a lesser extent in the adult heart, thymus and brain (Table 2). The mouse gene shows a single TSS with no additional 5′-UT exon and a single promoter region drives the expression of a unique mRNA.[181] Several recent studies have shown that the zebrafish *st8sia2* gene has conserved the early developmental stage specific expression described among vertebrates.[87,110,182,183]

The mouse *st8sia3* gene is expressed in adult brain and testis and during embryonic development (Table 2). It has probably the simplest sialyltransferase gene organization with only four coding and 5′-UT exons spanning 8 kb in the mouse genome.[184] Developmentally regulated expression is observed in brain and testis from a unique promoter region leading to the expression of three transcripts that differ in their 3′ polyadenylation sites. The human *ST8SIA3* gene shows similar organization and pattern of expression in brain and fetal brain.[93] A unique promoter region was shown to control the human gene expression from a unique TSS in glioblastoma cells.[185] The fish *st8sia3* gene identified in *D. rerio* genome[1] was found to be expressed in the embryonic developing brain,[87] in the developing myotomes[95] and in the adult brain.[5]

As summarized in Table 2, the mouse *st8sia4* gene is expressed mainly in the developing brain and also in various adult tissues such as spleen and bone marrow and lymphocytes, but not in testis.[119] The mouse gene organization and pattern of expression are fairly similar to those of the mouse *st8sia2* gene. It shows no additional upstream 5′-UT exons, a single TSS and a unique minimal promoter region has been delineated.[186,187] However, no obvious homology could be found between the promoter regions of these two duplicated genes suggesting that an alternative mechanism of subfunctionalization has taken place, by which both copies were prevented from nonfunctionalization resulting in a differential gene expression pattern. As its mammalian counterpart, the zebrafish *st8sia4* gene displays similar gene expression pattern in the developing central nervous system.[87,110,182,183]

The mouse *st8sia5* gene shows a very restricted expression profile in adult brain and in the developing

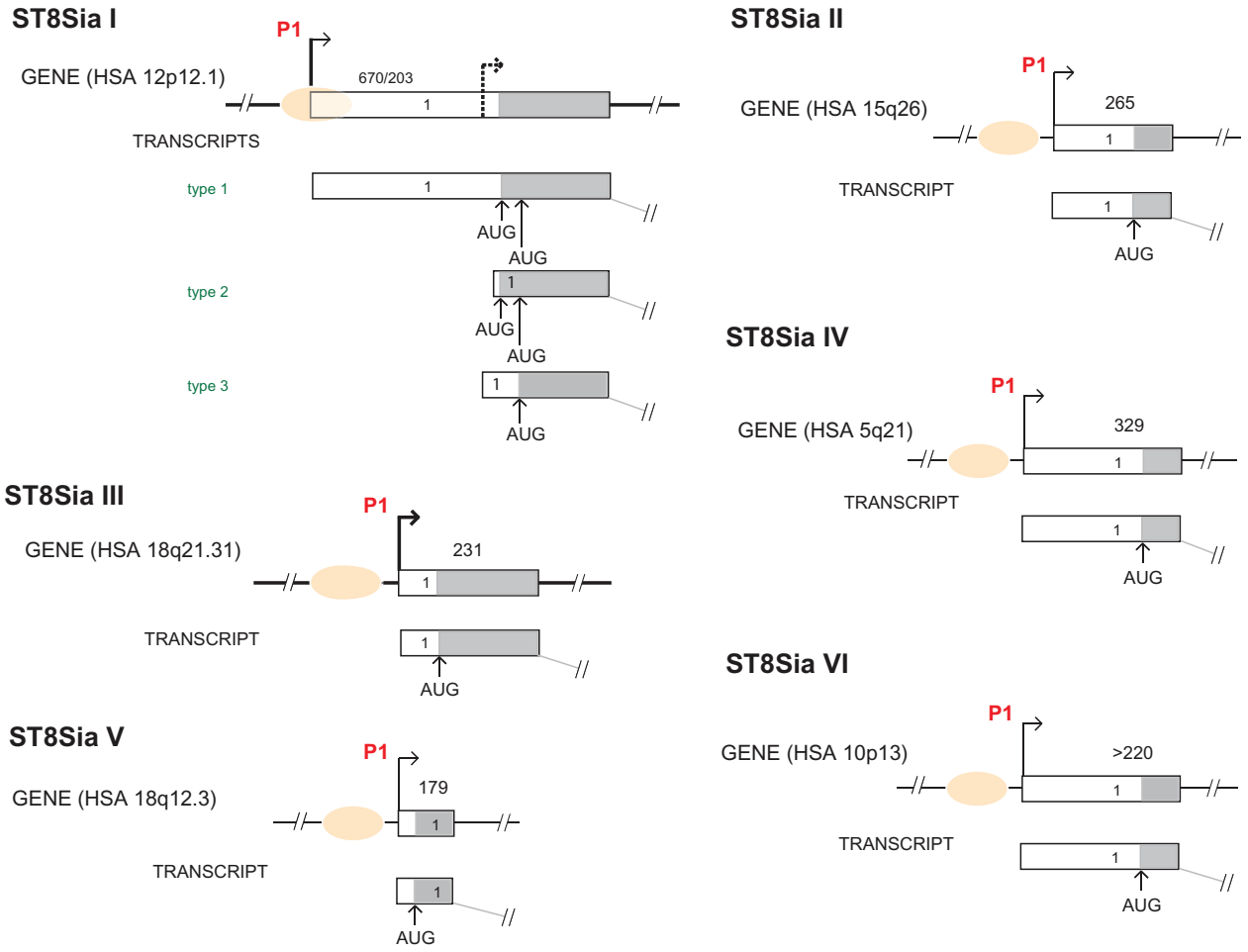

Figure 10. Schematic representation of the 5′-end of human ST8Sia genes and the major transcripts expressed. The boxes and lines connecting boxes indicate exons and introns respectively. Grey rectangles represent coding regions whereas open rectangles represent 5′-UT exon. Promoter regions are symbolized by colored ovals and TSS by bent arrows. Multiplicity of TSS is indicated by dashed bent arrows. Intron lengths are indicated in green below the gene schematic and the exon length in black above. The translation start codon AUG is indicated. Not drawn to scale.

brain[89] and a unique transcript has been described suggesting the occurrence of a unique promoter region. Similarly, the human *ST8SIA5* gene is expressed almost exclusively in fetal and adult brain.[88] No further transcriptional studies have been published yet.

The mouse *st8sia6* gene is expressed in all the adult tissues examined with higher expression levels detected in kidney, heart and spleen.[92] The human *ST8SIA6* gene shows also a basal level of expression in almost all the human adult tissues examined.[91] We have no clues indicating the use of alternative 5′-UT exons and no functional promoter studies are reported.

Promoter prediction studies for the mammalian sialyltransferase genes have further shown absence of the canonical TATA and CCAAT boxes coupled to the presence of several GC boxes and numerous AP2 and Sp1 binding sites in *cis*-regulatory regions. These features are typical of housekeeping genes expressed in nearly all tissues as it is the case for *st6gal1* gene and are often associated with limited promoter conservation across evolution.[188] On the other hand, several sialyltransferase genes exhibit highly tissue-restricted expression. *Cis*-regulatory negative elements silencing the expression of several of these sialyltransferase genes have been described as the Neuron Restrictive Silencing Factor in the human *ST6GAL2* gene promoter region[143] or the GT-CG repeat sequence described in the promoter region of the human *ST8SIA1* gene expressed in the central nervous system.[90] Despite the paucity of sialyltransferase gene expression data reported to date in vertebrates, it appears that changes in tissue-specific

sialyltransferase gene expression patterns between vertebrate are not related to levels of coding sequence divergence, but rather to divergence of *cis*-regulatory sequences of duplicates, use of tissue-specific transcription factor and epigenetic modifications. From an evolutionary point of view, *cis*-regulatory sequences of homologous sialyltransferase genes that originated from the duplication of an ancestral gene (paralogs like *st8sia1* and *st8sia5* genes) and also from speciation (orthologs like *st3gal5* and *st3gal5-r* genes) may evolve independently to ensure specific spatio-temporal expression and function.[189]

Concluding Remarks

Sialyltransferases represent a multigene superfamily characterized by conserved peptide sequences named sialylmotifs, which facilitate their identification in animal genomes. Classically, sialyltransferases are divided in four families ST6Gal, ST3Gal, ST6GalNAc and ST8Sia according to the glycosidic-linkage formed and the monosaccharide acceptor used. Currently twenty sialyltransferases sub-families are known in higher vertebrates. Orthologs of the 20 mammalian sialyltransferase sub-families are found in lower vertebrates, but also additional sub-families were maintained in fish genomes, but have disappeared in higher vertebrates. This variety of sialyltransferases might explain marked differences of sialylation observed between animals. Our knowledge of the enzymatic repertoire of most vertebrate is far from complete. One should remain cautious because of differences in biochemical activity and pattern of expression of sialyltransferases. These past years, molecular phylogeny and phylogenomic tools have been developed to study the origin and the fate of sialyltransferase duplicates to gain insights into the importance of multiple sialyltransferase sub-families in vertebrates. Indeed, biological function of these various vertebrate sialyltransferases *in vivo* is not yet elucidated and modern molecular genetics, mouse gene knock-outs in particular, have revealed extensive functional redundancy in sialyltransferase gene families. However, genetic studies in model organisms with null mutations in sialyltransferase genes have proved that sialylated glycans are required for proper vertebrate development as these mutations produce phenotypes ranging from embryonic lethality to growth defects to impaired morphogenesis and cognitive function, but some have no obvious effects under laboratory conditions.

Acknowledgements

I thank my lab and several collaborators, whose contributions made possible much of the work reported here. In particular, I gratefully acknowledge Philippe Delannoy, Rafael Oriol, Daniel Petit and Jean-Michel Petit for critical comments and helpful discussions, Anne-Marie Mir for her excellent technical assistance, Bernard and Christine Thisse for whole mount *in situ* hybridization analysis of the zebrafish sialyltransferase genes. Research support was from the Université des Sciences et Technologies de Lille 1, CNRS (Centre National de la Recherche Scientifique), the European Commission as part of the ZF-Models integrated project in the 6th Framework Program and the National Institute of health (to C. Thisse and B. Thisse) and the PPF bioinformatique de Lille.

Disclosure

This manuscript has been read and approved by the author. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The author reports no conflicts of interest.

References

1. Harduin-Lepers A, Mollicone R, Delannoy P, Oriol R. The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach. *Glycobiology*. 2005;15(8):805–17.
2. Harduin-Lepers A, Vallejo-Ruiz V, Krzewinski-Recchi MA, Samyn-Petit B, Julien S, Delannoy P. The human sialyltransferase family. *Biochimie*. 2001;83(8):727–37.
3. Takashima S. Characterization of mouse sialyltransferase genes: their evolution and diversity. *Biosci Biotechnol Biochem*. 2008;72(5):1155–67.
4. Tsuji S, Datta AK, Paulson JC. Systematic nomenclature for sialyltransferases. *Glycobiology*. 1996;6(7):v–vii.
5. Harduin-Lepers A, Petit D, Mollicone R, Delannoy P, Petit JM, Oriol R. Evolutionary history of the alpha2,8-sialyltransferase (ST8Sia) gene family: tandem duplications in early deuterostomes explain most of the diversity found in the vertebrate ST8Sia genes. *BMC Evol Biol*. 2008;8:258.
6. Hashimoto K, Tokimatsu T, Kawano S, Yoshizawa AC, Okuda S, Goto S, et al. Comprehensive analysis of glycosyltransferases in eukaryotic genomes for structural and functional characterization of glycans. *Carbohydr Res*. 2009;344(7):881–7.
7. Kikuchi N, Narimatsu H. Bioinformatics for comprehensive finding and analysis of glycosyltransferases. *Biochim Biophys Acta*. 2006;1760(4):578–83.
8. Fitch WM. Distinguishing homologous from analogous proteins. *Syst Zool*. 1970;19(2):99–113.

9. Coutinho PM, Deleury E, Davies GJ, Henrissat B. An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol*. 2003;328(2): 307–17.
10. Wang PH. Altered glycosylation in Cancer: Sialic acids and sialyltransferases. *Journal of Cancer Molecules*. 2005;1(2):73–81.
11. Kawamura YI, Toyota M, Kawashima R, Hagiwara T, Suzuki H, Imai K, et al. DNA hypermethylation contributes to incomplete synthesis of carbohydrate determinants in gastrointestinal cancer. *Gastroenterology*. 2008; 135(1):142–51 e3.
12. Paulson JC, Colley KJ. Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. *J Biol Chem*. 1989;264(30):17615–8.
13. Uemura S, Yoshida S, Shishido F, Inokuchi J. The cytoplasmic tail of G_{M3} synthase defines its subcellular localization, stability, and *in vivo* activity. *Mol Biol Cell*. 2009;20(13):3088–100.
14. Patel RY, Balaji PV. Length and composition analysis of the cytoplasmic, transmembrane and stem regions of human Golgi glycosyltransferases. *Protein Pept Lett*. 2007;14(6):601–9.
15. Hudgin RL, Schachter H. Porcine sugar nucleotide: glycoprotein glycosyltransferases. I. Blood serum and liver sialyltransferase. *Can J Biochem*. 1971;49(7):829–37.
16. Kitazume S, Oka R, Ogawa K, Futakawa S, Hagiwara Y, Takikawa H, et al. Molecular insights into beta-galactoside alpha2,6-sialyltransferase secretion *in vivo*. *Glycobiology*. 2009;19(5):479–87.
17. Datta AK. Comparative sequence analysis in the sialyltransferase protein family: analysis of motifs. *Curr Drug Targets*. 2009;10(6):483–98.
18. Livingston BD, Paulson JC. Polymerase chain reaction cloning of a developmentally regulated member of the sialyltransferase gene family. *J Biol Chem*. 1993;268(16):11504–7.
19. Datta AK, Paulson JC. Sialylmotifs of sialyltransferases. *Indian J Biochem Biophys*. 1997;34(1–2):157–65.
20. Datta AK, Chammas R, Paulson JC. Conserved cysteines in the sialyltransferase sialylmotifs form an essential disulfide bond. *J Biol Chem*. 2001;276(18):15200–7.
21. Jeanneau C, Chazalet V, Auge C, Soumpasis DM, Harduin-Lepers A, Delannoy P, et al. Structure-function analysis of the human sialyltransferase ST3Gal I: role of N-glycosylation and a novel conserved sialylmotif. *J Biol Chem*. 2004;279(14):13461–8.
22. Kapitonov D, Yu RK. Conserved domains of glycosyltransferases. *Glycobiology*. 1999;9(10):961–78.
23. Geremia RA, Harduin-Lepers A, Delannoy P. Identification of two novel conserved amino acid residues in eukaryotic sialyltransferases: implications for their mechanism of action. *Glycobiology*. 1997;7(2):v–vii.
24. Datta AK, Paulson JC. The sialyltransferase "sialylmotif" participates in binding the donor substrate CMP-NeuAc. *J Biol Chem*. 1995;270(4):1497–500.
25. Datta AK, Sinha A, Paulson JC. Mutation of the sialyltransferase S-sialylmotif alters the kinetics of the donor and acceptor substrates. *J Biol Chem*. 1998;273(16):9608–14.
26. Kitazume-Kawaguchi S, Kabata S, Arita M. Differential biosynthesis of polysialic or disialic acid Structure by ST8Sia II and ST8Sia IV. *J Biol Chem*. 2001;276(19):15696–703.
27. Rao FV, Rich JR, Rakic B, Buddai S, Schwartz MF, Johnson K, et al. Structural insight into mammalian sialyltransferases. *Nat Struct Mol Biol*. 2009.
28. Angata K, Fukuda M. Polysialyltransferases: major players in polysialic acid synthesis on the neural cell adhesion molecule. *Biochimie*. 2003; 85(1–2):195–206.
29. Fukumoto S, Miyazaki H, Goto G, Urano T, Furukawa K, Furukawa K. Expression cloning of mouse cDNA of CMP-NeuAc: Lactosylceramide alpha2,3-sialyltransferase, an enzyme that initiates the synthesis of gangliosides. *J Biol Chem*. 1999;274(14):9271–6.
30. Okajima T, Fukumoto S, Miyazaki H, Ishida H, Kiso M, Furukawa K, et al. Molecular cloning of a novel alpha2,3-sialyltransferase (ST3Gal VI) that sialylates type II lactosamine structures on glycoproteins and glycolipids. *J Biol Chem*. 1999;274(17):11479–86.
31. Sasaki K. Molecular cloning and characterization of sialyltransferases. *Trends in Glycoscience and Glycotechnology*. 1996;8(41):195–215.
32. Patel RY, Balaji PV. Identification of linkage-specific sequence motifs in sialyltransferases. *Glycobiology*. 2006;16(2):108–16.
33. Tsuji S. Sialyltransferases. In: Taniguchi N, Suzuki A, Ito Y, Narimatsu H, Kawasaki T, Hase S, editors. Experimental Glycoscience. Springer Japan ed: Springer; 2009. p. 42–6.
34. Kitagawa H, Paulson JC. Differential expression of five sialyltransferase genes in human tissues. *J Biol Chem*. 1994;269(27):17872–8.
35. Groux-Degroote S, Harduin-Lepers A, Delannoy P. Biosynthesis of mucin O-glycan chains in normal and pathological states. In: Van Seuningen I, editor. The epithelial Mucins: Structure/Function. Roles in Cancer and Inflammatory Diseases. Kerala, India: Research Signpost; 2008. p. 39–54.
36. Ito H, Hiraiwa N, Sawada-Kasugai M, Akamatsu S, Tachikawa T, Kasai Y, et al. Altered mRNA expression of specific molecular species of fucosyl- and sialyl-transferases in human colorectal cancer tissues. *Int J Cancer*. 1997;71(4):556–64.
37. Kono M, Ohyama Y, Lee YC, Hamamoto T, Kojima N, Tsuji S. Mouse beta-galactoside alpha 2,3-sialyltransferases: comparison of *in vitro* substrate specificities and tissue specific expression. *Glycobiology*. 1997;7(4): 469–79.
38. Kurosawa N, Hamamoto T, Inoue M, Tsuji S. Molecular cloning and expression of chick Gal beta 1,3GalNAc alpha 2,3-sialyltransferase. *Biochim Biophys Acta*. 1995;1244(1):216–22.
39. Lee YC, Kurosawa N, Hamamoto T, Nakaoka T, Tsuji S. Molecular cloning and expression of Gal beta 1,3GalNAc alpha 2,3-sialyltransferase from mouse brain. *Eur J Biochem*. 1993;216(2):377–85.
40. Giordanengo V, Bannwarth S, Laffont C, Van Miegem V, Harduin-Lepers A, Delannoy P, et al. Cloning and expression of cDNA for a human Gal(beta1-3) GalNAc alpha2,3-sialyltransferase from the CEM T-cell line. *Eur J Biochem*. 1997;247(2):558–66.
41. Kojima N, Lee YC, Hamamoto T, Kurosawa N, Tsuji S. Kinetic properties and acceptor substrate preferences of two kinds of Gal beta 1,3GalNAc alpha 2,3-sialyltransferase from mouse brain. *Biochemistry*. 1994;33(19): 5772–6.
42. Lehmann F, Kelm S, Dietz F, von Itzstein M, Tiralongo J. The evolution of galactose alpha2,3-sialyltransferase: Ciona intestinalis ST3GAL I/II and Takifugu rubripes ST3GAL II sialylate Galbeta1,3GalNAc structures on glycoproteins but not glycolipids. *Glycoconj J*. 2008; 25(4):323–34.
43. Kitagawa H, Paulson JC. Cloning and expression of human Gal beta 1,3(4) GlcNAc alpha 2,3-sialyltransferase. *Biochem Biophys Res Commun*. 1993; 194(1):375–82.
44. Kitagawa H, Paulson JC. Cloning of a novel alpha 2,3-sialyltransferase that sialylates glycoprotein and glycolipid carbohydrate groups. *J Biol Chem*. 1994;269(2):1394–401.
45. Sasaki K, Watanabe E, Kawashima K, Sekine S, Dohi T, Oshima M, et al. Expression cloning of a novel Gal beta (1-3/1-4) GlcNAc alpha 2, 3-sialyltransferase using lectin resistance selection. *J Biol Chem*. 1993; 268(30):22782–7.
46. Carvalho AS, Harduin-Lepers A, Magalhaes A, Machado E, Mendes N, Costa LT, et al. Differential expression of alpha-2,3-sialyltransferases and alpha-1,3/4-fucosyltransferases regulates the levels of sialyl Lewis a and sialyl Lewis x in gastrointestinal carcinoma cells. *Int J Biochem Cell Biol*. 2010;42:80–9.
47. Ishii A, Ohta M, Watanabe Y, Matsuda K, Ishiyama K, Sakoe K, et al. Expression cloning and functional characterization of human cDNA for ganglioside GM3 synthase. *J Biol Chem*. 1998;273(48):31652–5.
48. Richardson K, Jamieson JC. Release of sialyltransferases from rat liver Golgi membranes by a cathepsin D-like proteinase: comparison of the release of Gal beta 1-4GlcNAc alpha 2-6 sialyltransferase, Gal beta 1-3(4) GlcNAc alpha 2-3 sialyltransferase and lactosylceramide alpha 2-3 sialyltransferase (SAT-1). *Comp Biochem Physiol B Biochem Mol Biol*. 1995;110(2):445–50.
49. Preuss U, Gu X, Gu T, Yu RK. Purification and characterization of CMP-N-acetylneuraminic acid: lactosylceramide (alpha 2-3) sialyltransferase (GM3-synthase) from rat brain. *J Biol Chem*. 1993;268(35):26273–8.

50. Berselli P, Zava S, Sottocornola E, Milani S, Berra B, Colombo I. Human GM3 synthase: a new mRNA variant encodes an NH2-terminal extended form of the protein. *Biochim Biophys Acta*. 2006; 1759(7):348–58.
51. Chisada SI, Yoshimura Y, Sakaguchi K, Uemura S, Go S, Ikeda K, et al. Zebrafish and mouse alpha2,3-sialyltranserases responsible for synthesizing GM4 ganglioside. *J Biol Chem*. 2009;284(44): 30534–46.
52. Dall'Olio F. The sialyl-alpha2,6-lactosaminyl-structure: biosynthesis and functional role. *Glycoconj J*. 2000;17(10):669–76.
53. Paulson JC, Rearick JI, Hill RL. Enzymatic properties of beta-D-galactoside alpha2 leads to 6 sialytransferase from bovine colostrum. *J Biol Chem*. 1977;252(7):2363–71.
54. Nemansky M, Van den Eijnden DH. Bovine colostrum CMP-NeuAc: Gal beta(1—>4)GlcNAc-R alpha(2—>6)-sialyltransferase is involved in the synthesis of the terminal NeuAc alpha(2—>6)GalNAc beta(1—>4)GlcNAc sequence occurring on N-linked glycans of bovine milk glycoproteins. *Biochem J*. 1992;287(Pt 1):311–6.
55. Grundmann U, Nerlich C, Rein T, Zettlmeissl G. Complete cDNA sequence encoding human beta-galactoside alpha-2,6-sialyltransferase. *Nucleic Acids Res*. 1990;18(3):667.
56. Weinstein J, Lee EU, McEntee K, Lai PH, Paulson JC. Primary structure of beta-galactoside alpha 2,6-sialyltransferase. Conversion of membrane-bound enzyme to soluble forms by cleavage of the NH_2-terminal signal anchor. *J Biol Chem*. 1987;262(36):17735–43.
57. Hamamoto T, Kawasaki M, Kurosawa N, Nakaoka T, Lee YC, Tsuji S. Two step single primer mediated polymerase chain reaction. Application to cloning of putative mouse, beta-galactoside alpha 2,6-sialyltransferase cDNA. *Bioorg Med Chem*. 1993;1(2):141–5.
58. Mercier D, Wierinckx A, Oulmouden A, Gallet PF, Palcic MM, Harduin-Lepers A, et al. Molecular cloning, expression and exon/intron organization of the bovine beta-galactoside alpha2,6-sialyltransferase gene. *Glycobiology*. 1999;9(9):851–63.
59. Kurosawa N, Kawasaki M, Hamamoto T, Nakaoka T, Lee YC, Arita M, et al. Molecular cloning and expression of chick embryo Gal beta 1,4GlcNAc alpha 2,6-sialyltransferase. Comparison with the mammalian enzyme. *Eur J Biochem*. 1994;219(1–2):375–81.
60. Rohfritsch PF, Joosten JA, Krzewinski-Recchi MA, Harduin-Lepers A, Laporte B, Juliant S, et al. Probing the substrate specificity of four different sialyltransferases using synthetic beta-D-Galp-(1—>4)-beta-D-Glcp-NAc-(1—>2)-alpha-D-Manp-(1—>O) (CH(2))7CH3 analogues general activating effect of replacing N-acetylglucosamine by N-propionylglucosamine. *Biochim Biophys Acta*. 2006;1760(4):685–92.
61. Krzewinski-Recchi MA, Julien S, Juliant S, Teintenier-Lelievre M, Samyn-Petit B, Montiel MD, et al. Identification and functional expression of a second human beta-galactoside alpha2,6-sialyltransferase, ST6Gal II. *Eur J Biochem*. 2003;270(5):950–61.
62. Takashima S, Tsuji S, Tsujimoto M. Characterization of the second type of human beta-galactoside alpha 2,6-sialyltransferase (ST6Gal II), which sialylates Galbeta 1,4GlcNAc structures on oligosaccharides preferentially. Genomic analysis of human sialyltransferase genes. *J Biol Chem*. 2002;277(48):45719–28.
63. Takashima S, Tsuji S, Tsujimoto M. Comparison of the enzymatic properties of mouse beta-galactoside alpha2,6-sialyltransferases, ST6Gal I and II. *J Biochem*. 2003;134(2):287–96.
64. Laporte B, Gonzalez-Hilarion S, Maftah A, Petit JM. The second bovine {beta}-Galactoside-{alpha}2,6-sialyltransferase (ST6Gal II): genomic organization and stimulation of its *in vitro* expression by IL-6 in bovine mammary epithelial cells. *Glycobiology*. 2009.
65. Koles K, Irvine KD, Panin VM. Functional characterization of Drosophila sialyltransferase. *J Biol Chem*. 2004;279(6):4346–57.
66. Ikehara Y, Kojima N, Kurosawa N, Kudo T, Kono M, Nishihara S, et al. Cloning and expression of a human gene encoding an N-acetylgalactosamine-alpha2,6-sialyltransferase (ST6GalNAc I): a candidate for synthesis of cancer-associated sialyl-Tn antigens. *Glycobiology*. 1999; 9(11):1213–24.
67. Kono M, Tsuda T, Ogata S, Takashima S, Liu H, Hamamoto T, et al. Redefined substrate specificity of ST6GalNAc II: a second candidate sialyl-Tn synthase. *Biochem Biophys Res Commun*. 2000; 272(1):94–7.
68. Kurosawa N, Hamamoto T, Lee YC, Nakaoka T, Kojima N, Tsuji S. Molecular cloning and expression of GalNAc alpha 2,6-sialyltransferase. *J Biol Chem*. 1994;269(2):1402–9.
69. Kurosawa N, Kojima N, Inoue M, Hamamoto T, Tsuji S. Cloning and expression of Gal beta 1,3GalNAc-specific GalNAc alpha 2, 6-sialyltransferase. *J Biol Chem*. 1994;269(29):19048–53.
70. Samyn-Petit B, Krzewinski-Recchi MA, Steelant WF, Delannoy P, Harduin-Lepers A. Molecular cloning and functional expression of human ST6GalNAc II. Molecular expression in various human cultured cells. *Biochim Biophys Acta*. 2000;1474(2):201–11.
71. Asahina S, Sato C, Kitajima K. Developmental expression of a sialyltransferase responsible for sialylation of cortical alveolus glycoprotein during oogenesis in rainbow trout (Oncorhynchus mykiss). *J Biochem (Tokyo)*. 2004;136(2):189–98.
72. Harduin-Lepers A, Stokes DC, Steelant WF, Samyn-Petit B, Krzewinski-Recchi MA, Vallejo-Ruiz V, et al. Cloning, expression and gene organization of a human Neu5Ac alpha 2-3Gal beta 1-3GalNAc alpha 2, 6-sialyltransferase: hST6GalNAcIV. *Biochem J*. 2000;352 (Pt 1):37–48.
73. Lee YC, Kaufmann M, Kitazume-Kawaguchi S, Kono M, Takashima S, Kurosawa N, et al. Molecular cloning and functional expression of two members of mouse NeuAcalpha2,3Galbeta1,3GalNAc GalNAcalpha2, 6-sialyltransferase family, ST6GalNAc III and IV. *J Biol Chem*. 1999; 274(17):11958–67.
74. Sjoberg ER, Kitagawa H, Glushka J, van Halbeek H, Paulson JC. Molecular cloning of a developmentally regulated N-acetylgalactosamine alpha2, 6-sialyltransferase specific for sialylated glycoconjugates. *J Biol Chem*. 1996;271(13):7450–9.
75. Tsuchida A, Ogiso M, Nakamura Y, Kiso M, Furukawa K, Furukawa K. Molecular cloning and expression of human ST6GalNAc III: restricted tissue distribution and substrate specificity. *J Biochem*. 2005;138(3): 237–43.
76. Okajima T, Chen HH, Ito H, Kiso M, Tai T, Furukawa K, et al. Molecular cloning and expression of mouse GD1alpha/GT1aalpha/GQ1balpha synthase (ST6GalNAc VI) gene. *J Biol Chem*. 2000; 275(10):6717–23.
77. Okajima T, Fukumoto S, Ito H, Kiso M, Hirabayashi Y, Urano T, et al. Molecular cloning of brain-specific GD1alpha synthase (ST6GalNAc V) containing CAG/Glutamine repeats. *J Biol Chem*. 1999;274(43): 30557–62.
78. Senda M, Ito A, Tsuchida A, Hagiwara T, Kaneda T, Nakamura Y, et al. Identification and expression of a sialyltransferase responsible for the synthesis of disialylgalactosylgloboside in normal and malignant kidney cells: downregulation of ST6GalNAc VI in renal cancers. *Biochem J*. 2007;402(3):459–70.
79. Tsuchida A, Okajima T, Furukawa K, Ando T, Ishida H, Yoshida A, et al. Synthesis of disialyl Lewis a (Le(a)) structure in colon cancer cell lines by a sialyltransferase, ST6GalNAc VI, responsible for the synthesis of alpha-series gangliosides. *J Biol Chem*. 2003;278(25):22787–94.
80. Daniotti JL, Martina JA, Giraudo CG, Zurita AR, Maccioni HJ. G_{M3} alpha2,8-sialyltransferase (GD3 synthase): protein characterization and sub-golgi location in CHO-K1 cells. *J Neurochem*. 2000; 74(4):1711–20.
81. Luque ME, Crespo PM, Monaco ME, Aybar MJ, Daniotti JL, Sanchez SS. Cloning and functional characterization of two key enzymes of glycosphingolipid biosynthesis in the amphibian Xenopus laevis. *Dev Dyn*. 2008;237(1): 112–23.
82. Rimoldi S, Papis E, Bernardini G, Prati M, Gornati R. Molecular cloning and expression of alpha2,8-sialyltransferase (ST8Sia I, GD3 Synthase) in Xenopus. *Mol Cell Biochem*. 2007;301(1–2):143–53.
83. Watanabe Y, Nara K, Takahashi H, Nagai Y, Sanai Y. The molecular cloning and expression of alpha 2,8-sialyltransferase (GD3 synthase) in a rat brain. *J Biochem*. 1996;120(5):1020–7.

84. Zeng G, Gao L, Ariga T, Yu RK. Molecular cloning of cDNA for rat brain GD3-synthase. *Biochem Biophys Res Commun*. 1996; 226(2):319–23.
85. Nara K, Watanabe Y, Maruyama K, Kasahara K, Nagai Y, Sanai Y. Expression cloning of a CMP-NeuAc: NeuAcα2-3Galβ1-4Glcβ1-1'Cer α2,8-sialyltransferase (GD3 synthase) from human melanoma cells. *Proc Natl Acad Sci U S A*. 1994;91(17):7952–6.
86. Nakayama J, Fukuda MN, Hirabayashi Y, Kanamori A, Sasaki K, Nishi T, et al. Expression cloning of a human GT3 synthase. GD3 AND GT3 are synthesized by a single enzyme. *J Biol Chem*. 1996; 271(7):3684–91.
87. Chang LY, Mir AM, Thisse C, Guerardel Y, Delannoy P, Thisse B, et al. Molecular cloning and characterization of the expression pattern of the zebrafish alpha2,8-sialyltransferases (ST8Sia) in the developing nervous system. *Glycoconj J*. 2009;26(3):263–75.
88. Kim YJ, Kim KS, Do S, Kim CH, Kim SK, Lee YC. Molecular cloning and expression of human alpha2,8-sialyltransferase (hST8Sia V). *Biochem Biophys Res Commun*. 1997;235(2):327–30.
89. Kono M, Yoshida Y, Kojima N, Tsuji S. Molecular cloning and expression of a fifth type of α2,8-sialyltransferase (ST8Sia V). Its substrate specificity is similar to that of SAT-V/III, which synthesize GD1c, GT1a, GQ1b and GT3. *J Biol Chem*. 1996;271(46):29366–71.
90. Yu R, Bieberich E, Xia T, Zeng G. Regulation of ganglioside biosynthesis in the nervous system. *J Lipid Res*. 2004;45(5):783–93.
91. Teintenier-Lelievre M, Julien S, Juliant S, Guerardel Y, Duonor-Cerutti M, Delannoy P, et al. Molecular cloning and expression of a human hST8Sia VI (α2,8-sialyltransferase) responsible for the synthesis of the diSia motif on O-glycosylproteins. *Biochem J*. 2005;392(Pt 3):665–74.
92. Takashima S, Ishida HK, Inazu T, Ando T, Ishida H, Kiso M, et al. Molecular cloning and expression of a sixth type of α2,8-sialyltransferase (ST8Sia VI) that sialylates O-glycans. *J Biol Chem*. 2002;277(27):24030–8.
93. Lee YC, Kim YJ, Lee KY, Kim KS, Kim BU, Kim HN, et al. Cloning and expression of cDNA for a human Sia alpha 2,3Gal beta 1, 4GlcNA:alpha 2,8-sialyltransferase (hST8Sia III). *Arch Biochem Biophys*. 1998;360(1):41–6.
94. Yoshida Y, Kojima N, Kurosawa N, Hamamoto T, Tsuji S. Molecular cloning of Sia α2,3Galβ1,4GlcNAcα2,8-sialyltransferase from mouse brain. *J Biol Chem*. 1995;270(24):14628–33.
95. Bentrop J, Marx M, Schattschneider S, Rivera-Milla E, Bastmeyer M. Molecular evolution and expression of zebrafish St8SiaIII, an alpha-2, 8-sialyltransferase involved in myotome development. *Dev Dyn*. 2008.
96. Yu RK, Macala LJ, Taki T, Weinfield HM, Yu FS. Developmental changes in ganglioside composition and synthesis in embryonic rat brain. *J Neurochem*. 1988;50(6):1825–9.
97. Angata K, Suzuki M, McAuliffe J, Ding Y, Hindsgaul O, Fukuda M. Differential biosynthesis of polysialic acid on neural cell adhesion molecule (NCAM) and oligosaccharide acceptors by three distinct alpha 2, 8-sialyltransferases, ST8Sia IV (PST), ST8Sia II (STX), and ST8Sia III. *J Biol Chem*. 2000;275(24):18594–601.
98. Sato C, Matsuda T, Kitajima K. Neuronal differentiation-dependent expression of the disialic acid epitope on CD166 and its involvement in neurite formation in Neuro2A cells. *J Biol Chem*. 2002;277(47):45299–305.
99. Eckhardt M, Muhlenhoff M, Bethe A, Koopman J, Frosch M, Gerardy-Schahn R. Molecular characterization of eukaryotic polysialyltransferase-1. *Nature*. 1995;373(6516):715–8.
100. Kojima N, Yoshida Y, Tsuji S. A developmentally regulated member of the sialyltransferase family (ST8Sia II, STX) is a polysialic acid synthase. *FEBS Lett*. 1995;373(2):119–22.
101. Nakayama J, Fukuda MN, Fredette B, Ranscht B, Fukuda M. Expression cloning of a human polysialyltransferase that forms the polysialylated neural cell adhesion molecule present in embryonic brain. *Proc Natl Acad Sci U S A*. 1995;92(15):7031–5.
102. Scheidegger EP, Sternberg LR, Roth J, Lowe JB. A human STX cDNA confers polysialic acid expression in mammalian cells. *J Biol Chem*. 1995;270(39):22685–8.
103. Yoshida Y, Kojima N, Tsuji S. Molecular cloning and characterization of a third type of N-glycan alpha 2,8-sialyltransferase from mouse lung. *J Biochem*. 1995;118(3):658–64.
104. Kojima N, Tachida Y, Yoshida Y, Tsuji S. Characterization of mouse ST8Sia II (STX) as a neural cell adhesion molecule-specific polysialic acid synthase. Requirement of core alpha1,6-linked fucose and a polypeptide chain for polysialylation. *J Biol Chem*. 1996;271(32):19457–63.
105. Muhlenhoff M, Eckhardt M, Bethe A, Frosch M, Gerardy-Schahn R. Polysialylation of NCAM by a single enzyme. *Curr Biol*. 1996;6(9): 1188–91.
106. Nakayama J, Angata K, Ong E, Katsuyama T, Fukuda M. Polysialic acid, a unique glycan that is developmentally regulated by two polysialyltransferases, PST and STX, in the central nervous system: from biosynthesis to function. *Pathol Int*. 1998;48(9):665–77.
107. Hildebrandt H, Muhlenhoff M, Weinhold B, Gerardy-Schahn R. Dissecting polysialic acid and NCAM functions in brain development. *J Neurochem*. 2007;103 Suppl 1:56–64.
108. Kudo M, Takayama E, Tashiro K, Fukamachi H, Nakata T, Tadakuma T, et al. Cloning and expression of an alpha-2,8-polysialyltransferase (STX) from Xenopus laevis. *Glycobiology*. 1998;8(8):771–7.
109. Asahina S, Sato C, Matsuno M, Matsuda T, Colley K, Kitajima K. Involvement of the α2,8-polysialyltransferases II/STX and IV/PST in the biosynthesis of polysialic acid chains on the O-linked glycoproteins in rainbow trout ovary. *J Biochem (Tokyo)*. 2006;140(5):687–701.
110. Marx M, Rivera-Milla E, Stummeyer K, Gerardy-Schahn R, Bastmeyer M. Divergent evolution of the vertebrate polysialyltransferase Stx and Pst genes revealed by fish-to-mammal comparison. *Dev Biol*. 2007;306: 560–71.
111. Harduin-Lepers A, Krzewinski-Recchi MA, Hebbar M, Samyn-Petit B, Vallejo-Ruiz V, Julien S, et al. Sialyltransferases and breast cancer. *Recent Res Devel Cancer*. 2001;3:111–26.
112. Kaneko M, Nishihara S, Narimatsu H, Saitou N. The evolutionary history of glycosyltransferase genes. *Trends in glycoscience and glycotechnology*. 2001;13(70):147–55.
113. Ohno S. Evolution by gene duplication. Heidelberg: Springer-verlag; 1970.
114. Wang Y, Gu X. Evolutionary patterns of gene families generated in the early stage of vertebrates. *J Mol Evol*. 2000;51(1):88–96.
115. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 1999;151(4):1531–45.
116. Canestro C, Catchen JM, Rodriguez-Mari A, Yokoi H, Postlethwait JH. Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet*. 2009;5(5):e1000496.
117. Danchin EG, Gouret P, Pontarotti P. Eleven ancestral gene families lost in mammals and vertebrates while otherwise universally conserved in animals. *BMC Evol Biol*. 2006;6:5.
118. Altheide TK, Hayakawa T, Mikkelsen TS, Diaz S, Varki N, Varki A. System-wide genomic and biochemical comparisons of sialic acid biology among primates and rodents: Evidence for two modes of rapid evolution. *J Biol Chem*. 2006;281(35):25689–702.
119. Comelli EM, Head SR, Gilmartin T, Whisenant T, Haslam SM, North SJ, et al. A focused microarray approach to functional glycomics: transcriptional regulation of the glycome. *Glycobiology*. 2006;16(2): 117–31.
120. Nairn AV, York WS, Harris K, Hall EM, Pierce JM, Moremen KW. Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem*. 2008;283(25):17298–313.
121. Takashima S, Tsuji S. Comparison of genomic structures of four members of beta-galactoside alpha2,3-sialyltransferase genes in the mouse. *Cytogenet Cell Genet*. 2000;89(1–2):101–6.
122. Taniguchi A. Promoter structure and transcriptional regulation of human β-galactoside α2,3-sialyltransferase genes. *Current Drug Targets*. 2008;9: 310–6.
123. Zeng G, Yu R. Cloning and transcriptional regulation of genes responsible for synthesis of gangliosides. *Current Drug Targets*. 2008;9: 317–24.
124. Harduin-Lepers A, Shaper NL, Mahoney JA, Shaper JH. Murine beta 1, 4-galactosyltransferase: round spermatid transcripts are characterized by an extended 5′-untranslated region. *Glycobiology*. 1992;2(4):361–8.

125. Shaper NL, Harduin-Lepers A, Shaper JH. Male germ cell expression of murine beta 4-galactosyltransferase. A 796-base pair genomic region, containing two cAMP-responsive element (CRE)-like elements, mediates male germ cell-specific expression in transgenic mice. *J Biol Chem*. 1994;269(40):25165–71.
126. Koyama N, Miyoshi E, Ihara Y, Kang R, Nishikawa A, Taniguchi N. Human N-acetylglucosaminyltransferase III gene is transcribed from multiple promoters. *Eur J Biochem*. 1996;238(3):853–61.
127. Buckhaults P, Chen L, Fregien N, Pierce M. Transcriptional regulation of N-acetylglucosaminyltransferase V by the src oncogene. *J Biol Chem*. 1997;272(31):19575–81.
128. Koda Y, Soejima M, Kimura H. Changing transcription start sites in H-type alpha(1,2)fucosyltransferase gene (FUT1) during differentiation of the human erythroid lineage. *Eur J Biochem*. 1998; 256(2):379–87.
129. Svensson EC, Soreghan B, Paulson JC. Organization of the beta-galactoside alpha 2,6-sialyltransferase gene. Evidence for the transcriptional regulation of terminal glycosylation. *J Biol Chem*. 1990;265(34):20863–8.
130. Wang X, O'Hanlon TP, Young RF, Lau JT. Rat beta-galactoside alpha 2, 6-sialyltransferase genomic organization: alternate promoters direct the synthesis of liver and kidney transcripts. *Glycobiology*. 1990;1(1):25–31.
131. O'Hanlon TP, Lau KM, Wang XC, Lau JT. Tissue-specific expression of beta-galactoside alpha-2,6-sialyltransferase. Transcript heterogeneity predicts a divergent polypeptide. *J Biol Chem*. 1989;264(29):17389–94.
132. Wen DX, Svensson EC, Paulson JC. Tissue-specific alternative splicing of the beta-galactoside alpha 2,6-sialyltransferase gene. *J Biol Chem*. 1992;267(4):2512–8.
133. Wuensch SA, Huang RY, Ewing J, Liang X, Lau JT. Murine B cell differentiation is accompanied by programmed expression of multiple novel beta-galactoside alpha2,6-sialyltransferase mRNA forms. *Glycobiology*. 2000;10(1):67–75.
134. Dalziel M, Huang RY, Dall'Olio F, Morris JR, Taylor-Papadimitriou J, Lau JT. Mouse ST6Gal sialyltransferase gene expression during mammary gland lactation. *Glycobiology*. 2001;11(5):407–12.
135. Aas-Eng DA, Asheim HC, Deggerdal A, Smeland E, Funderud S. Characterization of a promoter region supporting transcription of a novel human beta-galactoside alpha-2,6-sialyltransferase transcript in HepG2 cells. *Biochim Biophys Acta*. 1995;1261(1):166–9.
136. Aasheim HC, Aas-Eng DA, Deggerdal A, Blomhoff HK, Funderud S, Smeland EB. Cell-specific expression of human beta-galactoside alpha 2,6-sialyltransferase transcripts differing in the 5′ untranslated region. *Eur J Biochem*. 1993;213(1):467–75.
137. Lo NW, Lau JT. Transcription of the beta-galactoside alpha 2, 6-sialyltransferase gene in B lymphocytes is directed by a separate and distinct promoter. *Glycobiology*. 1996;6(3):271–9.
138. Taniguchi A, Higai K, Hasegawa Y, Utsumi K, Matsumoto K. Differentiation elicits negative regulation of human beta-galactoside alpha2, 6-sialyltransferase at the mRNA level in the HL-60 cell line. *FEBS Lett*. 1998;441(2):191–4.
139. Wang X, Vertino A, Eddy RL, Byers MG, Jani-Sait SN, Shows TB, et al. Chromosome mapping and organization of the human beta-galactoside alpha 2,6-sialyltransferase gene. Differential and cell-type specific usage of upstream exon sequences in B-lymphoblastoid cells. *J Biol Chem*. 1993;268(6):4355–61.
140. Taniguchi A, Hasegawa Y, Higai K, Matsumoto K. Transcriptional regulation of human beta-galactoside alpha2,6-sialyltransferase (hST-6Gal I) gene during differentiation of the HL-60 cell line. *Glycobiology*. 2000;10(6):623–8.
141. Appenheimer MM, Huang RY, Chandrasekaran EV, Dalziel M, Hu YP, Soloway PD, et al. Biologic contribution of P1 promoter-mediated expression of ST6Gal I sialyltransferase. *Glycobiology*. 2003;13(8): 591–600.
142. Xu L, Kurusu Y, Takizawa K, Tanaka J, Matsumoto K, Taniguchi A. Transcriptional regulation of human beta-galactoside alpha2,6-sialyltransferase (hST6Gal I) gene in colon adenocarcinoma cell line. *Biochem Biophys Res Commun*. 2003;307(4):1070–4.
143. Lehoux S, Groux-Degroote S, Cazet A, Dhaenens CM, Maurage CA, Caillet-Boudin ML, et al. Transcriptional regulation of the human ST6GAL2 gene in cerebral cortex and neuronal cells. *Glycoconj J*. 2009.
144. Koles K, Repnikova E, Pavlova G, Korochkin LI, Panin VM. Sialylation in protostomes: a perspective from Drosophila genetics and biochemistry. *Glycoconj J*. 2009;26(3):313–24.
145. Kim K, Lawrence SM, Park J, Pitts L, Vann WF, Betenbaugh MJ, et al. Expression of a functional Drosophila melanogaster N-acetylneuraminic acid (Neu5Ac) phosphate synthase gene: evidence for endogenous sialic acid biosynthetic ability in insects. *Glycobiology*. 2002;12(2):73–83.
146. Fabini G, Wilson I. Stage and tissue specific expression of a sialyltransferase-like gene (CG4871) in *Drosophila melanogaster*. *Drosophila Information Service*. 2002;85:45–9.
147. 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*. 1995;5(3):319–25.
148. Taniguchi A, Yoshikawa I, Matsumoto K. Genomic structure and transcriptional regulation of human Galbeta1,3GalNAc alpha2,3-sialyltransferase (hST3Gal I) gene. *Glycobiology*. 2001;11(3):241–7.
149. Taniguchi A, Morishima T, Tsujita Y, Matsumoto Y, Matsumoto K. Genomic structure, expression, and transcriptional regulation of human Gal beta. 1,3 GalNAc alpha 2,3-sialyltransferase gene. *Biochem Biophys Res Commun*. 2003;300(2):570–6.
150. Grahn A, Barkhordar GS, Larson G. Identification of seven new alpha2, 3-sialyltransferase III, ST3Gal III, transcripts from human foetal brain. *Glycoconj J*. 2004;20(7–8):493–500.
151. 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*. 2003;1626(1–3):92–6.
152. 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*. 1996;271(2):931–8.
153. Taniguchi A, Hioki M, Matsumoto K. Transcriptional regulation of human Galbeta1,3GalNAc/Galbeta1, 4GlcNAc alpha2,3-sialyltransferase (hST3Gal IV) gene in testis and ovary cell lines. *Biochem Biophys Res Commun*. 2003;301(3):764–8.
154. Taniguchi A, Matsumoto K. Epithelial-cell-specific transcriptional regulation of human Galbeta1,3GalNAc/Galbeta1,4GlcNAc alpha2, 3-sialyltransferase (hST3Gal IV) gene. *Biochem Biophys Res Commun*. 1999;257(2):516–22.
155. Gassen U, Kelm S, Schauer R. Differential gene expression of a human alpha2,3-sialyltransferase in leukaemic cell lines and leucocytes. *FEBS Lett*. 1998;427(1):91–5.
156. Grahn A, Larson G. Identification of nine alternatively spliced alpha2, 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*. 2001;18(10): 759–67.
157. Kapitonov D, Bieberich E, Yu RK. Combinatorial PCR approach to homology-based cloning: cloning and expression of mouse and human GM3-synthase. *Glycoconj J*. 1999;16(7):337–50.
158. Zeng G, Gao L, Xia T, Tencomnao T, Yu RK. Characterization of the 5′-flanking fragment of the human GM3-synthase gene. *Biochim Biophys Acta*. 2003;1625(1):30–5.
159. Kim KW, Kim SW, Min KS, Kim CH, Lee YC. Genomic structure of human GM3 synthase gene (hST3Gal V) and identification of mRNA isoforms in the 5′-untranslated region. *Gene*. 2001;273(2):163–71.
160. Kim SW, Lee SH, Kim KS, Kim CH, Choo YK, Lee YC. Isolation and characterization of the promoter region of the human GM3 synthase gene. *Biochim Biophys Acta*. 2002;1578(1–3):84–9.
161. Chung TW, Choi HJ, Lee YC, Kim CH. Molecular mechanism for transcriptional activation of ganglioside GM3 synthase and its function in differentiation of HL-60 cells. *Glycobiology*. 2005;15(3):233–44.
162. Xia T, Zeng G, Gao L, Yu RK. Sp1 and AP2 enhance promoter activity of the mouse GM3-synthase gene. *Gene*. 2005;351:109–18.

163. Matsuhashi H, Horii Y, Kato K. Region-specific and epileptogenic-dependent expression of six subtypes of alpha2,3-sialyltransferase in the adult mouse brain. *J Neurochem*. 2003;84(1):53–66.
164. Stern CA, Braverman TR, Tiemeyer M. Molecular identification, tissue distribution and subcellular localization of mST3GalV/GM3 synthase. *Glycobiology*. 2000;10(4):365–74.
165. 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*. 2001;287(5):1148–56.
166. Kurosawa N, Takashima S, Kono M, Ikehara Y, Inoue M, Tachida Y, et al. Molecular cloning and genomic analysis of mouse GalNAc alpha2,6-sialyltransferase (ST6GalNAc I). *J Biochem*. 2000;127(5):845–54.
167. Li GS, Zhu L, Zhang H, Lv JC, Ding JX, Zhao MH, et al. Variants of the ST6GALNAC2 promoter influence transcriptional activity and contribute to genetic susceptibility to IgA nephropathy. *Hum Mutat*. 2007;28(10):950–7.
168. Kurosawa N, Inoue M, Yoshida Y, Tsuji S. Molecular cloning and genomic analysis of mouse Galbeta1, 3GalNAc-specific GalNAc alpha2, 6-sialyltransferase. *J Biol Chem*. 1996;271(25):15109–16.
169. Takashima S, Kurosawa N, Tachida Y, Inoue M, Tsuji S. Comparative analysis of the genomic structures and promoter activities of mouse Siaalpha2,3Galbeta1,3GalNAc GalNAc alpha2,6-sialyltransferase genes (ST6GalNAc III and IV): characterization of their Sp1 binding sites. *J Biochem*. 2000;127(3):399–409.
170. Harduin-Lepers A, Stokes DC, Steelant WF, Samyn-Petit B, Krzewinski-Recchi MA, Vallejo-Ruiz V, et al. Cloning, expression and gene organization of a human Neu5Ac α2-3Galβ1-3GalNAc-α2,6-sialyltransferase: hST6GalNAc IV. *Biochem J*. 2000;352 (Pt 1):37–48.
171. Kang NY, Park YD, Choi HJ, Kim KS, Lee YC, Kim CH. Regulatory elements involved in transcription of the human NeuAcalpha2,3Galbeta 1,3GalNAc alpha2,6-sialyltransferase (hST6GalNAc IV) gene. *Mol Cells*. 2004;18(2):157–62.
172. Kim SW, Kang NY, Lee SH, Kim KW, Kim KS, Lee JH, et al. Genomic structure and promoter analysis of human NeuAc alpha2,3Gal beta1,3GalNAc alpha2,6-sialyltransferase (hST6GalNAc IV) gene. *Gene*. 2003;305(1):113–20.
173. Daniotti JL, Rosales Fritz V, Kunda P, Nishi T, Maccioni HJ. Cloning, characterization and developmental expression of alpha2,8-sialyltransferase (GD3 synthase, ST8Sia I) gene in chick brain and retina. *Int J Dev Neurosci*. 1997;15(6):767–76.
174. Yamamoto A, Yamashiro S, Fukumoto S, Haraguchi M, Atsuta M, Shiku H, et al. Site restricted and neuron dominant expression of alpha 2,8-sialyltransferase gene in the adult mouse brain and retina. *Glycoconj J*. 1996;13(3):471–80.
175. Haraguchi M, Yamashiro S, Yamamoto A, Furukawa K, Takamiya K, Lloyd KO, et al. Isolation of GD3 synthase gene by expression cloning of GM3 α2,8-sialyltransferase cDNA using anti-GD2 monoclonal antibody. *Proc Natl Acad Sci U S A*. 1994;91(22):10455–9.
176. Okada M, Furukawa K, Yamashiro S, Yamada Y, Haraguchi M, Horibe K, et al. High expression of ganglioside alpha-2,8-sialyltransferase (GD3 synthase) gene in adult T-cell leukemia cells unrelated to the gene expression of human T-lymphotropic virus type I. *Cancer Res*. 1996; 56(12):2844–8.
177. Sasaki K, Kurata K, Kojima N, Kurosawa N, Ohta S, Hanai N, et al. Expression cloning of a GM3-specific α2,8-sialyltransferase (GD3 synthase). *J Biol Chem*. 1994;269(22):15950–6.
178. Dae HM, Kwon HY, Kang NY, Song NR, Kim KS, Kim CH, et al. Isolation and functional analysis of the human glioblastoma-specific promoter region of the human GD3 synthase (hST8Sia I) gene. *Acta Biochim Biophys Sin (Shanghai)*. 2009;41(3):237–45.
179. Furukawa K, Horie M, Okutomi K, Sugano S, Furukawa K. Isolation and functional analysis of the melanoma specific promoter region of human GD3 synthase gene. *Biochim Biophys Acta*. 2003; 1627(2–3):71–8.
180. Kang NY, Kang SK, Lee YC, Choi HJ, Lee YS, Cho SY, et al. Transcriptional regulation of the human GD3 synthase gene expression in Fas-induced Jurkat T cells: a critical role of transcription factor NF-kappaB in regulated expression. *Glycobiology*. 2006;16(5):375–89.
181. Yoshida Y, Kurosawa N, Kanematsu T, Kojima N, Tsuji S. Genomic structure and promoter activity of the mouse polysialic acid synthase gene (mST8Sia II). Brain-specific expression from a TATA-less GC-rich sequence. *J Biol Chem*. 1996;271(47):30167–73.
182. Rieger S, Volkmann K, Koster RW. Polysialyltransferase expression is linked to neuronal migration in the developing and adult zebrafish. *Dev Dyn*. 2008;237(1):276–85.
183. Thisse B, Pflumio S, Fürthauer M, Loppin B, Heyer V, Degrave A, et al. Expression of the zebrafish genome during embryogenesis. 2001.
184. Yoshida Y, Kurosawa N, Kanematsu T, Taguchi A, Arita M, Kojima N, et al. Unique genomic structure and expression of the mouse alpha 2, 8-sialyltransferase (ST8Sia III) gene. *Glycobiology*. 1996;6(6):573–80.
185. Kim SJ, Chung TW, Jin UH, Suh SJ, Lee YC, Kim CH. Molecular mechanisms involved in transcriptional activation of the human Sia-alpha2, 3-Gal-beta1,4-GlcNAc-R:alpha2,8-sialyltransferase (hST8Sia III) gene induced by KCl in human glioblastoma cells. *Biochem Biophys Res Commun*. 2006;344(4):1057–64.
186. Eckhardt M, Gerardy-Schahn R. Genomic organization of the murine polysialyltransferase gene ST8SiaIV (PST-1). *Glycobiology*. 1998;8(12): 1165–72.
187. Takashima S, Yoshida Y, Kanematsu T, Kojima N, Tsuji S. Genomic structure and promoter activity of the mouse polysialic acid synthase (mST8Sia IV/PST) gene. *J Biol Chem*. 1998;273(13):7675–83.
188. Farre D, Bellora N, Mularoni L, Messeguer X, Alba MM. Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biol*. 2007;8(7):R140.
189. Studer RA, Robinson-Rechavi M. How confident can we be that orthologs are similar, but paralogs differ? *Trends Genet*. 2009;25(5):210–6.
190. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14(6):1188–90.
191. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*. 1990;18(20):6097–100.
192. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4(4):406–25.
193. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*. 1985;39:783–91.
194. Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel H, editors. Evolving genes and proteins. Academic Press ed. New York; 1965. p. 97–166.
195. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007;24(8):1596–9.
196. Gblocks, server. http://molevol.ibmb.csic.es/Gblocks_server.html.
197. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003;52(5): 696–704.
198. Thisse B, Heyer V, Lux A, Alunni V, Degrave A, Seiliez I, et al. Spatial and temporal expression of the zebrafish genome by large-scale *in situ* hybridization screening. *Methods Cell Biol*. 2004;77:505–19.
199. Thisse C, Thisse B. High resolution whole-mount *in situ* hybridization. In: Zebrafish Science Monitor. Vol. 5. Eugene, OR 97403-5274: University of Oregon Press; 1998. p. http://zfin.org.
200. Thisse C, Thisse B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc*. 2008;3(1):59–69.
201. Lee YC, Kojima N, Wada E, Kurosawa N, Nakaoka T, Hamamoto T, et al. Cloning and expression of cDNA for a new type of Gal beta 1,3GalNAc alpha 2,3-sialyltransferase. *J Biol Chem*. 1994;269(13):10028–33.
202. Kono M, Takashima S, Liu H, Inoue M, Kojima N, Lee YC, et al. Molecular cloning and functional expression of a fifth-type alpha 2, 3-sialyltransferase (mST3Gal V: GM3 synthase). *Biochem Biophys Res Commun*. 1998;253(1):170–5.

203. Miura Y, Kainuma M, Jiang H, Velasco H, Vogt PK, Hakomori S. Reversion of the Jun-induced oncogenic phenotype by enhanced synthesis of sialosyllactosylceramide (G_{M3} ganglioside). *Proc Natl Acad Sci U S A*. 2004;101(46):16204–9.
204. Carninci P, Hayashizaki Y. High-efficiency full-length cDNA cloning. *Methods Enzymol*. 1999;303:19–44.
205. Klein SL, Strausberg RL, Wagner L, Pontius J, Clifton SW, Richardson P. Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. *Dev Dyn*. 2002;225(4):384–91.
206. Bruses JL, Rutishauser U. Roles, regulation, and mechanism of polysialic acid function during neural development. *Biochimie*. 2001;83(7):635–43.