

The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach

Anne Harduin-Lepers^{1,2,4}, Rosella Mollicone^{3,4},
Philippe Delannoy^{2,4}, and Rafael Oriol^{3,4}

²Glycobiologie Structurale et Fonctionnelle, UMR CNRS/USTL 8576, Laboratoire de Chimie Biologique, Bâtiment C9, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq cedex, France;

³Glycobiologie et signalisation cellulaire, INSERM U 504, Université de Paris Sud XI, 16 Avenue, P. Vaillant-Couturier, 94807 Villejuif cedex, France; and
⁴GDR CNRS 2590, Génomique et génie des glycosyltransférases, France

Received on January 31, 2005; revised on March 18, 2005; accepted on April 12, 2005

The animal sialyltransferases are Golgi type II transmembrane glycosyltransferases. Twenty distinct sialyltransferases have been identified in both human and murine genomes. These enzymes catalyze transfer of sialic acid from CMP-Neu5Ac to the glycan moiety of glycoconjugates. Despite low overall identities, they share four conserved peptide motifs [L (large), S (small), motif III, and motif VS (very small)] that are hallmarks for sialyltransferase identification. We have identified 155 new putative genes in 25 animal species, and we have exploited two lines of evidence: (1) sequence comparisons and (2) exon–intron organization of the genes. An ortholog to the ancestor present before the split of ST6Gal I and II subfamilies was detected in arthropods. An ortholog to the ancestor present before the split of ST6GalNAc III, IV, V, and VI subfamilies was detected in sea urchin. An ortholog to the ancestor present before the split of ST3Gal I and II subfamilies was detected in ciona, and an ortholog to the ancestor of all the ST8Sia was detected in amphioxus. Therefore, single examples of the four families (ST3Gal, ST6Gal, ST6GalNAc, and ST8Sia) have appeared in invertebrates, earlier than previously thought, whereas the four families were all detected in bony fishes, amphibians, birds, and mammals. As previously hypothesized, sequence similarities among sialyltransferases suggest a common genetic origin, by successive duplications of an ancestral gene, followed by divergent evolution. Finally, we propose predictions on these invertebrates sialyltransferase-related activities that have not previously been demonstrated and that will ultimately need to be substantiated by protein expression and enzymatic activity assays.

Key words: animal sialyltransferases/phylogeny/gene organization/ortholog to a common ancestor

Introduction

Sialyltransferases are a subset of glycosyltransferases that use CMP-Neu5Ac as an activated sugar donor to catalyze

the transfer of sialic acid residues to terminal nonreducing positions of oligosaccharide chains of glycoproteins and glycolipids. They catalyze the formation of different linkages (α 2–3, α 2–6, and α 2–8) and differ in their acceptor specificities (for reviews, see Harduin-Lepers *et al.*, 1995, 2001; Takashima *et al.*, 2002a,b).

All vertebrate sialyltransferases have a similar architecture. They are type II transmembrane glycoproteins that predominantly reside in the *trans*-Golgi compartment. They have a short N-terminal cytoplasmic tail, a unique transmembrane domain, and a stem region of variable length from 20 to 200 amino acids followed by a large C-terminal catalytic domain. The vertebrate sialyltransferase amino acid sequences described up to date show overall limited sequence identity (from 15 to 57% for human sialyltransferases), but share four peptide conserved motifs called the sialylmotifs: L (large), S (small) (Drickamer, 1993; Livingston and Paulson, 1993), motif III (Jeanneau *et al.*, 2004), and motif VS (very small) (Geremia *et al.*, 1997; Jeanneau *et al.*, 2004). These motifs are involved in the formation of essential disulfide bonds and are implicated in the recognition of both donor and acceptor substrates (Datta and Paulson, 1995; Datta *et al.*, 1998) and in the catalytic activity (Jeanneau *et al.*, 2004). The sialylmotifs are hallmarks for the identification of eukaryotic sialyltransferase genes (Harduin-Lepers *et al.*, 2001).

Sialyltransferase genes are made up of multiple exons, and as it has been reviewed recently, they are widely dispersed in human (Harduin-Lepers *et al.*, 2001) and mouse genomes (Takashima *et al.*, 2003). Genome sequencing programs offer a new route into understanding multi-gene families both within a single species and across different species. All the animal sialyltransferases belong to the same CAZy 29 glycosyltransferase family (Coutinho *et al.*, 2003) that is based on the detection of common modules in protein sequences. The evolution of complex organisms has been associated with the generation of gene families by successive duplications of an initial relatively small set of ancestral genes. Through this process, followed by subsequent mutation, duplication and exon shuffling between gene families, genes have evolved both discrete and partially redundant functions with the other family members.

Although sialic acid residues are detected mainly in the deuterostome lineage (vertebrates, ascidians, echinoderms), sialyltransferases enzymatic activities have only been documented in mammals, birds, amphibians, bony fishes, and very recently, in *Drosophila melanogaster* (Koles *et al.*, 2004). By contrast, the evidence for sialylation in plants and protostomes (annelids, arthropods, and mollusks) has been scarce and controversial.

In this study, we have used the data available from various genome-sequencing programs to start to build a

¹To whom correspondence should be addressed; e-mail: anne.harduin@university-lille1.fr

foundation for understanding the evolution of the animal sialyltransferase family. As a result, many novel sialyltransferases genes were found. The purpose of this article is to analyze the sialyltransferase sequences that are, as previously mentioned (Harduin-Lepers *et al.*, 2001; Kaneko *et al.*, 2001), secondary to many gene duplications, which occurred early in vertebrate evolution. In addition, we provide evidence that the appearance of the four main sialyltransferase families has occurred among the ancestors of the present invertebrates.

Results

The ST3Gal family

All the described enzymes of this family transfer Neu5Ac residues in α 2,3-linkage to terminal galactose (Gal) residues found in glycoproteins or glycolipids. In this ST3Gal family, the ST3Gal I and II subfamilies use exclusively the type 3 oligosaccharide structure Gal β 1-3GalNAc-R, whereas the ST3Gal III, IV, V, and VI use the oligosaccharide isomers Gal β 1-3/4Glc(NAc)-R. Within this last group, the ST3Gal V subfamily uses exclusively the lactosyl-ceramide (Gal β 1-4Glc β 1-Cer) as an acceptor substrate giving rise to the synthesis of the ganglioside G_{M3}.

The two main branches of the phylogenetic tree of the ST3Gal family clearly separate ST3Gal I and II from ST3Gal III, IV, V, and VI subfamilies (Figure 1). In the first branch, the determination of subfamily-specific positions for the ST3Gal I and II clearly confirmed 12 gene products in the ST3Gal I subfamily (>60% ST3Gal I specific conserved positions) and 16 gene products in the ST3Gal II subfamily (>70% ST3Gal II specific conserved positions) (Figure 2), but two invertebrate hypothetical proteins (from the sea squirts *Ciona intestinalis* and *Ciona savignyi*) were identified that we could not ascribe to either of the ST3Gal I or ST3Gal II subfamilies, because they contained half of the positions specific for ST3Gal I and half of the positions specific for ST3Gal II (between 45 and 55%). The phylogenetic tree of the ST3Gal family is in good agreement with these findings and suggests that the branching of these two urochordate potential sialyltransferases occurred before the duplication at the origin of the present ST3Gal I and II subfamilies that are present in all the vertebrates studied. Therefore, we propose that the two ciona genes constitute a new group of orthologs of the common ancestor present before the split of ST3Gal I and II subfamilies that we will call ST3Gal I/II (Figure 1).

No other intermediate genes were found among the other members of the ST3Gal family (ST3Gal III, IV, V, or VI), and they all could be clearly classified in one of the four subfamilies by both methods: classical phylogeny (Figure 1) and the determination of the relative proportion of subfamily-specific positions (see online supplement data). The relative positions of the duplication events at the origin of each of the four subfamilies, of this second branch, cannot be unequivocally defined, because of the low bootstrap values at the root of the subfamilies, but a genomic organization unique for ST3Gal V suggests that this subfamily might be different from the other three (Figure 3A). This last observation is in good agreement with the fact that the ST3Gal V

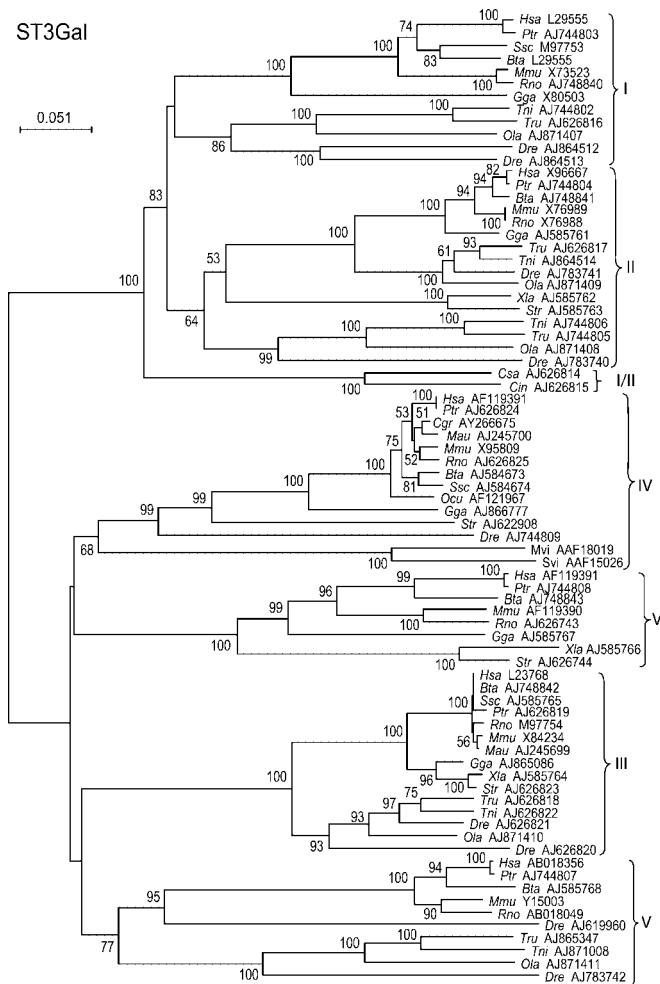


Fig. 1. Neighbor-joining phylogenetic tree of the 77 sialyltransferases of the ST3Gal family. One hundred and fifty-six of the 235 positions (60%) were selected in seven G-BLOCKS. Bootstrap values were calculated from 500 replicates, and values >50% are reported at the left of each divergence point. The scale bar represents the number of substitutions per site for a unit branch length. The two urochordate genes from *Ciona savignyi* (*Csa* AJ626814) and *Ciona intestinalis* (*Cin* AJ626815) are orthologs to the common ancestor (ST3Gal I/II) present before the split of ST3Gal I and II subfamilies.

subfamily is the only one able to use Gal β 1-4Glc β 1-Cer (lactosylceramide), whereas the other three subfamilies of this branch use Gal β 1-3/4GlcNAc-R as an acceptor substrate.

Two sialyltransferase genes were found in virus pathogenic for rabbits (*Mvi* AAF18019 and *Svi* AQAF15026). These genes encode α 2,3-sialyltransferase enzymes that have been shown to have catalytic activity toward the oligosaccharide Gal β 1-3/4GlcNAc-R (Jackson *et al.*, 1999). They belong to the ST3Gal IV subfamily (Figure 1), but they probably result from a retrotransposition followed by a horizontal transfer from the rabbit host to the virus, because the two are monoexonic sialyltransferase genes, and several complete genomes of the same family of virus have no other sialyltransferase genes detectable.

ST3Gal I		-----L-----		-----S-----		-----III-----		-----VS-----		%	% total																																																																									
<i>Bta</i>	L29555	RRC	AVV	SGN	GLR	ES	YGP	Q	IDS	HDF	VFL	RMN	K	A	P	T	A	G	F	E	D	V	G	R	K	T	H	H	L	V	P	E	S	R	E	L	A	N	V	S	M	V	L	V	P	F	K	T	D	L	E	W	V	I	S	A	T	T	G	I	S	H	T	Y	P	V	P	A	K	I														
<i>Mmu</i>	X73523	RRC	AVV	SGN	GLR	ES	YGP	Q	IDS	HDF	VFL	RMN	K	A	P	T	A	G	F	E	D	V	G	R	K	T	H	H	L	V	P	E	S	R	E	L	A	N	V	S	M	V	L	V	P	F	K	T	D	L	E	W	V	I	S	A	T	T	G	I	S	H	T	Y	P	V	P	A	K	I														
<i>Rno</i>	AJ748840	RRC	AVV	SGN	GLR	ES	YGP	Q	IDS	HDF	VFL	RMN	K	A	P	T	A	G	F	E	D	V	G	R	K	T	H	H	L	V	P	E	S	R	E	L	A	N	V	S	M	V	L	V	P	F	K	T	D	L	E	W	V	I	S	A	T	T	G	I	S	H	T	Y	P	V	P	A	K	I														
<i>Hsa</i>	L29555	RRC	AVV	SGN	GLR	ES	YGP	Q	IDS	HDF	VFL	RMN	K	A	P	T	A	G	F	E	D	V	G	R	K	T	H	H	L	V	P	E	S	R	E	L	A	N	V	S	M	V	L	V	P	F	K	T	D	L	E	W	V	I	S	A	T	T	G	I	S	H	T	Y	P	V	P	A	K	I														
<i>Ptr</i>	AJ744803	RRC	AVV	SGN	GLR	ES	YGP	Q	IDS	HDF	VFL	RMN	K	A	P	T	A	G	F	E	D	V	G	R	K	T	H	H	L	V	P	E	S	R	E	L	A	N	V	S	M	V	L	V	P	F	K	T	D	L	E	W	V	I	S	A	T	T	G	I	S	H	T	Y	P	V	P	A	K	I														
<i>\$Hsa</i>	AJ865084	WHC	AVV	SGN	GLR	Q	L	S	Y	R	P	E	A	L	X	H	N	F	M	L	R	M	N	K	A	P	T	A	G	F	E	A	A	A	G	S	K	T	A	H	L	V	P	E	S	R	E	L	A	N	V	S	M	V	L	V	P	L	K	T	M	N	L	E	W	V	S	T	T	T	G	A	I	S	H	T	Y	P	V	L	V	K	I	
<i>\$Ptr</i>	AJ865085	WHC	AVV	SGN	GLR	Q	L	S	Y	R	P	E	I	N	S	H	N	F	M	L	R	M	N	K	A	P	T	A	G	F	E	K	D	A	G	S	K	T	A	H	L	V	P	E	S	R	E	L	A	N	V	S	M	V	L	V	L	P	L	K	T	M	N	L	E	W	V	S	T	T	T	G	A	I	S	H	T	Y	P	V	L	V	K	I
<i>Ssc</i>	M97753	RRC	AVV	SGN	GLR	ES	YGP	Q	IDS	HDF	VFL	RMN	K	A	P	T	E	G	F	E	D	V	G	S	K	T	H	H	F	V	P	E	S	R	E	L	A	N	V	S	M	V	L	V	P	F	K	T	D	L	E	W	V	I	S	A	T	T	T	G	I	S	H	T	Y	P	V	P	A	K	I													
<i>Gga</i>	X80503	RRC	AVV	SGN	GLR	Q	S	Y	G	Q	D	I	S	H	D	F	V	L	R	M	N	K	A	P	T	I	G	E	S	D	V	G	S	K	T	H	H	F	V	P	E	S	R	E	L	A	N	V	S	M	V	I	P	F	K	T	D	L	R	V	I	T	A	L	T	T	G	I	N	F	T	Y	P	V	P	R	K							
<i>Dre</i>	AJ864512	RTC	AVV	SGN	GLR	L	S	N	Y	G	R	L	I	D	L	H	D	F	V	L	R	I	N	K	G	P	T	K	G	F	E	D	V	G	S	K	T	H	R	I	M	P	E	S	A	V	D	L	N	S	T	H	L	V	L	F	P	K	I	D	M	O	W	L	I	S	I	F	T	T	K	H	I	T	R	Y	M	S	V	E	S	T		
<i>Dre</i>	AJ864513	RIC	AVV	SGN	GLR	L	S	H	Y	G	L	I	D	S	H	D	F	V	I	R	I	N	K	G	P	T	K	G	F	E	D	V	G	S	K	T	H	R	I	M	P	E	S	A	V	D	L	N	S	T	H	L	V	L	F	P	K	I	D	M	O	W	L	I	S	I	F	T	T	K	H	I	T	R	Y	M	S	V	E	S				
<i>Ola</i>	AJ871407	RTC	AVV	SGN	GLR	K	S	N	Y	G	A	L	I	D	T	S	D	L	V	I	R	M	N	K	A	P	T	K	G	F	E	D	V	G	A	K	T	H	H	M	P	E	S	A	V	D	L	N	S	T	S	L	V	L	V	P	F	K	T	D	L	O	W	I	S	A	L	T	T	G	I	R	G	T	Y	M	P	V	K	S				
<i>Tru</i>	AJ626816	RTC	AVV	SGN	GLR	K	S	N	Y	G	R	L	I	D	S	S	D	F	I	R	M	N	K	A	P	T	S	G	F	E	D	V	G	T	R	T	H	H	M	P	E	S	A	K	D	L	N	S	T	S	L	V	L	I	P	F	K	T	D	L	O	W	I	S	A	L	T	T	G	A	I	K	H	T	Y	P	V	M	S					
<i>Tni</i>	AJ744802	RTC	AVV	SGN	GLR	K	S	N	Y	G	R	L	I	D	S	S	D	F	I	R	M	N	K	A	P	T	L	G	F	E	D	V	G	T	K	T	H	H	M	P	E	S	A	K	D	L	N	S	T	S	L	V	L	I	P	F	K	T	D	L	O	W	I	S	A	L	T	T	G	A	I	K	H	K	Y	V	P	V	M					
ST3Gal I/II		-----L-----		-----S-----		-----III-----		-----VS-----																																																																												
<i>Cin</i>	AJ626815	RSC	AVV	SGN	GLR	L	S	N	Y	G	N	V	I	D	G	H	D	F	V	I	R	M	N	K	A	P	T	Y	N	E	N	D	V	G	S	K	T	H	R	F	M	Y	P	T	A	S	S	L	P	G	K	L	F	V	L	V	P	F	Q	P	L	D	I	K	W	L	S	A	L	T	T	G	E	I	T	Y	Q	P	L	V	R	V		
<i>Csa</i>	AJ626814	RRC	AVV	SGN	GLR	L	S	K	Y	G	K	N	I	D	S	H	D	F	V	I	R	M	N	K	A	P	T	E	G	F	E	D	V	G	R	K	T	H	R	F	M	Y	P	A	T	A	S	S	L	P	G	V	S	L	V	L	P	F	Q	P	D	V	K	W	L	S	A	L	T	T	G	E	L	N	T	Y	Q	P	V	I	S			
ST3Gal II		-----L-----		-----S-----		-----III-----		-----VS-----																																																																												
<i>Ola</i>	AJ871408	RSC	AVV	SGN	GLR	L	S	H	G	E	R	I	D	S	H	S	F	V	I	R	M	N	K	A	P	T	R	G	F	E	K	D	V	G	N	R	T	H	H	F	L	P	E	S	A	V	D	V	D	G	G	V	S	L	V	L	P	F	K	L	R	D	E	W	L	T	S	A	L	S	T	G	H	I	K	T	Y	M	R	V	K	E	R	
<i>Tru</i>	AJ744805	HTC	AVV	SGN	GLR	L	S	H	S	V	L	R	M	N	K	A	P	T	R	G	F	E	K	D	V	G	N	R	T	H	H	F	L	P	E	S	A	V	D	V	D	G	G	V	S	L	V	L	P	F	K	L	R	D	E	W	L	T	S	A	L	S	T	G	K	V	K	M	Y	M	R	V	K	D										
<i>Tni</i>	AJ744806	HTC	AVV	SGN	GLR	L	S	H	S	V	L	R	M	N	K	A	P	T	R	G	F	E	K	D	V	G	N	R	T	H	H	F	L	P	E	S	A	V	D	V	D	G	G	V	S	L	V	L	P	F	K	L	R	D	E	W	L	T	S	A	L	S	T	G	K	V	K	M	Y	M	R	V	K											
<i>Dre</i>	AJ783740	RKC	AVV	SGN	GLR	L	S	K	Y	G	A	L	I	D	S	H	S	T	V	I	R	M	N	K	A	P	T	V	G	Y	D	E	D	V	G	R	T	H	H	F	L	P	E	S	A	I	H	L	R	P	G	V	H	L	V	L	P	F	K	L	K	D	M	O	W	L	S	S	A	L	S	T	G	E	I	K	M	T	Y	M	R	V		
<i>Str</i>	AJ585763	RRC	AVV	SGN	GLR	K	S	R	H	G	K	N	I	D	S	H	G	L	I	F	R	M	N	K	A	P	T	G	F	E	K	D	V	G	S	K	T	H	H	F	M	P	E	S	A	V	H	V	Q	P	G	V	H	L	V	L	I	P	F	K	L	Q	D	L	K	W	I	S	A	L	T	T	G	E	V	T	Y	T	R	M				
<i>Xla</i>	AJ585762	RRC	AVV	SGN	GLR	K	S	R	H	G	K	N	I	D	S	H	G	L	I	F	R	M	N	K	A	P	T	G	F	E	K	D	V	G	S	K	T	H	H	F	M	P	E	S	A	V	N	L	Q	P	G	V	H	L	V	L	I	P	F	K	L	Q	D	L	K	W	I	S	A	L	T	T	G	E	V	T	Y	T						
<i>Bta</i>	AJ748841	RRC	AVV	SGN	GLR	S	G	Y	G	E	D	V	D	G	H	N	F	I	R	M	N	K	A	P	T	V	G	F	E	D	V	G	S	R	T	H	H	F	M	P	E	S	A	K	N	L	P	A	N	V	S	F	V	L	V	P	F	K	A	L	D	L	L	W	I	A	S	A	L	S	T	G	I	R	F	T	Y	A						
<i>Tru</i>	AJ626817	RRC	AVV	SGN	GLR	S	G	Y	G	E	D	V	D	G	H	N	F	I	R	M	N	K	A	P	T	V	G	F	E	D	V	G	S	R	T	H	H	F	M	P	E	S	A	K	N	L	P	A	N	V	S	F	V	L	V	P	F	K	A	L	D	L	L	W	I	A	S	A	L	S	T	G	I	R	F	T	Y							
<i>Gga</i>	AJ585761	RRC	AVV</																																																																																	

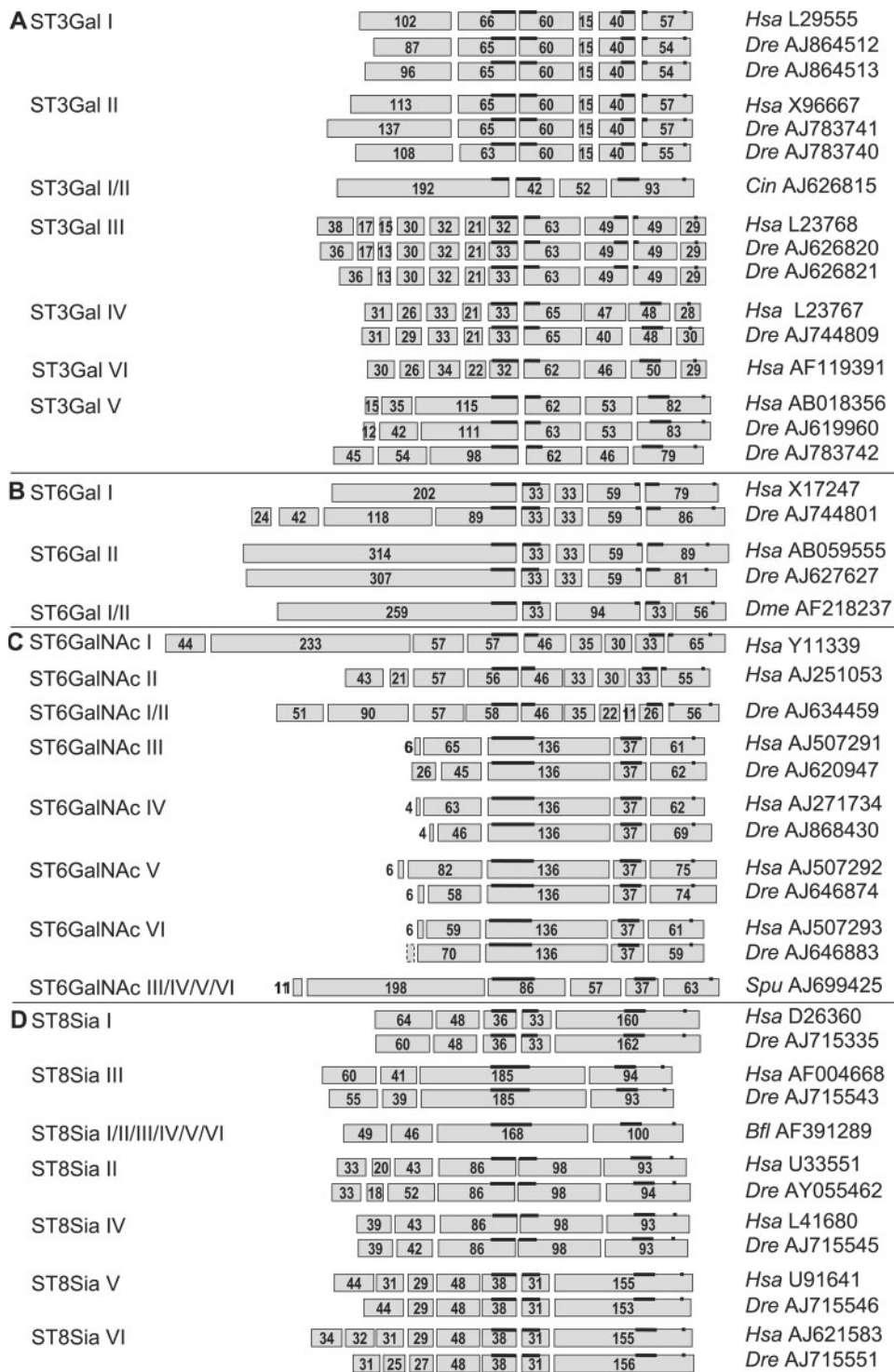


Fig. 3. Schematic showing the genomic organization of sialyltransferase genes. Coding exons are represented by rectangles with their relative sizes in amino acids indicated within. Each subfamily is indicated on the left side and the accession number in EMBL/GenBank on the right side. In the ST3Gal family (A), a similar genomic organization was found for the ST3Gal I and II subfamilies, and a similar genomic organization was also found for the ST3Gal III, IV, and VI subfamilies. The genomic organization of ST3Gal V is unique and different from its closest neighbors, the ST3Gal III, IV, and VI subfamilies. Two copies of *Danio rerio* genes are present in ST3Gal I, II, and III subfamilies. (ST3Gal I: *Dre* AJ864512, *Dre* AJ864513; ST3Gal II: *Dre* AJ783740, *Dre* AJ783741; and ST3Gal III: *Dre* AJ626820, *Dre* AJ626821). In the ST6Gal family (B), the ST6Gal I/II *Dme* AF218237 gene shares a similar genomic organization with the ST6Gal I and II subfamilies. In the ST6GalNAc family (C), the ST6GalNAc I/II *Dre* AJ634459 gene shares a similar genomic organization with ST6GalNAc I and II subfamilies and the ST6GalNAc III, IV, V, and VI subfamilies share also similar genomic organization. In the ST8Sia family (D), the ST8Sia I/II/III/IV/V/VI from amphioxus (*Bfl* AF391289) has similar genomic organization with the ST8Sia III subfamily. The ST8Sia II and IV subfamilies share similar genomic organization, and finally, the ST8Sia V and VI subfamilies share also similar genomic organization.

Two sialyltransferase pseudogenes with several premature stop codons, and with a continuous DNA sequence devoid of introns, were found in human chromosome 4 (*Hsa* AJ865084) and chimpanzee chromosome 3 (*Ptr* AJ865085) (Figure 2). These two pseudogenes were ascribed to the ST3Gal I subfamily, but were not included in the phylogeny analysis.

The ST6Gal family

The enzymes of this family comprise only two subfamilies, ST6Gal I and II, that both use the Gal β 1-4GlcNAc-R as the acceptor substrate. Members of both subfamilies were found in all vertebrates from fish to man (Figure 4). An ST6Gal cDNA cloned from *D. melanogaster* (Koles *et al.*, 2004) suggested that the ST6Gal family was present in insects. This fact is strongly supported by the finding of ST6Gal putative genes in two other flies, *Drosophila yakuba* (*Dya* AJ821848) and *Drosophila pseudoobscura* (*Dps* AJ821849), and in the mosquito *Anopheles gambiae* (*Aga* AJ821850). All these insect potential sialyltransferases branch out from the tree, before the split of ST6Gal I and II subfamilies, illustrating that the duplication at the origin of the present ST6Gal I and II subfamilies occurred after the separation of insects from the common evolutionary trunk, and before the appearance of vertebrates. This was

confirmed by the determination of the proportion of subfamily-specific positions in the 2 \times 2 clustal alignments (Figure 5), which show that the four insect gene products cannot be ascribed to either of ST6Gal I or II subfamilies, because they contain between 48 and 61% specific positions of both subfamilies. However, they belong to the ST6Gal family, because they are more closely related to this family than to any of the other sialyltransferase families, they constitute a new group ortholog to the common ancestor (ST6Gal I/II) that was present before the split of ST6Gal I and II subfamilies (Figure 4).

The ST6GalNAc family

The enzymes of this family catalyze the transfer of Neu5Ac residues in α 2-6 linkage to the *N*-acetylgalactosamine (GalNAc) residues found in *O*-glycosylproteins (ST6GalNAc I, II, and IV) or found in glycolipids (ST6GalNAc III, V, and VI). Interestingly, ST6GalNAc I and II catalyze the transfer of Neu5Ac onto Gal β 1-3GalNAc peptides (sialylated or not), and their activity greatly depends on the peptide moiety, whereas ST6GalNAc III, IV, V, and VI exhibit a more restricted substrate specificity, only utilizing sialylated acceptor substrates (Neu5Ac α 2-3Gal β 1-3GalNAc-R), found either in glycoproteins or glycolipids such as G_{M1b}. In good agreement with these different substrate specificities, the phylogenetic tree of the ST6GalNAc family shows two main branches, the first containing ST6GalNAc I and II and the second containing ST6GalNAc III, IV, V, and VI subfamilies (Figure 6).

Several hypothetical sialyltransferases of this family had intermediate proportions of subfamily conserved specific amino acids and could not be clearly ascribed to any of the six subfamilies. In the first branch of the ST6GalNAc family, the sialyltransferases of five species of bony fish: *Oncorhynchus mykiss* (*Omy* AB097943), *Danio rerio* (*Dre* AJ634459), *Takifugu rubripes* (*Tru* AJ634460 and *Tru* AJ634461), *Tetraodon nigroviridis* (*Tni* AJ634462), and *Oryzias latipes* (*Ola* AJ871602) branch out before the split of ST6GalNAc I and II subfamilies and have between 26 and 74% specific conserved positions of both subfamilies (online supplement data), suggesting that they are orthologs of the ancestor (ST6GalNAc I/II) present before the split of ST6GalNAc I and II. This allows us to assign the duplication at the origin of these two subfamilies after the appearance of fish and before the appearance of amphibians, because the two ST6GalNAc I and II subfamilies were clearly identified in amphibians, birds, and mammals.

In the second branch of the ST6GalNAc family containing the remaining four subfamilies, the sea urchin *Strongylocentrotus purpuratus* putative sialyltransferase gene (*Spu* AJ699425) is clearly branching out before the occurrence of the duplications at the origin of the four subfamilies and contains 42 and 58% of subfamily ST6GalNAc III and IV specific conserved positions (Figure 7). A similar result was obtained with the subfamily-specific conserved positions of ST6GalNAc V and VI (45 and 55%, online supplement data), suggesting that the *S. purpuratus* gene is an ortholog to the common ancestor (ST6GalNAc III/IV/V/VI) present before the separation of these four subfamilies.

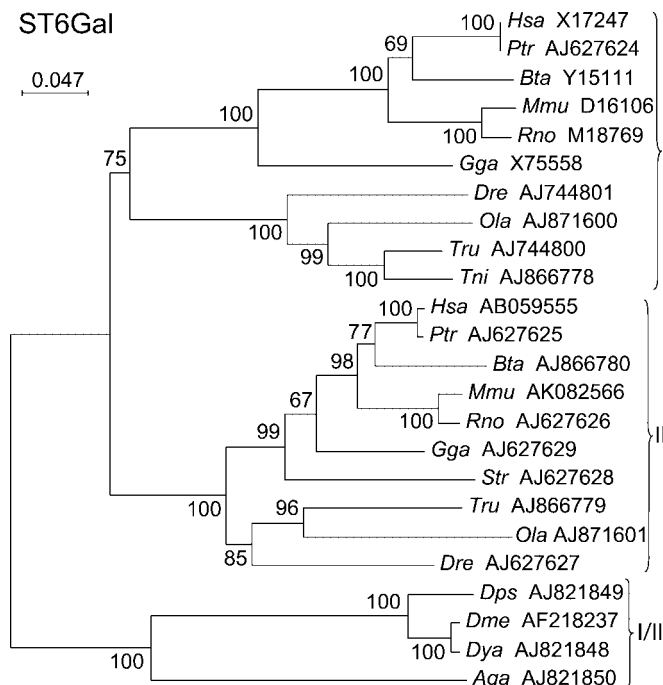


Fig. 4. Neighbor-joining phylogenetic tree of the 24 sialyltransferases of the ST6Gal family. Two hundred and seventy-five of the 338 positions (81%) were selected in 11 G-BLOCKS. Bootstrap values were calculated from 500 replicates, and values >50% are reported at the left of each divergence point. The scale bar represents the number of substitutions per site for a unit branch length. The four insect sialyltransferases from *Drosophila pseudoobscura* (*Dps* AJ821849), *Drosophila melanogaster* (*Dme* AF218237), *Drosophila yakuba* (*Dya* AJ821848), and *Anopheles gambiae* (*Aga* AJ821850) are orthologs to the common ancestor (ST6Gal I/II) present before the split of ST6Gal I and II subfamilies.

ST6Gal I		-L-																																																																																				
Hsa X17247	WGRC	CAVVS	SAGSLKSS																																																																																			
Bta Y15111	WGRC	CAVVS	SAGSLKSS																																																																																			
Ptr AJ627624	WGRC	CAVVS	SAGSLKSS																																																																																			
Dre AJ744801	FKTC	CAVVS	SAGSLKSS																																																																																			
Mmu D16106	CTRC	CAVVS	SAGSLKSS																																																																																			
Rno M18769	WQRC	CAVVS	SAGSLKSS																																																																																			
Tru AJ744800	YKSC	CAVVS	SAGSRSS																																																																																			
Hsa AJ866778	FRSC	CAVVS	SAGSLRSS																																																																																			
Gga X75558	LGRC	CAVVS	SAGSLRSS																																																																																			
Ola AJ871600	LKSC	CAVVS	SAGSLRSS																																																																																			
ST6Gal I/II		-L-																																																																																				
Aga AJ821850	NGSC	CVIVAS	SAGSLRSS																																																																																			
Dya AJ821848	IKTC	CAIVS	SAGSLRSS																																																																																			
Dps AJ821849	IKTC	CAIVS	SAGSLRSS																																																																																			
Dme AF218237	IKTC	CAIVS	SAGSLRSS																																																																																			
ST6Gal II		-L-																																																																																				
Ola AJ871601	FRTC	CAVVS	SAGAILHSS																																																																																			
Dre AJ627627	LKTC	CAVVS	SAGAILHSS																																																																																			
Tru AJ866779	FKSC	CAVVS	SAGAILRSS																																																																																			
Bta AJ627628	FRTC	CAVVS	SAGAILNSSL																																																																																			
Str AJ866780	LRTCA	CAVVS	SAGAILNSSL																																																																																			
Gga AJ627629	FGSC	CAVVS	SAGAILNSSL																																																																																			
Ptr AJ627625	LRSC	CAVVS	SAGAILNSSL																																																																																			
Rno AJ627626	LRSC	CAVVS	SAGAILNSSL																																																																																			
Hsa AB059555	LRSC	CAVVS	SAGAILNSSL																																																																																			
Mmu AK082566	LRSC	CAVVS	SAGAILNSSL																																																																																			
ST6Gal I		-S-		-III		--VS--		% total																																																																														
Hsa X17247	RKLL	HPQ	PPFY	ILK	QMP	EWEL	WDIL	Q	EIS	P	EE	-IQN	PPSS	GLG	I	I	M	T	L	C	D	Q	V	I	Y	E	F	L	P	S	R	R	T	D	V	C	Y	Y	Q	K	F	F	D	S	A	C	T	M	G	A	Y	H	P	L	L	E	K	100	0	32																										
Bta Y15111	RKLL	HPD	Q	PPFY	ILK	QMP	EWEL	WDI	I	Q	EIS	SE	-IQN	PPSS	GLG	I	A	I	M	S	L	C	D	Q	V	I	Y	E	F	L	P	S	R	R	T	D	V	C	Y	Y	Q	K	F	F	D	S	A	C	T	M	G	A	Y	H	P	L	L	E	K	100	0	30																								
Ptr AJ627624	RKLL	HPN	Q	PPFY	ILK	QMP	EWEL	WDI	L	Q	EIS	PE	-IQN	PPSS	GLG	I	I	M	T	L	C	D	Q	V	I	Y	E	F	L	P	S	R	R	T	D	V	C	Y	Y	Q	K	F	F	D	S	A	C	T	M	G	A	Y	H	P	L	L	E	K	100	0	32																									
Dre AJ744801	RRLH	P	Q	PPFY	I	VH	P	R	M	E	W	L	Q	R	I	X	D	N	M	G	E	A	-IQN	PPSS	GLG	I	G	T	V	L	M	S	L	C	E	V	H	V	Y	E	F	L	P	S	R	R	T	D	V	C	Y	Y	Q	K	F	F	D	S	A	C	T	M	G	A	Y	H	P	L	L	E	K	100	0	22												
Mmu D16106	RRLH	P	S	Q	PPFY	ILK	QMP	EWEL	WDI	I	Q	EIS	PD	-IQN	PPSS	GLG	I	I	M	T	L	C	D	Q	V	I	Y	E	F	L	P	S	R	R	T	D	V	C	Y	Y	Q	K	F	F	D	S	A	C	T	M	G	A	Y	H	P	L	L	E	K	97	3	33																								
Rno M18769	RRLN	P	S	Q	PPFY	ILK	QMP	EWEL	WDI	I	Q	EIS	AD	-IQN	PPSS	GLG	I	I	M	T	L	C	D	Q	V	I	Y	E	F	L	P	S	R	R	T	D	V	C	Y	Y	Q	K	F	F	D	S	A	C	T	M	G	A	Y	H	P	L	L	E	K	97	7	32																								
Tru AJ744800	RMLH	P	M	Q	PPFY	ILH	P	R	F	E	W	Q	R	I	Q	D	N	M	A	E	P	-IQN	PPSS	GLG	I	G	T	V	M	S	L	C	E	V	H	V	Y	E	F	L	P	S	R	R	T	D	V	C	Y	Y	Q	K	F	F	D	S	A	C	T	M	G	A	Y	H	P	L	L	E	K	91	9	22														
Hsa AJ866778	RMLH	P	L	Q	PPFY	IL	Q	R	F	E	W	Q	L	W	Q	D	N	M	A	E	P	-IQN	PPSS	GLG	I	G	T	V	M	S	L	C	E	V	H	V	Y	E	F	L	P	S	R	R	T	D	V	C	Y	Y	Q	K	F	F	D	S	A	C	T	M	G	A	Y	H	P	L	L	E	K	88	12	24														
Gga X75558	RIRH	P	Q	PPFY	ILN	P	K	M	E	W	Q	L	W	I	L	Q	E	N	S	L	E	H	-IQN	PPSS	GLG	I	G	I	V	M	T	L	C	E	V	H	V	Y	E	F	L	P	S	R	R	T	D	V	C	Y	Y	Q	K	F	F	D	S	A	C	T	M	G	A	Y	H	P	L	L	E	K	79	21	28													
Ola AJ871600	RRLH	P	Q	PPFY	ILH	P	R	F	E	W	Q	V	Q	V	E	N	M	A	E	S	-IQN	PPSS	GLG	I	G	T	V	L	M	S	L	C	E	V	H	V	Y	E	F	L	P	S	R	R	T	D	V	C	Y	Y	Q	K	F	F	D	S	A	C	T	M	G	A	Y	H	P	L	L	E	K	75	25	24														
ST6Gal I/II		-S-		-III		--VS--																																																																																
Aga AJ821850	RSSH	P	Q	S	N	F	H	I	D	P	R	S	I	W	R	A	T	A	L	Q	D	L	D	L	P	-IRK	N	P	P	T	S	G	F	I	G	L	G	L	L	P	V	C	R	I	D	V	E	Y	I	P	S	T	R	M	N	G	L	C	H	Y	D	D	Q	L	N	L	G	T	F	G	A	W	H	P	L	A	E	K	61	49	18					
Dya AJ821848	RRRY	P	K	S	R	A	F	L	I	D	P	H	S	V	W	R	L	Q	S	L	Q	M	F	A	G	N	R	P	I	S	R	N	P	P	S	S	G	F	I	G	L	A	L	L	P	H	C	P	Q	V	E	Y	I	P	S	T	R	L	N	G	R	C	H	Y	S	K	E	M	N	S	A	C	T	F	G	S	W	H	P	L	A	E	K	50	50	21
Dps AJ821849	RRRY	P	K	S	R	A	F	L	I	D	P	H	S	V	W	R	L	Q	S	L	Q	M	F	A	G	N	R	P	I	S	R	N	P	P	S	S	G	F	I	G	L	A	L	L	P	H	C	P	Q	V	E	Y	I	P	S	T	R	L	N	G	R	C	H	Y	S	K	E	M	N	S	A	C	T	F	G	S	W	H	P	L	A	E	K	48	52	20
Dme AF218237	RRRY	P	K	S	R	A	F	L	I	D	P	H	S	V	W	R	L	Q	S	L	Q	M	F	A	G	N	R	P	I	S	R	N	P	P	S	S	G	F	I	G	L	A	L	L	P	H	C	P	Q	V	E	Y	I	P	S	T	R	L	N	G	R	C	H	Y	S	K	E	M	N	S	A	C	T	F	G	S	W	H	P	L	A	E	K	48	52	21
ST6Gal II		-S-		-III		--VS--																																																																																
Ola AJ871601	RKNH	A	E	Q	L	F	Y	I	L	H	P	S	Y	L	W	L	W	D	L	I	Q	S	N	T	Q	E	K	-IQN	PPSS	GLG	I	G	I	L	T	M	A	L	C	D	K	L	H	V	Y	E	I	P	S	M	R	O	T	D	L	C	H	Y	H	E	N	Y	D	A	A	C	T	L	G	A	Y	H	P	L	L	E	K	25	75	28						
Dre AJ627627	RMRP	F	S	Q	PPFY	ILH	P	K	Y	I	W	L	Q	L	W	D	V	I	Q	A	N	N	L	E	N	-IQN	PPSS	GLG	I	G	I	L	M	S	L	C	E	V	H	V	Y	E	I	P	S	L	R	O	T	D	L	C	H	Y	H	E	R	Y	D	A	A	C	T	L	G	A	Y	H	P	L	L	E	K	19	81	27										
Tru AJ866779	RRAH	P	Q	PPFY	ILH	P	K	Y	I	W	L	Q	L	W	D	V	I	Q	A	N	N	L	E	N	-IQN	PPSS	GLG	I	G	I	L	M	S	L	C	E	V	H	V	Y	E	I	P	S	M	R	O	T	D	L	C	H	Y	H	E	R	Y	D	A	A	C	T	L	G	A	Y	H	P	L	L	E	K	13	87	23											
Str AJ627628	RKRN	P	Q	PPFY	ILH	P	K	F	I	W	L	Q	L	W	D	I	I	Q	E	N	S	N	E	K	-IQN	PPSS	GLG	I	G	I	L	M	S	L	C	R	T	V	H	V	Y	E	I	P	S	Y	R	O	T	D	L	C	H	Y	H	E	Q	Y	D	A	A	C	T	L	G	A	Y	H	P	L	L	E	K	4	96	27										
Bta AJ866780	RQRN	P	Q	PPFY	ILH	P	K	F	I	W	L	Q	L	W	D	I	I	Q	E	N	T	K	E	K	-IQN	PPSS	GLG	I	G	I	L	M	S	L	C	E	V	H	V	Y	E	I	P	S	V	R	O	T	D	L	C	H	Y	H	E	P	Y	D	A	A	C	T	L	G	A	Y	H	P	L	L	E	K	3	97	36											
Gga AJ627629	RKRN	P	Q	PPFY	ILH	P	K	F	I	W	L	Q	L	W	D	I	I	Q	E	N	T	K	E	K	-IQN	PPSS	GLG	I	G	I	L	M	S	M	C	R	E	V	H	V	Y	E	I	P	S	V	R	O	T	D	L	C	H	Y	H	E	L	Y	D	A	A	C	T	L	G	A	Y	H	P	L	L	E	K	0	100	35										
Ptr AJ627625	RQRN	P	Q	PPFY	ILH	P	K	F	I	W	L	Q	L	W	D	I	I	Q	E	N	T	K	E	K	-IQN	PPSS	GLG	I	G	I	L	M	S	M	C	R	E	V	H	V	Y	E	I	P	S	V	R	O	T	D	L	C	H	Y	H	E	L	Y	D	A	A	C	T	L	G	A	Y	H	P	L	L	E	K	0	100	37										
Rno AJ627626	RLKY	P	T	Q	PPFY	ILH	P	K	F	I	W	L	Q	L	W	D	I	I	Q	E	N	T	R	E	K	-IQN	PPSS	GLG	I	G	I	L	M	S	M	C	Q	E	V	H	V	Y	E	I	P	S	V	R	O	T	D	L	C	H	Y	H	E	L	Y	D	A	A	C	T	L	G	A	Y	H	P	L	L	E	K	0	100	36									
Hsa AB059555	RQRN	P	Q	PPFY	ILH	P	K	F	I	W	L	Q	L																																																																									

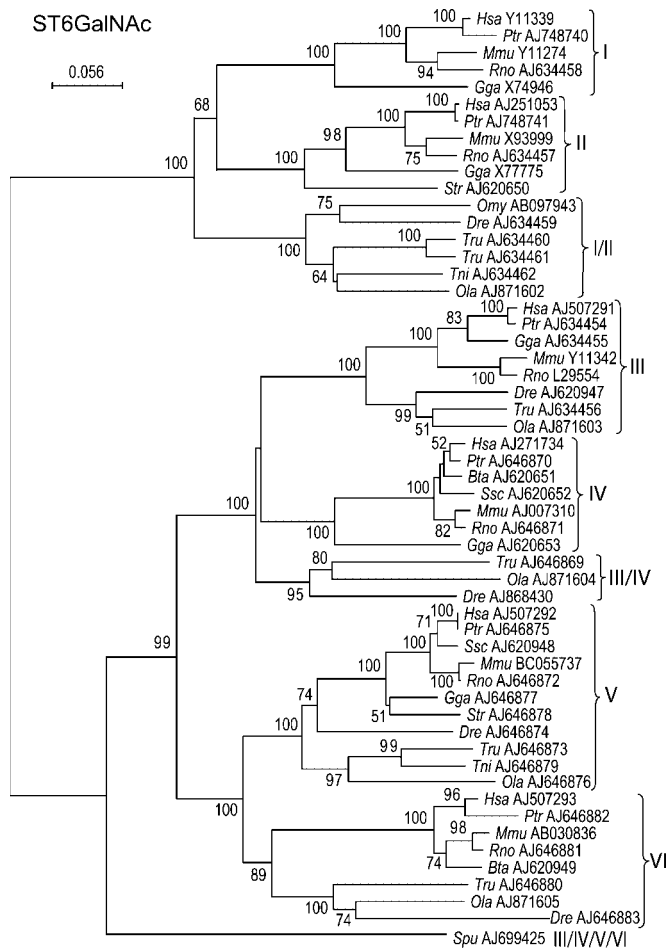


Fig. 6. Neighbor-joining phylogenetic tree of the 55 sialyltransferases of the ST6GalNAc family. One hundred and seventy-two of 247 positions (70%) were selected in eight G-BLOCKS. Bootstrap values were calculated from 500 replicates, and values above 50% are reported on the left of each divergence point. The scale bar represents the number of substitutions per site for a unit branch length. Six genes from the bony fish *Oncorhynchus mykiss* (Omy AB097943), *Danio rerio* (Dre AJ634459), *Takifugu rubripes* (Tru AJ634460 and Tru AJ634461), *Tetraodon nigroviridis* (Tni AJ634462), and *Oryzias latipes* (Ola AJ871602) are orthologs to the common ancestor (ST6GalNAc I/II) present before the split of ST6GalNAc I and II subfamilies. The gene from the sea urchin *Strongylocentrotus purpuratus* (Spu AJ699425) is ortholog to the common ancestor (ST6GalNAc III/IV/V/VI) present before the separation of ST6GalNAc III, IV, V, and VI subfamilies. Three genes from the bony fish *T. rubripes* (Tru AJ646869), *O. latipes* (Ola AJ871604), and *D. rerio* (Dre AJ868430) branch out just before the split of ST6GalNAc III and IV with a nonsignificant low bootstrap (<50%). This fact plus the lack of other ST6GalNAc IV genes from bony fish suggest that these three sialyltransferase sequences belong to the bony fish ST6GalNAc IV subfamily.

Discussion

Phylogeny calculations are based on the differences among sequences, as opposed to the percentages of conserved positions that are directly related to similarities among subfamily-specific conserved positions. However, these two approaches are complementary and point toward the same conclusions.

Sialyltransferase activities have been detected in prokaryotes (Gilbert *et al.*, 1996, 2000; Yamamoto *et al.*, 1998; Shen *et al.*, 1999; Hood *et al.*, 2001; Jones *et al.*, 2002), but these activities are carried out by enzymes devoid of the four sialylmotifs, and therefore, they are not directly related to the eukaryote sialyltransferases. Three *Arabidopsis thaliana* and three *Oryza sativa* sialyltransferase-like sequences are present in the CAZy family 29 of glycosyltransferases. They contain the sialylmotifs L and S, but the last two motifs (III and VS) are difficult to identify, and, even if the presence of sialylated structures has been reported in plants (Shah *et al.*, 2003), their hypothetical sialyltransferase activity is still a matter of debate (Séveno *et al.*, 2004). In contrast, all the known animal sialyltransferases have the four sialylmotifs suggesting that they have evolved by successive duplications from a common ancestral gene followed by divergent evolution, similar to the evolutionary model proposed for the fucosyltransferases (Oriol *et al.*, 1999).

As Figure 10 illustrates, the four main animal families of sialyltransferases were detected in invertebrates, but none of these invertebrate genes could be ascribed to a specific sialyltransferase subfamily, suggesting that the duplication events at the origin of the present 20 subfamilies of sialyltransferases had occurred early in the vertebrate lineage, before the emergence of amphibians because many subfamilies are detected in bony fish. No Neu5Ac and no sialyltransferase-like sequences were found in the genomes of nematodes (*Caenorhabditis elegans*), suggesting that the first precursor gene of the animal sialyltransferase series might be the ST6Gal from insects, and no other ST6Gal genes were found in other invertebrates. In a similar way, the first precursor gene of the ST6GalNAc family was found in *S. purpuratus*, and no other ST6GalNAc genes were found in other invertebrates. The first precursor gene of the ST3Gal family was found in ciona, and no other ST3Gal genes were found in other invertebrates. The first precursor gene of the ST8Sia family was found in *B. floridae* (amphioxus), and no other genes of the ST8Sia family were found in other invertebrates (Figure 10). All these findings suggest that the sialyltransferases have appeared earlier in evolution than previously thought (Schauer, 1982; Varki, 1993; Angata and Varki, 2002). In addition, there is an apparent trend for the appearance of the sialyltransferase activities, starting with α 2,6 and followed by α 2,3 and finally α 2,8 linkages, but we can not define a more precise order, probably because we have not yet found many missing links of this chain. This is particularly evident for the ST8Sia family, because the enzymes of this family catalyze the transfer of Neu5Ac residues on the terminal Neu5Ac of sialylated acceptors, and therefore, other α 2,3 and/or α 2,6 sialyltransferase activities have to be present in amphioxus to build the corresponding sialylated acceptors for this hypothetical ST8Sia enzyme. Another possibility would be a potential multifunctional nature of these sialyltransferases that would be capable of catalyzing the formation of various α 2,3, α 2,6, or α 2,8 linkages as described for a *Neisseria meningitidis* sialyltransferase (Wakarchuk *et al.*, 2001).

Traditionally, the animal sialyltransferase superfamily has been divided into four families: ST3Gal, ST6Gal,

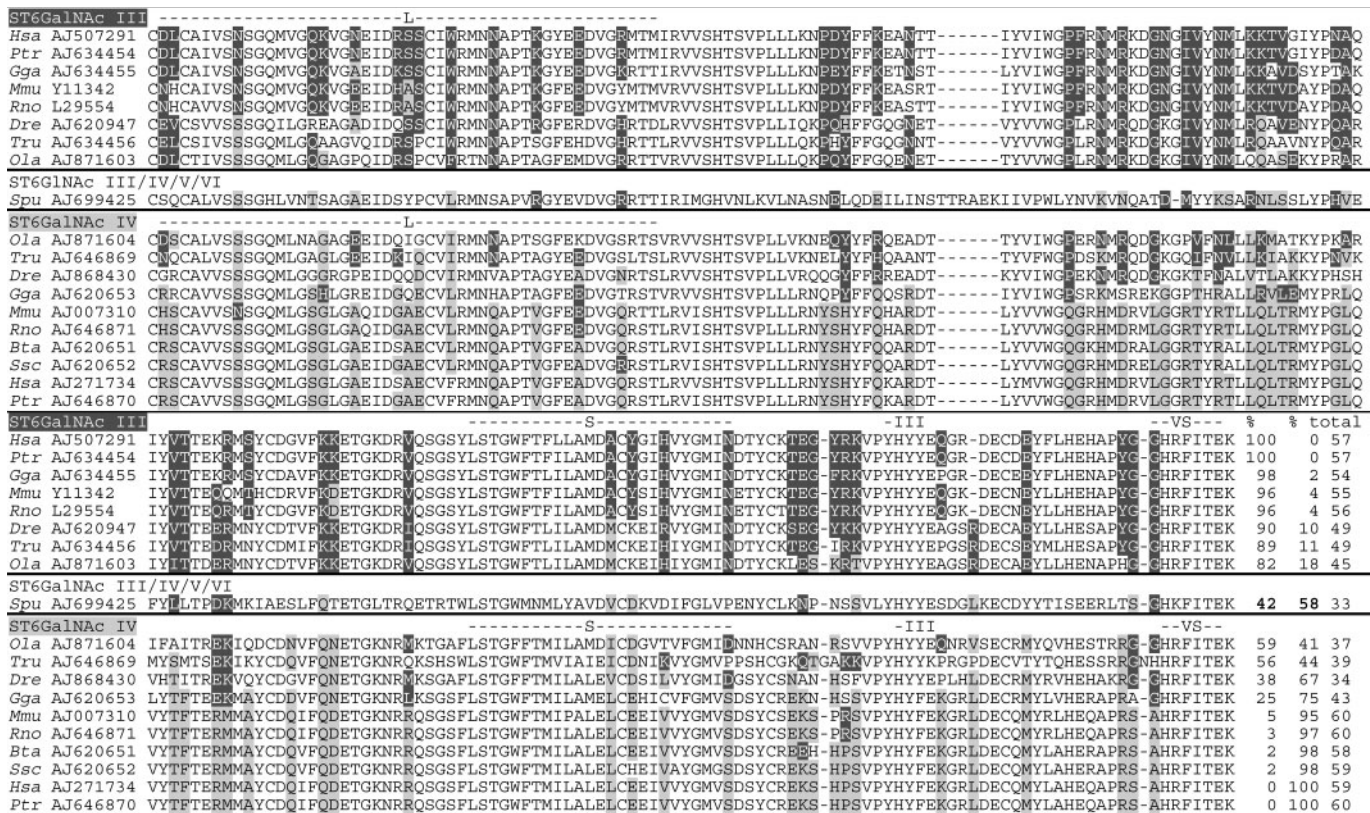


Fig. 7. ClustalW alignments of the sialylmotif regions of ST6GalNAc III and IV subfamilies. White letters with dark gray background represent conserved positions specific from ST6GalNAc III, and black letters on pale gray background represent conserved positions specific from ST6GalNAc IV. The total number of subfamily-specific conserved positions is represented in the last column, preceded by the relative proportions (%) of ST6GalNAc III and IV subfamily-specific positions, for each protein. The position of the sialylmotifs [L (large), S (small), motif III, and motif VS (very small)] is indicated in the first line of each subfamily. The gene from the sea urchin *Strongylocentrotus purpuratus* (Spu AJ699425) flanked by horizontal thick lines, could not be classified in either of the two subfamilies, because it has similar proportions of ST6Gal III and IV specific conserved positions (41 and 59%, bold characters). The three genes from bony fish *Oryzias latipes* (Ola AJ871604), *Takifugu rubripes* (Tru AJ646869), and *Danio rerio* (Dre AJ868430) also have intermediate values of subfamily-specific positions, but they probably belong to the ST6GalNAc IV (see Discussion and Figure 6).

ST6GalNAc, and ST8Sia according to the linkage formed and the acceptor substrate used by each enzyme (Harduin-Lepers *et al.*, 2001). In this study, molecular phylogeny suggests a correlation between the acceptor substrate specificity, the amino acid sequence, and the genomic organization of the animal sialyltransferase genes.

The mammalian ST3Gal I and II show a narrow acceptor substrate specificity by using exclusively the Galβ1-3GalNAc disaccharide sequence found onto O-glycosylproteins and glycolipids (Giordanengo *et al.*, 1997). The ciona ST3Gal I/II orthologs to the common precursor of ST3Gal I and II subfamilies may have similar substrate specificities.

As it has been demonstrated for other gene families (Robinson-Rechavi *et al.*, 2001; Abi-Rached *et al.*, 2002; McLysaght *et al.*, 2002), large gene duplication events appear to have occurred early after fish radiation. In good agreement with this concept, *D. rerio* has two genes from each of the ST3Gal I, II, III, V (Figure 3A) and ST6GalNAc I/II subfamilies (Figures 3C and 6). There are two genes of ST3Gal I (Figure 1), ST6GalNAc I/II (Figure 6), ST8Sia III, and ST8Sia VI subfamilies (Figure 8) in *T. rubripes*, and there are two genes of ST3Gal I (Figure 1)

and ST8Sia III (Figure 8) subfamilies in *T. nigroviridis*. These observations raise the possibility of additional new enzymatic specificities in fish sialyltransferases, because new unusual gangliosides have been found in fish (Y. Guerardel, personal communication). In addition, further studies are necessary to investigate the specificity of these new sialyltransferases, because besides Neu5Ac, other sialic acids such as Neu5Gc and KDN are present in high amounts in fish glycoconjugates (Inoue *et al.*, 1996).

The ST6Gal family is the simplest animal sialyltransferase family, containing only two subfamilies, ST6Gal I and II, and both use the Galβ1-4GlcNAc-R as an acceptor substrate in mammals. Before the identification of a sialyltransferase gene in *D. melanogaster*, the evidence for the presence of sialic acid in insects was scarce and controversial (for a review, see Schauer, 2000; Marchal *et al.*, 2001). Now the *D. melanogaster* cDNA has been cloned and its enzymatic specificity toward Gal(NAc)β1-4GlcNAc-R demonstrated (Koles *et al.*, 2004), suggesting that this protein has some enzymatic properties in common with the two ST6Gal subfamilies

Previous studies using recombinant enzymes, based on *in vitro* assays, have shown that the mammalian

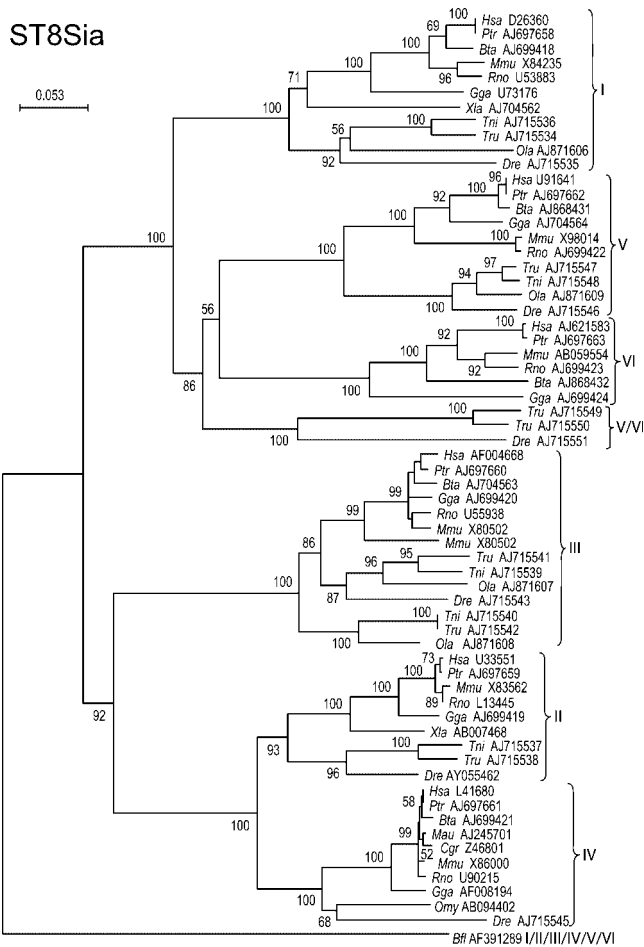


Fig. 8. Neighbor-joining phylogenetic tree of the 64 sialyltransferases of the ST8Sia family. Two hundred and four of the 292 positions (70%) were selected in 10 G-BLOCKS. Bootstrap values were calculated from 500 replicates, and values >50%, are reported at the left of each divergence point. The scale bar represents the number of substitutions per site for a unit branch length. The cephalochordate sialyltransferase sequence from *Branchiostoma floridae* (*Bfl* AJ15545) branched out at the root of the tree and is ortholog to the common ancestor present before the emergence of the six subfamilies of the ST8Sia family. Three genes from bony fish *Takifugu rubripes* (*Tru* AJ15549 and *Tru* AJ15550) and *Danio rerio* (*Dre* AJ15551), branched out before the split of ST8Sia V and VI subfamilies, but the low bootstrap (56%) and the lack of other bony fish sialyltransferases in the ST8Sia VI subfamily suggest that they are genes of the ST8Sia VI subfamily.

ST6GalNAc I and II subfamilies have a narrow acceptor substrate specificity requiring a GalNAc residue O-linked to a peptide as an acceptor substrate (Ikehara *et al.*, 1999; Kurosawa *et al.*, 2000; Samyn-Petit *et al.*, 2000; Marcos *et al.*, 2004). Therefore, we can speculate that the fish ST6GalNAc I/II sialyltransferases are involved in the sialylation of the core GalNAc residue of mucin-type O-glycosylproteins.

In the second branch of the ST6GalNAc family that encompasses the ST6GalNAc III, IV, V, and VI subfamilies, a new gene product named ST6GalNAc III/IV/V/VI, ortholog to the common ancestor present before the separation of the four subfamilies, has been identified in the sea urchin *S. purpuratus*. This hypothetical enzyme may also

share common enzymatic properties with the subfamilies using the trisaccharide Neu5Ac α 2-3Gal β 1-3GalNAc as an acceptor substrate, which are found either in glycopeptides of mucin-type O-glycosylproteins or in the α series of gangliosides such as G_{D1} α . Although it is difficult to predict the enzymatic specificity of this ortholog of a common ancestor, one can assume that it does also use the trisaccharide Neu5Ac α 2-3Gal β 1-3GalNAc as an acceptor substrate, independently of the aglycone moiety. It thus could be involved in the biosynthesis of the sialosphingolipids identified in sea urchin (Kochetkov *et al.*, 1976), or in the biosynthesis of the oligo/polyNeu5Gc chains O-linked via GalNAc residues of the polypeptide chain of the egg receptor for sperm in the same species (Kitazume *et al.*, 1996).

The fish ST6GalNAc sialyltransferases, *Tru* AJ646869, *Ola* AJ871604, and *Dre* AJ864430, are probably fish orthologs of the ST6GalNAc IV subfamily that have evolved from the common ancestor of ST6GalNAc III and IV at a slower rate, than its paralog genes found in the fish ST6GalNAc III subfamily, because there are no other fish ST6GalNAc IV genes, and the apparent branching out before the split of ST6GalNAc III and IV has a very low bootstrap (Figure 6). These new fish ST6GalNAc IV genes could be involved either in the sialylation of mucin-type O-glycosylproteins or in the biosynthesis of gangliosides, which are both expressed in fishes.

As hypothesized above for the ST6GalNAc III and IV, the fish *Tru* AJ15549, *Tru* AJ15550, and *Dre* AJ15551 are probably the fish orthologs of the ST8Sia VI subfamily that have evolved from a common ancestor of ST8Sia V and VI, with a slower rate than its paralog genes found in the fish ST8Sia V subfamily, because there are no other fish ST8Sia VI genes, and the apparent branching out before the split of ST8Sia V and VI has a low bootstrap (Figure 8).

In the ST8Sia family, a sialyltransferase gene was found in the cephalochordate *B. floridae* that is an ortholog to the common ancestor of all the ST8Sia subfamilies. However, to our knowledge, there is no data concerning the presence of sialylated structures in cephalochordata or about the presence of sialyltransferase activity in these species. The ST8Sia family contains two groups of subfamilies exhibiting common enzymatic properties. One branch contains ST8Sia II, III, and IV, which are known as polysialyltransferases involved in the elongation of linear chains of sialic acids found mainly in glycoproteins, whereas the known enzymes of the other branch mediate the transfer of a unique sialic acid residue onto specific sialylated substrates driving the synthesis of di-Sia structures. This last branch includes ST8Sia I (G_{D3} synthase), ST8Sia V (G_{T3} synthase), and ST8Sia VI, which transfer only one sialic acid residue to form G_{D3}, G_{T3}, or O-linked disialyl sequences, respectively.

Materials and methods

Nomenclature

In this study, we consider all the animal sialyltransferases sharing the four sialylmotifs as a super family containing four main families according to the substrate and the sugar linkage made (i.e., ST3Gal, ST6GalNAc, ST8Sia, and

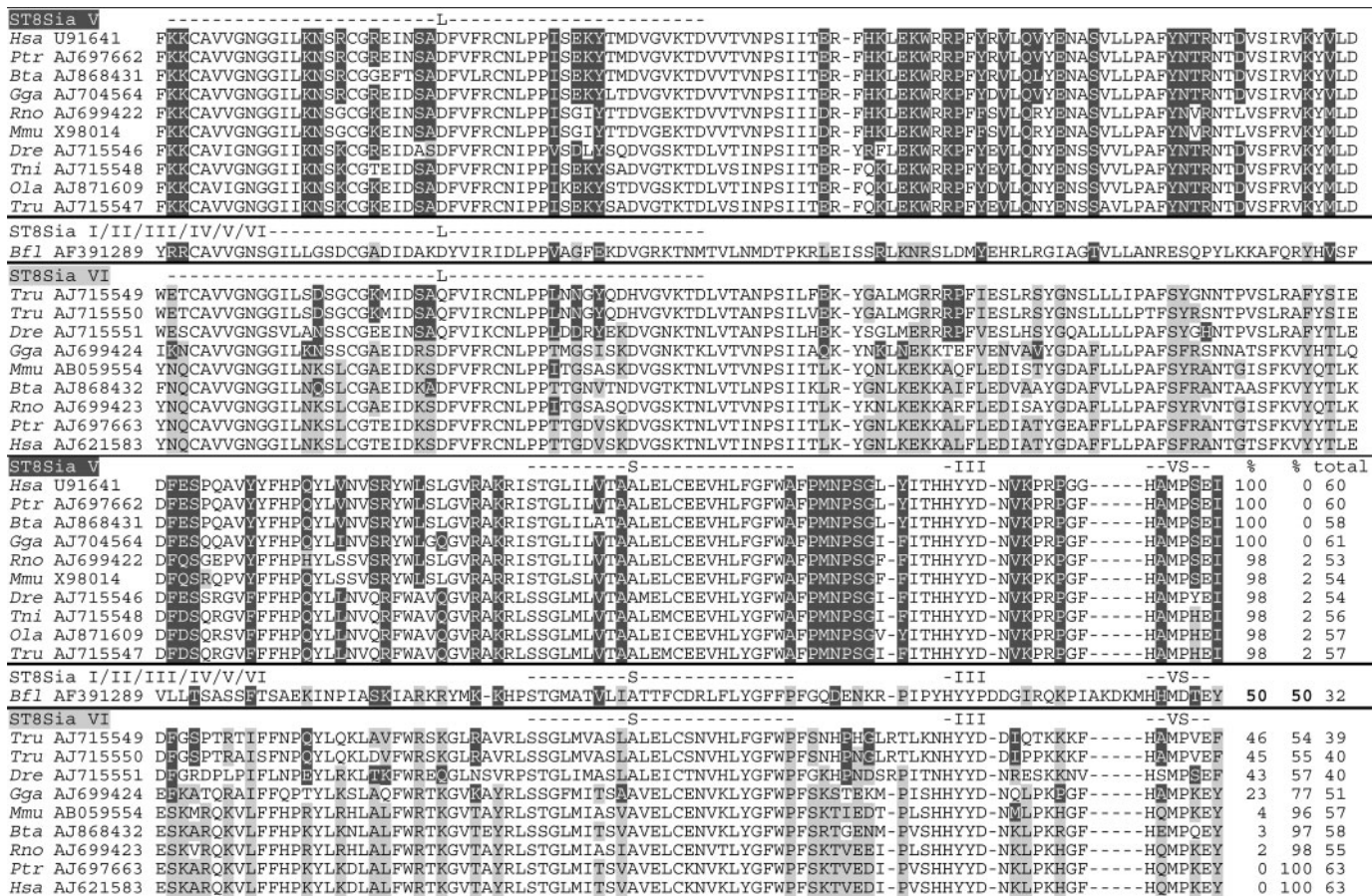


Fig. 9. ClustalW alignment of the sialylmotif regions of the ST8Sia V and VI subfamilies. White letters with dark gray background represent conserved positions specific from ST8Sia V, and black letters on pale gray background represent conserved positions specific from ST8Sia VI. The total number of subfamily-specific conserved positions is represented in the last column, preceded by the relative proportions (%) of ST8Sia V and ST8Sia VI specific positions, for each protein. The position of the sialylmotifs [L (large), S (small), motif III, and motif VS (very small)] is indicated in the first line of each subfamily. The cephalochordate gene from *Branchiostoma floridae* (*Bfl* AJ715545) cannot be classified in either of the two subfamilies, because it contains the same proportion of ST8Sia V and VI subfamily-specific positions (50%). Three genes from bony fish *Takifugu rubripes* (*Tru* AJ715549 and *Tru* AJ715550) and *Danio rerio* (*Dre* AJ715551) have also similar proportions of subfamily-specific positions (between 43 and 57%), but they probably belong to the ST8Sia VI subfamily (see Discussion and Figure 8).

ST6Gal). The first three contain six subfamilies each (I to VI) and the last one two subfamilies (I and II). Sialyltransferases are named according to Tsuji et al. (1996).

Sialyltransferase genes present in databanks

Human DNA and protein sequence were retrieved from the literature. Ortholog protein sequences from other species were searched with gapped-BLAST and PSI-BLAST (Altschul et al., 1997) with default parameters until convergence. Sixty-five complete open reading frames containing the four sialylmotifs and the transmembrane domain were retrieved from different animal species. Transmembrane domains were predicted by PHD-htm for each hypothetical protein (Roost et al., 1995).

New sialyltransferase genes

Two different strategies were followed for the searches of new putative sialyltransferase genes: expressed sequence tag (EST)

assembly and genomic reconstruction. The other animal EST databanks (nonhuman, nonmouse) were searched by TBLASTN with the known sialyltransferases. EST contigs for each gene of each animal species were built with cap3 (Huang and Madan, 1999) and LALIGN (Huang and Miller, 1991). The different general and species specific genomic banks (WGS) (Table I) were also searched by TBLASTN with the known sialyltransferases. Putative exons were identified, and the best intron/exon boundaries were searched by using the Internet *Drosophila melanogaster* site (Table I). Identification of gene and determination of gene structure were carried out with the use of the specialized Internet sites (Table I), following the AG/GT rule. The EST contig sequences and the genomic sequences corresponding to each hypothetical enzyme of each species were compared, to correct eventual splicing errors, and the open reading frames of the complete new sialyltransferases containing the four sialylmotifs and the transmembrane domain were submitted to European Molecular Biology Laboratory (EMBL) (online supplement data, Supplementary

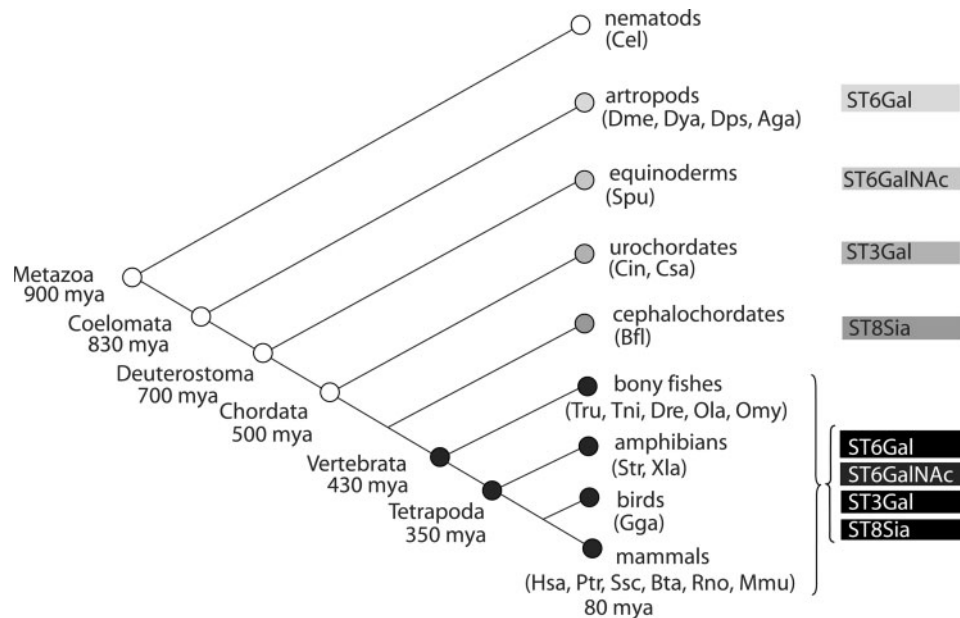


Fig. 10. Scheme of the divergent evolutionary model of the four families of sialyltransferases. The first detectable member of each of the four main families of sialyltransferases is shown, suggesting a general trend for the appearance of sialyltransferase activities: starting with $\alpha 2,6$, followed by $\alpha 2,3$, and finally, $\alpha 2,8$ linkages. However, some links are probably missing in the invertebrate portion of this chain. In contrast, the four families of sialyltransferases were found in vertebrates and most of the duplications, giving rise to the present 20 subfamilies of sialyltransferases, have probably occurred very early in the vertebrate lineage, because most subfamily-specific proteins were already detected in bony fishes.

Table II). One hundred and fifty-five sialyltransferases were obtained by this double approach. Overall the 65 sialyltransferases present in databanks and the 155 new sialyltransferases, described in this article, constitute the pool of 220 proteins from 25 different animal species used for phylogeny. Survey of the databanks was completed in January 2005.

Phylogeny

Protein sequences were aligned with ClustalW 1.8 (Thomson *et al.*, 1994), the selection of informative sites was made by G-BLOCKS (Castresana, 2000), and the output saved in PHYLIP format. Phylogeny analysis was carried out with PHYLWIN (Galtier *et al.*, 1996), using neighbor joining, observed distances, and 500 bootstrap replicates (Felsenstein, 1985).

Determination of subfamily-specific conserved positions

New sequences were a priori considered to belong to the subfamily of the sialyltransferase used for the search and giving the highest BLAST scores. However, the subfamily of each hypothetical sialyltransferase was further confirmed by the determination of the relative proportions of subfamily-specific conserved positions in ClustalW alignments. The alignments of all the subfamilies of each family (taken 2×2) were carried on with the region comprised between the first position before the first motif (L) and the first position after the last motif (VS). Shading of alignments was based on a chemical alphabet comprising five groups: acidic or amide (E, D, Q, N); hydrophobic (I, L, V, M); aromatic (F, Y, W); basic (R, H, K); and hydroxyl (S, T), and the remaining four amino acids, A, G, P, and C, were analyzed separately. This alphabet is based on frequencies of evolutionary replacement

among amino acids, chemical characterizations, and minimal base differences between codons. Amino acids of the same group were considered equivalent for the definition of conserved positions. A threshold of $>50\%$ of conserved amino acids in each position was used for the definition of the subfamily-specific amino acids. The total number of shaded amino acids specific from either one or the other subfamily is reported at the end of each 2×2 clustal line, preceded by the relative proportion (%) of conserved amino acids specific from each of the two subfamilies. Finally, the different lines of each 2×2 clustal were ordered by decreasing proportion of specific positions for one subfamily (from 100 to 0%) and increasing proportion of specific positions for the second subfamily (from 0 to 100%). This simple test helps to find intermediate sialyltransferases that have equivalent proportions of specific positions for each of the two subfamilies and therefore may constitute a new intermediate group, which has similarities with both subfamilies, but cannot be considered to belong to one or the other of the two original subfamilies. One example of the 2×2 subfamily alignments from each sialyltransferase family is included in the text, and the other alignments are available as online supplement data.

Conclusion

The molecular phylogeny analysis of the animal sialyltransferase super family indicates that there are many sialyltransferase sequences containing the sialylmotifs (signatures of the sialyltransferases). In addition, we have defined five new distinct evolutionary groups (ST3Gal I/II, ST6Gal I/II, ST6GalNAc I/II, ST6GalNAc III/IV/V/VI, and ST8Sia I/II/III/IV/V/VI) that correspond to

Table I. URL sites consulted for the *in silico* search of new sialyltransferase genes

Databanks	URL
General banks	
CAZy	http://afmb.cnrs-mrs.fr/CAZY/index.html
DDBJ	http://www.ddbj.nig.ac.jp/search/blast-e.html
NCBI	http://www.ncbi.nlm.nih.gov/BLAST/
UK MRC HGMP-RC	http://www.hgmp.mrc.ac.uk/Registered/Menu/
TIGR	http://tigrblast.tigr.org/tgi/
Ensembl	http://www.ensembl.org
PHD-htm	http://pbil.ibcp.fr/htm/index.php
JGI	http://genome.jgi-psf.org
Animal species-specific banks	
IMCB fugu genome project	http://www.fugu-sg.org/Blast2.html
<i>Takifugu rubripes</i>	http://fugu.hgmp.mrc.ac.uk/blast/
<i>Tetraodon nigroviridis</i>	http://www.genoscope.cns.fr/
<i>Oryzias latipes</i>	http://dolphin.lab.nig.ac.jp/medaka/index.php
<i>Ciona savignyi</i>	http://www.broad.mit.edu/annotation/ciona/
<i>Ciona intestinalis</i>	http://genome.jgi-psf.org/cgi-bin/runAlignement?db=ciona4
<i>Ciona intestinalis</i>	http://ghost.zool.kyoto-u.ac.jp/indexr1.html
<i>Drosophila melanogaster</i>	http://www.fruitfly.org/seq_tools/splice.html
<i>Anopheles gambiae</i>	http://www.genoscope.cns.fr/

The initial searches were performed between 2000 and 2004, and they were all updated by a last search performed in January 2005.

orthologs of the common ancestors of different sialyltransferase subfamilies that might provide a useful new foundation for understanding the structure–function relatedness of the sialyltransferase super family members.

Supplementary material

Supplementary files of Table II and Figures 11, 12, and 13 are available at *Glycobiology* online (<http://glycob.oupjournals.org>).

Acknowledgments

The work was partially supported by the “Centre National de la Recherche Scientifique” (CNRS), the “Institut National de la Santé et de la Recherche Médicale” (INSERM), and the “Association pour la Recherche sur le Cancer” (ARC) grant 3611.

Abbreviations

Aga, *Anopheles gambiae*; *Bfl*, *Branchiostoma floridae*; *Bta*, *Bos taurus*; *Cel*, *Caenorhabditis elegans*; *Cgr*, *Cricetulus*

griseus; *Cin*, *Ciona intestinalis*; *Csa*, *Ciona savignyi*; *Dme*, *Drosophila melanogaster*; *Dps*, *Drosophila pseudoobscura*; *Dya*, *Drosophila yakuba*; *Dre*, *Danio rerio*; *EST*, expressed sequence tag; *Gal*, galactose; *Gga*, *Gallus gallus*; *Hsa*, *Homo sapiens*; *Mau*, *Mesocricetus auratus*; *Mmu*, *Mus musculus*; *mya*, millions of years ago; *Neu5Ac*, *N*-acetylneuraminic acid; *Ocu*, *Oryctolagus cuniculus*; *Ola*, *Oryzias latipes*; *Omy*, *Oncorhynchus mykiss*; *Ptr*, *Pan troglodytes*; *Rno*, *Rattus norvegicus*; *Spu*, *Strongylocentrotus purpuratus*; *Ssc*, *Sus scrofa*; *Str*, *Silurana tropicalis*; *Tni*, *Tetraodon nigroviridis*; *Tru*, *Takifugu rubripes*; *Xla*, *Xenopus laevis*.

References

- Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P., and Inoko, H. (2002) Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet.*, **31**, 100–105.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Angata, T. and Varki, A. (2002) Chemical diversity in the sialic acids and related α -keto acids: an evolutionary perspective. *Chem. Rev.*, **102**, 439–469.
- Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.*, **17**, 540–552.
- Coutinho, P.M., Deleury, E., Davies, G.J., and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.*, **328**, 307–317.
- Datta, A.K. and Paulson, J.C. (1995) The sialyltransferase “sialylmotif” participates in binding the donor substrate CMP-NeuAc. *J. Biol. Chem.*, **270**, 1497–1500.
- Datta, A.K., Sinha, A., and Paulson, J.C. (1998) Mutation of the sialyltransferase S-sialylmotif alters the kinetics of the donor and acceptor substrates. *J. Biol. Chem.*, **273**, 9608–9614.
- Drickamer, K. (1993) A conserved disulphide bond in sialyltransferases. *Glycobiology*, **3**, 2–3.
- Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstraps. *Evolution*, **39**, 783–791.
- Galtier, N., Gouy, M., and Gautier, C. (1996) Two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.*, **12**, 543–548.
- Geremia, R.A., Harduin-Lepers, A., and Delannoy, P. (1997) Identification of two novel conserved amino acid residues in eukaryotic sialyltransferases: implications for their mechanism of action. *Glycobiology*, **7**, v–vii.
- Gilbert, M., Watson, D.C., Cunningham, A.M., Jennings, M.P., Young, N.M., and Wakarchuk, W.W. (1996) Cloning of the lipooligosaccharide α -2,3-sialyltransferase from the bacterial pathogens *Neisseria meningitidis* and *Neisseriae gonorrhoeae*. *J. Biol. Chem.*, **271**, 28271–28276.
- Gilbert, M., Brisson, J.R., Karwaski, M.F., Michniewicz, J., Cunningham, A.M., Wu, Y., Young, N.M., and Wakarchuk, W.W. (2000) Biosynthesis of ganglioside mimics in *Campylobacter jejuni* OH4384. Identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-mhZ (1)h and (13)c NMR analysis. *J. Biol. Chem.*, **275**, 3896–3906.
- Giordanengo, V., Bannwarth, S., Laffont, C., Van Miegem, V., Harduin-Lepers, A., Delannoy, P., and Lefebvre, J.C. (1997) Cloning and expression of cDNA for a human Gal β 1–3GalNAc α 2,3-sialyltransferase from the CEM T-cell line. *Eur. J. Biochem.*, **247**, 558–566.
- Harduin-Lepers, A., Recchi, M.A., and Delannoy, P. (1995) 1994, the year of sialyltransferases. *Glycobiology*, **5**, 741–758.
- Harduin-Lepers, A., Vallejo-Ruiz, V., Krzewinski-Recchi, M.A., Samyn-Petit, B., Julien, S., and Delannoy, P. (2001) The human sialyltransferase family. *Biochimie*, **83**, 727–737.
- Hood, D.W., Cox, A.D., Gilbert, M., Makepeace, K., Walsh, S., Deadman, M.E., Cody, A., Martin, A., Mansson, M., Schweda, E., and others. (2001) Identification of a lipopolysaccharide α -2,3-sialyltransferase from *Haemophilus influenzae*. *Mol. Microbiol.*, **39**, 341–350.

- Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.*, **9**, 868–877.
- Huang, X. and Miller, W. (1991) A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.*, **12**, 337–357.
- Ikehara, Y., Kojima, N., Kurosawa, N., Kudo, T., Kono, M., Nishihara, S., Issiki, S., Morozumi, K., Itzkowitz, S., Tsuda, T., and others. (1999) Cloning and expression of a human gene encoding an N-acetylgalactosamine α 2,6-sialyltransferase (ST6GalNAc I): a candidate for synthesis of cancer-associated sialyl-Tn antigens. *Glycobiology*, **9**, 1213–1224.
- Inoue, S., Kitajima, K., and Inoue, Y. (1996) Identification of, 2-keto-3-deoxy-D-glycero – galactononic acid (KDN, deaminoneuraminic acid) residues in mammalian tissues and human lung carcinoma cells. Chemical evidence of the occurrence of KDN glycoconjugates in mammals. *J. Biol. Chem.*, **271**, 24341–24344.
- Jackson, R.J., Hall, D.F., and Kerr, P.J. (1999) Myxoma virus encodes an α 2,3-sialyltransferase that enhances virulence. *J. Virol.*, **73**, 2376–2384.
- Jeanneau, C., Chazalet, V., Auge, C., Soumpasis, D.M., Harduin-Lepers, A., Delannoy, P., Imbert, A., and Breton, C. (2004) Structure-function analysis of the human sialyltransferase ST3Gal I: role of N-glycosylation and a novel conserved sialylmotif. *J. Biol. Chem.*, **279**, 13461–13468.
- Jones, P.A., Samuels, N.M., Phillips, N.J., Munson, R.S.J., Bozue, J.A., Arseneau, J.A., Nichols, W.A., Zaleski, A., Gibson, B.W., and Apicella, M.A. (2002) *Haemophilus influenzae* type b strain A2 has multiple sialyltransferases involved in lipooligosaccharide sialylation. *J. Biol. Chem.*, **277**, 14598–14611.
- Kaneko, M., Nishihara, S., Narimatsu, H., and Saitou, N. (2001) The evolutionary history of glycosyltransferase genes. *Trends Glycosci. Glycotechnol.*, **13**, 147–155.
- Kitazume, S., Kitajima, K., Inoue, S., Haslam, S.M., Morris, H.R., Dell, A., Lennarz, W.J., and Inoue, Y. (1996) The occurrence of novel 9-O-sulfated N-glycolylneuraminic acid-capped α 2–5-O-glycolyl-linked oligo/polyNeu5Gc chains in sea urchin egg cell surface glycoprotein. Identification of a new chain termination signal for polysialyltransferase. *J. Biol. Chem.*, **271**, 6694–66701.
- Kochetkov, N.K., Smirnova, G.P., and Chekareva, N.V. (1976) Isolation and structural studies of a sulfated sialosphingolipid from the sea urchin *Echinocardium cordatum*. *Biochim. Biophys. Acta.*, **424**, 274–283.
- Koles, K., Irvine, K.D., and Panin, V.M. (2004) Functional characterization of *Drosophila* sialyltransferase. *J. Biol. Chem.*, **279**, 4346–4357.
- Kurosawa, N., Takashima, S., Kono, M., Ikehara, Y., Inoue, M., Tachida, Y., Narimatsu, H., and Tsuji, S. (2000) Molecular cloning and genomic analysis of mouse GalNAc α 2,6-sialyltransferase (ST6GalNAc I). *J. Biochem.*, **127**, 845–854.
- Livingston, B.D. and Paulson, J.C. (1993) Polymerase chain reaction cloning of a developmentally regulated member of the sialyltransferase gene family. *J. Biol. Chem.*, **268**, 11504–11507.
- Marchal, I., Jarvis, D.L., Cacan, R., and Verbert, A. (2001) Glycoproteins from insect cells: sialylated or not? *Biol. Chem.*, **382**, 151–159.
- Marcos, N.T., Pinho, S., Grandela, C., Cruz, A., Samyn-Petit, B., Harduin-Lepers, A., Almeida, R., Silva, F., Morais, V., Costa, J., and others. (2004) Role of the human ST6GalNAc I and ST6GalNAc II in the synthesis of the cancer associated sialyl-Tn antigen. *Cancer Res.*, **64**, 7050–7057.
- McLysaght, A., Hokamp, K., and Wolfe, K.H. (2002) Extensive genomic duplication during early chordate evolution. *Nat. Genet.*, **31**, 200–204.
- Oriol, R., Mollicone, R., Cailleau, A., Balanzino, L., and Breton, C. (1999) Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. *Glycobiology*, **9**, 323–334.
- Robinson-Rechavi, M., Marchand, O., Escriva, H., Bardet, P.L., Zelus, D., Hughes, S., and Laudet, V. (2001) Euteleost fish genomes are characterized by expansion of gene families. *Genome Res.*, **11**, 781–788.
- Roost, B., Casadio, R., Fariselli, P., and Sander, C. (1995) Prediction of helical transmembrane segments at 95% accuracy. *Prot. Sci.*, **4**, 521–533.
- Samyn-Petit, B., Krzewinski-Recchi, M.A., Steelant, W.F., Delannoy, P., and Harduin-Lepers, A. (2000) Molecular cloning and functional expression of human ST6GalNAc II. Molecular expression in various human cultured cells. *Biochim. Biophys. Acta.*, **1474**, 201–211.
- Schauer, R. (1982) Chemistry, metabolism, and biological functions of sialic acids. *Adv. Carbohydr. Chem. Biochem.*, **40**, 131–234.
- Schauer, R. (2000) Achievements and challenges of sialic acid research. *Glycoconj. J.*, **17**, 485–499.
- Séveno, M., Bardor, M., Paccalet, T., Gomord, V., Lerouge, P., and Faye, L. (2004) Glycoprotein sialylation in plants? *Nat. Biotechnol.*, **22**, 1351–1352.
- Shah, M.M., Fujiyama, K., Flynn, C.R., and Joshi, L. (2003) Sialylated endogenous glycoconjugates in plant cells. *Nat. Biotechnol.*, **21**, 1470–1471.
- Shen, G.J., Datta, A.K., Izumi, M., Koeller, K.M., and Wong, C.H. (1999) Expression of α 2,8/2,9-polysialyltransferase from *Escherichia coli* K92. Characterization of the enzyme and its reaction products. *J. Biol. Chem.*, **274**, 35139–35146.
- Takashima, S., Ishida, H.K., Inazu, T., Ando, T., Ishida, H., Kiso, M., Tsuji, S., and Tsujimoto, M. (2002a) Molecular cloning and expression of a sixth type of α 2,8-sialyltransferase (ST8Sia VI) that sialylates O-glycans. *J. Biol. Chem.*, **277**, 24030–24038.
- Takashima, S., Tsuji, S., and Tsujimoto, M. (2002b) Characterization of the second type of human β -galactoside α 2,6-sialyltransferase (ST6Gal II), which sialylates Gal β 1,4GlcNAc structures on oligosaccharides preferentially. Genomic analysis of human sialyltransferase genes. *J. Biol. Chem.*, **277**, 45719–45728.
- Takashima, S., Tsuji, S., and Tsujimoto, M. (2003) Comparison of the enzymatic properties of mouse β -galactoside α 2,6-sialyltransferases, ST6Gal I and II. *J. Biochem. (Tokyo)*, **134**, 287–296.
- Thomson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
- Tsuji, S., Datta, A.K., and Paulson, J.C. (1996) Systematic nomenclature for sialyltransferases. *Glycobiology*, **6**, v–vii.
- Varki, A. (1993) Biological roles of oligosaccharides, all of the theories are correct. *Glycobiology*, **3**, 97–130.
- Wakarchuk, W.W., Watson, D., St. Michael, F., Li, J., Wu, Y., Brisson, J.R., Young, N.M., and Gilbert, M. (2001) Dependence of the bi-functional nature of a sialyltransferase from *Neisseria meningitidis* on a single amino acid substitution. *J. Biol. Chem.*, **276**, 12785–12790.
- Yamamoto, T., Nakashizuka, M., and Terada, I. (1998) Cloning and expression of a marine bacterial β -galactoside α 2,6-sialyltransferase gene from *Photobacterium damsela* JT0160. *J. Biochem. (Tokyo)*, **123**, 94–100.