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 ($\alpha 2$ -3, $\alpha 2$ -6, and $\alpha 2$ -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 multigene 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

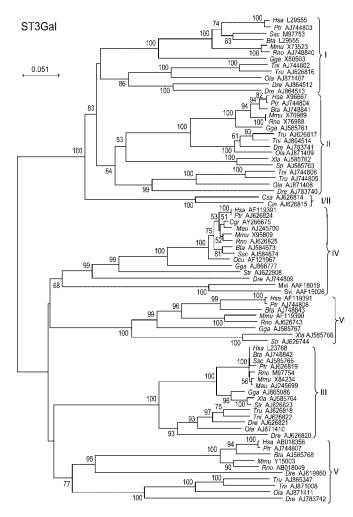


Fig. 1. Neighbor-joining phylogenetic tree of the 77 sialyltransferases of the ST3Gal family. One hundred and fifty-six of the 235 positions (60%) were selected in seven G-BLOCKS. Bootstrap values were calculated from 500 replicates, and values >50% are reported at the left of each divergence point. 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 $\alpha 2,3$ -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.

ST3Ga	al I	L			
Bta	L29555	RRCAVVGNSGNLRESWYGEQIDSHDFVLRMNKAPTAGFEEDVGRKTTHHEVYPESERELAENVSMVLVPFKTVDLEWVESATTKGEI	HTYVI	PVPAKI	Ι
Mmu	X73523	RRCAVVGNSGNLKDSSYGEEIDSHDFVLRMNKAPTVGFEADVGSRTTHHLVYPESFRELGENVNMVLVPFKTTDLGWVISATTTGII	THTYVI	PVPPKI	Γ
Rno		RRCAVVGNSGNLKDSSYGPEIDSHDFVLRMNRAPTVGFEADVGSRTTHHLVYPESBRELGENVNMVLVPFKITDLWVISATTTGII			
Hsa	L29555	RRCAVVGNSGNLRSSYGEEIDSHDFVLRMNKAPTAGFEADVGTKTTHHLVYPESRELGENVSMILVPFKTIDLEWVVSAITTGII			
Ptr		RRCAVVGNSGNLRSSYGBEIDSHDFVLRMNKAPTAGFEADVGTKTTHHLVYPESFRELGDNVSMILVPFKTIDLEWVVSAITTGII			
		WHCAVVGNSGNLROLSYREAIXXHNFMLRMNKAPTAGFEAAAGSKTAHHLVYPESFRELGENVSMVLVPLKTMNLEWVVSTTTTGAI			
		WHCAVMGNSGNLROLSYREEINSHNFMLRMNKAPTAGFKADAGSKTAHHLVYPESFRELGENVSMVLVPLKTMNLEWVVSTTTTGAI			
Ssc	M97753	RRCAVVGNSGNLKESYYGEQIDSHDFVLRMNKAPTEGFEADVGSKTTHHFVYPESERELAGEVSMILVPFKTTDLEWVISATTTGRI			
Gga	X80503	RRCAVVGNSGNLROSQYGQDIDSHDFVLRMNRAPTIGYESDVGSKTTHHFVYPESYKELAENVSMIVIPFKTLDLRWIWTALTTG IN			
Dre		RTCAVVGNSGNLLRSNYGRLIDLHDFVLRINKGPTKGFEKDVGSKTTHR MYPESAVDLDNSTHLVLFPYKILDMQWLISIFTTKHI	TRTYM	SVPSTI	T T
Dre		RICAVVGNSGNILGSHYGQLIDSHDFVIRINKGPTKGYETDVGSKTTHRIMYPESAMDLDNSTHLVLLPFKVKDMCWLLSVFTTKHI			
Ola		RTCSVVGNSGNLKGSNYGALIDTSDLVIRMNKAPTKGFEKDVGAKTTHHVMYPESAVDLSNTTSLVLVPFKTLDLCWI SALTTG I			
Tru		RTCAVVGNSGNLKGSQYGRLIDSSDFIIRMNQAPTSGFEDDVGTRTTHHIMYPESAKDLDNGTSLVLIPFKTLDLOWI SALTTGAI			
Tni		RTCAVVGNSGNLKGSQYGRLIDSSDFIIRMNQAPTLGFEEDVGTKTTHHVMYPESAKDLDNGTSLVLIPFKTLDLCWI SALTTGII			
-	al I/II	T			
Cin		RSCAVVGNSGNILMSNYGNVIDGHDFVIRMNMGPTYNYENDVGSKTTHRFMYPTTASSLPGGKLFVLVPFQPLDIKWLESALTTGEI	PRTYO	PL.VR.	7
Csa		RCAVVGNSGNLINSKYGKNIDSHDFVIRLN:GPTEGFENDVGRKTTHRFMYPATASSLAGGVSLVLLPFOPODVKNLISALTTGEL			
	al II			U TOR	
ola		RSCAVVGNSGNLRESGHGERIDSHSFVIRMNKALTRGFEKDVGNRTTHHFLYPESAVDVDGGVSLVLLPFKLRDLEWLTSALSTGHI	ZTTTVM		T
Tru		KSCAV GNSGNLRGSGNGER I DSHDSVIRMINALI RGFAKDVGNRTTHHFI I FDSAVDVDHGVSLVLLPFKLRDIDWLTSALSIGKU			
Tni		HTCAVVGNSGRLRSSGNGKLIDSHISVIRMNAVIRGFEKDVGRRTHHRJYPESAVDVGGGVSLVLLPFKLRDIDWLISALSIGKV			
Dre		RICAVORSGREAS SINGLIDSHSVIAMAVITAVITAVITAVITAL SKALAVITAL SALAVGAVITAL SALAVGAVITAL SALAVGAVITAVITAVITAVITAVITAVITAVITAVITAVITAVIT			
Str		RRCAVVGNSGNLKGSRHGKNIDSHGLIFMNGARTYGFEKDVGSKTTHHTMIPESATHKGYNDVGHULIPFKLODLKWITSALTGEV			
Xla		RRCAVVGNSGNLKGSRIGKNIDSHGLIFRMNGARIIGFENDVGSRIIHHFMIPESAVNVQPGVHLVLIPFKLQDLKWIISALIIGEV			
Bta		RRCAVVGNSGNLAGSRIGRRIDSHGLIFRMNGARIIGFENDVGSRIIHHMIPESAVNLOPGVHLVDIPFRLQDLKWIIGALIIGEV			
Tru		LRCAVVGNSGNLRGAGYGATIDGHNYIMRINLAPTVGFEDDUGAGHTTHHTMYPESARNLAANVSFVLVPFKTDLUWITSALSTGOI			
Gga		ARCAVORNSGNLRGSGYGHEIDGHDFINNINNAFIYGFEDDAOGRTTHHTMIPESARNILPANVSFVLVPFKTDDLWIIASALSTGOI			
Ola		LRCAVVGNSGNLRGSGYGKIIDEHHFIMRINLAPTVGYEEDVGSRTTHOFMYPESAKNLAPNISFVLVPFKTLDLVWIISALSTGOI			
Tni		LRCAVVGNSGNLRGAGYGPTIDGHDHVMRINLAPTVGFEEDAGSRTTHHFMYPESAKNLAANVSFVLVPFKTLDLVWITSALSTGOI			
Dre		LRCAVVGNSGNLRGAGYGPVIDGHDFIMRMNLAPTVGYEEDAGSRTTHHFMYPESAKNLAANVSFVLVPFKTLDLLWITSALSTGOI			
Rno	X76988	RCAVVGNSGNLRGSGYGQEVDSHNFIMRMNQAPTVGFEKDVGSRTTHHFMYPESAKNLPANVSFVLVPFKALDLMWIASALSTGOI			
Hsa	X96667	RCAVVGNSGNLRGSGYGQDVDGHNFIMRMNQAPTVGFEQDVGSRTTHHFMYPESAKNLPANVSFVLVPFKVLDLLWIASALSTGOI			
Mmu	X76989	RRCAVVGNSGNLRGSGYGQeVDSHNFIMRMNQAPTVGFEKDVGSRTTHHFMYPESAKNLPANVSFVLVPFKALDLMWIASALSTGQIIINAUNAUNAUNAUNAUNAUNAUNAUNAUNAUNAUNAUNAUN			
Ptr		RRCAVVGNSGNLRGSGYGQDVDGHNFIMRMNQAPTVGFEQDVGSRTTHHFMYPESAKNLPANVSFVLVPFKALDLLWIASALSTGQI	RETYA	PVKSFI	L.
101					
ST3G			olo	% tot	tal
ST3Ga Bta	L29555		% 100	% tot 0 27	tal 7
ST3Ga Bta Mmu	L29555 X73523		% 100 100	% tot 0 25 0 29	tal 7 9
ST3Ga Bta Mmu Rno	L29555 X73523 AJ748840		% 100 100 100	% tot 0 27 0 29 0 29	tal 7 9
ST3Ga Bta Mmu Rno Hsa	L29555 X73523 AJ748840 L29555		% 100 100 100 96	% tot 0 25 0 29 4 28	tal 7 9 9
ST3Ga Bta Mmu Rno Hsa Ptr	L29555 X73523 AJ748840 L29555 AJ744803		% 100 100 100 96 96	% tot 0 22 0 22 4 28 4 26	tal 7 9 9 8
ST3Ga Bta Mmu Rno Hsa Ptr §Has	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084		% 100 100 96 96 96	% tot 0 22 0 29 4 28 4 26 4 23	tal 7 9 9 8 5 3
ST3Ga Bta Mmu Rno Hsa Ptr §Has §Ptr	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 AJ865085		% 100 100 96 96 96 96	% tot 0 22 0 22 4 28 4 26 4 22 4 23	tal 7 9 9 8 5 3 3
ST3G Bta Mmu Rno Hsa Ptr §Has §Ptr Ssc	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 AJ865085 M97753		% 100 100 96 96 96 96 93	% tot 0 22 0 22 4 28 4 26 4 23 4 23 7 27	tal 7 9 9 3 5 3 3 7
ST3G Bta Mmu Rno Hsa Ptr §Has §Ptr Ssc Gga	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 AJ865085 M97753 X80503		% 100 100 96 96 96 93 83	<pre>% tot 0 22 0 22 4 28 4 22 4 22 7 22 17 22 </pre>	tal 7 9 9 3 5 3 3 7 3
ST3G Bta Mmu Rno Hsa Ptr §Has §Ptr Ssc Gga Dre	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 AJ865085 M97753 X80503 AJ864512		% 100 100 96 96 96 93 83 76	<pre>% tot 0 22 0 22 4 28 4 20 4 22 7 22 17 22 24 21 </pre>	tal 7 9 9 8 5 3 3 7 3 1
ST3G Bta Mmu Rno Hsa Ptr §Has §Ptr Ssc Gga Dre Dre	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 AJ865085 M97753 X80503 AJ864512 AJ864512		% 100 100 96 96 96 93 83 76 78	<pre>% tot 0 2? 0 29 4 28 4 20 4 22 7 2? 17 23 24 21 22 18</pre>	tal 7 9 9 8 5 3 3 7 3 1 8
ST3G Bta Mmu Rno Hsa Ptr §Has §Ptr Ssc Gga Dre Dre Ola	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 AJ865085 M97753 X80503 AJ864512 AJ864512 AJ864512 AJ864512		* 100 100 96 96 96 93 83 76 78 71	<pre>% tot 0 2? 0 29 4 28 4 20 4 22 7 2? 17 23 24 21 22 18 29 21</pre>	tal 7 9 9 8 5 3 3 7 3 1 8 1
ST3Ga Bta Mmu Rno Hsa Ptr Ssc Gga Dre Ola Tru	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 AJ865085 M97753 X80503 AJ864512 AJ864513 AJ871407 AJ626816		<pre>% 100 100 96 96 96 93 83 76 78 71 67</pre>	<pre>% tot 0 22 0 22 4 28 4 20 4 22 4 22 7 22 17 22 24 21 22 18 29 21 33 22</pre>	tal 7 9 9 3 3 3 7 3 1 3 1 1
ST3Ga Bta Mmu Rno Hsa Ptr §Has §Ptr Ssc Gga Dre Dre Ola Tru Tni	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 AJ865085 M97753 X80503 AJ864512 AJ864513 AJ871407 AJ626816 AJ744802		* 100 100 96 96 96 93 83 76 78 71	<pre>% tot 0 2? 0 29 4 28 4 20 4 22 7 2? 17 23 24 21 22 18 29 21</pre>	tal 7 9 9 3 3 3 7 3 1 3 1 1
ST3Ga Bta Mmu Rno Hsa Ptr §Has §Ptr Scc Gga Dre Dre Ola Tru Tni ST3Ga	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864512 AJ864512 AJ864512 AJ864513 AJ871407 AJ626816 AJ744802 AJ I/II		<pre>% 100 100 96 96 96 93 83 76 78 71 67 63</pre>	<pre>% tot 0 22 0 29 4 28 4 20 4 22 7 22 17 22 24 21 29 21 33 22 37 24</pre>	tal 7 9 3 3 3 7 3 1 1 1 4
ST3G Bta Mmu Rno Hsa Ptr SHas SPtr Ssc Gga Dre Ola Tru Tru ST3G Cin	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864513 AJ864513 AJ871407 AJ626816 AJ744802 al I/II AJ626815		<pre>% 100 100 96 96 96 93 83 76 78 71 67 63 48</pre>	<pre>% tot 0 22 0 29 4 28 4 23 4 23 4 23 7 23 17 23 24 21 22 18 29 21 33 23 37 24</pre>	tal 7 99 93 55 53 33 1 1 1 1 1 1 1
ST3G Bta Mmu Rno Hsa Ptr \$Has \$Ptr Ssc Gga Dre Dre Ola Tru Tni ST3G Cin Csa	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864513 AJ864513 AJ871407 AJ626816 AJ744802 al I/II AJ626815 AJ626814		<pre>% 100 100 96 96 96 93 83 76 78 71 67 63</pre>	<pre>% tot 0 22 0 29 4 28 4 20 4 22 7 22 17 22 24 21 29 21 33 22 37 24</pre>	tal 7 99 93 55 53 33 1 1 1 1 1 1 1
ST3G Bta Mmu Rno Hsa Ptr Shas Sptr Sptr Sptr Gga Dre Dre Ola Tru Tni ST3G Cin Csa ST3G	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 AJ865085 AJ865085 AJ864512 AJ864512 AJ864512 AJ864512 AJ864513 AJ871407 AJ626816 AJ744802 AJ626815 AJ626815 AJ626814 AJ II		<pre>% 100 100 96 96 96 93 83 76 78 71 67 63 48 45</pre>	<pre>% tot 0 22 0 29 4 226 4 22 4 22 17 22 24 21 22 18 29 21 33 22 37 24 55 20</pre>	tal 7 9 8 3 3 3 1 1 1 1 1 1 1 0
ST3Ga Bta Mmu Rno Hsa Ptr §Has §Ptr Ssc Gga Dre Dre Ola Tru Tni ST3Ga Cin ST3Ga ST3Ga Ola	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 AJ865085 X80503 AJ864512 AJ864513 AJ864513 AJ871407 AJ626816 AJ744802 al I/II AJ626815 AJ626814 AJ II AJ871407		<pre>% 100 100 100 96 96 96 93 83 76 78 71 67 63 48 45 29</pre>	<pre>% tot 0 25 0 29 4 28 4 26 4 23 7 25 17 23 24 21 22 18 29 21 33 22 337 24 52 22 55 20 71 21</pre>	tal 7 9 3 3 3 3 3 1 1 1 1 4 1 0
ST3G Bta Mmu Rno Hsa Ptr \$Ptr \$Sc Gga Dre Ola Tru Tru ST3G Cin Csa ST3G Ola Tru ST3G Cin Csa ST3G	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864512 AJ864513 AJ871407 AJ626816 AJ744802 AJ626815 AJ626814 AJ744802		* 100 100 96 96 96 93 83 76 78 71 67 63 48 45 29 27	<pre>% tot 0 22 0 22 4 22 4 26 4 23 7 21 17 23 24 21 22 12 23 22 37 24 55 20 71 21 73 22</pre>	tal 7 9 3 3 3 3 3 1 1 1 4 4 1 2
ST3G Bta Mmu Rno Hsa Ptr \$Has \$Ptr Ssc Gga Dre Dre Ola Tru Tni ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864512 AJ864512 AJ864512 AJ864513 AJ871407 AJ626816 AJ744802 AJ626815 AJ626814 AJ666814 AJ666814 AJ666814 AJ66		* 100 100 100 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25	<pre>% tot 0 22 0 29 4 28 4 26 4 23 7 22 17 23 24 21 22 18 29 21 33 22 37 24 52 22 55 20 71 21 73 22 75 24</pre>	tal 99 93 55 33 33 11 11 14 10 11 22 4
ST3G Bta Mmu Rno Hsa Ptr §Has §Ptr Ssc Gga Dre Ola Tru Tni ST3G Cin ST3G Cin ST3G Ola Tru Tru Dre	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 AJ865085 AJ865085 AJ864512 AJ864512 AJ864512 AJ864512 AJ744802 AJ744802 AJ626815 AJ626815 AJ626815 AJ626815 AJ626815 AJ744806 AJ783740	VKKKNKILIYHPAF KYVEDSWIOCHGRYPSTGILSVIFSLHVCDEVDLYGFGADSKGNWHHYWENNES -AGAFRKTGVHDCDFES KVKQEKILIYHPAF KYVEDSWIOCHGRYPSTGILSVIFSLHVCDEVDLYGFGADSKGNWHHYWENNES -AGAFRKTGVHDCDFES KVKQEKILIYHPAF KYVEDSWIOCHGRYPSTGILSVIFSILICDEVDLYGFGADSKGNWHHYWENNES -AGAFRKTGVHDCDFES RVKQDKILIYHPAF KYVEDSWIOCHGRYPSTGILSVIFSMHVCDEVDLYGFGADSKGNWHHYWENNES -AGAFRKTGVHDADFES RVKQDKILIYHPAF KYVEDSWIOCHGRYPSTGILSVIFSMHVCDEVDLYGFGADSKGNWHHYWENNES -AGAFRKTGVHDADFES RVKQDKILIYHPAF KYVEDSWIOCHGRYPSTGILSVIFSMHVCDEVDLYGFGADSKGNWHHYWENNES -AGAFRKTGVHDADFES RVKQDKILIYHPAF KYVEDSWIOCHGRYPSTGILSVIFSMHVCD*VDLYSF*ADSKGN*HHYWENNES -AGAFRKTGVHDADFES RVKQDKILIYHPAF KYVEDSWIOCHGRYPSTGILSVIFSMHVCD*VDLYSF*ADSKGN*HHYWENNES -AGAFRKTGVHDAGFES RVKQDKILIYHPAF KYVEDSWIOCHGRYPSTGILSVIFSLHICDEVDLYGFGADSGONHHYWENNES -AGAFRKTGVHDCDFES KVKKEKULIYHPAF KYVEDSWIOCHGRYPSTGLISVIFSLHICDEVDVYGFGADSGOWHHYDDSMIOS -AGAFRKTGVHDCDFES KVKKEKULIYHPAF KYVEDSWIOCHGRYPSTGLISVIFSLHICDEVVVGFGADSGOWHHYDDSMIOS -AGAFRKTGVHDCDFES KVKKEKULIYHPAF KYVEDSWIOCHGYPSTGLISVIFSLHICDEVVVGFGADSKGNWHHYWENNES -AGAFRKTGVHDCDFES KVKKEKULIYHPAF KYVEDSWIOCHGYPSTGLISVIFSLIV FSMHVCDEVVVYGFGADSGOWHYPDSSMION	<pre>% 100 100 100 96 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25 22</pre>	<pre>% tot 0 22 0 29 4 226 4 22 4 22 17 23 24 21 22 18 29 21 33 22 37 24 55 22 71 21 73 22 75 24 75 24 78 23</pre>	tal 7 9 9 3 3 3 3 1 1 1 1 4 4 1 0 1 2 2 4 3
STIG Bta Mmu Rno Hsa Ptr §Htr Ssc Gga Dre Ola Tru STIG Cin Csa STIG Ola Tru Tri STIG Cin Csa STIG Str Str	L29555 X73523 AJ748800 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864512 AJ864512 AJ864512 AJ864513 AJ871407 AJ626816 AJ744802 AJ744802 AJ71407 AJ744805 AJ744805 AJ744805 AJ744805 AJ744805 AJ744805 AJ783740 AJ783740	<pre></pre>	<pre>% 100 100 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25 22 14</pre>	<pre>% tot 0 22 0 22 4 26 4 26 4 23 7 27 17 23 24 21 22 16 29 21 33 22 37 24 55 20 71 21 73 22 75 24 71 21 73 22 75 24 78 23 86 23</pre>	tal 7 9 9 9 3 3 5 5 3 3 1 1 1 1 1 2 2 4 3 3 1
ST3G Bta Mmu Rno Hsa Ptr §Has §Ptr Ssc Gga Dre Ola Tru Tru Tru ST3G Cin Csa ST3G Cin Csa ST3G Cin Csa ST3G Cin Csa ST3G ST3G ST3G Cin ST3G ST3G Cin ST3G ST3G ST3G ST3G ST3G ST3G ST3G ST3G	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864513 AJ864513 AJ871407 AJ626816 AJ744802 AJ744802 AJ744805 AJ744806 AJ744806 AJ744806 AJ744806 AJ783740 AJ585763	<pre></pre>	* 100 100 96 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25 22 14 12	<pre>% tot 0 22 0 22 4 26 4 26 4 23 7 22 17 23 24 21 22 16 29 21 33 22 37 24 52 22 55 20 71 21 73 22 75 24 78 23 86 21 88 24</pre>	tal 7 9 9 3 5 5 3 3 7 3 1 1 1 1 1 2 4 4 3 1 1 2 4 4 3 1 4
ST3G Bta Mmu Rno Hsa Ptr \$Has \$Ptr Ssc Gga Dre Dre Ola Tru Tni ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G St St St St St St St St St St St St St	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864512 AJ864513 AJ871407 AJ626816 AJ744802 AJ626815 AJ626815 AJ626815 AJ626815 AJ626815 AJ626815 AJ744806 AJ744806 AJ744805 AJ744805 AJ7485763 AJ585762 AJ748841	XVKKNKILIY PAFI KYVDDSWLOCHGRYPSTGILSVI FSLHVCDEVDLYGFGADSKGNWHHYWENNPS AGAFRKTGVHDCDFES KVKQEKILIY PAFI KYVDDSWLOCHGRYPSTGILSVI FSIHICDEVDLYGFGADSKGNWHHYWENNPS AGAFRKTGVHDCDFES KVKQEKILIY PAFI KYVDDSWLOCHGRYPSTGILSVI FSIHICDEVDLYGFGADSKGNWHHYWENNPS AGAFRKTGVHDCDFES RVKQDKILIY PAFI KYVDDSWLOCHGRYPSTGILSVI FSIHICDEVDLYGFGADSKGNWHHYWENNPS AGAFRKTGVHDADFES RVKQDKILIY PAFI KYVDDSWLOCHGRYPSTGILSVI FSMHVCDEVDLYGFGADSKGNWHHYWENNPS AGAFRKTGVHDADFES RVKQDKILIY PAFI KYVDDSWLOCHGRYPSTGILSVI FSMHVCDEVDLYGFGADSKGNWHHYWENNPS AGAFRKTGVHDAGFES RVKQDKILIY PAFI KYVDDSWLOCHGRYPSTGILSVI FSMHVCD*VDLYSF*ADSKGN*HHY*ENNLS AGSFHKTGVHDAGFES RVKQDKILIY PAFI KYVDDSWLOCHGRYPSTGLSVI FSMHVCD*VDLYSF*ADSKGN*HHY*ENNLS AGAFRKTGVHDCDFES NADREXVMILIY PAFI KYVDDSWLOCHGRYPSTGLSVI FSLHICDEVDLYGFGADSKGNWHYWENNS AGAFRKTGVHDCDFES NADREXVNILIY PAFI KYVDDSWLOCHGRYPSTGLLSVI FALHVCDEVDVYGFGADSKGNWHYWENNS AGAFRKTGVHDCDFES NADREXVNILIY PAFI KYVDDSWLOCHGRYPSTGLLSVI FALHVCDEVDVYGFGADSKGNWHYWENNS AGAFRKTGVHDCDFES NADREXVNILIY PAFI KYVDDSWLOCHGRYPSTGLISVI FALHVCDEVDVYGFGADSKGNWHYWENNS AGAFRKTGVHDCDFES NADREXVNILIY PAFI KYVDDSWLOCHGRYPSTGLISVI FALHICDEVSVFGFGADQHGNWHYWENNS AGAFRKTGVHDCDFES NADREXVNILIY PAFI KYVDDSWLOCHGRYPSTGLISVI FALHICDEVSVFGFGADQHGNWHYWENNS AGAFRKTGVHDCDFES NADREXVNILIYNPFF KYVDSWLCHGRYPSTGFLSUL LAIH	<pre>% 100 100 100 96 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25 22 14 12 7</pre>	<pre>% tot 0 22 0 29 4 28 4 20 4 20 4 20 17 20 22 18 22 20 23 22 29 21 33 22 37 24 52 20 55 20 71 20 71 20 75 24 78 20 88 24 88 24 88 24 88 20 88 20 71 20 71 20 75 24 78 20 75 24 78 20 75 24 78 20 78 20 71 20 75 24 78 20 75 24 78 20 75 24 78 20 75 24 78 20 75 24 78 20 75 24 78 20 79 30 70 20 70 20 70 70 20 70 70 20 70 70 70 20 70 70 70 20 70 70 70 70 70 70 70 70 70 70 70 70 70</pre>	tal 7 9 9 3 5 5 3 3 7 3 1 1 1 1 1 2 2 4 3 1 1 2 2 4 3 1 1 2 2 4 3 1 1 2 2 4 3 1 1 2 2 4 3 1 1 2 2 4 3 3 3 7 7 3 1 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
ST3G Bta Mmu Rno Hsa Ptr §Has §Ptr Ssc Gga Dre Ola Tru Tni ST3G Cia ST3G Cia ST3G Cia ST3G St3G St3G St3G St3G St3G St3G St3G St	L29555 X73523 AJ748800 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864512 AJ864512 AJ864512 AJ626816 AJ744802 AJ744802 AJ744805 AJ744805 AJ744806 AJ7847480 AJ783740 AJ585763 AJ748841 AJ626817	<pre></pre>	<pre>% 100 100 100 96 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25 22 14 12 7 7 0</pre>	<pre>% tot 0 22 0 29 4 226 4 22 4 22 17 23 24 21 22 18 23 2 29 21 33 2 37 24 55 22 55 22 71 21 73 22 75 24 78 23 86 24 93 30 88 24 93 30 100 2</pre>	tal 7 9 9 8 5 5 3 3 1 1 1 4 1 0 0 1 2 2 4 4 3 1 1 2 7 7 7 7 7 7 7 7 7 7 7 9 9 9 9 9 9 8 5 5 3 3 1 1 1 1 9 9 9 9 9 9 9 9 9 9 9 9 9 9
ST3G Bta Mmu Rno Hsa Ptr SHas SFtr Ssc Gga Dre Ola Tru Tru ST3G Cin Csa ST3G Cin Csa ST3G Cin Csa ST3G ST3G ST3G Cin Csa ST3G St3 St3 St3 St3 St3 St3 St3 St3 St3 St3	L29555 X73523 AJ748800 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864512 AJ864513 AJ871407 AJ626816 AJ744802 AJ744802 AJ744805 AJ744805 AJ744805 AJ744805 AJ744805 AJ744806 AJ783740 AJ783740 AJ585763 AJ585762 AJ748841 AJ626817 AJ585761	<pre>- IIIVS KVKKNKLLIY PAFT KYVDSWLOCHGRYPSTGILSVIFSLHVCDEVDLYGFGADSKGNWHHYWENNSAGAFRKTGVHDCDPES KVKQEKILIY PAFT KYVDDWLOCHGRYPSTGILSVIFS IHICDEVDLYGFGADSKGNWHHYWENNSAGAFRKTGVHDCDPEX KVKQEKILIY PAFT KYVDDWLOCHGRYPSTGILSVIFS IHICDEVDLYGFGADSKGNWHHYWENNSAGAFRKTGVHDDPEX KVKQEKILIY PAFT KYVDDWLOCHGRYPSTGILSVIFS MHVCDEVDLYGFGADSKGNWHHYWENNSAGAFRKTGVHDDPES RVKQDKILIY PAFT KYVDDWLOCHGRYPSTGILSVIFSMHVCDEVDLYGFGADSKGNWHHYWENNSAGAFRKTGVHDDPES RVKQDKILIY PAFT KYVDDWLOCHGRYPSTGILSVIFSMHVCD*VDLYSF*ADSKGNWHHYWENNSAGAFRKTGVHDDPES RVKQDKILIY PAFT KYVDDWLOSHRYPLTSILSVIFSMHVCD*VDLYSF*ADSKGNWHHYWENNSAGAFRKTGVHDDFES RVKQDKILIY PAFT KYVDDWLOSHRYPLTSILSVIFSMHVCD*VDLYSF*ADSKGNWHHYWENNSAGAFRKTGVHDDGFES RVKQDKILIY PAFT KYVDDWLOSHRYPLTSILSVIFSMHVCD*VDLYSF*ADSKGNWHHYWENNSAGAFRKTGVHDGFES RVKQDKILIY PAFT KYVDDWLOHGRYPSTGLLSVIFSLHICDEVDLYGFGADSKGNWHHYPENSLAGSFHKTGVHDGFES KVKKEKULIYNPSFT KYVDENWLOHGRYPSTGLLSVIFSLHLCDEVDLYGFGADSKGNWHHYPDSNASFSNLHGCDFEN NADREKVMIL PAFT KYVDDWLOHGRYPSTGFIT LIFALHICDQVSVFGFGADLGOWHHYPDESNA</pre>	* 100 100 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25 22 14 12 7 0	<pre>% tot 0 22 0 22 4 22 4 26 4 23 7 22 17 22 24 21 22 12 23 32 37 24 55 20 71 21 73 22 75 24 78 23 86 21 88 24 93 30 100 25</pre>	tal 7 9 9 3 3 3 1 1 1 1 2 4 4 1 2 2 4 3 1 1 2 2 4 3 1 1 2 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
ST3G Bta Mmu Rno Hsa Ptr §Has Ssc Gga Dre Ola Tru Tni ST3G Cin Csa ST3G Cin Sta ST3G Cin ST3G Cin Sta Sta Sta ST3G Cin ST3G Cin Sta Sta Sta Sta Sta Sta Sta Sta Sta Sta	L29555 X73523 AJ748840 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864513 AJ864513 AJ864513 AJ871407 AJ626815 AJ626815 AJ626815 AJ744805 AJ744805 AJ744805 AJ744805 AJ744806 AJ783740 AJ585762 AJ74841 AJ626817 AJ626817 AJ585761 AJ585761 AJ585761 AJ871409	<pre></pre>	<pre>% 100 100 100 96 96 96 93 83 76 93 83 76 78 71 67 63 48 45 29 27 25 22 14 12 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</pre>	<pre>% tot 0 22 0 22 4 26 4 26 4 26 4 27 17 23 24 21 22 18 29 21 33 22 37 24 52 22 55 20 71 21 73 22 75 24 78 23 86 21 88 24 93 30 100 22 100 22</pre>	tal 7 9 9 3 3 3 3 1 3 1 1 2 4 4 1 0 7 7 9 7 7 7
ST3G Bta Mmu Rno Hsa Ptr SHas SPtr Sptr Sptr Sptr Sptr Sptr Sptr Sta Dre Ola Tru Tni ST3G Cin ST3G ST3G ST3G Cin ST3G Ci	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 AJ865085 X80503 AJ864512 AJ864512 AJ864512 AJ626816 AJ744802 AJ626815 AJ626815 AJ626815 AJ626815 AJ744806 AJ7871407 AJ7871407 AJ7871407 AJ788714 AJ78876 AJ78876 AJ78876 AJ78876 AJ78876 AJ78876 AJ78871409 AJ864514	<pre>S</pre>	<pre>% 100 100 100 96 96 96 96 93 83 76 78 71 67 63 29 27 25 22 14 12 7 0 0 0</pre>	<pre>% tot 0 22 0 29 4 28 4 26 4 22 17 22 17 22 24 21 22 18 23 22 29 21 33 22 55 20 71 21 73 22 75 24 78 23 86 21 88 24 88 24 93 30 100 25 100 25 100 25</pre>	tal 79 99 55 33 77 31 1 1 1 1 2 4 31 1 4 00 77 97 79 79
STIG Bta Mmu Rno Hsa Ptr §Har Ssc Gga Dre Ola Tru STIG Cin Csa STIG Ola Tru STIG Cin Csa Stra Str Xla Bta Str Xla Bta	L29555 X73523 AJ748800 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864512 AJ864512 AJ864512 AJ626816 AJ744802 AJ744802 AJ744802 AJ744805 AJ744805 AJ744805 AJ744805 AJ744805 AJ744805 AJ744805 AJ785762 AJ748841 AJ626817 AJ585761 AJ585761 AJ585761 AJ781409 AJ864514 AJ783741	LIIIVS LIIIVS KVKQKIKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF FLH/CDEVDLYGFGADSKGNWHYWENNPSAGAFRKTGVHDCDFEX KVQEKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SIHICDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDCDFEX KVQEKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SIHICDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDCDFEX KVQEKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SHICDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDADFES RVKQDKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SMHVCDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDADFES RVKQDKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SMHVCDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDADFES RVKQDKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SHVCD*VDLYSF*ADSKGN*HYWENNISAGSFKTGVHDADFES RVKQKEKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SLHICDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDAFES RVKQKKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SLHICDEVDLYGFGADSKGNWHYWENNSAGAFRGTGVHD CPES KVKKEKULIYIPSF KYVDSDWI CHGRYPSTGILS IF SLHICDEVDVYGFGADSKGNWHYWENNSAGAFRGTGVHD CPES KVKKEKULIYNSFF KYVDSBWI CHGRYPSTGFI LIFALHICDQVSVFGFGADSKGNWHYWENNSAGAFRGTGVHD CPES KVKKEKVIIYSFFFKYV ESWI EHGGYPSTGFI LIFALHICDQVSVFGFGADKLDGWHHYWEENSLAGAFRHTGVHD CPES KANKKVVVINPFH KYV ESWI EHGRYPSTGFI LIFALHICDQVSVFGFGADHGNHHYWEENSLAGAFRHTGVHD CPEY KANKKVVVINPFH KYV ESWI DCHGRYPSTGFI LIFALHICDQVSVFGFGADHGNHHYWEENSLAGAFRHTGVHD CPYEY KANKKVVVINPFH KYVESWI DCHGRYPSTGFI LIFALHICDQVSVFGFGADHGNHHYWEENSLAGAFRHTGVHD CPYEY KANKKVVVINPFF KYVESWI DCHGRYPSTGFI LIFALHICDQVSVFGFGADHGNMHYWEENSLAGAFRHTGVHD CPYEY KANKKVVVINPFF KYVHDRWTCHHGRYPSTGLIAI YALHECDEVDVYGFGADAGGNWHHYWEENSLAGAFRHTGVHD CPYEY CCKSKITIISPTF RYVHDRWTCHHGRYPSTGMLAI FALHICDQVSVFGGADAGGNWHHYWEENSLAGAFRHTGVHD CPYEY CANKKUVVNPAFF KYVHDRWTCHHGRYPSTGMLAI FALHICDQVSVFGGADAGGNWHHYWEENSLAGAFRHTGVHDSAQEN TCVKKVUVNPAFF KYVHDRWTCHHGRYPSTGMLAI FALHICDQVSVFGGADAGGNWHHYWEENSLAGAFRKTGVHADFET ADKNKVLVNPAFF KYVHDRWTCHHGRYPSTGMLAI FALHICDQVSVFGGADAGGNWHHYWENRYAGAFRKTGVHADFET ADKNKVLVNPAFF KYVHDRWTCHHGRYPSTGMLAI FALHICDQVSVFGGADAGGNWHHYWENRYAGAFRKTGVHADFET ADKNKVLVNPAFF KYVHDRWTCHHGRYPSTGMLAI FALHUCDEVNVFGGADBGNWHHYWENRYA	<pre>% 100 100 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25 22 14 12 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</pre>	<pre>% tot 0 22 0 22 4 22 4 22 4 22 17 22 22 12 22 12 22 2 33 22 37 24 29 21 33 22 37 24 29 21 33 22 37 24 55 20 71 21 73 22 75 24 75 24 71 21 73 22 75 24 71 21 75 24 71 21 75 24 75 24 7</pre>	tal 7 9 9 3 5 5 3 3 1 1 1 4 1 1 2 2 4 3 1 1 2 4 4 0 0 7 9 9 7 9 9 7 9 9 9 9 9 9 9 9 9 9 9
ST3G Bta Mmu Rno Hsa Ptr SHas SFtr Ssc Gga Dre Ola Tru Tru ST3G Cin Csa ST3G Ola Tru Tni Dre St7 ST3G Ola Tru Tni Dre St7 ST3G Ola Tru Csa ST3G Ola Tru Csa ST3G Ola Tru Dre ST3G Ola Tru Dre ST3G Ola Tru Dre ST3G Ola Tru Dre ST3G Ola Tru Dre ST3G Ola Tru Dre ST3G Ola Tru Dre ST3G Cin Csa ST3G Ola Tru Dre ST3G Cin Csa ST3G Ola Tru Dre ST3G Cin Dre ST3G ST3G ST3G ST3G Cin Dre ST3G ST3G ST3G ST3G ST3G ST3G ST3G ST3G	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 AJ865085 X80503 AJ864512 AJ864512 AJ864512 AJ626816 AJ744802 AJ626815 AJ626815 AJ626815 AJ626815 AJ744806 AJ7871407 AJ7871407 AJ7871407 AJ788714 AJ78876 AJ78876 AJ78876 AJ78876 AJ78876 AJ78876 AJ78871409 AJ864514	<pre></pre>	<pre>% 100 100 96 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25 22 14 12 77 0 0 0 0 0 0 0 0</pre>	<pre>% tot 0 22 0 22 4 26 4 26 4 26 4 27 7 27 17 22 24 21 22 18 29 21 33 22 37 24 55 20 71 21 73 22 75 24 78 23 86 21 88 24 93 30 100 25 100 25 1000 25 100 25 100 100 25 100 100 25 100 100 100 100 1000 100 100000000000</pre>	tal 7 9 9 3 5 5 3 3 7 7 3 1 1 1 4 1 0 7 7 9 9 7 7 9 9 8 1 1 2 4 4 0 7 7 9 9 8 5 5 3 3 7 7 3 1 1 2 9 9 8 5 5 3 3 7 7 9 9 9 8 5 5 5 3 3 7 7 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
ST3G Bta Mmu Rno Hsa Ptr §Has Spc Gga Dre Ola Tru Tni ST3G Cin ST3G ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G ST3G ST3G ST3G ST3G ST3G ST3G ST3G	L29555 X73523 AJ748840 L29555 AJ744803 AJ865084 M97753 X80503 AJ864512 AJ864513 AJ871407 AJ626816 AJ744802 AJ744802 AJ744805 AJ7485762 AJ74845 AJ748740 AJ787400 AJ78740 AJ78740	LIIIVS LIIIVS KVKQKIKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF FLH/CDEVDLYGFGADSKGNWHYWENNPSAGAFRKTGVHDCDFEX KVQEKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SIHICDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDCDFEX KVQEKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SIHICDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDCDFEX KVQEKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SHICDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDADFES RVKQDKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SMHVCDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDADFES RVKQDKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SMHVCDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDADFES RVKQDKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SHVCD*VDLYSF*ADSKGN*HYWENNISAGSFKTGVHDADFES RVKQKEKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SLHICDEVDLYGFGADSKGNWHYWENNSAGAFRKTGVHDAFES RVKQKKILIYIPAFI KYVDDSWI CHGRYPSTGILS IF SLHICDEVDLYGFGADSKGNWHYWENNSAGAFRGTGVHD CPES KVKKEKULIYIPSF KYVDSDWI CHGRYPSTGILS IF SLHICDEVDVYGFGADSKGNWHYWENNSAGAFRGTGVHD CPES KVKKEKULIYNSFF KYVDSBWI CHGRYPSTGFI LIFALHICDQVSVFGFGADSKGNWHYWENNSAGAFRGTGVHD CPES KVKKEKVIIYSFFFKYV ESWI EHGGYPSTGFI LIFALHICDQVSVFGFGADKLDGWHHYWEENSLAGAFRHTGVHD CPES KANKKVVVINPFH KYV ESWI EHGRYPSTGFI LIFALHICDQVSVFGFGADHGNHHYWEENSLAGAFRHTGVHD CPEY KANKKVVVINPFH KYV ESWI DCHGRYPSTGFI LIFALHICDQVSVFGFGADHGNHHYWEENSLAGAFRHTGVHD CPYEY KANKKVVVINPFH KYVESWI DCHGRYPSTGFI LIFALHICDQVSVFGFGADHGNHHYWEENSLAGAFRHTGVHD CPYEY KANKKVVVINPFF KYVESWI DCHGRYPSTGFI LIFALHICDQVSVFGFGADHGNMHYWEENSLAGAFRHTGVHD CPYEY KANKKVVVINPFF KYVHDRWTCHHGRYPSTGLIAI YALHECDEVDVYGFGADAGGNWHHYWEENSLAGAFRHTGVHD CPYEY CCKSKITIISPTF RYVHDRWTCHHGRYPSTGMLAI FALHICDQVSVFGGADAGGNWHHYWEENSLAGAFRHTGVHD CPYEY CANKKUVVNPAFF KYVHDRWTCHHGRYPSTGMLAI FALHICDQVSVFGGADAGGNWHHYWEENSLAGAFRHTGVHDSAQEN TCVKKVUVNPAFF KYVHDRWTCHHGRYPSTGMLAI FALHICDQVSVFGGADAGGNWHHYWEENSLAGAFRKTGVHADFET ADKNKVLVNPAFF KYVHDRWTCHHGRYPSTGMLAI FALHICDQVSVFGGADAGGNWHHYWENRYAGAFRKTGVHADFET ADKNKVLVNPAFF KYVHDRWTCHHGRYPSTGMLAI FALHICDQVSVFGGADAGGNWHHYWENRYAGAFRKTGVHADFET ADKNKVLVNPAFF KYVHDRWTCHHGRYPSTGMLAI FALHUCDEVNVFGGADBGNWHHYWENRYA	<pre>% 100 100 100 96 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25 22 14 12 7 0 0 0 0 0 0 0 0 0</pre>	<pre>% tot 0 22 0 22 4 22 4 22 4 22 4 22 17 23 24 21 22 18 29 21 33 22 37 24 52 22 55 20 71 21 73 22 75 24 78 23 86 21 88 24 93 30 100 22 100 22 100 22 100 22 100 22 100 22</pre>	tal 79998 5533 1114 100 12243 11400779998 899
ST3G Bta Mmu Rno Hsa Ptr §Has Spc Gga Dre Ola Tru Tni ST3G Cin ST3G ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G ST3G Cin ST3G Cin ST3G Cin ST3G Cin ST3G ST3G ST3G ST3G ST3G ST3G ST3G ST3G	L29555 X73523 AJ748800 L29555 AJ744803 AJ865085 M97753 X80503 AJ864512 AJ864512 AJ864512 AJ864512 AJ626816 AJ744802 AJ744802 AJ744802 AJ744805 AJ744806 AJ78474806 AJ78474806 AJ783740 AJ585763 AJ74881 AJ626817 AJ585761 AJ78871407 AJ585761 AJ78874841 AJ626817 AJ585761 AJ783740 AJ585761 AJ783741 X76988 X96667 X76989	<pre></pre>	<pre>% 100 100 96 96 96 93 83 76 78 71 67 63 48 45 29 27 25 22 14 12 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</pre>	<pre>% tot 0 22 0 22 4 26 4 26 4 26 4 27 7 27 17 22 24 21 22 18 29 21 33 22 37 24 55 20 71 21 73 22 75 24 78 23 86 21 88 24 93 30 100 25 100 25 1000 25 100 25 100 100 25 100 100 25 100 100 100 100 1000 100 100000000000</pre>	Lal 7799985533377311411111111111111111111111111111

Fig. 2. ClustalW alignment of the peptide sequences of the sialylmotif regions of the ST3Gal I and II subfamilies. White letters with dark gray background represent conserved positions specific from ST3Gal I, and black letters on pale gray background represent conserved positions specific from ST3Gal I. The total number of subfamily-specific conserved positions is represented in the last column preceded by the relative proportions (%) of ST3Gal I and II 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 two urochordate genes from *Ciona savignyi* (*Csa* AJ626814) and *Ciona intestinalis* (*Cin* AJ626815), flanked by horizontal thick lines, cannot be classified in either of the two subfamilies because they have similar proportions of ST3Gal I and II specific conserved positions (between 45 and 55%, bold characters).

A ST3Gal I	102 66 60 15 40 57	Hsa L29555
	87 65 60 15 40 54 96 65 60 15 40 54	Dre AJ864512 Dre AJ864513
0700-11		
ST3Gal II	113 65 60 15 40 57 137 65 60 15 40 57	<i>Hsa</i> X96667 <i>Dre</i> AJ783741
		Dre AJ783740
ST3Gal I/II		Cin AJ626815
ST3Gal III	38 17 15 30 32 21 32 63 49 49 29	Hsa L23768
	36 17 13 30 32 21 33 63 49 49 29	Dre AJ626820
	36 13 30 32 21 33 63 49 49 29	Dre AJ626821
ST3Gal IV	31 26 33 21 33 65 47 48 28	Hsa L23767
	31 29 33 21 33 65 40 48 30	Dre AJ744809
ST3Gal VI	30 26 34 22 32 62 46 50 29	Hsa AF119391
ST3Gal V	15 35 115 62 53 82	Hsa AB018356
orodurv	12 42 111 63 53 83	Dre AJ619960
	45 54 98 62 46 79	Dre AJ783742
B ST6Gal I	202 33 33 59 79	Hsa X17247
	24 42 118 89 33 33 59 86	Dre AJ744801
ST6Gal II	314 33 33 59 89	Hsa AB059555
	307 33 33 59 81	Dre AJ627627
ST6Gal I/II	259 33 94 33 56	Dme AF218237
C ST6GalNAc I 44	233 57 57 46 35 30 33 65	Hsa Y11339
ST6GalNAc II	43 21 57 56 46 33 30 33 55	Hsa AJ251053
ST6GalNAc I/II	51 90 57 58 46 35 22 1 26 56	Dre AJ634459
ST6GalNAc III	6 65 136 37 61	Hsa AJ507291
	26 45 136 37 62	Dre AJ620947
ST6GalNAc IV	4 63 136 37 62	Hsa AJ271734
	4 46 136 37 69	Dre AJ868430
ST6GalNAc V	6 82 136 37 75	Hsa AJ507292
OTOGAINAG V	6 58 136 37 74	Dre AJ646874
ST6GalNAc VI	6 59 136 37 61	Hsa AJ507293
	70 136 37 59	Dre AJ646883
ST6GalNAc III/IV/V/VI	11 198 86 57 37 63	Spu AJ699425
D ST8Sia I	64 48 36 33 160	Hsa D26360
Boroolar	60 48 36 33 162	Dre AJ715335
ST8Sia III	60 41 185 94	Hsa AF004668
	55 39 185 93	Dre AJ715543
ST8Sia I/II/III/IV/V/VI	49 46 168 100	Bfl AF391289
ST8Sia II	33 20 43 86 98 93	Hsa U33551
	33 18 52 86 98 94	Dre AY055462
ST8Sia IV	39 43 86 98 93	<i>Hsa</i> L41680
	39 42 86 98 93	Dre AJ715545
ST8Sia V	44 31 29 48 38 31 155	Hsa U91641
ST8Sia V	44 31 29 48 38 31 155 44 29 48 38 31 153	Dre AJ715546
ST8Sia V ST8Sia VI		

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 V is unique and different from its closest neighbors, 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, and VI subfamilies share also similar genomic organization. In the ST8Sia family (**D**), the ST8Sia I/II/III/IV/V/I from amphioxus (*Bfl* AF391289) has similar genomic organization with the ST8Sia III 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 GalB1-4GlcNAc-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 AJ821848), 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

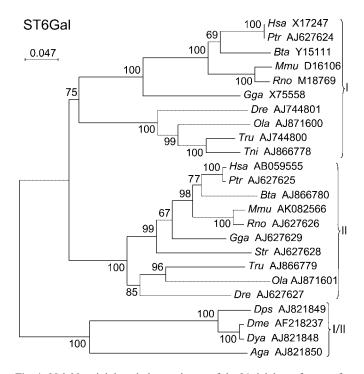


Fig. 4. Neighbor-joining phylogenetic tree of the 24 sialyltransferases of the ST6Gal family. Two hundred and seventy-five of the 338 positions (81%) were selected in 11 G-BLOCKS. Bootstrap values were calculated from 500 replicates, and values >50% are reported at the left of each divergence point. 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.

confirmed by the determination of the proportion of subfamily-specific positions in the 2×2 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 $\alpha 2$ -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
β1-3GalNAc 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 (Neu5Aca2-3GalB1-3GalNAc-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.

ST6Gal I	LL			
Hsa X17247	WGRCAVVSSAGSLKSSØLGREIDDHDAVLRFNGAPTANFQØDVGEKTTIRLMNSQLVTURKR-FLKDSLYNEGILIWWDPSVYHSDIEKWYØNP	DYNFI	NNY	KTY
<i>Bta</i> Y15111	WGRCAVVSSAGSLKSSRLGREIDDHDAVLRFNCAPTVKFQQDVGKTTIRLVNSQLVTHEAG-FLKDSLYNECILIWDPSVYHSDIEKWYRNP			
	WGRCAVVSSAGSIKSSQLGREIDDHDAVLRFNSAPTANFQQDVGTKTTIRLMNSQLVTTEKR-FLKDSLYNEGILIJWDPSVYHSDIEKWYQNP			
	. FKTCAVVSSAGSLKNSGLGKEIDSHDAVIRFNAAPTAGFETDVGSKTTVRLINSQLMASEDHHFLSSSLYSAGILVSWDPSPYSSDLWEWFNKT	DYPIF	FKQY	QRY
Mmu D16106	CTKCAVVSSAGSLKNSQLGREIDNHDAVLRFNGAPTDNFOODVGNKTTIRLVNSQLVTNEKR-FLKDSLYNEGILIIWDPSVYHADIPQWYOKP	DYNFF	FETY	KSY
Rno M18769	WQRCAVVSSAGSLKNSSLGREIDNHDAVLRFNSAPTDNFQCDVGSKTTIRLMNSQLVTTERR-FLKDSLYREGILIWWDPSVYHADIEKWYGKP	DYNFI	PDTY	KSY
Tru AJ744800	YKSCAVVISAGSMRSSGLGKEIDSHDAVLRFNAAPTSGYENDVGSKTTIRLVNSQVMASBAHRFLSSSLYSSGTLVAWDPAPFSADLTOWFNRT FRSCAVVSSAGSLRSSGLGKEIDSHDAVLRFNAAPTSGFENDVGSKTTIRLVNSQVMASDAHRFLSSSLNSSGTLVAWDPAPFSADLREWYNRT	DYPIE	TOY	2RY
Gga X75558	LGRCAVVSSAGSLKSSHLGPEIDSHDAVLRFNCAPVKGFODVGOKTTIRLVNSQLVTVDEQQFLKDALYNTGILIWWDPAPYHAEIHDWYRKP LKSCAVVSSAGSLRHSGLGKEIDSHDAVMRFNAAPTSGFEKDVGSKTTMRLINSQVMASDEYRFLSSSLYSSVLVAWDPAPFSSDLTOWLWRT	DIKFI	DAI	NOI
	TRACKIVE BROADELDSTDAVIKTINAFI SETEKDVOKTINKUTASQVIMADELKTISSSUIGETVAMDIATISSDUMIKT	DIFIL	Agr	2141
ST6Gal I/II	NGSCVIVASAGSLERSOLGSFIDEHDIVMRFNHAPTEGYEADVGSKTTIRVVNSQVVTKPEYQLLTAPLFRNVTIAAWDPGKFDQTLAEWLATP	DENTI	DMIR	ALC IN
Dva 1.1921949	NGCVIVASAGSLAGSLAGSLAGSIDHUVMRNHAPIGGHADVGSKTIRVVNSQVVIVEFDFDTAPIFRNVILAAWDEGKTDGILAAWAAIP	DYDLI	TIME	TT.V
Drg 1.1821849	INICATVSSAGSLAGSKLGAFIDIADIVARFNAAFIQGHEVDVGSKTIIRVVASQVVIKPEFDFAAFIFRAVITAAWDEGKINGIDEDWUTSA IKTCAIVSSAGSLAGSKLGRFIDTHDIVARFNAAFIQGHEVDVGSKTTIRVVASQVVTKPEFDFAHAPIFRAVITAAWDEGKINGIDEDWUTSA	DYDLI	SNIV	TV
Dme AF218237	IKTCAIVSSACSLAGSKLGFIDTHDIVMRFNHAPTOGHEVDVGSKTTIRVVNSQVVTKPEPDFTRAPIFRNVTIAAWDPGKVNGTLDDUITSA	DYDL	SNY	RLY
ST6Gal II				
	FRTCAVVSSAGAILHSGLGKEIDSHDAVLRFNAAPTEGYEODVGLKTTIRIINSQILANPKHEFKTSSIYKNITLVAWDPAPYTLNLDEWFASP	DYDL	CPV	ZEH
Dre AJ627627	LKTCAVVISAGAMLHSGLGKEIDSHDAVLRFNTAPTVGYERDVGNKTTIRIINSGILANPMHRFNRSSLYKNVTLVAWDPAPYTLNLH, WYSNP	DYNL	FTPY	MEY
	FKSCAVVUSAGAILRSGLGREIDAHDAVLRFNAAPTEGYERDVGNKTTIRIINSQVLANPNHRFNTSSLYKDVVLVAWDPAPYTLDLHKWYASP			
	FSTCAVVSSAGAILNSSLGAEIDSHDAVLRFNSAPTRNYEKDVGNKTTLRIINSQILTNPNHHFTDSSLYKDVTLIAWDPSPYYADLHMWYHKP			
	LRTCAVVESAGAILNSSLGEEIDSHDAVLRFNSAPTRGYEKDVGNKTTVRIINSQILTNPSYHFMDSALYKDVILVAWDPAPYSANLNLRYKKP			
	FGSCAVVMSAGAILNSSLGDEIDSHDAVLRFNSAPTRGYEKDVGNKTTMRIINSQILTNPNHHFVDSSLYKDVILVAWDPAPYSANLN-WYKKP			
	LRSCAVVMSAGAILNSSLGEEIDSHDAVLRFNSAPTRGYEKDVGNKTTVRIINSQILTNPSHHFVDSSLYKDVILVAWDPAPYSANLNLWYKKP			
	$\label{eq:linear} LRSCAVVMSAGAILNSSLGEEIDSHDAVLRFNSAPTRGYEKDVGNKTTVRIINSQILANPSHHFIDSSLYKDVILVAWDPAPYSANLNLWYKKPFAPARAMANANANANANANANANANANANANANANANANANAN$			
	LRSCAVVMSAGAILNSSLGFEIDSHDAVLRFNSAPTRGYEKDVGNKTTIRIINSQILTNPSHHFIDSSLYKDVILVAWDPAPYSANLNLWYKKP			
and the second se	LSSCAVVMSAGAILNSSLGEEIDSHDAVLRFNSAPTRGYEKDVGNKTTVRIINSQILANPSHHFIDSALYKDVILVAWDPAPYSANLNLWYKKP			
ST6Gal I		8		otal
Hsa X17247	RK <mark>LH</mark> PNQPFYILKPQMPWELWDILQEISPEE-IQPNPPSSG <mark>W</mark> LGIIIMMTLCDQVDIYEFLPS <mark>KRKT</mark> DVCYYYQ K FFDSACTMGAYHPLLYEK RKLHPDQPFYILKPOMPWELWDIIQEISSE U -IQPNPPSSGWLGIAIMMSLCDQVDIYEFLPSKRKTDVCYYYO R YFDSACTMGAYHPLLFEK		0	
Bta Y15111		100	0	
		100	0	
Mmu D16106	RRL PSOPFYILKPONEWELWDIIOEISPD - IOPNPPSSG/LGIIIMMTLCDOV/IYEFLPS/RRKTDVCYYHO/FFDSACTMGAYHPLLFEK	97	3	
Rno M18769	RRLNPSQPFYILKPOMPWELWDIIQEISADL-IQPNPPSSGMLGIIIMMTLCDQVLIYEFLPSKRKTDVCYYHQKFFDSACTMGAYDPLLFEK	97	7	
	RML PMOPFYILHPR FEWOVWORIODNMAEP-IOKNPPSSGLGTVMMMSLCEVVHVYEFLPSRR TELCHYYORFFDAACTLGAYHPLLYEK	91	9	
Tni AJ866778		88	12	
Gga X75558	RIRHPEOPFYILNPKWOWQLWDILOENSLEH-IQPNPPSSGWLGIVIMMTLCDEVWVYEFLPSKROTDICHYYOKFHDHACTMGAYHPLLFEK	79	21 :	28
Ola AJ871600	RREPOOPFFILHPRFEWQVWQQVQENMAES-IQKNPPSSGFLGTVLMMSLCEVVHVYEFLPSKRKTELCHYYQFFYDAACTLGAYHPLLYEK	75	25	24
ST6Gal I/II	S			
	RSSEPQSNFHIIDPRSIWRAWTALQDLTDLP-IRKNPPTSGFIGLGLLLPVCRYIEVVEYIPSTRMNGLCHYYDDQLNLGCTFGAWHPLAAEK	61	49	
	RRRYPKSRAFLIDPHSVWRLWQSLQMFAGNRPISRNPPSSGFIGLALLLPHCPQVDFVPSTRLNGRCHYYS <mark>K</mark> EMNSACTFGSWHPLAAEK	50	50	
Dps AJ821849	${\tt RRRYPKSRAFLIDPHSVWRLWQTLQMFAGNRSIRRNPPSSGFIGLALLLPHCPQVDFVEYIPSTRLNGRCHYYSKEMNAACTFGSWHPLAAEK}$	48	52	
	$\mathbf{RR}\mathbf{Y} \mathbf{p} \mathbf{K} \mathbf{S} \mathbf{R} \mathbf{F} \mathbf{L} \mathbf{D} \mathbf{P} \mathbf{S} \mathbf{S} \mathbf{V} \mathbf{M} \mathbf{F} \mathbf{A} \mathbf{G} \mathbf{R} \mathbf{P} \mathbf{I} \mathbf{S} \mathbf{K} \mathbf{P} \mathbf{P} \mathbf{S} \mathbf{S} \mathbf{G} \mathbf{F} \mathbf{I} \mathbf{G} \mathbf{L} \mathbf{L} \mathbf{L} \mathbf{P} \mathbf{D} \mathbf{P} \mathbf{O} \mathbf{V} \mathbf{T} \mathbf{F} \mathbf{S} \mathbf{V} \mathbf{P} \mathbf{I} \mathbf{S} \mathbf{A} \mathbf{C} \mathbf{T} \mathbf{F} \mathbf{G} \mathbf{S} \mathbf{M} \mathbf{P} \mathbf{L} \mathbf{A} \mathbf{E} \mathbf{K} \mathbf{S} \mathbf{S} \mathbf{T} \mathbf{T} \mathbf{S} \mathbf{S} \mathbf{S} \mathbf{T} \mathbf{S} \mathbf{S} \mathbf{S} \mathbf{S} \mathbf{S} \mathbf{S} \mathbf{S} S$	48	52	21
ST6Gal II				
<i>Ola</i> AJ871601	RKNHAEQLFYILHPSYLWQLWDLIQSNTQEK-IQPNPPSSGFIGILTMMALCDKLHVYEYIPSMRQTDLCHYHENYYDAACTLGAYHPLIYEK	25	75	
	RMEPSQPFYILHEKYIWQLWDVIQANNLEN-IQPNPPSSGFIGILLWWELCEVHVYEYIPSRCTDLCHYHERYYDAACTLGAYHPLLYEK	19	81	
	RRAHPDQPFYILHPRYVWRLWDVIQGNTQEN-IQPNPPSSGFIGILLMMTLCEQVHVYEYIPSMRQSDLCHYHERYYDAACTLGAYHPLLYEK RKRNPDOPFYILHPKFTWELWKIIOENSNEK-IOPNPPSSGFIGILIMMSMCRTVHVYEYIPSYROTDLCHYHEOYYDAACTLGAYHPLLYEK	13 4	87 : 96 :	
	RKRNPDQPFYILHPKFTWELWKIIQENSNEK-IQPNPPSSGFIGILIMMSMCRTVHVYEYIPSYRQTDLCHYHEQYYDAACTLGAYHPLLYEK RQRNPNQPFYILHPKFIWQLWDIIQENTKEK-IQPNPPSSGFIGILLMMNLCGEVHVYEYVPSVRQTDLCHYHEPYHDAACTLGAYHPLLYEK	4	96 .	
	RKNPNOFFILHERFIWQUWDIIOENTKE-IOPNPPSSGFIGILIMMSKCNEVHVYEIVFSWRJDLCHYHEIYYDAACTLGAYHPLIEK		100	
	RORNPOPFYILHPKFIWQLWDIIOENTKEK-IOPPPSSOFIGILLMMSMCREVHVIEYIPSVROTELCHVHELYYDAACTLGAYHPLLYEK		100 1	
	RLKYPTOPFYILHPKFIWOLWDIIOENTREK-IOPNPPSSGFIGILVMMSMCOEVHVYEYIPSVROTELCHYHELYYDAACTLGAYHPLLYEK		100	
	RORNPNOPFYILHPKFIWOLWDIIOENTKEK-IOPNPPSSGFIGILIMMSMCREVHVYEYIPSVROTELCHYHELYYDAACTLGAYHPLLYEK		100	
	RRKYPTOPFYILHPKFIWOLWDIIOENTREK-IOPNPPSSGFIGILIMMSMCKEVHVYEYIPSVROTELCHYHELYYDAACTLGAYHPLLYEK		100	

Fig. 5. ClustalW alignments of the sialylmotif regions of the ST6Gal I and II subfamilies. White letters with dark gray background represent conserved positions specific from ST6Gal I, and black letters on pale gray background represent conserved positions specific from ST6Gal II. The total number of subfamily-specific conserved positions is represented in the last column preceded by the relative proportions (%) of ST6Gal I and II 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 four insect sialyltransferases from *Drosophila pseudoobscura (Dps* AJ821849), *Drosophila melanogaster (Dme* AF218237), *Drosophila yakuba (Dya* AJ821848), and *Anopheles gambiae (Aga* AJ821850), flanked by thick horizontal lines, cannot be classified in either of the two subfamilies, because they have similar proportions of ST6Gal I and II specific conserved positions (between 48 and 61%, bold characters).

All the remaining gene products could be clearly ascribed to one of the four subfamilies of this branch with the exception of three (*Tru* AJ646869, *Ola* AJ871604, and *Dre* AJ868430) that had intermediate proportions of subfamilyspecific amino acid positions for ST6GalNAc III and IV, and branched out just before the ST6GalNAc III and IV split, but with a very low bootstrap (not shown because it is lower than the threshold of 50%) (Figure 6).

The ST8Sia family

Enzymes of this ST8Sia family mediate the transfer of Neu5Ac residues in $\alpha 2,8$ -linkage to other Neu5Ac residues found in glycoproteins and glycolipids. The two main branches of this family tree contain three subfamilies each:

ST8Sia I, V, and VI in the first branch and ST8Sia II, III, and IV in the second branch (Figure 8). Three fish potential sialyltransferases of the first branch had a proportion of subfamily-specific conserved positions intermediate between ST8Sia V and ST8Sia VI (Figure 9), and they apparently branch out before the split of ST8Sia V and VI, but with a low bootstrap (56%) (Figure 8). Finally, one hypothetical protein from the cenhalocher

Finally, one hypothetical protein from the cephalochordata *Branchiostoma floridae* (*Bfl* AF391289) had intermediate values of subfamily-specific conserved positions in all the 2×2 alignments of the ST8Sia family (Figure 9) and appeared at the root of this family tree, suggesting that it is an ortholog to the common ancestor present before the occurrence of the duplication events that lead to the emergence of the six subfamilies of ST8Sia (Figure 8).

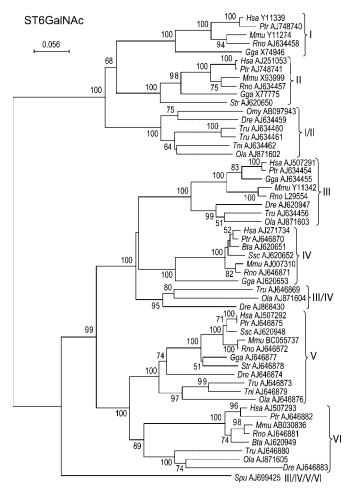


Fig. 6. Neighbor-joining phylogenetic tree of the 55 sialyltransferases of the ST6GalNAc family. One hundred and seventy-two of 247 positions (70%) were selected in eight G-BLOCKS. Bootstrap values were calculated from 500 replicates, and values above 50% are reported on the left of each divergence point. The scale bar represents the number of substitutions per site for a unit branch length. Six genes from the bony fish Oncorhvnchus mykiss (Omy ABO97943), Danio rerio (Dre AJ634459), Takifugu rubripes (Tre AJ634460 and Tre 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 $\alpha 2,6$ and followed by $\alpha 2.3$ and finally $\alpha 2.8$ 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 $\alpha 2.3$ and/or $\alpha 2.6$ sialyltransferase activities have to be present in amphioxus to built 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 $\alpha 2, 3, \alpha 2, 6$, or $\alpha 2, 8$ linkages as described for a Neisseiria meningitidis sialyltransferase (Wakarchuk et al., 2001).

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

ST6GalNAC III	
Hsa AJ507291 CDICAIVSNSGQMVGQEVGNEIDRSSCIMMNNAPTEGYEEDVGRMTMIRVVSHTSVPLLKNDDYFFEANTTIYVIWGPERNREDGIYYNNLKKTVGIYE	PNAQ
Ptr AJ634454 CDLCAIVSNSGQMVGQKVGNEIDRSSCIMRMNNAPTKGYEEDVGRMTMIRVVSHTSVPLLLKNPDYFFKEANTTIYVIWGPFRNMRKDENGIVYNMLKKTVGIYE	
Gga AJ634455 CDLCAIVSNSGQMVGDKVGAEIDKSSCIMRMNNAPTKGYEEDVGKRTTIRVVSHTSVPLLLKNPEYFFKETNSTLYVIWGPFRNMRKDENGIVYNMLKKAVDSYE	TAK
Mmu Y11342 CEHCAIVSNSGOMVGOKVGEIDHASCIMRMNNAPTEGFEDVGYMTMVRVVSHTSVPLLLKNPDYFFEASRTIYVIWGPFRNMRKDENGIVYNMLKKTVDAYE	DAO
Mmu Y11342 CNHCAIVSNSGQMVGQEVGBEIDHASCIMRMNNAPTKGFEBDVGYMTMVRVVSHTSVPLLLKNPDYFFKEASTTIYVIWGPFRNMRKDGNGIVYNNLKKTVDAYE Rno L29554 ONHCAVVSNSGQMVGQEVGBEIDRASCIMRMNNAPTKGFEBDVGYMTMVRVVSHTSVPLLLKNPDYFFKEASTTIYVIWGPFRNMRKDGNGIVYNNLKKTVDAYE	DAO
Dre AJ620947 CENCSVVSSSGQILGREAGADIDOSSCIMRMNNAPTRGFERDVGHRTDLRVVSHTSVPLLIOKPOHFFGQGNETVVVVWGPLRNMRODCKGIVYNMLROAVENYE	
Tru AJ634456 CELCSIVSSSGOMLGEAAGVQIDRSPCIMRMNAPTSGFEHDVGHRTTLRVVSHTSVPLLLQKEHMFFGQGMNTVVVVWGPLRMRKDEKGIVYNVLRQAAVNYE	OAR
Ola AJ871603 CDLCTIVSSSGOMLGOGAGPOIDRSPCVERTNNAPTAGFEMDVGRRTTVRVVSHTSVPLLLOKPOVFFGOENETTYVVWGPLRNMRKDCKGIVYNNLOGASEKYE	RAR
STEGINAC III/IV/V/VI	
Spu AJ699425 CSOCALVSSSGHLVNTSAGAEIDSYPCVLRMNSAPVRGYEVDVGRTTIRIMGHVNLKVLNASNELODEILINSTTRAEKIIVPWLYNVKVNOATE - WYYKSARNLSSLYF	UNE
ST6GalNAC IV	and the second s
Ola AJ871604 CESCALVSSSGQMLNAGAGEEIDQIGCVIRMNNAPTSGFEKDVGSRTSVRVVSHTSVPLLVKNEWYFEQEADTTYVIWGERNMRQDEKGFWFWLLKMATKYF	DKAD
Tru AJ646869 CNQCALVSSSGMLGAGLGBEIDKIQCVIRMNNAPTAGYEBDVGSLTSLRVVSHTSVPLLVKNELYYFBQAANTTYVFWGBDSKMRODCKGCIFNVLLKIAKKYF	
Dre AJ868430 CGRCAVVSSSGMLGGGRGPEIDOODCVIRMNVAPTAGYEADVGNRTSLRVVSHTSVPLLVROGGYFFEREADTKYVIWGEKNMRODCKGKTFNALVTLAKKYI	
Gqa AJ626653 CRCAVVSSSGMLGBLGREIDGCBCVLRMNHAPTAGFEBDVGTRSTVRVVSHTSVPLLLRNOPVFFQOSRDTIYVING-SRKMSREKGGPTHRALLRVDMYF	
Mmu AJ007310 CHSCAVVSJSGOMLGSGLGA0IDGAECVLRMNOAPTVGFBEDVGORTTLRVISHTSVPLLLRNYSHTPOHARDTLYVVWGORHMDRVLGGRTYRTLLOTTRMYI	
RIO A J646871 CHSCAVVSSSGOMLGSGLGAOIDGABCVLRMNOAPTVGFD=DVGORSTLRVISHTSVPLLLRNYSHYPOHARDTLYVVWGOGRHMDRMLGGRTYRTLLOLTRMYI	
Bta AJ620651 CRSCAVVSSSGOMLGSGLGAEIDSAECVLRMNOAPTVGFEADVGORSTLRVISHTSVPLLLRNYSHYPOOARDTLYVVWGOGKHMDRALGGRTYRALLOITRMYI	
SSC AJ620652 CRSCAVVSSSGOMLGSGLGAEIDGAECVLRMNOAPTVGFEADVGRRSTLRVISHTSVPLLLRNYSHYPOOARDTLYVVWGOGRHMDRELGGRTYRALLOLTRMYI	
Hsa AJ271734 CRSCAVVSSSGOMLGSGLGAEIDSAECVFRMNOAPTVGFEADVGORSTLRVVSHTSVPLLLRNYSHYFOKARDTLYMVWGOGRHMDRVLGGRTYRTLLOLTRMYF	
Ptr AJ646870 CRSCAVVSSSGOMLGSGLGAEIDGAECVFRMNOAPTVGFEADVGORSTLRVISHTSVPLLLRNYSHYPOKARDTLYVVWGOGRHMDRVLGGRTYRTLLOLTRMYF	
ST6CAINAG 111 VS % % to	
	57
	57
	54
	55
	56
Dre AJ620947 IVVITEBERNNYCDTVFKKETGKDR OSGSYLSTGWFTLILANDMCKEIRVYGMINDTYCKSEG-VKKVPYHYYEAGSRDECABYLLHESAPYG-CHRFITEK 90 10	
Tru AJ634456 IYUITEDEMNYCDMIFKKETGKDR OSGSYLSTGWFTIILANDMCKEIHIYGMINDTYCKIEG-IRKVPYHYYEPGSRDECSBYMLHESAPYG-CHRFITEK 89 11	
Ola AJ871603 IYITTDERMNYCDTVFKKETGKDRWOSGSYLSTGWFTLILANDMCKEIHIYGMINDTYCKLES-KRTVPYHYYEAGSEDECAEYLLHENAPHE-CHRFITEK 82 18	
ST6GalNAC III/IV/V/VI	*0
STORATING TIT/1/1///TERESLFOTETGLTROETRTWLSTGWMNMLYAVDWCDKVDIFGLVPENYCLKNP-NSSVLYHYYESDGLKECDYYTISEERLTS-EHKFITEK 42 58	22
	55
	27
Ola AJ871604 IFAITREKIQDCDNVFQNETGKNRKKTGAFLSTGFFTMILAMDICDGVTVFGMIDNNHCSRAN-RSVVPYHYYEQNRVSECRMYQVHESTRRE-EHRFITEK 59 41 Tru AJ646869 MYSMTSEKIKYCDQVFQNETGKNRQKSHSWLSTGWFTMVIAIEICDNIKVYGMVPPSHCGKOTGAKKVPYHYYKPRGPDECVTYTQHESSRRENHHRFITEK 56 44	
Tru AJ646655 MISHISHIKICDVPORETGANRONSANDSIGWFIMVALE CONTRYTGMPPSHC562 GARAVPIHIKPROPDEVITIOHDSSKRUMHKTIEN 36 44 Dre AJ868430 VHITRBRVOYCDGVPORETGANRUSSGRISTGFFTMILALEVCDSILVYGMIDGSYCSNAN-HSFVPYHYPEDHLDECRMYRVHEHAKKE-EHRFITEK 38 67	
Gga AJ620653 LYTFFEEKWAYCDDVFONETGKNRUSSSFLSTGWFTMILAMELCEHICVFGMVSDSYCREKN-HSSVPYHYFEKGRLDECRMYLVHERAPRA-EHRFITEK 25 75	
Mmu AJ007310 VYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTMIPALELCEEIVVYGMVSDSYCSEKS-PRSVPYHYFEKGRLDECQMYRLHEQAPRS-AHRFITEK 5 95 Rno AJ646871 VYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTMILALELCEEIVVYGMVSDSYCSEKS-PRSVPYHYFEKGRLDECQMYRLHEQAPRS-AHRFITEK 3 97	
Ssc AJ620652 VYTFTERMMAYCDQVFQDETGKNRRQSGSFLSTGWFTMILALELCHEIVAYGMGSDSYCREKS-HPSVPYHYFEKGRLDECQMYLAHERAPRS-AHRFITEK 2 98 Hsa AJ271734 VYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTMILALELCEEIVVYGMVSDSYCREKS-HPSVPYHYFEKGRLDECQMYLAHEQAPRS-AHRFITEK 0 100	
HSA AJ271734 VYTETERMMAYCDQIFQUETGKNRKQSGSFLSTGWFTMILALELCEEIVVJGMVSDSYCREKS-HPSVPJHYFEKGRLDECQMYLAHEQAPKS-AHRFITEK 0 100 Ptr Aj646870 VYTETERMMAYCDQIFDDETGKNRGSGSFLSTGWFTMILALELCEEIVVJGMVSDSYCREKS-HPSVPJHYFEKGRLDECQMYLAHEQAPKS-AHRFITEK 0 100	
FLT AUG400/U VIITEIERMMAICLQUFQUEIGANKKQSGSELSIGWFIMILALEDCEEIVVIGMVEDSICKEKS-HPSVPIHIFEKGKUDECQMILAHEQAPKS-AHKFITEK U 100	60

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 III. 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* AJ86430) 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 Gal β 1-3GalNAc 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 sialyl-transferases, 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 Gal β 1-4GlcNAc-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) β 1-4GlcNAc-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

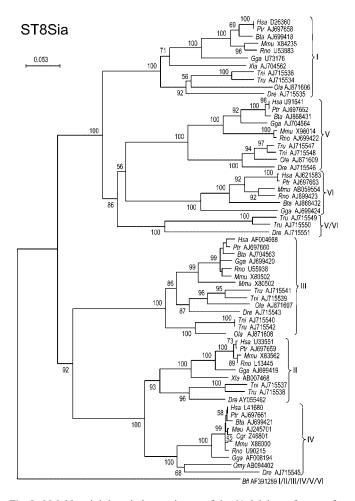


Fig. 8. Neighbor-joining phylogenetic tree of the 64 sialyltransferases of the ST8Sia family. Two hundred and four of the 292 positions (70%) were selected in 10 G-BLOCKS. Bootstrap values were calculated from 500 replicates, and values >50%, are reported at the left of each divergence point. The scale bar represents the number of substitutions per site for a unit branch length. The cephalochordate sialyltransferase sequence from *Branchiostoma floridae* (*Bf* AJ715545) 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* AJ715549 and *Tru* AJ715550) and *Danio rerio* (*Dre* AJ715551), 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 $G_{D1\alpha}$. 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* AJ715549, *Tru* AJ715550, and *Dre* AJ715551 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 cephalochordata *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 disially 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

ST8Sia VL		
Hsa U91641 FKKCAVVGNGGILKNSRCGREINSADFV	FRCNLPPISEKYTMDVGVKTDVVTVNPSIITER-FHKLEKWRRPFYRVLQVYENASVLLP	AFYNTRNTDVSIRVKYVLD
Ptr AJ697662 FKKCAVVGNGGILKNSRCGREINSADFV	FRCNLPPISEKYTMDVGVKTDVVTVNPSIITER-FHKLEKWRRPFYRVLOVYENASVLLP.	AFYNTRNTDVSIRVKYVLD
Bta AJ868431 FKKCAVVGNGGILKNSRCGGEFTSADFV	LRCNLPPISEKYTMDVGVKTDVVTVNPSIITER-FHKLEKWRREFYRVLOLYBNASVLLP. FRCNLPPISEKYLTDVGVKTDVVTVNPSIITER-FHKLEKWRKEFYDVLOYYBNASVLLP.	AFYNTRNTDVSIRVKYVLD
Gga AJ704564 FKKCAVVGNGGILKNSRCGREIDSADFV	FRCNLPPISEKYLTDVGVKTDVVTVNPSIITER-FHKLEKWRKPFYDVLQVYENASVLLP	AFYNTRNTDVSIRVKYVLD
Rno AJ699422 FKACAVVGNGGILKNSGCGKEINSADFV	FRCNLPPESGITTDVGEKTDVVTVNPSIIIER-FHKLEKWRREFESWLERYENASVLLP	AFYNVRNTLVSFRVNYMLD
Mmu X98014 FKKCAVVGNGGILKNSGCGKEINSADFV	FRCNLPPISGIYTTDVGEKTDVVTVNPSIIIDR-FHKLEKMRREFESVLORYENASVLLP. FRCNIPPYS <mark>D</mark> LYSQDVGSKTDLVTINPSIITER-YRFLEKMRKEFYEVLONYENSSVVLP.	AFYNVRNTLVSFRVKYMLD
Dre AJ715546 FKKCAVIGNGGIIKNSKCGREIDASDFV	FRCNIPPWSDLYSQDVGSKTDLVTINPSIITER-YRFLEKWRKPFYEVLONYENSSVVLP	AFYNTRNTDVSFRVKYMLD
Tni AJ715548 FKKCAVVGNGGIIKNSKCGTEIDSADFV	FRCNIPPISEKYSADVGTKTDLVSINPSIITER-FOKLEKWRREFYEVLONYENSSVVLP	AFYNTRNTDVSFRVKYMLD
Ola AJ871609 FKKCAVIGNGGIIKNSKCGKEIDSADFV	FRCNIPPIKEKYSTDVGSKTDLVTINPSIITER-FOKLEKWRRPFYDVLONYENSSVVLP	AFYNTRNTDVSFRVKYMLD
Tru AJ715547 FKKCAVVGNGGIIKNSKCGKEIDSADFV	FRCNIPPISEKYSADVGTKTDLVSINPSIITER-FQKLEKWRRPFYEVLQNYENSSAVLP	AFYNTRNTDVSFRVKYMLD
ST8Sia I/II/III/IV/V/VIL		
	IRIDLPPWAGFEKDVGRKTNMTVLNMDTPKRLEISSRLKNRSLDMYEHRLRGIAGULLA	NRESOPVIKKAFORVHUSE
ST8Sia VI		in bogi i biddin girini or
	IRCNLPPLNNGYQDHVGVKTDLVTANPSILFEK-YGALMGRRRPFIESLRSYGNSLLLIP	A FEVENNTDUCL DA FVETF
Tru AJ715550 WETCAWGNGGILSDSGCGMIDSDOFV	IRCNLPPINNGYODHVGVKTDLVTANPSILVEK-YGALMGRRRPFIESLRSYGNSLLLLP	TECVENTEUCLEAFUCIE
	IKCNLPPLDRYEKDVGNKTNLVTANPSILLEK-YSGLMERRRPFVESLHSYGQALLLLP	
Cap A.1699424 TKNCAVVGNGGTLKNSSCGEEINDEDEV	FRCNLPPTMGSISKDVGNKTKLVTVNPSIIACK-YNKLNEKKTEFVENVAWYGDAFLLLP	AFGERGNNATGERVVUTLO
Mmii AB059554 VNOCAVVGNGGTLNKSLCCAFIDKSDFV	FRCNLPPTGSASKDVGSKTNLVTVNPSIITLK-YONLKEKKAOFLEDISTYGDAFLLLP	AFSTRANTGISEKUVOTI.K
	FRCNLPPTTGNVTNDVGTKTNLVTLNPSIIKLR-YGNLKEKKAIFLEDVAAYGDAFVLLP	
REG AJ699423 VNOCAVVGNGGTLNKSLCGAEIDKSDEV	FRCNLPPTTGSASQDVGSKTNLVTVNPSIITLK-YKNLKEKKARFLEDISAYGDAFLLLP	AFSYRVNTGISEKVVOTLK
	FRCNLPPTTGDVSKDVGSKTNLVTINPSIITLK-YGNLKEKKALFLEDIATYGEAFFLLP	
	FRCNLPPTTGDVSKDVGSKTNLVTINPSIITLK-YGNLKEKKALFLEDIATYGDAFFLLP	
100 00 000 00 00000		- BENEFECTURE DIS
CTRCia V		
ST8Sia V Hea 191641 DEESDOAVWYEHDOVLWNVSEYWESLOW		VS % % total
Hsa U91641 DFESPOAVWYFHPOYLWNVSRYWLSLGV	RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-YITHHYYD-NVKPREGG	HAMPSEI 100 0 60
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV	RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL- <mark>W</mark> ITHHYYD-NVWPRPGG) RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-WITHHYYD-NVWPRPGF)	HAMPSEI 100 0 60 HAMPSEI 100 0 60
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLWNVSRYWLSLGV	RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL- <mark>WITHHYYD-NVKPRP</mark> GG RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-WITHHYYD-NVKPRPGF RAKRISTGLILATAALELCEEVHLFGFWAFPMNPSGL-WITHHYYD-NVKPRPGF	HAMPSEI 100 0 60 HAMPSEI 100 0 60 HAMPSEI 100 0 58
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLWNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPOYLLNVSRYWLCGV	RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYYD-NVKPRFGG RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYYD-NVKPRFGF RAKRISTGLILATAALELCEEVHLFGFWAFPMNPSGL-TITHHYYD-NVKPRFGF RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGI-TITHHYYD-NVKPRFGF	HAMPSEI 100 0 60 HAMPSEI 100 0 60 HAMPSEI 100 0 58 HAMPSEI 100 0 51
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLWNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPOYLLNVSRYWLCGV Rno AJ699422 DFOSGEPVYFFHPHYLSSVSRYWLSLGV	RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYYD-NVKPRPGGJ RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYYD-NVKPRPGFJ RAKRISTGLILATAALELCEEVHLFGFWAFPMNPSGL-TITHHYYD-NVKPRPGFJ RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGF-FITHHYYD-NVKPRPGFJ	HAMPSEI 100 0 60 HAMPSEI 100 0 60 HAMPSEI 100 0 58 HAMPSEI 100 0 61 HAMPSEI 98 2 53
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLWNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPOYLLNVSRYWLCGV Rno AJ699422 DFOSEBPVYFFHPHYLSSVSRYWLSLGV Mmu X98014 DFOSRQPVYFFHPOYLSSVSRYWLSLGV	RAKRISTGLILVTÄALELCEEVHLFGFWAFPMNPSGL-VITHHYYD-NVKPRPGG RAKRISTGLILVTÄALELCEEVHLFGFWAFPMNPSGL-VITHHYYD-NVKPRPGF RAKRISTGLILATAALELCEEVHLFGFWAFPMNPSGL-VITHHYYD-NVKPRPGF RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGF-FITHHYYD-NVKPRPGF RARRISTGLILVTAALELCEEVHLFGFWAFPMNPSGF-FITHHYYD-NVKPRPGF RARRISTGLILVTAALELCEEVHLFGFWAFPMNPSGF-FITHHYYD-NVKPRPGF	HAMPSEI 100 0 60 HAMPSEI 100 0 60 HAMPSEI 100 0 58 HAMPSEI 100 0 61 HAMPSEI 98 2 53 HAMPSEI 98 2 54
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLWNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPOYLWNVSRYWLSLGV Rno AJ699422 DFOSGPPVFFHPOYLSVSRYWLSLGV Mmu X98014 DFOSGQPVFFHPOYLSSVSRYWLSLGV Dre AJ715546 DFESSGVFFFHPOYLLNVORFWAVGGV	RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGG RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF RAKRISTGLILATAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGI-FITHHYD-NVKPRPGF RARRISTGLILVTAALELCEEVHLFGFWAFPMNPSGI-FITHHYD-NVKPKPGF RARRISTGLSLVTAALELCEEVHLFGFWAFPMNPSGI-FITHHYD-NVKPKPGF RARRISTGLSLVTAALELCEEVHLFGFWAFPMNPSGI-FITHHYD-NVKPKPGF	HAMPSEI 100 0 60 HAMPSEI 100 0 60 HAMPSEI 100 0 58 HAMPSEI 100 0 61 HAMPSEI 98 2 53 HAMPSEI 98 2 54 HAMPSEI 98 2 54
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLWNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPOYLLNVSRYWLSLGV Rno AJ699422 DFOSGEPVYFHPOYLSVSRYWLSLGV Mmu X98014 DFOSGPVYFFHPYLSSVSRYWLSLGV Dre AJ715546 DFESSRGVFFHPOYLLNVQRFWAVGGV Tni AJ715548 DFDSQR0VFFHPOYLLNVQRFWAVGV	RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGG RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF RAKRISTGLILATAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGI-TITHHYD-NVKPKPGF RARRISTGLILVTAALELCEEVHLFGFWAFPMNPSGF-TITHHYD-NVKPKPGF RAKRISTGLSVTAALELCEEVHLFGFWAFPMNPSGF-TITHHYD-NVKPKPGF	HAMPSET 100 0 60 HAMPSET 100 0 60 HAMPSET 100 0 58 HAMPSET 100 0 61 HAMPSET 98 2 53 HAMPSET 98 2 54 HAMPSET 98 2 54 HAMPET 98 2 54 HAMPET 98 2 54
Hsa U91641 DFESPQAVYYFHPOYLUNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLUNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLUNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPOYLUNVSRYWLSGV Rno AJ699422 DFOSGEPVYFFHPOYLLSVSRYWLSGV Mmu X98014 DFQSROPVYFFHPOYLSVSRYWLSLGV Dre AJ715546 DFESSRGVFFFHPOYLUNVQRFWAVGGV Tni AJ715548 DFDSQRGVFFFHPOYLUNVQRFWAVGGV Ola AJ871609 DFDSQRSVFFFHPOYLUNVQRFWAVGGV	RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGG RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGI-TITHHYD-NVKPRPGF RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGI-TITHHYD-NVKPKPGF RARRISTGLILVTAALELCEEVHLFGFWAFPMNPSGI-TITHHYD-NVKPKPGF RARRISTGLLVTAALELCEEVHLFGFWAFPMNPSGI-TITHHYD-NVKPKPGF	HAMPSEI 100 0 60 HAMPSEI 100 0 60 HAMPSEI 100 0 58 HAMPSEI 100 0 61 HAMPSEI 98 2 53 HAMPSEI 98 2 54 HAMPSEI 98 2 54 HAMPSEI 98 2 56 HAMPHEI 98 2 57
Hsa U91641 DFESPOAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPOAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPOAVYYFHPOYLWNVSRYWLSLGV Gga AJ704564 DFESQOAVYFHPOYLWNVSRYWLSGV Gma AJ699422 DFOSGBPVYFHPOYLLNVSRYWLSGV Mmu Y98014 DFOSGPVYFHPOYLSVSRYWLSGV Dre AJ715546 DFESSRGVFFHPOYLSVSRYWLSGV Dri AJ715548 DFDSQRSVFFHPOYLLNVQRFWAVGV Ola A3871609 DFDSQRSVFFHPOYLLNVQRFWAVGV Tru AJ715547 DFDSQRSVFFHPOYLLNVQRFWAVGV	RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGG RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF RAKRISTGLILVTAALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF RARRISTGLILVTAALELCEEVHLFGFWAFPMNPSGI-TITHHYD-NVKPKPGF RARRISTGLILVTAALELCEEVHLFGFWAFPMNPSGI-TITHHYD-NVKPKPGF RAKRISSGLMLVTAALELCEEVHLFGFWAFPMNPSGI-TITHHYD-NVKPKPGF RAKRLSSGLMLVTAALELCEEVHLYGFWAFPMNPSGI-TITHHYD-NVKPRPGF	HAMPSEI 100 0 60 HAMPSEI 100 0 60 HAMPSEI 100 0 58 HAMPSEI 100 0 61 HAMPSEI 98 2 53 HAMPSEI 98 2 54 HAMPSEI 98 2 54 HAMPHEI 98 2 56 HAMPHEI 98 2 57
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLWNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPOYLWNVSRYWLSGV Mmu X98014 DFOSGPVYFFHPOYLLNVSRYWLSLGV Mmu X98014 DFOSGPVYFFHPOYLSVSRYWLSLGV Tni AJ715546 DFESGRVFFFHPOYLLNVQRFWAVGV Cla AJ871609 DFDSQRSVFFFHPOYLLNVQRFWAVGV Tru AJ715547 DFDSQRSVFFFHPOYLLNVQRFWAVGV ST8Sia I/II/II/II/V/V/VI	RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGG RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF	HAMPSET 100 0 60 HAMPSET 100 0 60 HAMPSET 100 0 58 HAMPSET 100 0 61 HAMPSET 98 2 53 HAMPSET 98 2 54 HAMPYET 98 2 54 HAMPHET 98 2 57 HAMPHET 98 2 57 HAMPHET 98 2 57 - VS
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLWNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPOYLLNVSRYWLSLGV Mmu X98014 DFOSRQPVYFFHPOYLSVSRYWLSLGV Dre AJ715546 DFESSRVFFFHPOYLLNVQRFWAVGV Tni AJ715548 DFDSQRGVFFFHPOYLLNVQRFWAVGV Ola AJ871609 DFDSQRSVFFFHPOYLLNVQRFWAVGV Tru AJ715547 DFDSQRSVFFFHPOYLLNVQRFWAVGV ST8Sia I/II/III/IV/V/VI Bfl AF391289 VLLISASSFTSAEKINPIASKIARKRYM	RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGG RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPRPGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPRPGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPRPGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPRPGF RAKRISSGLMLVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALELCEEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF LII RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF	HAMPSET 100 0 60 HAMPSET 100 0 60 HAMPSET 100 0 58 HAMPSET 98 2 53 HAMPSET 98 2 54 HAMPSET 98 2 54 HAMPSET 98 2 56 HAMPHET 98 2 56 HAMPHET 98 2 57 HAMPHET 98 2 57
Hsa U91641 DFESPQAVYYFHPOYLUNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLUNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLUNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPOYLUNVSRYWLSLGV Rno AJ699422 DFOSGEPVYFHPOYLLSVSRYWLSLGV Mmu X98014 DFOSQCPVYFHPOYLSVSRYWLSLGV Dre AJ715546 DFESSRGVFFHPOYLLNVQRFWAVGV Tni AJ715546 DFDSQCSVFFHPOYLUNVQRFWAVGV Ola AJ871609 DFDSQCSVFFHPOYLUNVQRFWAVGV Tru AJ715547 DFDSQCSVFFHPOYLUNVQRFWAVGV ST8Sia I/II/III/IV/V/VI Bfl AF391289 VLUSASSFTSAEKINPIASKIARKRYM ST8Sia VI	RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRGG RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPRGF RARRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPKPGF RARRISTGLSLVTALELCEEVHLFGFWAFPMNPSGF- ITHHYD-NVKPKPGF RAKRISSGLMLVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPKPGF RAKRLSSGLMLVTALELCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPRGF RAKRLSSGLMLVTALELCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPRGF	HAMPSET 100 0 60 HAMPSET 100 0 60 HAMPSET 100 0 58 HAMPSET 100 0 58 HAMPSET 98 2 53 HAMPSET 98 2 54 HAMPHET 98 2 54 HAMPHET 98 2 56 HAMPHET 98 2 57 VS VS
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLWNVSRYWLSGV Gga AJ704564 DFESQQAVYYFHPOYLWNVSRYWLSGV Mmu X98014 DFQSGPVYFFHPYLSSVSRYWLSLGV Mmu X98014 DFQSGPVYFFHPOYLLNVQRFWAVGV Dre AJ715546 DFESSGSVFFFHPOYLLNVQRFWAVGV Ola AJ871609 DFDSQRSVFFFHPOYLLNVQRFWAVGV Ola AJ871609 DFDSQRSVFFFHPOYLLNVQRFWAVGV ST8Sia I/II/III/IV/V/VI Bfl AF391289 VLLISASSFTSAEKINPIASKIARKRYM ST8Sia VI Tru AJ715549 DFGSPTRTIFFNPOYLQKLAVFWRSKGL	RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPREGG RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPREGF RAKRISTGLILVTALELCEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RARRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RARRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RARRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RARRISSGLMLVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RARRISSGLMLVTALELCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RARRISSGLMLVTALEECEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF	HAMPSEI 100 0 60 HAMPSEI 100 0 60 HAMPSEI 100 0 58 HAMPSEI 100 0 61 HAMPSEI 98 2 53 HAMPSEI 98 2 54 HAMPYEI 98 2 54 HAMPYEI 98 2 54 HAMPHEI 98 2 57 VS HAMPHEI 98 2 57 VS HAMPHEI 98 3 2 57 VS HAMPHEI 98 3 2 57 VS HAMPHEI 98 3 2 57
Hsa U91641 DFESPQAVYYFHPYLVNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPYLVNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPYLVNVSRYWLSGGV Gga AJ704564 DFESQQAVYYFHPYLLNVSRYWLSGGV Rno AJ699422 DFOSGBPVYFFHPYLSSVSRYWLSLGV Mmu X98014 DFOSQQVYFFHPYLSSVSRYWLSLGV Dre AJ715546 DFESGRVFFHPYLLNVQRFWAVGV Tni AJ715548 DFDSQRVFFFHPYLLNVQRFWAVGV Ola AJ871609 DFDSQRSVFFFHPYLLNVQRFWAVGV ST8Sia I/II/III/IV/V/VI Bfl AF391289 VLLTSASSTSAEKINPIASKIARKRYM ST8Sia VI Tru AJ715549 DFGSPTRTIFFNPYLQKLAVFWRSKGL Tru AJ715550 DFGSPTRAISFNPQYLQKLAVFWRSKGL	RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRFGG RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRFGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPRFGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPKPGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPKPGF RAKRLSSGLMLVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPKPGF RAKRLSSGLMLVTALELCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPRFGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPRFGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPRFGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPRFGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPRFGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPRFGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPRFGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI- ITHHYDD-NVKPRFGF RAKRLSSGLMLVTALEMCEEVHLYGFWAFPMNPSGI- ITHHYDD-NVKPRFGF 	HAMPSET 100 0 60 HAMPSET 100 0 60 HAMPSET 100 0 58 HAMPSET 100 0 58 HAMPSET 98 2 53 HAMPSET 98 2 54 HAMPYET 98 2 54 HAMPHET 98 2 57 HAMPHET 98 2 57 -VS HAMPHET 98 2 57 -VS HAMPYEF 46 54 39 HAMPYEF 45 55 40
Hsa U91641 DFESPQAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPOYLWNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPOYLWNVSRYWLSLGV Rno AJ699422 DFOSGEPVYFFHPOYLWNSRYWLSLGV Mmu X98014 DFOSQDAVFFHPOYLSVSRYWLSLGV Dre AJ715546 DFESSGVFFFHPOYLLNVQRFWAVGV Tni AJ715548 DFDSQRGVFFFHPOYLLNVQRFWAVGV Gla AJ871609 DFDSQRSVFFFHPOYLUNVQRFWAVGV Tru AJ715547 DFDSQRGVFFFHPOYLUNVQRFWAVGV ST8Sia I/II/II/IV/VI Bfl AF391289 VLLHSASSFTSAEKINPIASKIARKRYM ST8Sia VI Tru AJ715549 DFGSPTRTIFFNPOYLQKLAVFWRSKGL Tru AJ715550 DFGSPTRTIFFNPOYLQKLAVFWRSKGL Tru AJ715551 DFGSPTRTIFFNPOYLQKLAVFWRSKGL Tru AJ715551 DFGSPTRTIFFNPOYLQKLAVFWRSKGL	RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRGG RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPKPGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPKPGF RAKRISGLMVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPKPGF RAKRLSSGLMVTALELCEEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMVTALELCEEVHLGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMVTALELCEEVHLGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMVTALEMCEEVHLGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMVTALEMCEEVHLGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEEVHLGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEEVHLGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEEVHLGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEEVHLGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEVHLGFWAFPMNPSGI ITHHYD-NVKPRGF 	HAMPSET 100 0 60 HAMPSET 100 0 60 HAMPSET 100 0 58 HAMPSET 100 0 58 HAMPSET 98 2 53 HAMPSET 98 2 54 HAMPSET 98 2 54 HAMPHET 98 2 57 VS HAMPHET 98 2 57 VS HAMPHET 98 2 57 VS HAMPHET 98 2 57 VS HAMPYEF 46 54 39 HAMPVEF 45 55 40 HAMPYEF 43 57 40
Hsa U91641 DFESPOAVYYFHPOYLWNVSRYWLSLGV Ptr AJ697662 DFESPOAVYYFHPOYLWNVSRYWLSLGV Bta AJ868431 DFESPOAVYYFHPOYLWNVSRYWLSLGV Gga AJ704564 DFESQOAVYYFHPOYLWNVSRYWLSGGV Rno AJ699422 DFOSGBPVYFFHPYLSSVSRYWLSLGV Mmu X98014 DFOSGOPVFFHPOYLLNVORFWAVGV Dre AJ715546 DFESGRVFFHPOYLLNVORFWAVGV Tri AJ715548 DFDSQRVFFHPOYLLNVORFWAVGV Tru AJ715547 DFDSQRVFFHPOYLLNVORFWAVGV Tru AJ715547 DFDSQRVFFHPOYLLNVORFWAVGV ST8Sia I/II/III/IV/V/VI Bfl AF391289 VLISASSFTSAEKINPIASKIARKRYM ST8Sia VI Tru AJ715549 DFGSPTRTIFFNPOYLQKLAVFWRSKGL Tru AJ715550 DFGSPTRAISFNPOYLQKLAVFWRSKGL Dre AJ715551 DFGSPTRAISFNPOYLQKLAVFWRSKGL Dre AJ715551 DFGRDPLPIFLNPYLRKLTKFWREGL Gga AJ699424 EFKATORAIFFOPTYLKSLAQFWATKGV	RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPREGG RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL-TITHHYD-NVKPREGF RAKRISTGLILATALELCEEVHLFGFWAFPMNPSGI-FITHHYD-NVKPREGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI-FITHHYD-NVKPREGF RAKRISTGLSUTALELCEEVHLFGFWAFPMNPSGI-FITHHYD-NVKPREGF RAKRLSSGLMLVTALELCEEVHLFGFWAFPMNPSGI-FITHHYD-NVKPREGF	HAMPSEI 100 0 60 HAMPSEI 100 0 60 HAMPSEI 100 0 61 HAMPSEI 100 0 61 HAMPSEI 98 2 53 HAMPSEI 98 2 54 HAMPSEI 98 2 54 HAMPHEI 98 2 56 HAMPHEI 98 2 57 VS HAMPHEI 98 2 57 VS HAMPHEI 98 2 57 HAMPHEI 35 40 HSMPSEI 43 57 40 HSMPSEI 43 57 40
Hsa U91641 DFESPQAVYYFHPYLVNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPYLVNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPYLNVSRYWLSGV Gga AJ704564 DFESQQAVYFHPYLNVSRYWLSGV Mmu X98014 DFOSQPVYFFHPYLSSVSRYWLSGV Mmu X98014 DFOSQPVYFFHPYLSSVSRYWLSGV Tni AJ715548 DFDSQRSVFFHPYLLNVQRFWAVGV Tni AJ715548 DFDSQRSVFFHPYLLNVQRFWAVGV Tru AJ715547 DFDSQRSVFFHPYLLNVQRFWAVGV ST8Sia I/II/III/IV/V/VI Bfl AF391289 VLLISASSTSAEKINPIASKIAKKRYM ST8Sia VI Tru AJ715551 DFGSPTRTIFFNPYLQKLAVFWRSKGL Tru AJ715551 DFGSPTRAISFNPOYLQKLAVFWRSKGL Tru AJ715551 DFGSPTRAISFNPOYLQKLAVFWRSKGL Dre AJ715551 DFGSPTRAISFNPOYLQKLAVFWRSKGL Dre AJ715554 ESKMRCKVLFFHPPYLRKLAFWRTKGV	RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRFGG RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRFGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRFGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPRFGF	HAMPSET 100 0 60 HAMPSET 100 0 60 HAMPSET 100 0 58 HAMPSET 100 0 61 HAMPSET 98 2 53 HAMPSET 98 2 54 HAMPYET 98 2 54 HAMPYET 98 2 57 VS HAMPHET 98 2 57 VS HAMPHET 98 2 57 VS HAMPYEF 45 50 32 VS HAMPYEF 45 54 39 HAMPVEF 45 55 40 HSMPSEF 43 57 40 HAMPYEF 43 57 51 HOMPKEY 4 96 57
Hsa U91641 DFESPQAVYYFHPYLVNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPYLVNVSRYWLSLGV Bta AJ868431 DFESPQAVYYFHPYLVNVSRYWLSLGV Gga AJ704564 DFESQQAVYYFHPYLNVSRYWLSLGV Mmu X98014 DFESQQAVYFHPYLSSVSRYWLSLGV Mmu X98014 DFOSRQPVFFHPYLSSVSRYWLSLGV Mmu X98014 DFOSRQPVFFHPYLSSVSRYWLSLGV Mmu X98014 DFOSRQPVFFHPYLSSVSRYWLSLGV Mmu X98014 DFOSRQPVFFHPYLSSVSRYWLSLGV Mmu X98014 DFOSRQPVFFHPYLSVSRYWLSLGV Tri AJ715548 DFDSQRSVFFFHPYLLNVQRFWAVGV ST85ia J111/J11/V/V/VI Bf1 AF391289 VLLISASSFTSAEKINPIASKIARKRYM ST8Sia VI Tru AJ715550 DFGSPTRAISFNPOYLQKLAVFWRSKGL Tru AJ715550 DFGSPTRAISFNPOYLQKLAVFWRSKGL Dre AJ715550 DFGSPTRAISFNPOYLQKLAVFWRSKGL Dre AJ715551 DFGRDPLPIFLNPEYLRKLKVFWRSKGL Dre AJ715551 DFGRDPLPIFLNPEYLRWLAUFWRSKGL Dre AJ715551 DFGRDPLPIFLNPEYLRHLALFWRTKGV Mmu	RAKRISTGLILVTALELCEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRGG RAKRISTGLILVTALELCEVHLFGFWAFPMNPSGL- ITHHYD-NVKPRGF RAKRISTGLILVTALELCEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRISTGLILVTALELCEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRISTGLILVTALELCEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRISTGLILVTALELCEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALELCEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALELCEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALELCEVHLFGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF RAKRLSSGLMLVTALEMCEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF NVKPRGF RAKRLSSGLMLVTALEMCEVHLYGFWAFPMNPSGI ITHHYD-NVKPRGF NVKPRGF NII RAKRLSSGLMLVTALEMCEVHLYGFWFFSMFPMNPSGI ITHHYD-NVKPRGF NUT RAKRLSSGLMVASLALELCSNVHLFGFWFFSMFPMCLRTLKNHYDD-NVKPRGF NII RAKRLSSGLMVASLALELCSNVHLFGFWFFSMFPMCLRTLKNHYDD-NKFKKF NSVRPSTGLIMASLALELCSNVHLYGFWFSKTEM-FISHHYDD-NKPKGF NAVRLSSGFMITSAVELCENVKLYGFWFSKTEM-FISHHYDD-NMLPKGF TAYRLSTGLMISVAVELCENVKLYGFWFSKTEM-FISHHYDD-NMLPKGF	HAMPSET 100 0 60 HAMPSET 100 0 60 HAMPSET 100 0 58 HAMPSET 100 0 58 HAMPSET 98 2 53 HAMPSET 98 2 54 HAMPSET 98 2 54 HAMPYET 98 2 54 HAMPYET 98 2 57 VS HAMPHET 98 2 57 VS HAMPYEF 45 50 32 VS HAMPYEF 45 54 39 HAMPYEF 45 55 40 HSMPSEF 43 57 40 HSMPSEF 43 57 40 HSMPSEF 43 57 51 HOMPKEY 23 97 58
Hsa U91641 DFESPOAVYYFHPYLVNVSRYWLSLGV Ptr AJG97662 DFESPOAVYYFHPYLVNVSRYWLSLGV Bta AJ868431 DFESPOAVYFHPYLVNVSRYWLSGV Gga AJ704564 DFESQQAVYFHPYLNVSRYWLSGV Rno AJG99422 DFOSGBPVYFHPHYLSSVSRYWLSGV Dre AJ715546 DFESGRVFFHPYLSVSRYWLSGV Dre AJ715548 DFDSQRVFFHPYLLNVQRFWAVGV Tru AJ715548 DFDSQRVFFHPYLLNVQRFWAVGV Tru AJ715547 DFDSQRVFFHPYLLNVQRFWAVGV Tru AJ715549 DFDSQRSVFFHPYLLNVQRFWAVGV ST8Sia I/II/III/IV/V/VI Bfl AF391289 VLLSASSFTSAEKINPIASKIARKRYM ST8Sia VI Tru AJ715550 DFGSPTRIFFNPYLQKLAVFWRSKGL Dre AJ715551 DFGSPTRIFFNPYLRKLAKFWREGL Gga AJ699424 ESKARQKVLFFHPRYLRHLALFWRTKGV Mu AB059554 ESKARQKVLFFHPRYLRHLALFWRTKGV Bta AJ868432 ESKARQKVLFFHPRYLRHLALFWRTKGV	RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPREGG RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPREGF RAKRISTGLILATALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISTGLSUTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRLSSGLMLVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRLSSGLMLVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRLSSGLMLVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRLSSGLMLVTALELCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRLSSGLMLVTALENCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRLSSGLMLVTALENCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRLSSGLMLVTALENCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF	HAMPSEI 100 0 60 HAMPSEI 100 0 60 HAMPSEI 100 0 58 HAMPSEI 100 0 61 HAMPSEI 98 2 53 HAMPSEI 98 2 54 HAMPSEI 98 2 54 HAMPSEI 98 2 54 HAMPHEI 98 2 56 HAMPHEI 98 2 57 VS HAMPHEI 98 2 57 VS HAMPVEF 45 55 40 HAMPVEF 45 55 40 HAMPVEF 45 57 40 HAMPVEF 43 57 40 HAMPVEF 49 57 HAMPVEF 49 57 HAMPVEF 49 57 HAMPVEF 3 57 40 HAMPVEF 3 57 40 HAMPVEF 3 57 58 HOMPKEY 3 97 58 HOMPKEY 2 98 55
Hsa U91641 DFESPQAVYYFHPYLVNVSRYWLSLGV Ptr AJ697662 DFESPQAVYYFHPYLNVVSRYWLSLGV Bta AJ688431 DFESPQAVYYFHPYLNVVSRYWLSLGV Gga AJ704564 DFESQQAVYFHPYLNVSRYWLSGV Mmu X98014 DFOSGPVYFFHPYLSSVSRYWLSLGV Mmu X98014 DFOSGPVYFFHPYLSSVSRYWLSLGV Dre AJ715546 DFESGRVFFFHPYLLNVQRFWAVGV Dri AJ715548 DFDSQRSVFFHPYLLNVQRFWAVGV Ola AJ871609 DFDSQRSVFFHPYLLNVQRFWAVGV ST8Sia I/II/III/IV/V/VI Bfl AF391289 VLLSASSFTSAEKINPIASKIARKRYM ST8Sia VI Tru AJ715549 DFGSPTRTIFFNPYLLNVQRFWAVGV Tru AJ715550 DFGSPTRAISFNPYLQKLAVFWRSKGL Tru AJ715519 DFGSPTRAISFNPYLQKLAVFWRSKGL Dre AJ715551 DFGSPTRAISFNPYLKLAFFWREGL Gga AJ699424 ESKARQKVLFFHPYLRHLALFWRTKGV Mmu AB059554 ESKWRQKVLFFHPYLRHLALFWRTKGV Rno AJ699423 ESKARQKVLFFHPKYLKNLALFWRTKGV Ptr AJ697663 ESKARQKVLFFHPKYLKNLALFWRTKGV	RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPREGG RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGL- ITHHYD-NVKPREGF RAKRISTGLILATALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISTGLILVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALELCEEVHLFGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALELCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALEKCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALEKCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALEKCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALEKCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALEKCEEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALEKCEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALEKCEVHLYGFWAFPMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALEKCEVHLYGFWFFFMNPSGI- ITHHYD-NVKPREGF RAKRISSGLMLVTALEKCEVHLYGFWFFFFGO RAKRISSGLMLVTALEKCEVHLYGFWFFFFGO RAKRISSGLMLVTALEKCEVHLYGFWFFFFGO RAKRISSGLMLVTALEKCEVHLYGFWFFFFGO RAKRISSGLMLVTALECENVLYGFWFFSKTEKT- 	HAMPSET 100 0 60 HAMPSET 100 0 60 HAMPSET 100 0 58 HAMPSET 100 0 58 HAMPSET 98 2 53 HAMPSET 98 2 54 HAMPSET 98 2 54 HAMPYET 98 2 54 HAMPYET 98 2 57 VS HAMPHET 98 2 57 VS HAMPYEF 45 50 32 VS HAMPYEF 45 54 39 HAMPYEF 45 55 40 HSMPSEF 43 57 40 HSMPSEF 43 57 40 HSMPSEF 43 57 51 HOMPKEY 23 97 58

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 gaped-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: espressed 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

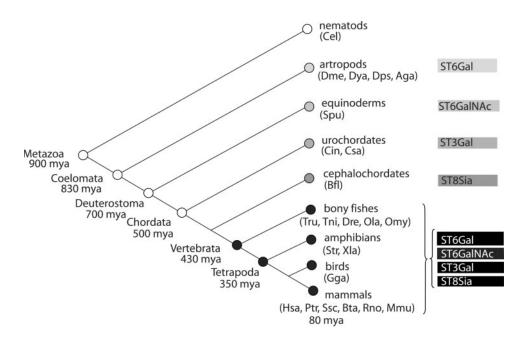


Fig. 10. Scheme of the divergent evolutionary model of the four families of sialyltransferases. The first detectable member of each of the four main families of sialyltransferases is shown, suggesting a general trend for the appearance of sialyltransferase activities: starting with $\alpha 2,6$, followed 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 PHYLOWIN (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 subfamilyspecific 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 <i>in silico</i> 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.htm		
Takifugu rubripes	http://fugu.hgmp.mrc.ac.uk/blast/		
Tetraodon nigroviridis	http://www.genoscope.cns.fr/		
Oryzias latipes	http://dolphin.lab.nig.ac.jp/medaka/ index.php		
Ciona savignyi	http://www.broad.mit.edu/annotation/ciona/		
Ciona intestinalis	http://genome.jgi-psf.org/cgi-bin/runAlign- ment?db=ciona4		
Ciona intestinalis	http://ghost.zool.kyoto-u.ac.jp/indexr1.html		
Drosophila melanogaster	http://www.fruitfly.org/seq_tools/splice.html		
Anopheles gambiae	http://www.genoscope.cnrs.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 savigny; 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 Galβ1–3GalNAc α2,3-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 Nacetylgalactosamine α2,6-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-3deoxy-D-glycero – galactonononic 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., Imberty, A., and Breton, C. (2004) Structurefunction 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 glycosyltransferases 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-Osulfated N-glycolylneuraminic acid-capped α2–5-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 α2,6-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 $\alpha 2,8/2,9$ -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 α2,8-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 α2,6-sialyltransferase (ST6Gal II), which sialylates Galβ1,4GlcNAc 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 α2,6-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.