

Published in final edited form as:

*Curr Opin Struct Biol.* 2010 October ; 20(5): 536–542. doi:10.1016/j.sbi.2010.07.004.

## Structure-based Evolutionary Relationship of Glycosyltransferases: A Case Study of Vertebrate $\beta$ 1, 4-Galactosyltransferase, Invertebrate $\beta$ 1,4-*N*-acetylgalactosaminyltransferase and $\alpha$ -Polypeptidyl-*N*-acetylgalactosaminyltransferase

Boopathy Ramakrishnan<sup>1,2</sup> and Pradman K. Qasba<sup>1,\*</sup>

<sup>1</sup> Structural Glycobiology Section, Center for Cancer Research Nanobiology Program, Center for Cancer Research, NCI-Frederick, Frederick, MD 21702

<sup>2</sup> Basic Research Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, Center for Cancer Research, NCI-Frederick, Frederick, MD 21702

### Summary

Cell surface glycans play important cellular functions and are synthesized by glycosyltransferases. Structure and function studies show that the donor sugar specificity of the invertebrate  $\beta$ 1,4-*N*-acetylgalactosaminyltransferase ( $\beta$ 4GalNAc-T) and the vertebrate  $\beta$ 1,4-galactosyltransferase I ( $\beta$ 4Gal-T1) are related by a single amino acid residue change. Comparison of the catalytic domain crystal structures of the  $\beta$ 4Gal-T1 and the  $\alpha$ -polypeptidyl-GalNAc-T (*app*GalNAc-T) shows that their protein structure and sequences are similar. Therefore, it seems that the invertebrate  $\beta$ 4GalNAc-T and the catalytic domain of *app*GalNAc-T might have emerged from a common primordial gene. When vertebrates emerged from invertebrates, the amino acid that determines the donor sugar specificity of the invertebrate  $\beta$ 4GalNAc-T might have mutated, thus converting the enzyme to a  $\beta$ 4Gal-T1 in vertebrates.

### Introduction

Glycosyltransferases exist as a super family of enzymes and constitute 1 to 2% of the human genome [1]. These transferases have been thought to have evolved by gene duplication and divergence as evidenced from their high protein sequence and gene structure (exon/intron organization) similarity among the subfamily members [2–4]. They generally transfer a donor sugar from an activated donor substrate such as a nucleotide-diphosphate-sugar to an acceptor sugar, extending the glycan chain of glyco-proteins and glyco-lipids, one sugar moiety at a time. The complex glycans, synthesized by involving many glycosyltransferases, are known to have very important cellular functions [5]. Therefore, absence or lethal mutation in any glycosyltransferase gene and its corresponding protein might lead to a total absence of a sugar epitope, thus changing the cellular physiology. For example, in humans the  $\alpha$ Gal1-3 $\beta$ Gal

\*Corresponding author: Structural Glycobiology Section, CCRNP, CCR, NCI-Frederick, Building 469, Room 221, Frederick, Maryland 21702; qasba@helix.nih.gov. Phone: 301-846-1934; Fax: 301-846-7149.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

epitope is absent and has an anti- $\alpha$ Gal IgG antibody against this epitope [6]. Sequencing of the human  $\alpha$ 1,3-galactosyltransferase gene that is responsible for the synthesis of this epitope has revealed that because of a number of lethal mutations, the active  $\alpha$ 3 galactosyltransferase enzyme can not be expressed in humans, thus explaining the absence of the epitope [7]. Similarly, sequencing of the human CMP-*N*-acetylneuraminic acid hydroxylase gene showed an inactivating mutation that is responsible for the lack of the *N*-glycolylneuraminic acid (Neu5Gc) in humans; Neu5Gc is otherwise widely found in most mammalian tissues [8]. The loss of Neu5Gc and access of its precursor *N*-acetylneuraminic acid (Neu5Ac) has been implicated in the evolution of their binding proteins, Siglecs, and in the evolutionary changes in the human lineage [9]. Similarly, other glycosyltransferases are known to have been involved in the evolution and diversification of species [10].

### **Crystal structure of bovine $\beta$ 4Gal-T1 revealed that a single amino acid, Tyr289, determines the donor sugar specificity**

In vertebrate cells, the most prevalent epitope, LacNAc ( $\beta$ Gal1-4 $\beta$ GlcNAc), is recognized by asialoglycoprotein receptors [11]. Furthermore, a LacNAc moiety with a terminal sialic acid linked in different configurations is involved in immunological and cell signaling functions [12–14]. Thus, the galactose moiety in the LacNAc epitope seems to have played an important role in vertebrate evolution. In the presence of manganese, the  $\beta$ 1,4-galactosyltransferase I ( $\beta$ 4Gal-T1) enzyme transfers the donor sugar, galactose (Gal), from UDP-Gal to an acceptor sugar  $\beta$ -*N*-acetylglucosamine ( $\beta$ GlcNAc) synthesizing LacNAc,  $\beta$ Gal1-4 $\beta$ GalNAc [15]. From the enzyme kinetic studies it has been known that the  $\beta$ 4Gal-T1 poorly transfers *N*-acetylgalactosamine (GalNAc) from UDP-GalNAc to the acceptor sugar GlcNAc [16,17]. In the crystal structure of the bovine  $\beta$ 4Gal-T1•Mn<sup>2+</sup>•UDP-GalNAc complex, the side-chain hydroxyl group of the Tyr289 residue forms a hydrogen bond with the carbonyl oxygen atom of the *N*-acetyl group of the bound UDP-GalNAc molecule; due to the lack of enough space between them, the *N*-acetyl moiety is found in a unfavorable orientation (Figure 1a) [17]. It was reasoned that this steric hindrance could be responsible for the poor catalytic activity by the  $\beta$ 4Gal-T1 enzyme with UDP-GalNAc as the donor substrate. Therefore, it was thought that a substitution of the Tyr289 residue with an amino acid with a less bulky side chain, such as Leu, Ile, or Asn, might not cause such steric hindrance and thereby would be expected to enhance the catalytic activity with UDP-GalNAc. Indeed, it was found that such substitution with any of these three amino acids enhanced the catalytic activity using UDP-GalNAc. The mutant enzyme with Leu substitution exhibited catalytic activity that was as efficient with the UDP-GalNAc as it was with the UDP-Gal. Therefore, the single amino acid, Tyr289, determines the donor sugar specificity of bovine  $\beta$ 4Gal-T1 [17].

### **The $\beta$ 4Gal-T1 homolog in the invertebrates is $\beta$ 4GalNAc-T enzyme and has a mutation at the Tyr289 residue**

In contrast to vertebrate cells that have LacNAc moiety invertebrate cells have mainly a LactiNAc moiety where  $\beta$ -*N*-acetylglucosamine ( $\beta$ GalNAc) is 1–4 linked to  $\beta$ GlcNAc [18]. Thus, it seems that in the invertebrate cells the vertebrate  $\beta$ 4Gal-T1, that is responsible for the synthesis of LacNAc moiety, is replaced by a  $\beta$ 1,4-*N*-acetyl-galactosaminyltransferase ( $\beta$ 4GalNAc-T) that synthesize the LactiNAc moiety. Indeed in several invertebrates a  $\beta$ 4GalNAc-T enzyme has been found that is homologous to the vertebrate  $\beta$ 4Gal-T1 protein [19–22]. These two enzymes not only exhibit reasonable protein sequence similarity but also similar gene structure [23]. However, the corresponding Y289 residue of bovine  $\beta$ 4Gal-T1 in the invertebrate  $\beta$ 4GalNAc-T proteins is found either as Leu or Ile residue, suggesting that this amino acid change may be responsible for the altered donor sugar, UDP-GalNAc, specificity of these enzymes (Figure 1b) [20,21].

To provide support for this hypothesis, we have expressed *Drosophila*  $\beta$ 4GalNAc-TA protein, a homolog protein of vertebrate  $\beta$ 4Gal-T1, that transfers GalNAc from UDP-GalNAc to the acceptor substrate GlcNAc; it transfers Gal from UDP-Gal to GlcNAc very poorly [23]. When the Ile289 residue is mutated to a Tyr residue, the mutant enzyme loses the catalytic activity with UDP-GalNAc; instead, it exhibits high catalytic activity with UDP-Gal transferring Gal to GlcNAc, thus confirming that this single amino acid determines the sugar donor specificity of *Drosophila*  $\beta$ 4GalNAc-TA enzyme [23]. It seems that when the vertebrates emerged from the invertebrates nearly 500 million years ago, the vertebrate  $\beta$ 4Gal-T1 enzyme might have also emerged from invertebrate  $\beta$ 4GalNAc-T, by the single amino acid mutation of Leu/Ile in the  $\beta$ 4GalNAc-TA to Tyr, thus converting the  $\beta$ 4GalNAc-T to  $\beta$ 4Gal-T (Figure 2) [23].

### The invertebrate $\beta$ 4GalNAc-T enzymes have an $\alpha$ -lactalbumin binding site

Lysozyme and  $\alpha$ -lactalbumin ( $\alpha$ -LA) share high similarity between their protein sequences and gene structures but have completely different biological functions [24]. These proteins are thought to have evolved from a primordial gene by a gene duplication nearly 300 million years ago.  $\alpha$ -LA is a calcium binding protein that is only expressed in the mammary gland during lactation. The evolution of mammals has been associated with the appearance of  $\alpha$ -LA in the mammary gland where it plays a role in the synthesis of lactose, a hallmark of mammals.  $\alpha$ -LA modulates  $\beta$ 4Gal-T1 by altering its acceptor specificity from GlcNAc to glucose (Glc), thus synthesizing the disaccharide lactose,  $\beta$ Gal1-4 $\beta$ Glc [25]. Also,  $\alpha$ -LA is known to inhibit the transfer of Gal to a GlcNAc acceptor by  $\beta$ 4Gal-T1. Interestingly, we have found that  $\alpha$ -LA binds to the wild-type *Drosophila*  $\beta$ 4GalNAc-TA enzyme and inhibits the catalytic activity of this protein but it does not alter its acceptor specificity from GlcNAc to Glc [23]. Earlier, a similar property was also observed for the *C. elegans*  $\beta$ 4GalNAc-T enzyme [20]. These observations suggest that the  $\alpha$ -LA binding site existed even before the appearance of the vertebrate  $\beta$ 4Gal-T1 protein, and that during the evolution of mammals, this site was utilized by  $\alpha$ -lactalbumin to modulate the acceptor specificity of the  $\beta$ 4Gal-T1 in mammals (Figure 2). This is further evidenced from the fact that the  $\alpha$ -LA can bind and modulate the acceptor specificity of a non-mammal vertebrate  $\beta$ 4Gal-T1 enzyme, like the one from chicken [26].

Although the evolution of  $\beta$ 1,4-galactosyltransferase appears to be coincidental with the appearance of vertebrates, the evolution of this enzyme played an important role in vertebrate development. For example, only in vertebrates several sialyltransferases are found to synthesize variety of sialylated LacNAc containing glycans, and during the evolution of mammals  $\alpha$ -LA was recruited for the synthesis of lactose. Similarly, different lectins had to be recruited to recognize these new glycoconjugates. Thus, the appearance of  $\beta$ 1,4-galactosyltransferases in vertebrates created a boundary between vertebrate and invertebrate glycans and made it possible to expand the variety of glycoconjugates that were needed for vertebrate development.

### The $\beta$ 4Gal-T1 and the catalytic domain of $\alpha$ ppGalNAc-T enzymes show structural similarity

Having observed an evolutionary relationship between the vertebrate  $\beta$ 4Gal-T1 and invertebrate  $\beta$ 4GalNAc-T enzymes, it was important to find out if there are any other GalNAc-T enzymes that might be related to them. In all species,  $\alpha$ -polypeptidyl-GalNAc-T ( $\alpha$ ppGalNAc-T) exists as a family of enzymes, and in the presence of manganese, they transfer GalNAc to the side-chain hydroxyl group of a Thr/Ser amino acid in a polypeptide acceptor [27]. In these transferases the presence of a lectin domain C-terminal to the catalytic domain and separated by a linker polypeptide had been predicted well before the crystal structure of  $\alpha$ ppGalNAc-T was available [28]. However, its structural relationship with the  $\beta$ 4Gal-T1 was first shown by Kubota et al. [29] who determined the crystal structure of an  $\alpha$ ppGalNAc-T10•Mn<sup>2+</sup>•UDP-GalNAc complex, which showed that the binding of UDP-GalNAc to  $\alpha$ ppGalNAc-T10 molecules was similar to UDP-GalNAc binding to  $\beta$ 4Gal-T1 enzyme [29]. The crystal structure studies on  $\alpha$ ppGalNAc-T2 revealed that upon the binding of manganese

and the donor sugar substrate, the enzyme undergoes conformational changes involving separate short and long flexible loops, similar to the  $\beta$ 4Gal-T1 enzyme (Figure 3a and b) [30]. Interestingly, in the short loop of these proteins, the side chain of a Trp residue undergoes conformational changes from outside to inside of the catalytic pocket allowing it to interact with the  $\beta$ -phosphate oxygen atom of the bound UDP-sugar substrate molecule. The superposition of the crystal structures of the bovine  $\beta$ 4Gal-T1 and the catalytic domain of  $\alpha$ ppGalNAc-T10 shows similarity in their crystal structures, protein sequences, and gene structures (Figure 3e and f). Interestingly, the Tyr289 residue that determines the donor sugar specificity of the bovine  $\beta$ 4Gal-T1 is present as Ala in the  $\alpha$ ppGalNAc-Ts. Therefore, the catalytic domains of  $\alpha$ ppGalNAc-T and the vertebrate  $\beta$ 4Gal-T1 seem to be related. Since vertebrate  $\beta$ 4Gal-T1 emerged from the invertebrate  $\beta$ 4GalNAc-T, the catalytic domain of the  $\alpha$ ppGalNAc-T and the invertebrate  $\beta$ 4GalNAc-T might have evolved from a common primordial glycosyltransferase; and during the evolution the  $\alpha$ ppGalNAc-T acquired the lectin domain (Figure 2). Based on their DNA sequences, this gene duplication and divergence was predicted to have occurred 1.3 billion years ago [2].

An additional protein domain, as either a contiguous or a non-contiguous polypeptide, has been observed in many other glycosyltransferases, and such domains have been defined as “add-on” domains [31,32]. For example, in the lactose synthase complex,  $\alpha$ -LA alters the acceptor sugar specificity of  $\beta$ 4Gal-T1 to Glc, thus acting as a non-contiguous “add-on” domain, similar to the lectin domain in the  $\alpha$ ppGalNAc-Ts (Figure 4). Recently, the crystal structure of  $\alpha$ 1,6-fucosyltransferase-8 revealed that this enzyme also has at its C-terminal end a SH3-like “add-on” domain that is usually found in proteins that interact with other proteins and mediate assembly of specific protein complexes, typically via binding to proline-rich peptides in their respective binding partner [31]. Interestingly, the N-terminal tandem-repeat, an “add-on” domain in O- $\beta$ GlcNAc-T, is known to impart acceptor substrate specificity of the enzyme [33]. Thus, it seems that during evolution these enzymes have acquired these “add-on” domains to impart new functions [32].

### Evolution of $\beta$ 4Gal-T7 enzyme

In all vertebrates, the  $\beta$ 1,4-galactosyltransferase family has seven members,  $\beta$ 4Gal-T1 to -T7, and they all transfer Gal from UDP-Gal to different acceptor sugar substrates [4]. Interestingly,  $\beta$ 4Gal-T7 differs from the rest by its gene structure [34]. However, all the known  $\beta$ 4Gal-T7 protein homologs in invertebrates and vertebrates have the conserved Phe/Tyr residue as in  $\beta$ 4Gal-T1 [34,36–40]. The  $\beta$ 4Gal-T7 is involved in the synthesis of the tetrasaccharide linker sequence of proteoglycans for heparin/heparan sulfate or chondroitin sulfate or dermatan sulfates, which are widely found on the cell surface and in the extracellular matrix of various tissues and are known to play important roles in several cellular functions.  $\beta$ 4Gal-T7 is an important enzyme for the species viability [38]; mutation in the human  $\beta$ 4Gal-T7 gene has been linked to Ehlers-Danlos syndrome [41,42]. Although recent crystal structure of the *Drosophila*  $\beta$ 4Gal-T7 shows that the overall structure of this enzyme is similar to  $\beta$ 4Gal-T1, with a similar catalytic pocket, its evolutionary relationship with  $\beta$ 4Gal-T1 is still not clear [43].

Though high protein sequence similarity between the vertebrate  $\beta$ 4Gal-T1 and invertebrate  $\beta$ 4GalNAc-T enzymes suggests that these proteins are related by evolution, the difference in their donor sugar specificity is mainly due to a single amino acid change. Furthermore, superposition of the three-dimensional structure of the catalytic domain of  $\beta$ 4Gal-T1 with the corresponding catalytic domain of ppGalNAc-T reveals a similarity in their protein sequence and gene structure, suggesting that these proteins are also evolutionarily related. In addition to protein sequence and structure similarity in their gene structure, exon/intron organization is also observed among the proteins that have been thought to have evolved from a common

primordial gene. This is in contrast to the observation of a common structural motif among the glycosyltransferases that have DxD metal binding motif and use UDP-sugar as donor substrates without any similarity in their gene structure. Thus, the structural and functional studies of these enzymes have provided deeper insight into the evolution of these glycosyltransferases.

## Acknowledgments

We thank Dr. Nathan Sharon from Weizmann Institute of Science, Israel, for encouraging us to write this review and for his valuable comments on the manuscript. The writing of this review was funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E. The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government. This research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

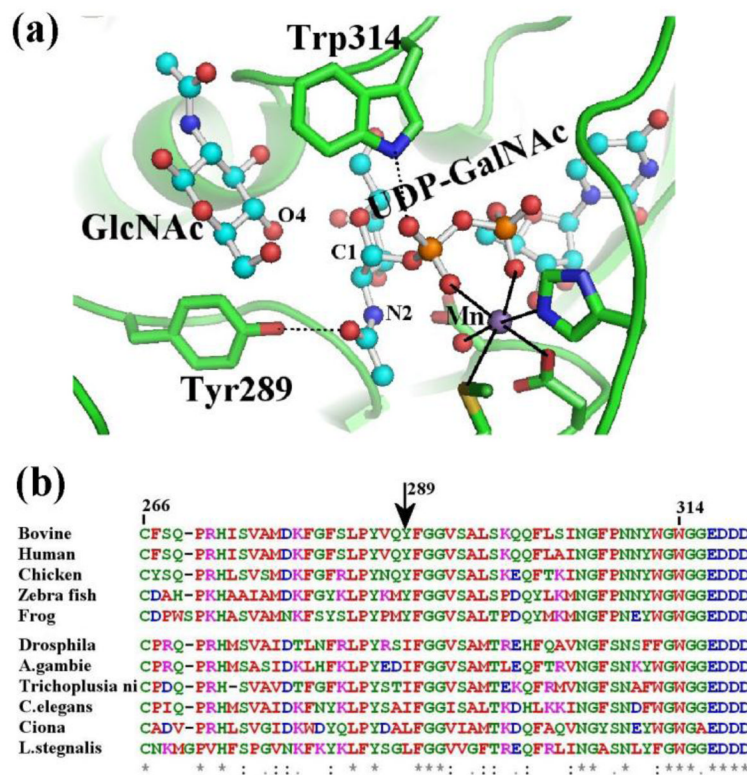
## References

1. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active EnZymes database (CAZy): An expert resource for glycomics. *Nucleic Acids Res* 2008;37:D233–238. [PubMed: 18838391]
- 2\*\*. Kaneko M, Nishihara S, Narimatsu H, Saitou N. The evolutionary history of glycosyltransferase genes. *Trends Glycosci Gycotechnol* 2001;13:147–155. Molecular evolutionary analysis on the glycosyltransferases that are involved in N- or O-glycan synthesis. The phylogenetic trees showed that these glycosyltransferases not only have evolved through gene duplications but also evolved more slowly than other genes. Interestingly they have predicted that the  $\beta$ 4Gal-T and  $\alpha$ ppGalNAc-T might have evolved from the same primordial gene.
3. Togayachi A, Sato T, Narimatsu H. Comprehensive enzymatic characterization of glycosyltransferases with a beta3GT or beta4GT motif. *Methods Enzymol* 2006;416:91–102. [PubMed: 17113861]
4. Hennet T. The galactosyltransferase family. *Cell Mol Life Sci* 2002;59:1081–1095. [PubMed: 12222957]
- 5\*. Varki, A.; Lowe, JB. *Essentials of Glycobiology*. 2. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 2009. Biological role of glycans; p. 75-88. A book on topics in glycobiology
6. Galili U, Macher BA, Buehler J, Shohet SB. Human natural anti- $\alpha$ -galactosyl IgG. II. The specific recognition of  $\alpha$ (1–3)-linked galactose residues. *J Exp Med* 1985;162:573–582. [PubMed: 2410529]
7. Larsen RD, Rivera-Marrero CA, Ernst LK, et al. Frameshift and nonsense mutations in a human genomic sequence homologous to a murine UDP-Gal: $\beta$ -D-Gal(1,4)-D-GlcNAc  $\alpha$ (1,3)-galactosyltransferase cDNA. *J Biol Chem* 1990;265:7055–7061.
8. Varki A. Multiple changes in sialic acid biology during human evolution. *Glycoconj J* 2009;26:231–245. [PubMed: 18777136]
- 9\*\*. Varki A. Uniquely human evolution of sialic acid genetics and biology. *Proc Natl Aca Sci U S A* 2010;107:8939–8946. This review discusses the evolution of the proteins that are involved in the binding of sialic acid and its implications in human evolution.
10. Turcot-Dubois AL, Le Moullac-Vaidye B, Despiau S, Roubinet F, Bovin N, Le Pendu J, Blancher A. Long-term evolution of the CAZY glycosyltransferase 6 (ABO) gene family from fishes to mammals—a birth-and-death evolution model. *Glycobiology* 2007;17:516–528. [PubMed: 17298992]
11. Ashwell G, Morell AG. *Adv. Enzymol. Relat. Areas. Mol Biol* 1974;41:99–128.
12. Crocker PR, Paulson JC, Varki A. Siglecs and their roles in the immune system. *Nat Rev Immunol* 2007;7:255–266. [PubMed: 17380156]
13. Liu FT, Rabinovich GA. Galectins: regulators of acute and chronic inflammation. *Ann N Y Acad Sci* 2010;1183:158–182. [PubMed: 20146714]
14. Vasta GR. Roles of galectins in infection. *Nat Rev Microbiol* 2009;7:424–438. [PubMed: 19444247]
15. Brew K, Vanaman TC, Hill RL. The role of alpha-lactalbumin and the A protein in lactose synthetase: a unique mechanism for the control of a biological reaction. *Proc Natl Acad Sci U S A* 1968;59:491–497. [PubMed: 5238979]

16. Do KY, Do SI, Cummings RD. Alpha-lactalbumin induces bovine milk beta 1,4-galactosyltransferase to utilize UDP-GalNAc. *J Biol Chem* 1995;270:18447–18451. [PubMed: 7629170]
- 17\*\*. Ramakrishnan B, Qasba PK. Structure-based design of beta-1,4-Galactosyl-transferase (beta 4Gal-T1) with equally efficient *N*-Acetylgalactosaminyltransferase activity. *J Biol Chem* 2002;277:20833–20840. Crystal structure and the enzyme kinetic studies showed that the in bovine beta 4Gal-T1 enzyme a single amino acid residue (Tyr289), found in the vicinity of the sugar donor-binding site that is not involved in any direct interactions with the donor sugar, determines the donor sugar specificity of the enzyme. The first successful structure based design of a glycosyltransferase to transfer unnatural substrates. [PubMed: 11916963]
18. Betenbaugh MJ, Tomiya N, Narang S, Hsu JT, Lee YC. Biosynthesis of human-type N-glycans in heterologous systems. *Curr Opin Struct Biol* 2004:14601–606.
19. Haines N, Irvine KD. Functional analysis of *Drosophila* beta 1,4-N-Acetylgalactosaminyltransferases. *Glycobiology* 2004;15:335–346. [PubMed: 15563714]
20. Kawar ZS, van Die I, Cummings RD. Molecular cloning and enzymatic characterization of a UDP-GalNAc:GlcNAc-R beta 1,4-N-Acetylgalactosaminyltransferase from *Caenorhabditis elegans*. *J Biol Chem* 2002;277:34924–34932. [PubMed: 12167666]
21. Vadaie N, Jarvis DL. Molecular cloning and functional characterization of a lepidopteran insect beta 4-N-Acetylgalactosaminyltransferase with broad substrate specificity, a functional role in glycoprotein biosynthesis, and a potential functional role in glycolipid biosynthesis. *J Biol Chem* 2004;279:33501–33518. [PubMed: 15173167]
22. Bakker H, Van Tetering A, Agterberg M, Smit AB, Van den Eijnden DH, Van Die I. Deletion of two exons from the *Lymnaea stagnalis* beta1—>4-N-acetylglucosaminyltransferase gene elevates the kinetic efficiency of the encoded enzyme for both UDP-sugar donor and acceptor substrates. *J Biol Chem* 1997;272:18580–18585. [PubMed: 9228024]
- 23\*\*. Ramakrishnan B, Qasba PK. Role of a single amino acid in the evolution of glycans of invertebrates and vertebrates. *J Mol Biol* 2007;365:570–576. This demonstrated that the donor-sugar specificity of *Drosophila* beta 4GalNAc-TA and mammalian beta 4Gal-T1 are related by a single amino acid change. Furthermore, based on the binding of alpha-lactalbumin, it has been proposed that during the evolution of vertebrates from invertebrates, this single amino acids change must have taken place to convert the invertebrate beta 4GalNAc-T enzyme to a beta 4Gal-T enzyme in vertebrates. [PubMed: 17084860]
- 24\*\*. Qasba PK, Kumar S. Molecular divergence of lysozymes and alpha-lactalbumin. *Crit Rev Biochem Mol Biol* 1997;32:255–306. Review on the evolutionary relationship between the lysozyme and alpha-lactalbumin and demonstrating that these proteins not only have similarity in their protein sequence but also in their gene structures. [PubMed: 9307874]
25. Brodbeck U, Denton WL, Tanahashi N, Ebner KE. The isolation and identification of the B protein of lactose synthetase as alpha-lactalbumin. *J Biol Chem* 1967;242:1391–1397. [PubMed: 6023212]
26. Shaper NL, Meurer JA, Joziassse DH, Chou TD, Smith EJ, Schnaar RL, Shaper JH. The chicken genome contains two functional nonallelic beta 1,4-Galactosyltransferase genes. Chromosomal assignment to syntenic regions tracks fate of the two gene lineages in the human genome. *J Biol Chem* 1997;272:31389–31399. [PubMed: 9395470]
27. Ten Hagen KG, Fritz TA, Tabak LA. All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology* 2003;13:1R–16R. [PubMed: 12634318]
28. Imberty A, Piller V, Piller F, Breton C. Fold recognition and molecular modeling of a lectin-like domain in UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Protein Eng* 1997;10:1353–1356. [PubMed: 9542995]
- 29\*\*. Kubota T, Shiba T, Sugioka S, Furukawa S, Sawaki H, Kato R, Wakatsuki S, Narimatsu H. Structural basis of carbohydrate transfer activity by human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase (pp-GalNAc-T10). *J Mol Biol* 2006;359:708–727. The x-ray crystal structure of appGalNAc-T10 in complex with UDP-GalNAc shows that the binding of UDP-GalNAc to appGalNAc-T10 enzyme is similar to the binding of UDP-GalNAc to beta 4Gal-T1. [PubMed: 16650853]
- 30\*. Fritz TA, Raman J, Tabak LA. Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-2. *J Biol Chem* 2006;281:8613–8619. The x-ray crystal structure and enzyme kinetics studies on human

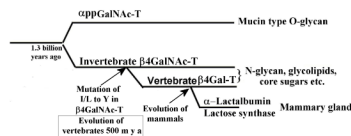
αppGalNAc-T2. This study showed that the catalytic domain of the enzyme alone exhibits catalytic activity. [PubMed: 16434399]

31. Ihara H, Ikeda Y, Toma S, Wang X, Suzuki T, Gu J, Miyoshi E, Tsukihara T, Honke K, Matsumoto A, et al. Crystal structure of mammalian alpha1,6-fucosyltransferase. *FUT8 Glycobiology* 2007;17:455–466.
32. Qasba PK, Ramakrishnan B. Letter to the Glyco-Forum: Catalytic domains of glycosyltransferases with 'add-on' domains. *Glycobiology* 2007;17:7G–9G.
33. Martinez-Fleites C, He Y, Davies GJ. Structural analyses of enzymes involved in the O-GlcNAc modification. *Biochim Biophys Acta* 2010;1800:122–133. [PubMed: 19647042]
34. Almeida R, Lavery SB, Mandel U, Kresse H, Schwientek T, Bennett EP, Clausen H. Cloning and expression of a proteoglycan UDP-galactose:beta-xylose beta1,4-galactosyltransferase I. A seventh member of the human beta4-galactosyltransferase gene family. *J Biol Chem* 1999;274:26165–26171. [PubMed: 10473568]
35. Stolz A, Haines N, Pich A, Irvine KD, Hokke CH, Deelder AM, Gerardy-Schahn R, Wührer M, Bakker H. Distinct contributions of beta 4GalNAcTA and beta 4GalNAcTB to *Drosophila* glycosphingolipid biosynthesis. *Glycoconj J* 2008;25:167–175. [PubMed: 17876704]
36. Vadaie N, Hulinsky RS, Jarvis JD. Identification and characterization of a *Drosophila melanogaster* ortholog of human beta1,4-galactosyltransferase VII. *Glycobiology* 2002;12:589–597. [PubMed: 12244071]
37. Nakamura Y, Haines N, Chen J, Okajima T, Furukawa K, Urano T, Stanley P, Irvine KD, Furukawa K. Identification of a *Drosophila* gene encoding xylosylprotein beta4-galactosyltransferase that is essential for the synthesis of glycosaminoglycans and for morphogenesis. *J Biol Chem* 2002;277:46280–46288. [PubMed: 12215432]
38. Takemae H, Ueda R, Okubo R, Nakato H, Izumi S, Saigo K, Nishihara S. Proteoglycan UDP-galactose:beta-xylose beta 1,4-galactosyltransferase I is essential for viability in *Drosophila melanogaster*. *J Biol Chem* 2003;278:15571–15578. [PubMed: 12590131]
39. Herman T, Horvitz HR. Three proteins involved in *Caenorhabditis elegans* vulval invagination are similar to components of a glycosylation pathway. *Proc Natl Acad Sci USA* 1999;96:974–979. [PubMed: 9927678]
40. Okajima T, Yoshida K, Kondo T, Furukawa K. Human homolog of *Caenorhabditis elegans* sqv-3 gene is galactosyltransferase I involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *J Biol Chem* 1999;274:22915–22918. [PubMed: 10438455]
41. Okajima T, Fukumoto S, Furukawa K, Urano T. Molecular basis for the progeroid variant of Ehlers-Danlos syndrome. Identification and characterization of two mutations in galactosyltransferase I gene. *J Biol Chem* 1999;274:28841–28844. [PubMed: 10506123]
42. Seidler DG, Faiyaz-Ul-Haque M, Hansen U, Yip GW, Zaidi SH, Teebi AS, Kiesel L, Götte M. Defective glycosylation of decorin and biglycan, altered collagen structure, and abnormal phenotype of the skin fibroblasts of an Ehlers-Danlos syndrome patient carrying the novel Arg270Cys substitution in galactosyltransferase I (beta4GalT-7). *J Mol Med* 2006;84:583–594. [PubMed: 16583246]
43. Ramakrishnan B, Qasba PK. Crystal structure of the catalytic domain of *Drosophila* β1,4-galactosyltransferase-7. *J Biol Chem* 2010;285:15619–15626. [PubMed: 20236943]



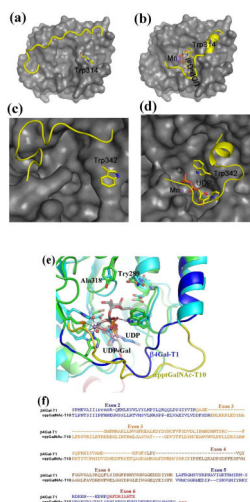
**Figure 1.** (a) The crystal structure of the bovine  $\beta 4\text{Gal-T1} \cdot \text{Mn}^{2+} \cdot \text{UDP-GalNAc}$  complex (pdb 1OQM) with a GlcNAc molecule modeled in the acceptor binding site. The UDP-GalNAc and GlcNAc molecules are shown in a ball and stick diagram. The side-chain hydroxyl group of Tyr289 residue forms a hydrogen bond (shown as black dotted line) with the carbonyl oxygen atom of the N-acetyl group of the GalNAc moiety. (b) Protein sequence comparison of the bovine  $\beta 4\text{Gal-T1}$  shown near the vicinity of the Y289 residues (black arrow), along with its homolog proteins from various species. Only in the vertebrate homolog proteins is the Y289 residue conversed as a Tyr residue, whereas in the invertebrates it is either Leu or Ile.





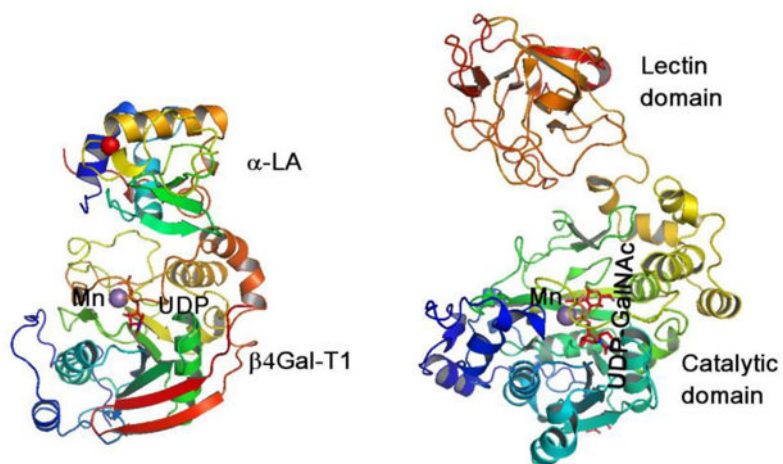
**Figure 2.**

Since the donor sugar specificity of the invertebrate  $\beta$ 4GalNAc-T and vertebrate  $\beta$ 4Gal-T are related by a single amino acid change, it is possible that when the vertebrate emerged from the invertebrate 500 million years ago, this single amino acid change might have happened, thus converting the invertebrate  $\beta$ 4GalNAc-T enzyme to a  $\beta$ 4Gal-T enzyme in the vertebrates. Furthermore, since the catalytic domains of  $\beta$ 4Gal-T1 and  $\alpha$ pptGalNAc-T enzymes exhibit protein structure, sequence, and gene structure similarity, the vertebrate  $\beta$ 4GalNAc-T and the  $\alpha$ pptGalNAc-T might have evolved from a common primordial gene by gene duplication and, further, during evolution the  $\alpha$ pptGalNAc-T might have picked up the lectin domain to define its acceptor specificity. Since  $\alpha$ -lactalbumin binds and inhibits but does not modulate the invertebrate  $\beta$ 4GalNAc-T enzyme, its binding site existed before the evolution of vertebrates. Only during the evolution of mammals did  $\alpha$ -lactalbumin evolve to bind to this site in the vertebrate  $\beta$ 4Gal-T1 and modulate its acceptor sugar specificity to synthesize lactose during lactation of mammals.



**Figure 3.**

The crystal structure of the bovine  $\beta$ 4Gal-T1 in the open (a) and closed (b) conformation. Upon the binding of manganese and UDP-Gal, the enzyme undergoes conformational changes involving a short and a long flexible loop. In the short loop, the side chain of the Trp314 residue moves from outside to inside the catalytic pocket to bind to the bound UDP-Gal molecule, while the long loop moves over the bound UDP-GalNAc molecule to cover it. The crystal structure of the human  $\alpha$ GalNAc-T2 is also found in an open (c) and a closed (d) conformation. Similar to  $\beta$ 4Gal-T1, upon the binding of manganese and UDP-GalNAc molecule, this enzyme also undergoes conformational changes involving two flexible loops. Also, in the short loop, the Trp342 residue moves from outside to inside the catalytic pocket to bind to the bound UDP-GalNAc molecule, while the long flexible loop covers the bound UDP-GalNAc molecule.. (e) The superposition of the catalytic domains of bovine  $\beta$ 4Gal-T1•Mn<sup>2+</sup>•UDP-Gal complex with the  $\alpha$ GalNAc-T10•Mn<sup>2+</sup>•UDP-GalNAc complex and (f) the corresponding protein sequence comparison. The Trp residue in the short flexible loops of these structures show good agreement, and the long flexible loop of  $\beta$ 4Gal-T1 and  $\alpha$ GalNAc-T10 are shown in blue and yellow, respectively. The Tyr289 residue in the bovine  $\beta$ 4Gal-T that determines the donor sugar specificity is naturally present as Ala318 residue in  $\alpha$ GalNAc-T10 in order to accommodate the *N*-acetyl moiety of the donor sugar GalNAc, and it is conserved in all  $\alpha$ GalNAc-T enzymes. In the protein-protein sequence comparison (f), the amino acids are colored, based on their exons. Since the acceptor substrate for  $\beta$ 4GalNAc-T1 is a sugar residue, while it is a linear peptide with a Thr/Ser amino acid for the  $\alpha$ GalNAc-T10, these enzymes are expected to have different acceptor binding sites. The additional amino acids in the exons 3 and 4 of  $\alpha$ GalNAc-T10 are found as a part of the acceptor substrate binding site in the  $\alpha$ GalNAc-T10 crystal structure.



**Figure 4.**

In the lactose synthase complex (left),  $\alpha$ -LA modulates the sugar acceptor specificity of the  $\beta$ 4Gal-T1 enzyme from GlcNAc to Glc, while in the  $\alpha$ pptGalNAc-Ts (right), although the catalytic domain alone exhibits catalytic activity, the lectin domain defines acceptor substrate specificity to the enzyme. Thus, in these proteins  $\alpha$ -LA and the lectin domain are considered as “add-on” domains. There are many other transferases found where such a domain brings about substrate specificity to the enzyme.