NIH Public Access
Author Manuscript
J Mol Biol. Author manuscript; available in PMC 2008 January 19.

Published in final edited form as:
J Mol Biol. 2007 January 19; 365(3): 570–576.

Role of a single amino acid in the evolution of glycans of invertebrates and vertebrates

Boopathy Ramakrishnan[§,¶] and **Pradman K. Qasba**[§,*]

§ *Structural Glycobiology Section§, CCR Nanobiology Program, Center for Cancer Research, National Cancer Institute at Frederick, Frederick, MD*

¶ *Basic Research Program¶, SAIC-Frederick, Inc., Center for Cancer Research, National Cancer Institute at Frederick, Frederick, MD*

Abstract

Structures of glycoconjugate *N*-glycans and glycolipids of invertebrates show significant differences from those of vertebrates. These differences are due largely to the vertebrate β1,4-galactosyltransferase-1 (β4Gal-T1), which is found as a β1,4-*N*-acetylgalactosaminyltransferase (β4GalNAc-T1) in invertebrates. Mutation of Tyr285 to Ile or Leu in human β4Gal-T1 converts the enzyme into an equally efficient β4GalNAc-T1. A comparison of all the human β4Gal-T1 ortholog enzymes shows that this Tyr285 residue in human β4Gal-T1 is conserved either as Tyr or Phe in all vertebrate enzymes, while in all invertebrate enzymes it is conserved as an Ile or Leu. We find that mutation of the corresponding Ile residue to Tyr in *Drosophila* β4GalNAc-T1 converts the enzyme to a β4Gal-T1 by reducing its *N*-acetylgalactosaminyltransferase activity by nearly 1000-fold, while enhancing its galactosyltransferase activity by 80-fold. Furthermore we find that, similar to the vertebrate/mammalian β4Gal-T1 enzymes, the wild-type *Drosophila* β4GalNAc-T1 enzyme binds to a mammary gland-specific protein, α-lactalbumin (α-LA). Thus, it would seem that, during the evolution of vertebrates from invertebrates over 500 million years ago, β4Gal-T1 appeared as a result of the single amino acid substitution of Tyr or Phe for Leu or Ile in the invertebrate β4GalNAc-T1. Subsequently, the preexisting α-LA-binding site was utilized during mammalian evolution to synthesize lactose in the mammary gland during lactation.

Keywords

Vertbrate β4Gal-T1; invertbrate β4GalNAc-T1; single mutation; evolutionary relationship; donor substrate specificity

---

Glycoconjugates in a cell play important roles in several cellular processes.[1] The invertebrate glycoconjugates differ from those in vertebrates in that they lack β1-4-linked galactose residues in their N-glycans and glycolipids.[2] In the vertebrates, a family of β1,4-galactosyltransferases (β4Gal-T1 to T7) is responsible for the addition of galactose in β1-4-linkage, from the donor UDP-galactose (UDP-Gal) to various glycans.[3, 4] For example, β4Gal-T6 is involved in the synthesis of lactosylceramide,[3, 4] the building block for the synthesis of the GM, GD, and Gb series of glycolipids; β4Gal-T7 adds galactose to the O-linked xylose of proteins for the

---

*Corresponding author. Structural Glycobiology Section, CCRNP, CCR, NCI-Frederick, Building 469, Room 221, Frederick, Maryland 21702; e-mail: 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.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

synthesis of the linker region of glycosaminoglycans such as heparin, heparan sulfate, chondroitin, and dermatan sulfate to make proteoglycans.[4, 5] These glycosaminoglycans are essential for the normal development of the organism. Most importantly, β4Gal-T1 is involved not only in the synthesis of the *N*-glycan of glycoproteins, but also in Notch signaling.[5] This enzyme, in complex with α-lactalbumin (α-LA),[6–8] also plays an important role in the synthesis of lactose in the mammary gland during lactation.

The β4Gal-T1 orthologs present in all the invertebrates sequenced to date transfer *N*-acetylgalactosamine (GalNAc) from UDP-GalNAc instead of galactose from UDP-Gal *in vitro*, and hence are named *N*-acetylgalactosaminyltransferases (GalNAc-T).[9–11] This finding accounts for the low amount of β1-4-linked galactose moieties in their glycoconjugates. Among the invertebrates, the *Drosophila* genome sequence is known, and three orthologs of β1,4-galactosyltransferase that correspond to β4Gal-T1, β4Gal-T6, and β4Gal-T7 of vertebrates[9, 12, 13] have been identified. The β4Gal-T1 ortholog in *Drosophila* has been shown to be an *N*-acetlylgalactosaminyltransferase, but the β4Gal-T6 ortholog has not yet been characterized. The β4Gal-T7 ortholog is a galactosyltransferase, which, like its vertebrate counterpart, transfers galactose to the O-linked xylose of proteins involved in proteoglycan synthesis.[12, 13] Although the null mutation of β4GalNAc-T1 has been shown to disrupt the normal neuromuscular physiology and development of *Drosophila,*[9] it does not affect the survival of the species. On the other hand, the β4Gal-T7 ortholog is essential for survival and maturation to adulthood.[13] By contrast, the β4Gal-T1 gene knockout in the mouse results in severe disability and illness, indicating the importance of β4Gal-T1 in vertebrate wellness.[14]

We have previously established that the sugar-donor specificity of bovine or human β4Gal-T1 towards UDP-Gal is determined by a single amino acid residue, Tyr (or Phe), in the catalytic pocket at position 289 or 286, respectively (Figure 1).[15–17] When the residue is mutated to Ile or Leu, the specificity is broadened, and the mutant enzyme transfers *N*-acetylgalactosamine from UDP-GalNAc as efficiently as it transfers galactose from UDP-Gal. In human β4Gal-T homologs and in all known β4Gal-T orthologs from such vertebrates as mammals, birds, amphibians, and fish, this residue has been conserved as either Tyr or Phe (Figure 2(a) (I)). In contrast, the invertebrate worms, fly, snail, mosquito, and moth have either Ile or Leu in the corresponding position in their β4Gal-T orthologs (Figure 2(a) (II)) and show *N*-acetylgalactosaminyltransferase (β4GalNAc-T) activity.[10, 11] Interestingly, the β4Gal-T7 ortholog in *Drosophila* has a Phe residue and exhibits the galactosyltransferase activity involved in the proteoglycan synthesis that is essential for its survival.[12, 13]

Although *Drosophila* β4GalNAc-T1 exhibits only 56% sequence similarity with the human β4Gal-T1 in its catalytic domain (Figure 2(b)), it exhibits the same sugar-acceptor specificity for *N*-acetylglucosamine. We show that the recombinant *Drosophila* β4GalNAc-T1 expressed in *Escherichia coli* and folded *in vitro* from the inclusion bodies,[8, 15, 18] exhibits high GalNAc-T activity but low Gal-T activity (Figure 3(a)). By contrast, the bovine and human β4Gal-T1 exhibit high Gal-T activity and low GalNAc-T activity[15]. When Tyr is substituted for the active-site residue Ile289 in *Drosophila* β4GalNAc-T1, its GalNAc-T activity is reduced dramatically (Figure 3(b)) while its Gal-T activity is greatly enhanced (Figure 3(c)), making it similar to human β4Gal-T1.[15] Similar results are obtained if Phe is substituted for the Ile[15] in mammalian (bovine) enzyme. Clearly, the sugar-donor specificity of both invertebrate and vertebrate enzymes is determined by a single amino acid at the catalytic site. The substitution of this amino acid is associated with a single-point mutation of the codon AUU for Ile to UUU for Phe.

Since vertebrate and invertebrate species diverged 500 million years ago, the vertebrate β4Gal-T1 gene must have emerged from the primordial invertebrate β4GalNAc-T1 as a result of a single amino acid substitution (Figure 4). Although one cannot rule out that there were other

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

changes in the structure of these enzymes that played equally important role which might account for the differences in the primary structure, α-LA interactions and catalytic efficiency (k_{cat}) of these enzymes. The locations of introns in the gene sequence of the catalytic domain of human β4Gal-T1 and *Drosophila* β4GalNAc-T1 at similar positions (Figure 2(b)) support this hypothesis. However, exon 5 in *Drosophila* is split into exon 5 and exon 6 in the human β4Gal-T1. The introns in the *Drosophila* gene are generally around 60 bases, whereas in the human gene they are in the thousands of bases. In both genes, however, the exon boundaries are in the vicinity of the functional residues of the protein molecule (Figure 2(b)). The similar gene architecture, thus, confirms that these two proteins are evolutionarily related. On the other hand the β4Gal-T7 or its orthologs are found in all the invertebrates and vertebrates and show no gene structure similarity but exhibit low protein sequence similarity with their corresponding β4Gal-T1/ β4GalNAc-T1 ortholog proteins. Also interestingly the gene structure of human β4Gal-T7 differs very much from its *Drosophila* ortholog, suggesting that these two proteins might have emerged from a common primordial gene.

The appearance of a mammary gland-specific protein, α-LA, during evolution has often been associated with the emergence of mammals.[19, 20] α-LA has very high sequence,[19, 20] structure[21] and gene architecture[22] similarities with lysozyme. It has been postulated that nearly 300 million years ago, α-LA evolved from a primordial lysozyme gene by gene duplication and divergence (Figure 4).[19, 20] Lysozyme, an enzyme that catalyzes the cleavage of the sugar glycosidic bond, is ubiquitously expressed in most living cells. By contrast, α-LA, a Ca^{2+}-binding protein, is expressed only in the mammary gland during lactation. It modulates the sugar-acceptor specificity of the β4Gal-T1 in such a way that instead of transferring galactose to *N*-acetylglucosamine, β4Gal-T1 transfers galactose to glucose,[6, 7] to form lactose, the milk sugar.

A protein sequence comparison shows that α-LA-binding site residues are present not only in all the vertebrate β4Gal-T1 proteins, but also in the invertebrate β4GalNAc-T1 proteins (Figure 2(b)). The recombinant *Drosophila* β4GalNAc-T1 binds very well to an α-LA-affinity column (Figure 5(a)) and also inhibits the transfer to *N*-acetylglucosamine *in vitro* (Figure 5(b)), confirming the presence of an α-LA- binding site; however, it transfers neither *N*-acetylgalactosamine nor galactose to glucose (Figure 5(b)). Similarly, *in vitro*, the GalNAc-T activity of β4GalNAc-T from *Caenorhabditis elegans* has been shown to be inhibited by α-LA.[10] Although the presence of αLA binding site in the invertebrate β4GalNAc-T1 is observed *in vitro*, it is not used for α-LA binding *in vivo*, since α-LA is found only in mammals. Thus, it appears that the α-LA-binding site was indeed present in the primordial β4GalNAc-T , 500 million years ago, even before the divergence of vertebrates and invertebrates (Figure 4). After the emergence of vertebrates and the appearance of β1,4-galactosyltransferase, and later of α-LA in mammals, this binding site was utilized to make lactose in the lactating mammary gland.

Since the invertebrate β4GalNAc-T1 and β4Gal-T7 ortholog diverged from the common ancestral gene, and the fact that β4Gal-T7 ortholog does not bind to α-LA (unpublished result), while β4GalNAc-T1 binds to α-LA, is not coincidence. Later the same enzyme, when evolved into β4Gal-T1 in vertebrates, this site is used in mammals by α-LA to form lactose synthase complex for the synthesis of lactose. During the evolution of the mammals, α-LA evolved to bind and modulate β4Gal-T1. This is further evidenced from the marsupial and monotreme mammals which evolved in parallel to eutherian (placental mammals), where α-LA works better as a lactose synthase with the β4Gal-T1 from the same species, and they do not make an efficient lactose synthase with β4Gal-T1 from eutherian.[23,24] Whereas α-LA from any eutherian forms an efficient lactose synthase complex with any β4Gal-T1, not only from eutherians but also from the other vertebrate species like chicken,[25] suggesting that the well evovled α-LA-binding site existed on the β4Gal-T1 orthologs of vertebrates before the divergence of birds and mammals. Therefore, it is expected that even the primordial β4Gal-T1

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

ortholog enzymes in invertebrates, such as *Drosophila* β4GalNAc-T1, to posses at least a partial α-LA binding site.

The β1,4-GalNAc-containing glycoconjugates in invertebrates are not abundant[2, 11] and their function in development is so far not well studied. These conjugates are synthesized by GalNAc-Ts, which are β4Gal-T orthologs containing Leu/Ile in their catalytic pockets. On the other hand, the glycosaminoglycans of proteoglycans have a β1,4-linked Gal moiety in the linker region that is synthesized by the β4Gal-T7 ortholog with a Phe residue in the catalytic pocket. These molecules are needed for the survival and growth of invertebrates,[12, 13], as well as morphological development, such as vulval morphogenesis in *Caenorhabditis elegans.*[26, 27]

In contrast to invertebrates, the β1,4-linked Gal-containing glycoconjugates are abundant in vertebrates and participate in a variety of cellular functions. Although the evolution of β1,4-galactosyltransferase appears to be coincidental with that of vertebrates, the evolution of this enzyme played an important role in vertebrate development. Like the recruitment of α-LA during mammalian evolution, several sialyltransferases had to be recruited to complete the *N*-glycan and other glycoconjugate syntheses. Similarly, different lectins had to be recruited to recognize the new glycoconjugates. Thus, the appearance of β1,4-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.

Acknowledgements

We thank Professor Soma Kumar of Georgetown University, and Drs. Elizabeth Boeggeman and Maria Manzoni of the Structural Glycobiology Section, CCR, NCI, for critically reading the manuscript and for helpful discussions. This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. 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 project has been funded in part with Federal funds from the NCI, NIH, under contract No. N01-C0-12400.

References

1. Varki A. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 1993;3:97–130. [PubMed: 8490246]
2. Betenbaugh MJ, Tomiya N, Narang S, Hsu JT, Lee YC. Biosynthesis of human-type N-glycans in heterologous systems. Curr Opin Struct Biol 2004;14:601–606. [PubMed: 15465322]
3. Amado M, Almeida R, Schwiente T, Clausen H. Identification and characterization of large galactosyltransferase gene families: galactosyltransferases for all functions. Biochim Biophys Acta 1999;1473:35–53. [PubMed: 10580128]
4. Hennet T. The galactosyltransferase family. Cell Mol Life Sci 2002;59:1081–1095. [PubMed: 12222957]
5. Chen J, Moloney DJ, Stanley P. Fringe modulation of Jagged1-induced Notch signaling requires the action of β4galactosyltransferase-1. Proc Natl Acad Sci USA 2001;98:13716–13721. [PubMed: 11707585]
6. Brodbeck U, Denton WL, Tanahashi N, Ebner KE. The isolation and identification of the B protein of lactose synthetase as α-lactalbumin. J Biol Chem 1967;242:1391–1397. [PubMed: 6023212]
7. Brew K, Vanaman TC, Hill RL. The role of α-lactalbumin and the A protein in lactose synthetase. A unique mechanism for the control of a biological reaction. Proc Natl Acad Sci USA 1968;59:491–497. [PubMed: 5238979]
8. Ramakrishnan B, Qasba PK. Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the β1,4-galactosyltransferase-I. J Mol Biol 2001;310:205–218. [PubMed: 11419947]
9. Nicola H, Irvine KD. Functional analysis of *Drosophila* β1,4-N-Acetlygalactosaminyltransferases. Glycobiology 2005;15:335–346. [PubMed: 15563714]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

10. Kawar ZS, van Die I, Cummings RD. Molecular cloning and enzymatic characterization of a UDP-GalNAc:GlcNAc-R β1,4-*N*-Acetylgalactosaminyltransferase from *Caenorhabditis elegans.* J Biol Chem 2002;277:34924–34932. [PubMed: 12167666]
11. Vadaie N, Jarvis DL. Molecular cloning and functional characterization of a lepidopteran insect β4-*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]
12. Vadaie N, Hulinsky RS, Jarvis DL. Identification and characterization of a *Drosophila melanogaster* ortholog of human β1,4-galactosyltransferase VII. Glycobiology 2002;12:589–597. [PubMed: 12244071]
13. Takemae H, Ueda R, Okubo R, Nakato H, Izumi S, Saigo K, Nishihara S. Proteoglycan UDP-galactose:beta-xylose β1,4-galactosyltransferase I is essential for viability in *Drosophila melanogaster*. J Biol Chem 2003;278:15571–15578. [PubMed: 12590131]
14. Asano M, Furukawa K, Kido M, Matsumoto S, Umesaki Y, Kochibe N, Iwakura Y. Growth retardation and early death of β1,4-galactosyltransferase knockout mice with augmented proliferation and abnormal differentiation of epithelial cells. EMBO J 1997;16:1850–1857. [PubMed: 9155011]
15. Ramakrishnan B, Qasba PK. Structure-based design of beta-1,4-Galactosyl-transferase-I (beta 4Gal-T1) with equally efficient *N*-Acetylgalactosaminyltransferase Activity. J Biol Chem 2002;277:20833–20840. [PubMed: 11916963]
16. Ramakrishnan B, Boeggeman E, Ramasamy V, Qasba PK. Structure and catalytic cycle of beta-1,4-galactosyltransferase. Curr Opin Struct Biol 2004;14:593–600. [PubMed: 15465321]
17. Qasba PK, Ramakrishnan B, Boeggeman E. Substrate induced conformational changes in glycosyltransferases. Trends Biochem Sci 2005;30:53–62. [PubMed: 15653326]
18. Boeggeman E, Ramakrishnan B, Qasba PK. The N-terminal stem region of bovine and human β1,4-galactosyltransferase-I increases the in vitro folding efficiency of their catalytic domain from inclusion bodies. Protein Expr and Purif 2003;30:219–229.
19. Qasba PK, Kumar S. Molecular divergence of lysozymes and α-lactalbumin. Crit Rev Biochem Mol Biol 1997;32:255–306. [PubMed: 9307874]
20. Grobler JA, Rao KR, Pervaiz S, Brew K. Sequences of two highly divergent canine type C lysozymes: implications for the evolutionary origins of the lysozyme/α-lactalbumin superfamily. Arch Biochem Biophys 1994;313:360–366. [PubMed: 8080284]
21. Browne WJ, North ACT, Phillips DC, Brew K, Vanaman TC, Hill RL. A possible three-dimensional structure of bovine α-lactalbumin based on hen's egg-white lysozyme. J Mol Biol 1969;42:65–89. [PubMed: 5817651]
22. Qasba PK, Safaya SK. Similarities of the nucleotide sequences of rat α-lactalbumin and chicken lysozyme genes. Nature 1984;308:377–380. [PubMed: 6709045]
23. Shaw DC, Messer M, Scrivener AM, Nicholas KR, Griffiths M. Isolation, partial characterisation, and amino acid sequence of alpha-lactalbumin from platypus (Ornithorhynchus anatinus) milk. Biochim Biophys Acta 1993;1161:177–86. [PubMed: 8431467]
24. Piotte CP, Marshall CJ, Hubbard MJ, Collet C, Grigor MR. Lysozyme and alpha-lactalbumin from the milk of a marsupial, the common brush-tailed possum (Trichosurus vulpecula). Biochim Biophys Acta 1997;1336:235–42. [PubMed: 9305795]
25. Shaper NL, Meurer JA, Joziasse DH, Chou TD, Smith EJ, Schnaar RL, Shaper JH. The chicken genome contains two functional nonallelic β1,4-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]
26. 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]
27. 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]

28. Khidekel N, Arndt S, Lamarre-Vincent N, Lippert A, Poulin-Kerstien KG, Ramakrishnan B, Qasba PK, Hsieh-Wilson LC. A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc modifications. J Am Chem Soc 2003;125:16162–16163. [PubMed: 14692737]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

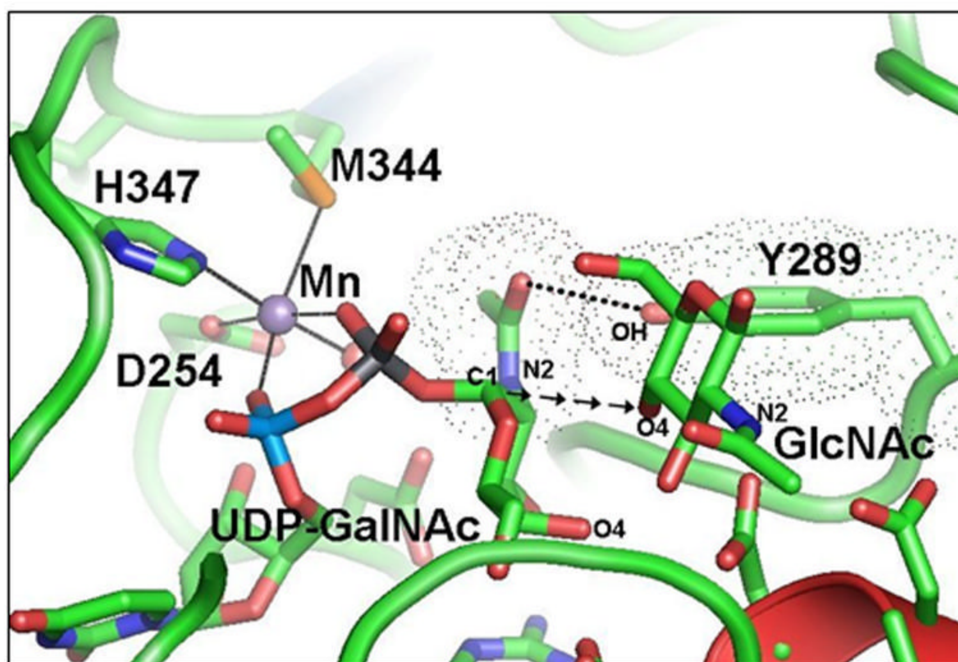


Figure 1.
The catalytic pocket of β4Gal-T1 in the presence of bound Mn^{2+}-UDP-GalNAc from the β4Gal-T1-Mn^{2+}-UDP-GalNAc-α-LA crystal structure (PDB entry 1OQM) and the GlcNAc molecule is modeled based on its binding in the crystal structure of the β4Gal-T1-α-LA-GlcNAc complex (PDB entry 1NQI). The wild-type β4Gal-T1 efficiently transfers Gal from UDP-Gal to GlcNAc, forming a glycosidic bond between the C1 of Gal to the O4 atom of GlcNAc (shown here as the arrows from the C1 atom of GalNAc moiety of UDP-GalNAc rather than from UDP-Gal) in a β-configuration, thus synthesizing the disaccharide moiety LacNAc (Galβ1-4GlcNAc). However, it also transfers GalNAc from UDP-GalNAc to GlcNAc, but this transfer is nearly 1000-fold less efficient than the transfer of Gal moiety from the UDP-Gal. The previous crystal structure studies show that the UDP-GalNAc binding to β4Gal-T1 is similar to the UDP-Gal binding: however, the *N*-acetyl moiety causes a severe steric hindrance with the side chain hydroxyl group of Try289, as shown by their van der Walls surface diagram (shown as dotted spheres), in which the spheres nearly merge together. In contrast to the N-acetyl moiety of the acceptor GlcNAc, the N-acetyl moiety of GalNAc of the donor UDP-GalNAc is in the same plane as the hexose moiety, a high-energy conformation. The hydroxyl group of the residue Tyr289 makes a hydrogen bond with the oxygen of the *N*-acetyl moiety of GalNAc (dotted line), applying a molecular brake during the transfer reaction. Mutation of the Tyr289 to Leu or Ile releases the brake and creates a space in the catalytic cavity, resulting in an efficient transfer of the GalNAc residue from UDP-GalNAc to the acceptor GlcNAc, and making the enzyme equally efficient as β1,4GalNAc-transferase as β1,4Gal-trasferase.[15] In addition, the mutant enzyme also transfers 2’-keto-galactose from its UDP derivative to the acceptor GlcNAc.[28]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

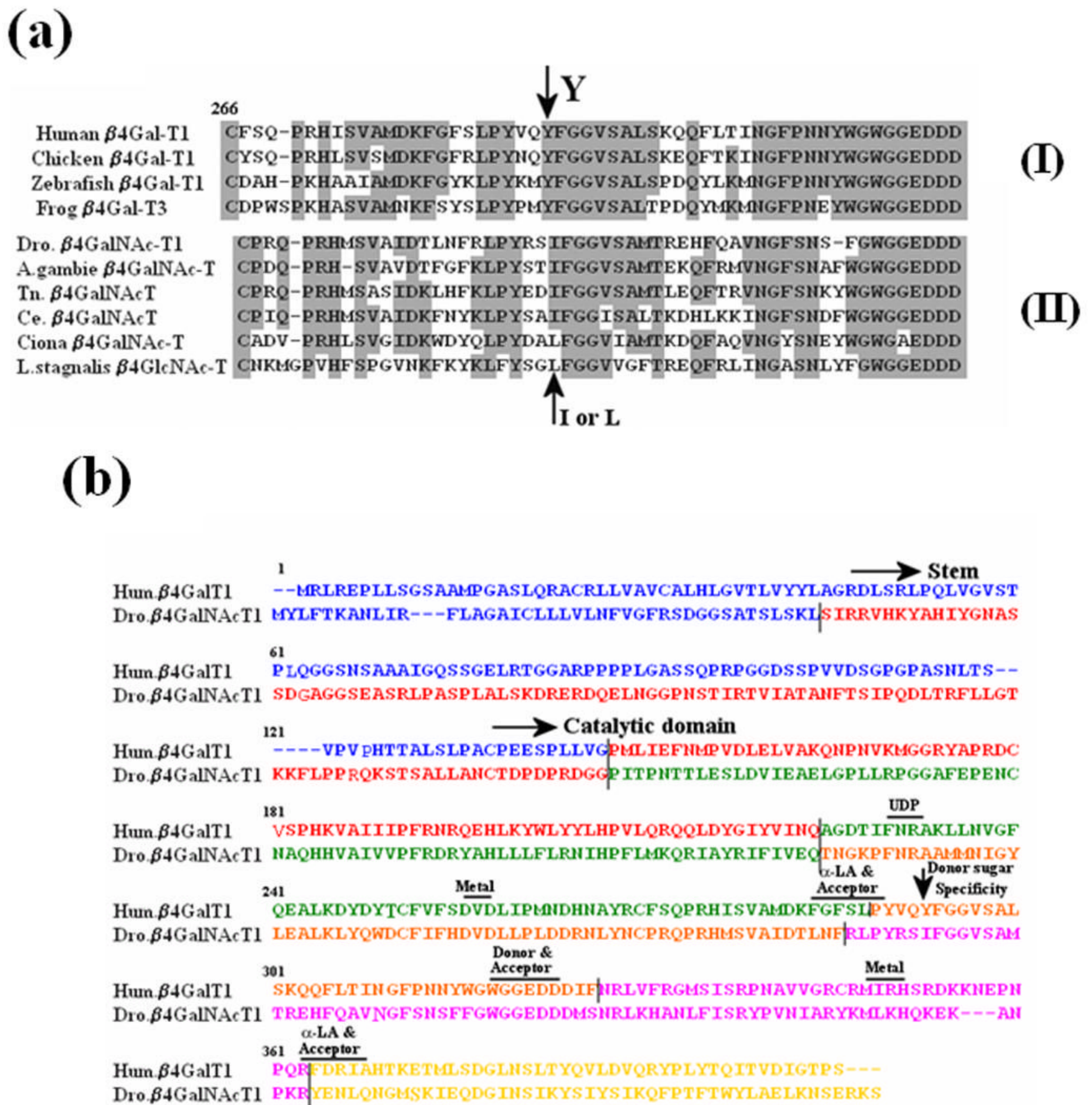


Figure 2.
The sequence comparison of a region of vertebrate and invertebrate ortholog genes of β1,4Gal-T. (a) The sequence comparison of the region containing Tyr286 in human β1,4Gal-T1 with the corresponding region in the ortholog proteins of vertebrates (I) and invertebrates (II). In some β1,4Gal-T homologs in vertebrates, Phe is substituted for the residue Tyr at position 286. (b) The sequence comparison of the functional regions of human-β1,4Gal-T1 with *Drosophila CG8536* (Dro.) protein. The two proteins show about 49% overall similarity and about 56% similarity in the catalytic domain. Ile at position 289 of Dro. protein (arrow), which corresponds

NIH-PA Author Manuscript

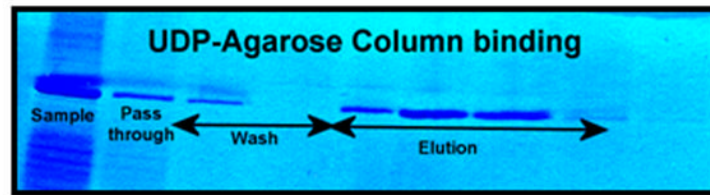
NIH-PA Author Manuscript

NIH-PA Author Manuscript

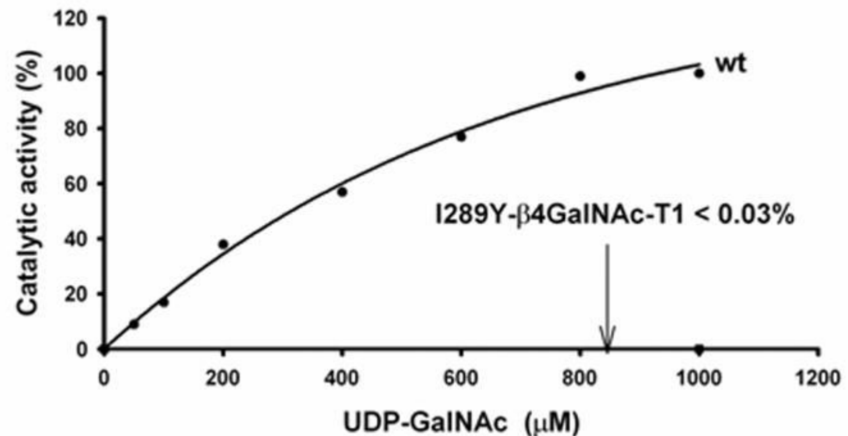
to residue Tyr at position 289 or 286 in bovine or human β1,4-Gal-T1, respectively, that imparts the β4GalNAc-T1 activity in the Dro. protein. The residues involved in the binding of metal ion, acceptor, sugar donor, and α-LA,[16, 8] are conserved in the two proteins. The sequence regions with different colors represent the regions coded by corresponding exons, and the line represents the intron-exon junctions.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

(a)



(b)



(c)

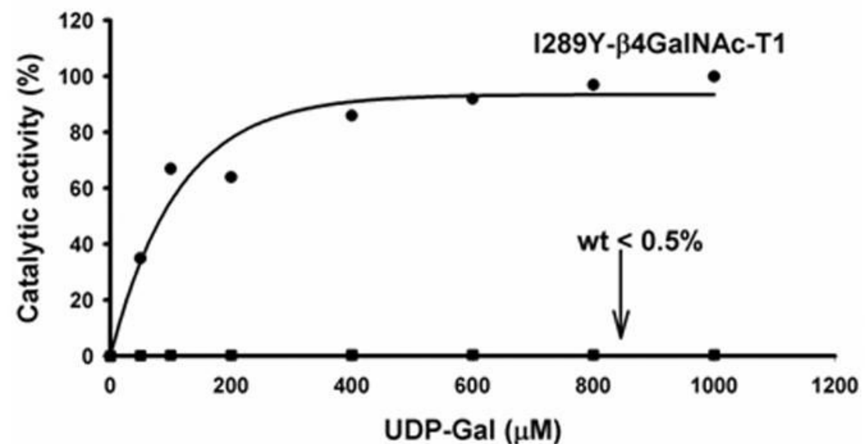


Figure 3.

The GalNAc-T and Gal-T activities of the *in vitro* folded recombinant *Drosophila* CG8536 protein and of the I289Y mutant protein. The cDNA, coding the soluble domain of the *Drosophila CG8536* protein, residues 29 to 403, was inserted into the pET23a vector and expressed in *E. coli*. The recombinant protein accumulates as inclusion bodies, which is purified and folded *in vitro* as described previously;[8, 18] the folded protein was further purified on UDP-agarose columns. The SDS gel analysis of bound, washed, and eluted fractions from the UDP-agarose column is shown in (a). The recombinant protein (wt) transferred mainly

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

GalNAc from UDP-GalNAc (GalNAc-T activity) (b), whereas the transfer of Gal from UDP-Gal (Gal-T activity) was less than 0.5% that of GalNAc (c). The mutation of the residue Ile to Tyr (I289Y) at position 289 produced an enzyme that exhibited nearly 80-fold enhanced Gal-T activity (c), while the GalNAc activity was reduced by more than 99% and comprised less than 0.03% of Gal-T activity (b). Thus, mutation of Ile289 to Tyr changes the sugar-donor specificity of β4GalNAc-T1 toward UDP-Gal. The 100% GalNAc-T or Gal-T activities correspond to a specific activity of 32 pmol/min/μg or 12.8 pmol/min/μg of protein, respectively. The average estimated error in these GalNAc-T and Gal-T activities based on the assays without the acceptor substrate are 1 pmol/min/μg and 0.5 pmol/min/μg, respectively.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

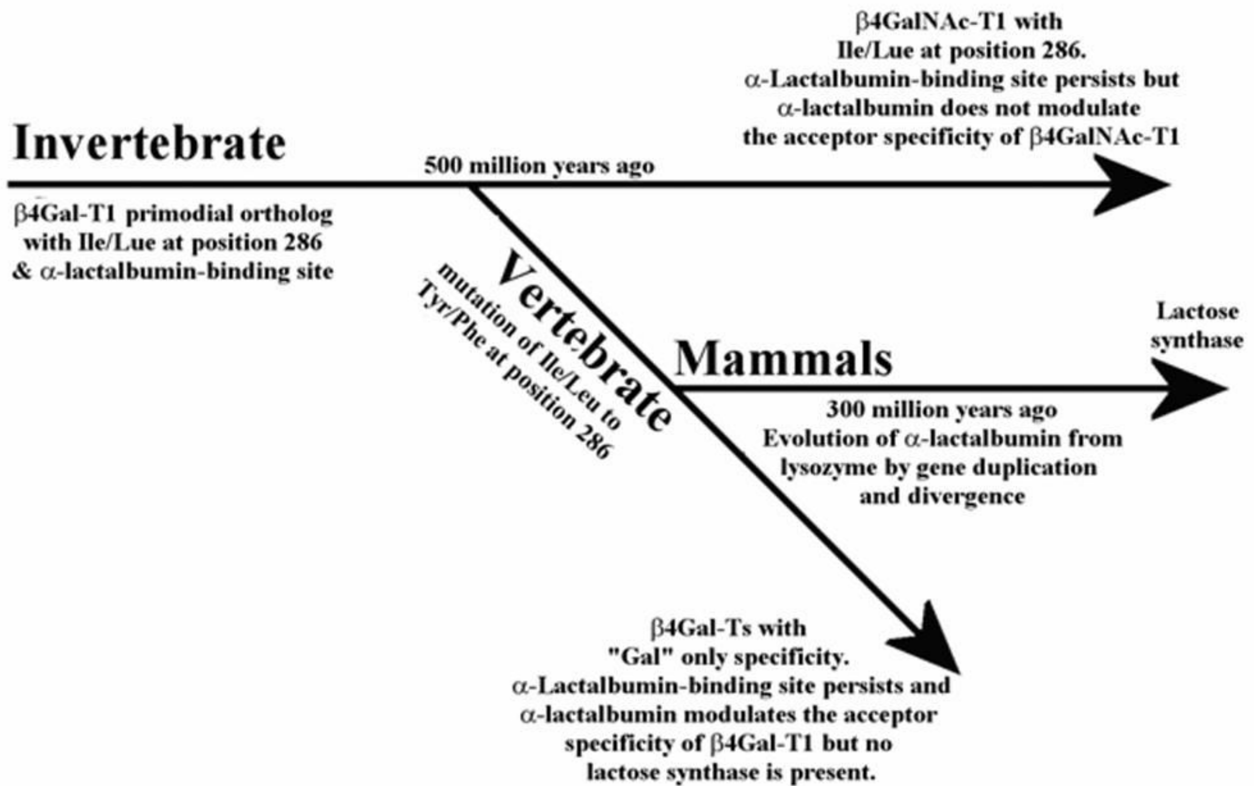


Figure 4.

Schematics of the divergence of vertebrate and invertebrate glycoconjugates and utilization of the preexisting α-LA-binding site in β4Gal-T ortholog during mammalian evolution. Nearly 500 million years ago, when vertebrates diverged from invertebrates, a substitution of the Leu/Ile residue to Tyr/Phe occurred in the catalytic pocket of the invertebrate β4GalNAc-T1, which already had an α-LA-binding site. This substitution generated a β4Gal-T1 enzyme that had sugar-donor specificity towards UDP-Gal, changing the glycosylation patterns of vertebrate glycoproteins and glycolipids. However, the β4Gal-T7 ortholog of invertebrates has a Phe residue in the catalytic pocket, keeping the specificity toward the sugar donor UDG-Gal to synthesize proteoglycans, which are essential for the species survival. The evolution of α-LA from lysozyme nearly 300 million years ago resulted in the emergence of the lactose synthase enzyme complex in which α-LA utilized the preexisting α-LA-binding site in the β4Gal-T1 ortholog to modulate the acceptor specificity toward glucose to produce lactose.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

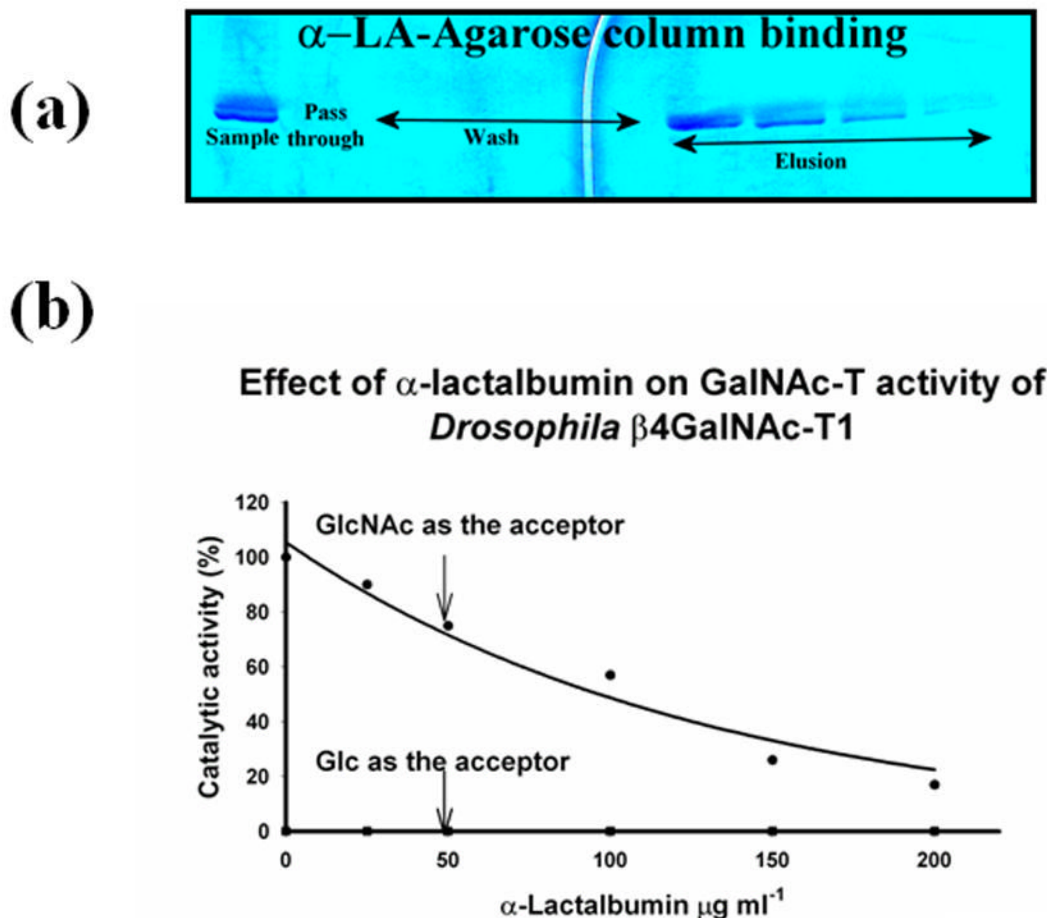


Figure 5.

The binding of the *in vitro* folded recombinant *Drosophila* CG8536 protein to the α-LA-agarose column and inhibition of the protein's GalNAc-T activity by α-LA. As shown by the SDS gel analysis (a), the *in vitro* folded protein also bound to the α-LA-agarose column in the presence of 10 mM GlcNAc and was eluted from the column with the buffer without GlcNAc. This property indicates that *Drosophila CG8536* protein has an α-LA-binding site that can be accessed only in the presence of its substrate, a characteristic of mammalian β4Gal-T1.[8, 16, 17] α-LA is known to bind mammalian β4Gal-T1 at the extended sugar-binding site,[16] where it inhibits the transfer of galactose from UDP-Gal to GlcNAc or to an oligosaccharide. At the same time, α-LA enhances the transfer of galactose to glucose to make lactose. In the presence of recombinant bovine α-lactalbumin, the *Drosophila CG8536* protein also inhibited the transfer of GalNAc from UDP-GalNAc to GlcNAc (b), or to chitobiose; however, it failed to transfer GalNAc to glucose (b). The 100% GalNAc-T activity corresponds to a specific activity of 32 pmol/min/μg of protein. GlcNAc and Glc concentrations were 25 mM each.