**The EMBO Journal** Vol. 20 No. 4 pp. 638–649, 2001

# Bovine α1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases

**Louis N.Gastinel[1], Christophe Bignon, Anup K.Misra[2,3], Ole Hindsgaul[2,3], Joel H.Shaper[4] and David H.Joziasse[5]**

Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 6098, CNRS and Universités d'Aix-Marseille I and II, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France, [2]Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada, [3]The Burnham Institute, La Jolla, CA 92037, [4]The Cell Structure and Function Laboratory, The Oncology Center and Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21231-1000, USA and [5]Department of Medical Chemistry, Vrije Universiteit, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands

[1]Corresponding author
e-mail: gastinel@afmb.cnrs-mrs.fr

**α1,3-galactosyltransferase (α3GalT, EC 2.4.1.151) is a Golgi-resident, type II transmembrane protein that transfers galactose from UDP-α-galactose to the terminal *N*-acetyllactosamine unit of glycoconjugate glycans, producing the Galα1,3Galβ1,4GlcNAc oligosaccharide structure present in most mammalian glycoproteins. Unlike most other mammals, humans and Old World primates do not possess α3GalT activity, which is relevant for the hyperacute rejection observed in pig-to-human xenotransplantation. The crystal structure of the catalytic domain of substrate-free bovine α3GalT, solved and refined to 2.3 Å resolution, has a globular shape with an α/β fold containing a narrow cleft on one face, and shares a UDP-binding domain (UBD) with the recently solved inverting glycosyltransferases. The substrate-bound complex, solved and refined to 2.5 Å, allows the description of residues interacting directly with UDP-galactose. These structural data suggest that the strictly conserved residue E317 is likely to be the catalytic nucleophile involved in galactose transfer with retention of anomeric configuration as accomplished by this enzyme. Moreover, the α3GalT structure helps to identify amino acid residues that determine the specificities of the highly homologous ABO histo-blood group and glycosphingolipid glycosyltransferases.**

*Keywords*: ABO histo-blood group/Forssman antigens/glycosphingolipids/nucleotide-binding protein/xenotrans-plantation

## Introduction

Mammalian glycosyltransferases form a group of 100 or more enzymes that collectively participate in the biosynthesis of the glycans of glycoproteins, proteoglycans and glycolipids. α1,3-galactosyltransferase (α3GalT, EC 2.4.1.151) catalyzes the transfer of galactose (Gal) from UDP-Gal to glycoconjugate acceptors having LacNAc (Galβ1,4GlcNAc) as the non-reducing terminal disaccharide in the presence of $Mn^{2+}$ as cofactor, according to the reaction: UDP-Gal + Galβ1,4GlcNAc-R → Galα1,3Galβ1,4GlcNAc-R + UDP in which R may be a glycoprotein or a glycolipid (Blanken and Van den Eijnden, 1985). Both α3GalT and its enzymatic product, the Galα1,3Gal glycan structure, are expressed by New World primates (platyrrhines) and many non-primate mammals, but are absent from the tissues of Old World primates (catarrhines), including *Homo sapiens* (Galili *et al*., 1988). The molecular basis for the species-specific absence of the enzyme involves the inactivation of the locus encoding the α3GalT gene in the primate taxa that do not express the Galα1,3Gal epitope (Larsen *et al*., 1990; Joziasse *et al*., 1991; Joziasse and Oriol, 1999).

A cDNA encoding α3GalT was first isolated from a bovine cDNA library (Joziasse *et al*., 1989). Other mammalian α3GalT orthologs have been cloned from mouse (Larsen *et al*., 1989; Joziasse *et al*., 1992), marmoset (Henion *et al*., 1994) and pig (Strahan *et al*., 1995). The α3GalT cDNA sequence predicts a type II transmembrane protein showing a structural domain organization similar to that of the other mammalian glycosyltransferases (Joziasse, 1992). The bovine enzyme contains a six amino acid N-terminal cytoplasmic tail linked to a single transmembrane domain (16 amino acids), which is connected by the 'stem region' to the lumenal, C-terminal catalytic domain. The stem region, rich in proline, glycine and polar/charged amino acids, is not conserved among α3GalT family members. N- and C-terminal truncations of recombinant, soluble α3GalT indicated that the bovine catalytic domain encompasses amino acids 87–368 (Henion *et al*., 1994).

The amino acid sequence E80–V368 of α3GalT shares a large degree of homology with α1,3-glycosyltransferases responsible for the synthesis of Forssman and *iso*-globoside glycosphingolipids (44–50% identity; Haslam *et al*., 1996; Xu *et al*., 1999; Keusch *et al*., 2000) and the ABO histo-blood group antigens (45% identity; Yamamoto and Hakomori, 1990; reviewed in Hakomori, 1999), suggesting a common phylogenetic origin. Moreover, the human ABO and Forssman glycosyltransferase genes share the same chromosomal localization (chromosome 9q34) with the human α3GalT homolog HGT-10, and the genomic organization with murine α3GalT, which demonstrates a close evolutionary relationship (Joziasse *et al*., 1992; Yamamoto *et al*., 1995; Xu *et al*., 1999). Together, the various genes constitute an α1,3-glycosyltransferase gene family, which probably arose from a series of gene duplications. Subsequent divergence has produced enzymes that use different donor and acceptor substrates from those of bovine α3GalT (Table I). The high degrees of identity present all along the catalytic domain of the



Published by Oxford University Press

α1,3-glycosyltransferases suggest a common fold. The fact that catalytic properties have diverged makes this gene family a useful system for analyzing the role of the various amino acids of the catalytic domain in determining substrate preference.

The elucidation of the mechanism of enzymatic glycosyl transfer has been another impetus to our efforts to derive the three-dimensional structure of α3GalT. α3GalT transfers galactose from the donor UDP-α-D-Gal to the acceptor Galβ1,4GlcNAc-R, while retaining the galactose in the α-anomeric configuration. Earlier, crystal structures were derived for β4GalT1 (Gastinel *et al*., 1999), glucuronyltransferase I (Pedersen *et al*., 2000) and GlcNAc-transferase I (Unligil *et al*., 2000), all of which transfer a sugar via an inverting mechanism. Comparison of the catalytic center of these enzymes with that of α3GalT may produce insight into the mechanism that determines whether inversion or retention occurs.

The α1,3GalT enzyme has recently attracted considerable attention because it synthesizes the Galα1,3Gal epitope. Naturally occurring anti-Galα1,3Gal antibodies in human serum (Galili *et al*., 1987) present a major barrier to the use of porcine and other non-primate organs for xenotransplantation in humans. Antibody binding to the Galα1,3Gal epitopes present on the vascular endothelium of the xenotransplants produces hyperacute graft rejection. Efforts are now in progress to overcome this difficulty by modifying the donor animal, the pig (reviewed in Cooper, 1998; Joziasse and Oriol, 1999). One strategy to this end could be the pre-treatment of pigs using an α3GalT-specific enzyme inhibitor. The structure-based design of such a drug will benefit from a knowledge of the three-dimensional structure of α3GalT.

Here, we report on the crystal structure of the bovine α3GalT catalytic domain in both the absence and presence of UDP-Gal. This is the first described structure for a 'retaining' glycosyltransferase. The crystal structure reveals a globular α/β fold, and allows the description of the donor-binding site at atomic resolution. The structure also suggests amino acids involved in the acceptor-binding site. The amino acids responsible for the substrate specificity of strongly related enzymes such as the ABO histo-blood group glycosyltransferases and certain glycosphingolipid synthases are identified. Moreover, a hypothesis is proposed to explain the glycosyl transfer mechanism by retention of anomeric configuration.

## Results and discussion

### *Overall protein structure*

Truncated bovine α3GalT (residues 80–368) produced as a seleniated molecule was crystallized and solved by the multiwavelength anomalous diffraction (MAD) phasing method using data sets collected at three wavelengths (Hendrickson *et al*., 1990) (Table II). The seleniated structure was refined to 2.8 Å and the seleniated model was used to refine the structural parameters against the 2.3 Å data collected from the native substrate-free α3GalT crystals (Table II). The substrate-bound α3GalT, obtained by soaking α3GalT crystals in the presence of Hg-UDP-Gal and $Mn^{2+}$, was solved to 2.5 Å resolution using the refined native coordinates of the substrate-free α3GalT structure.

**Table I.** Donor and acceptor substrate preference of the α1,3-glycosyltransferase family

| | Donor substrate | Acceptor substrate |
|---|---|---|
| α3GalT | UDP-Gal | **Gal**β1,4GlcNAc-R |
| A transferase | UDP-GalNAc | Fucα1,2**Gal**β1,3/4-R |
| B transferase | UDP-Gal | Fucα1,2**Gal**β1,3/4-R |
| Forssman synthase | UDP-GalNAc | **GalNAc**β1,3Galα1,4Galβ1,4Glcβ1-ceramide |
| Iso-globoside synthase | UDP-Gal | **Gal**β1,4Glcβ1-ceramide |

All enzymes produce α1,3-glycosyl linkages. The acceptor sugar is indicated in bold.

The substrate-free α3GalT (Ser81–Asn367) and the substrate-bound α3GalT (Lys82–Thr358) form a globular protein with overall dimensions of $50 \times 43 \times 58$ Å$^3$ (Figure 1). The structure consists of 10 β-strands, six α-helices and six $3_{10}$ helices, based on the PROMOTIF program (Hutchinson *et al*., 1996). The folding of the protein is that of an α/β protein with a central mixed twisted β-sheet of eight β-strands surrounded by four long α-helices. The structure starts at Ser81 with a short N-terminal α-helix (α1), followed by a β-hairpin containing one β-strand (β1), and then by a long connecting α-helix (α2). The central β-sheet can be divided into two portions. The first portion runs from Val129 to Met224, defining an N-terminal subdomain. It contains a β-sheet formed by four parallel β-strands in the strand order 4, 3, 2 and 5, surrounded by two long α-helices (α3 and α4). This N-terminal subdomain accounts for the binding of the nucleotide moiety of the nucleotide-sugar donor substrate because of the presence of unambiguous electronic density signatures of UMP and UDP-Gal occurring in the substrate-free and substrate-bound α3GalT crystals, respectively (Figures 1 and 2). The UMP molecule results from the elution step in the final affinity purification procedure of the enzyme using UDP-hexanolamine–Sepharose resin. The second portion of the central β-sheet consists of two parallel β-strands (β7 and β9) flanked by two antiparallel β-strands (β8 and β1) and with two long α-helices (α5 and α6) on one side. A second small β-sheet running almost parallel to the central β-sheet consists of two short antiparallel β-strands (β6 and β10). A structural homology search using the entire α3GalT structure and the DALI database revealed that its overall fold is unique (Holm and Sander, 1983).

Strand β10 is followed by a loop including the C-terminal residues Thr358–Asn367, which completes the substrate-free structure. It has been shown previously that removing the last three residues, K374NV, of marmoset α3GalT inactivates the enzyme (Henion *et al*., 1994). The substrate-free α3GalT structure, in contrast to the substrate-bound structure, shows electron density for the C-terminal segment containing residues Lys359–Asn367. This segment is stabilized by hydrophobic interactions involving mainly residues V363, V364 and the tryptophans W249, W250 and W314. The absence of visible electron density for the C-terminal segment in the substrate-bound structure is not clearly understood. This region of the protein is presumed to be disordered in the substrate-bound crystal, being exposed to a large solvent channel, and is not involved in crystal contacts.

**Table II.** Crystallographic data and refinement statistics

| | Sel-α3GalT | Sel-α3GalT | Sel-α3GalT | α3GalT | α3GalT + UDPG |
|---|---|---|---|---|---|
| Wavelength (Å) | 0.9796 Peak (W1) | 0.9800 inflection (W2) | 0.9324 remote (W3) | 0.9324 | 0.9324 |
| Energy (Kev) | 12.6566 | 12.6515 | 13.2971 | 13.2971 | 13.271 |
| Resolution (Å) | 30.0–2.8 | 30.0–2.8 | 30.0–2.8 | 30.0–2.0 | 30.0–2.5 |
| Unit cell, $a = b$, $c$ (Å) | 95.55, 112.71 | 95.55, 112.71 | 95.55, 112.71 | 95.56, 112.71 | 95.6, 110.72 |
| No. of unique reflections | 10 366 | 10 366 | 10 366 | 31 286 | 18 010 |
| $[I/\sigma_{(I)}]$ | 5.7 (1.8) | 5.6 (2.0) | 7.2 (1.7) | 7.5 (1.8) | 9.3 (2.4) |
| $R_{sym}$ (%)[a] | 7.9 (20) | 7.7 (19.5) | 7.5 (12) | 5.8 (38.3) | 4.3 (22.7) |
| $R_{anomal}$ (%) | 8.0 (15.5) | 5.9 (13.2) | 6.7 (9.5) | | |
| Completeness (%) | 97.6 (97.6) | 98.2 (98.2) | 95.7 (95.7) | 96.5 (96) | 99.5 (99.5) |
| Anomalous completeness (%) | 84.2 (85) | 88 (87) | 80.3 (79.2) | | |
| Multiplicity | 3.4 (3.4) | 3.7 (3.6) | 3.5 (3.4) | 5.6 (2.3) | 6.7 (6.3) |
| Resolution for the refinement (Å) | | | 15.0–2.8 | 15.0–2.3 | 15.0–2.5 |
| $R_{cryst}$ (%)/$R_{free}$ (%)[b] | | | 23/27 | 21/25 | 22/27 |
| R.m.s.d. (bonds) (Å)/(angles (°) | | | 0.0095/1.52 | 0.016/1.8 | 0.015/1.8 |
| No. of atoms | | | | | |
| protein/water | | | 2308/43 | 2393/130 | 2308/127 |
| cofactors[c] | | | | 20/1 | 36/1/1/11 |
| Average $B$-factor (Å$^2$) | | | | | |
| protein/water | | | 53.5/60 | 50/50 | 66/68 |
| cofactors[c] | | | | 41/44 | 59/78/77/60 |
| No. of ϕ/ψ angles (%) | | | | | |
| most favoured/allowed | | | 80/16 | 86/12.8 | 87.1/12.5 |

The values in parentheses refer to data in the high resolution shell.
[a]$R_{sym} = \Sigma_{hkl} \Sigma_i |I_i - (I)|/\Sigma(I)$
[b]$R_{cryst} = \Sigma(||Fp(obs)| - |Fp(calc)||/\Sigma|Fp(obs)|$ and $R_{free}$ = $R$-factor for a randomly selected subset (9.5%) of data that were not used to minimize the crystallographic residual.
[c]Cofactors: UMP/$Mn^{2+}$ in the case of α3GalT and UDP-Gal/-Hg/$Mn^{2+}$/galactose bound to E317 in the case of α3GalT + UDPG.

None of the three cysteines of α3GalT bovine catalytic domain, C223, C298 and C338, is engaged in disulfide bridges (Figure 1), which is consistent with a recent analysis of bovine α3GalT structure using liquid chromatography and electrospray ionization-tandem mass spectrometry (Yen *et al.*, 2000). The distances between the two strictly conserved cysteines, C223 and C298, is 13 Å. Each of the two conserved cysteine residues is located in a cluster of hydrophobic residues enclosed within the central part of the α3GalT protein. In this central position, changes are likely to affect protein fold integrity, which could explain their strict conservation among all the α1,3-glycosyltransferase family members.

The packing of α3GalT crystals shows a close interaction between two symmetrical α3GalT molecules mediated by their respective N-terminal subdomains rotated over ~90°. This hydrophobic interaction involves mainly residues V170, P174 and L175 from the long connecting loop between β3 and β4, and residues F184, V186 and F187 from strand β4. The total buried surface represents 840 Å$^2$ (calculated with TURBO software; A.Roussel and C.Cambillau, personal communication). So far, dimerization of soluble α3GalT has not been reported; it has been proposed that different glycosyltransferases may form non-covalent oligomers during Golgi sorting/trafficking. It is conceivable that the homologous UDP-binding subdomain shared by Golgi glycosyltransferases of different specificities (see below) might support their non-covalent association, increasing their efficiency in glycan biosynthesis.

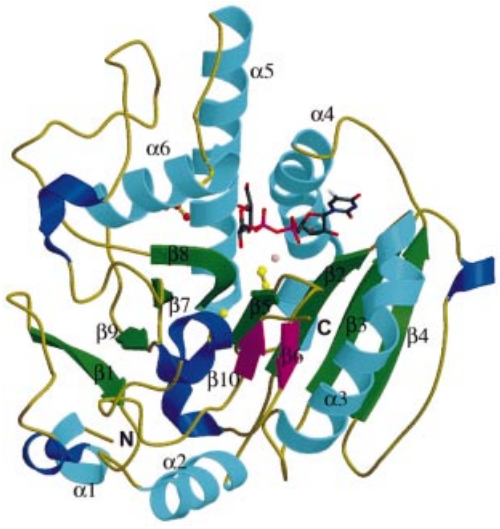



**Fig. 1.** Overall view of the bovine substrate-bound α3GalT catalytic domain structure. Ribbon diagram of the molecule viewed down the open pocket. The three cysteine residues Cys223, Cys298 and Cys338 are shown in ball-and-stick form in yellow. The only Asn293 potentially available for N-glycosylation is shown in ball-and-stick form in red. N and C indicate the N- and C-termini of the molecule. Secondary structure elements are color coded as follows: α-helices in cyan, $3_{10}$ helices in blue, the β-strands of the central twisted eight-stranded β-sheet in green, and the β-strands of the small two-stranded antiparallel β-sheet in magenta. Hg-UDP-Gal is shown in stick form color coded according to the nature of the atom.

### *α3GalT catalytic pocket*

The catalytic pocket of α3GalT was identified by the presence of: (i) a narrow cleft measuring ~14 × 14 Å on its molecular surface, formed between the N-terminal subdomain and the C-terminal portion of the protein; (ii) a large surface portion where amino acid residues that are invariant within the α1,3-glycosyltransferase family are

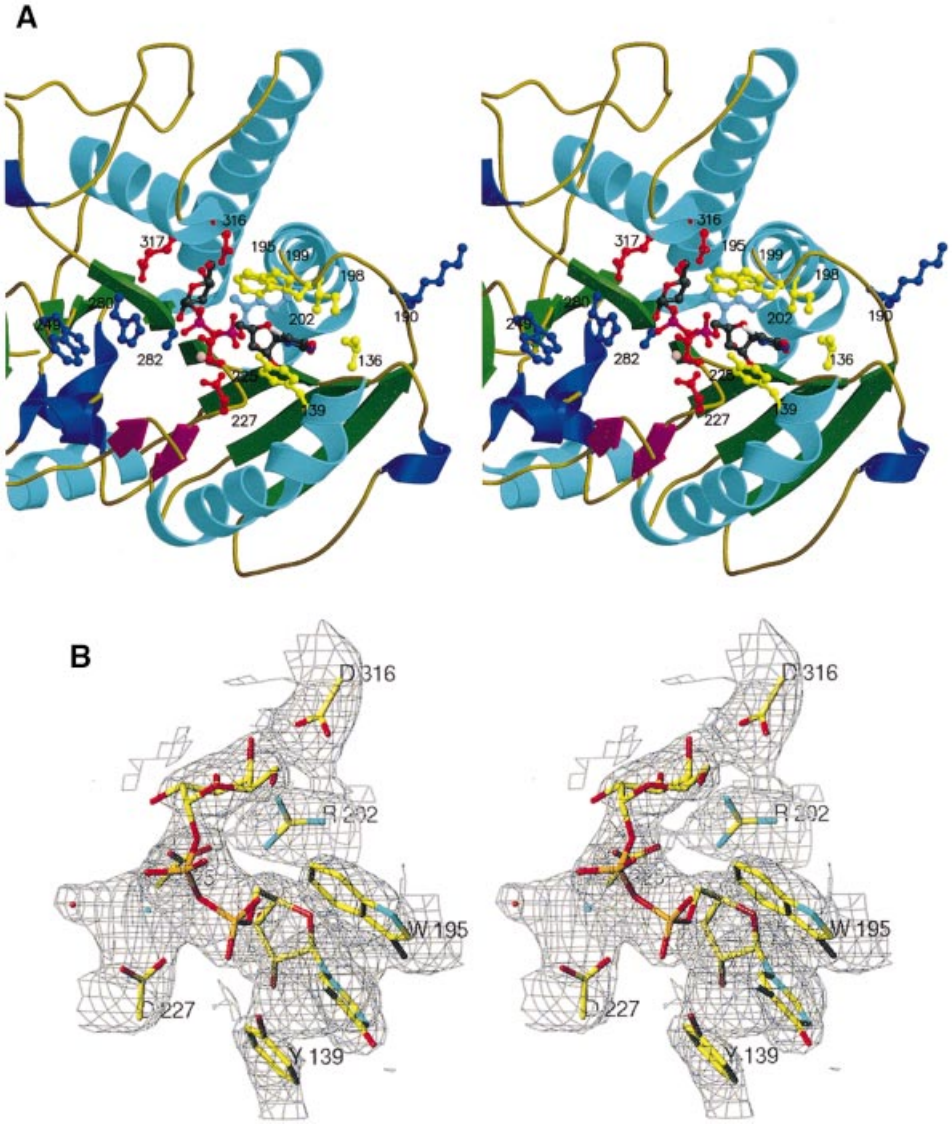



**Fig. 2.** Close-up stereoview of the α3GalT UDP-Gal-binding site. (**A**) Hg-UDP-Gal is shown in ball-and-stick form and color coded depending on the nature of the atoms; the $Mn^{2+}$ ion is shown as a pink sphere. Amino acid side chains interacting with Hg-UDP-Gal are shown in ball-and-stick form in yellow. The acidic residues from the motifs D225VD227 and the D316E317 are shown in ball-and-stick form in red. The four amino acid side chains of α3GalT residues at positions equivalent to the residues distinguishing human A-GT from B-GT are shown in ball-and-stick form in blue. (**B**) Stereoview of the electron density map ($2F_o - F_c$, 1σ) of the Hg-UDP-Gal-binding site.

concentrated; and (iii) one UMP molecule found in the substrate-free, or one Hg-UDP-Gal molecule found in the respective substrate-bound structures (Figures 1 and 2). The catalytic pocket contains the solvent-exposed sequence D225VD227, which was identified as the DXD motif, known to be well conserved in many mammalian glycosyltransferases (Wiggins *et al*., 1998). The α3GalT pocket is made up of 19 solvent-exposed residues, two hydrophobic, 12 polar (six tryptophan residues) and five ionizable side chain residues. The bottom of the cavity is formed from the β-strand β8 containing residues H280, A281 and A282, and its sides are lined with the following residues: Y139, W195, S199, R202, D225, D227, Q228, Q247, W249, W250, T259, W314, D316, E317 and W356 (Figure 2). Eight of these residues, Y139, W195, S199, R202, D225, D227, W314 and E317, are invariant in the α1,3-glycosyltransferase family. They are involved in the binding of the UDP moiety of the UDP-sugar donor substrate and may contain the acidic residue responsible for the nucleophilic attack on the C1 atom of the transferred sugar. The other residues, Q228, Q247, W249, W250, T259, H280, A281, A282, D316 and W356, are conserved only in all α3GalT orthologs but are different in A-GT, B-GT, $iGB_3$ synthase and Forssman glycosyltransferase. This list of residues might include some of the amino acids that interact with the donor substrate sugar (galactose or *N*-acetylgalactosamine) and with the acceptor substrate.

### *UDP-galactose-binding site*

Hg-UDP-Gal binds across the depression found on the α3GalT molecular surface, with its uridine and ribose moieties maintained by conserved residues that form a small pocket in the N-terminal subdomain. The uracil base

**Table III.** Interatomic distances between Hg-UDP-Gal, $Mn^{2+}$, bound water molecules, UMP and α3GalT protein atoms in substrate-free and substrate-bound structures

| Substrate-free structure | | | Substrate-bound structure | | |
|---|---|---|---|---|---|
| Interacting atoms | | Distance (Å) | Interacting atoms | | Distance (Å) |
| Uracil N3 | V136 O | 3.0 | Uracil N3 | V136 O | 3.0 |
| Uracil O2 | V136 O | 3.0 | Uracil O2 | V136 N | 2.9 |
| Uracil O2 | V136 N | 2.9 | Ribose O2′ | F134 O | 3.0 |
| Uracil O4 | R138 NH2 | 2.7 | Ribose O3′ | R202 NH1 | 2.7 |
| Ribose O2′ | V226 N | 2.7 | Ribose O3′ | Wat 89 | 2.5 |
| Ribose O2′ | Wat 2 | 2.7 | Ribose O3′ | V226 N | 3.4 |
| Ribose O3′ | D227 N | 3.3 | O2Pα | Y139 OH | 2.6 |
| O3Pα | Y139 OH | 2.7 | $Mn^{2+}$ | D225 OD2 | 2.5 |
| $Mn^{2+}$ | D225 OD2 | 2.6 | $Mn^{2+}$ | D227 OD1 | 2.4 |
| $Mn^{2+}$ | D227 OD2 | 2.7 | $Mn^{2+}$ | D227 OD2 | 2.4 |
| $Mn^{2+}$ | D227 OD1 | 2.4 | $Mn^{2+}$ | Wat 82 | 2.9 |
| $Mn^{2+}$ | Wat 97 | 2.3 | $Mn^{2+}$ | O2Pβ | 2.9 |
| $Mn^{2+}$ | O1Pα | 2.1 | $Mn^{2+}$ | O3Pα | 2.5 |
| | | | Gal O2 | A282 N | 3.2 |
| | | | Gal O3 | Wat 84 | 3.2 |
| | | | Gal O3 | A281 O | 3.1 |
| | | | Gal O3 | D225 OD1 | 3.2 |
| | | | Gal O3 | D225 OD2 | 3.4 |
| | | | Gal O3 | R202 NH2 | 2.9 |
| | | | Gal O4 | D316 OD1 | 2.9 |
| | | | Gal O4 | D316 OD2 | 2.8 |
| | | | Gal O4 | E317 N | 3.4 |
| | | | Hg | Y139 OH | 2.5 |

binds via aromatic stacking involving Y139, W195 and I198, and its N3 and O2 atoms hydrogen-bond to V136 (Table III; Figure 2A). The ribose O3′ forms water-mediated and direct hydrogen bonds with the NH1 of R202 and the N atom of V226.

The galactose moiety is involved in several interactions with the protein, in particular with R202, D225, A281, A282, D316 and E317 (Table III; Figure 2).

An $Mn^{2+}$ atom is found in an approximately octahedral coordination state in which two of the coordination atoms, O3Pα and O2Pβ, are from the α- and β-phosphate of the UDP molecule (Table III). D227 forms a bidentate interaction through OD1 and OD2, and D225 interacts through OD2. The last ligand is from a water molecule. Three of the six $Mn^{2+}$ coordination sites are from direct interaction with α3GalT protein, which might produce substantial affinity of the protein for $Mn^{2+}$ even in the absence of donor substrate. In the GnT I and GlcAT-1 structures, only one and two direct interactions, respectively, have been observed between $Mn^{2+}$ and protein, which excludes the existence of an independent metal-binding site in these enzymes.

Only a few significant differences are observed between the binding of UMP in the substrate-free structure and that of Hg-UDP-Gal in the substrate-bound structure (Table III). In the case of the substrate-free structure, the uracil N3, O2 and O4 atoms hydrogen-bond to V136 and the R138 NH2 atom, respectively. The two ribose oxygens O2′ and O3′ form one water-mediated and two direct hydrogen bonds to V226 and D227 main chain nitrogen atoms. These differences may be due to the presence of the mercury atom bound to the C5 atom of the uracil ring. Five of the six possible coordination sites of $Mn^{2+}$ are used. One coordination atom, O1Pα, is from the UMP molecule; the other four coordination atoms are equivalent to those in the substrate-bound structure (Table III).

### *The α1,3-glycosyltransferase family: donor substrate specificity*

The glycosyltransferases responsible for the synthesis of ABO histo-blood group carbohydrates and certain glycosphingolipids belong to the same α1,3-glycosyltransferase family as bovine α3GalT, because they show a high level of identity in the catalytic domain (Figure 3). The human blood group transferases A-GT and B-GT, which transfer *N*-acetylgalactosamine and galactose, respectively, to α1,2-fucosylated complex-type glycans, share strict amino acid sequence identity except at four positions, from A-GT to B-GT: Arg146→Gly146, Gly235→Ser235, Leu266→Met266 and Gly268→Ala268. Chimeric enzymes have been constructed at these positions between A-GT and B-GT, and only two positions, 266 (Leu→Met) and 268 (Gly→Ala), were found to be crucial for defining the donor preference, GalNAc versus Gal (Seto *et al.*, 1999).

In the α3GalT structure, the two equivalent positions are His280 (Leu266 or Met) and Ala282 (Gly268 or Ala), respectively, both of which are located in the catalytic pocket of the enzyme. The residues His280 and Ala282 are located in β-strand β8 in the middle of the central β-sheet forming the bottom of the catalytic pocket (Figure 2A). His280 does not interact directly with galactose, but replacing it by either of the hydrophobic residues methionine (B-GT) or leucine (A-GT) will result in a local change in the shape or size of the cavity in proximity to the chemical group attached to the donor sugar C2 atom. The side chain of Ala282, which is replaced by glycine (Gly268) in the human A-GT, is pointing upward from the bottom of the catalytic cavity and could restrict by steric

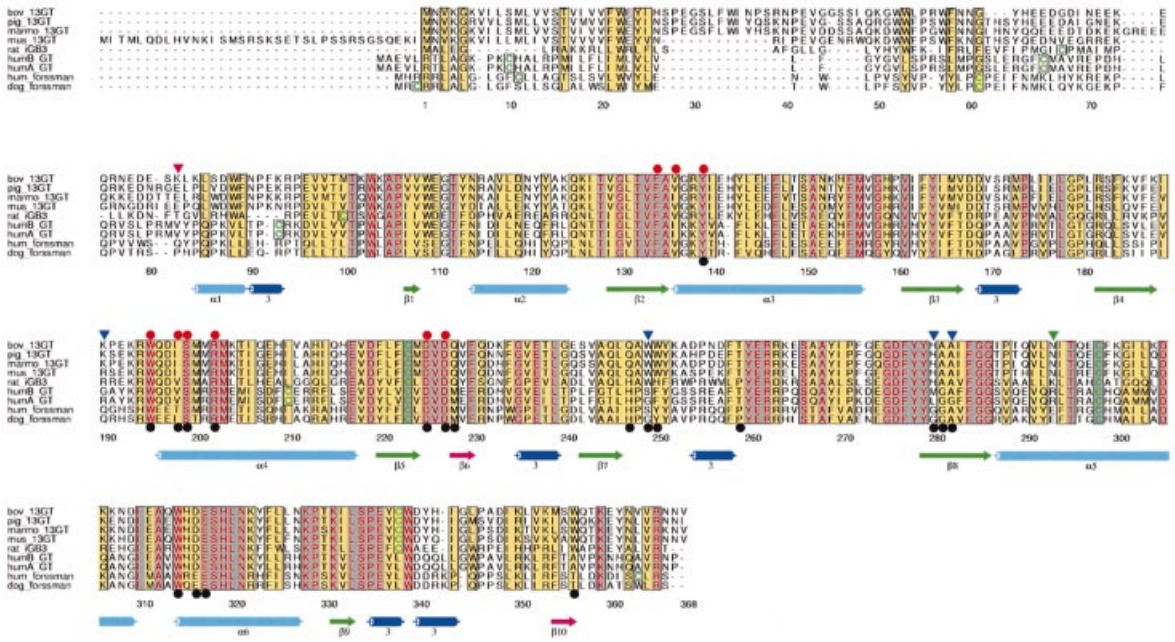


**Fig. 3.** Sequence alignment of bovine α3GalT and homologous proteins belonging to the family of related α1,3-glycosyltransferases. The invariant residues are highlighted in red with a gray background, whereas conserved residues have an orange background. Cysteine residues are displayed in green. Secondary structure elements of bovine α3GalT are indicated beneath the sequences color coded as in Figure 1A. Amino acid sequences used for the alignment are: bovine α3GalT (P14769; Joziasse *et al*., 1989), pig_13GT (P50127; Strahan *et al*., 1995), marmoset_13GT (Q28855; Henion *et al*., 1994), mouse_13GT (P23336; Larsen *et al*., 1989), rat_iGb3 synthase (AF248543; Keusch *et al*., 2000), humB_GT and humA_GT (AF134414 and AF134412; Yamamoto *et al*., 1990), hum_forssman (DDBJ/EMBL/GenBank 163572; Xu *et al*., 1999) and dog_forssman (Q95158; Haslam *et al*., 1996). Residues exposed at the surface in the pocket of the substrate-bound α3GalT structure are indicated by closed black circles. Residues that interact with the UDP portion of Hg-UDP-Gal are indicated by closed red circles. The four residues that differentiate between the humA_GT and humB_GT sequences are indicated by closed blue triangles. The first well defined residue in the electron density map of the substrate-bound α3GalT structure, Lys82, is indicated by a closed magenta triangle. The Asn293 involved in the only potential N-glycosylation site is indicated by a closed green triangle. The bovine α3GalT sequence is numbered every tenth residue.

hindrance the access of the *N*-acetyl moiety of the UDP-*N*-acetylgalactosamine, the sugar donor of the human A-GT. A cDNA coding for the rare $O^2$ allele of the ABO histo-blood group has been described recently as containing a point mutation whereby Ala268 is replaced by Arg268, resulting in the complete inactivation of the enzyme (Hakomori, 1999). In the bovine α3GalT structure, a long and charged arginine side chain, replacing the alanine residue at the bovine α1,3GalT equivalent position 282, should completely block the donor sugar-binding site by steric hindrance and electrostatic incompatibility.

### *Acceptor specificity of α3GalT*

The members of the α1,3-glycosyltransferase family differ in the nature of the acceptor substrate that they use (Table I). In the absence of X-ray data on acceptor substrate bound to α3GalT crystals, we can only make assumptions as to the probable position of the acceptor substrate in the protein. Surface hydrophobic residues, particularly solvent-exposed tryptophans, have been suggested to interact with the sugar ring by hydrophobic stacking interactions. Among the five solvent-accessible tryptophan residues found in the α3GalT catalytic pocket (W195, W249, W250, W314 and W356), only two, W249 and W356, are specific to α1,3-glycosyltransferases that transfer galactose to non-fucosylated LacNAc acceptors. Trp249 of α3GalT is replaced by either serine or glycine in the A-GT, B-GT and Forssman synthases, whose acceptor substrate is either a β1,3/4-linked α1,2-fucosylated galactose or a β1,3-linked *N*-acetylgalactosamine (Table I). Trp356 of α3GalT is replaced by alanine (Ala343) in the ABO histo-blood group glycosyltransferases. This modification indicates that Ala343 might be close to the fucose-binding site of α1,2-fucosylated acceptors. A tryptophan residue replacing alanine at this position might block the binding of α1,2-fucosylated LacNAc by steric interaction with the fucose. This might explain the inability of α3GalT to transfer galactose to α1,2-fucosylated acceptors.

### *The UDP-binding domain of glycosyltransferases*

Few X-ray crystal structures are available for the large number of nucleotide–hexose glycosyltransferases that are currently known. Only three crystal structures have been published so far, with their coordinates already available in the Protein Data Bank: the β-glucosyltransferase from phage T4 (1bgt, 1bgu; Vrielink *et al*., 1994; Morera *et al*., 1999), SpsA from *Bacillus subtilis* (1qgq; Charnock and Davies, 1999) and the bovine β4GalT1 (1fgx, 1fr8; Gastinel *et al*., 1999). Quite recently, the structures of two more inverting glycosyltransferases were reported: β1,3GlcAT-1, the human glucuronyltransferase involved in GAG biosynthesis (1fgg; Pedersen *et al*., 2000), and the rabbit GnT I, a β1,2-*N*-acetylglucosaminyltransferase (1fo8; Unligil *et al*., 2000).

A distant evolutionary relationship between T4 phage UDP-α-D-glucose glucosyltransferase (βGlcT) and glycogen phosphorylase was recognized earlier, based on

topological similarities in the core secondary structure elements as well as the similar positioning of their respective ligands UDP-Glc and PLP (Holm and Sander, 1995). Modeling of the α3GalT nucleotide-binding site was accomplished recently using the coordinates of βGlcT, but no biochemical or biological data have yet been reported to confirm or disprove this model (Imberty *et al.*, 1999). Rigid body superimposition of βGlcT (residues T179–D318) and α3GalT (residues V129–N231) results in a root-mean-square deviation (r.m.s.d.) of 4.4 Å on 102 aligned $C_\alpha$ atoms. Moreover, in this superimposition, a 15 Å minimum distance separates both ligands, i.e. UDP bound to βGlcT and Hg-UDP-Gal bound to α3GalT, indicating that the location of their UDP-binding sites is not conserved. Our crystal structure does not support the hypothesis that the α1,3-glycosyltransferase family belongs to the βGlcT and GP superfamily.

Although no clear structural resemblance between β-GlcT and α3GalT has been observed, a comparison between the three-dimensional structures of the other glycosyltransferases reveals a certain similarity in the spatial arrangement of those secondary structure elements that are involved in the binding of the UDP moiety (Figure 4). These similarities, in the absence of sequence homology, are particularly significant in the case of the N-terminal 100 amino acid portion of SpsA (Val4–Tyr101), the N-terminal portion of human GlcAT-1 between Thr76 and Ser200, that of GnT I between Ile110 and Leu217, and, to a lesser extent, the middle portion of the bovine β4GalT1 (Lys181–Pro257) (Figure 4G). A structure of similar topology is now found in the bovine α3GalT polypeptide between Val129 and Asn231. The topology (Figure 4) of the domain that binds to the nucleotide-sugar is quite well conserved, with a general central 3, 2, 1, 4 parallel β-strand arrangement surrounded by a variable number of helices on both sides, reminiscent of the mononucleotide-binding motif or Rossmann fold. The position of the short DXD sequence that stabilizes the $Mn^{2+}$ ion and thus indirectly binds the diphosphate moiety of UDP is equivalent in all structures. The D98D99 sequence of SpsA, the D194DD196 sequence of GlcAT-1, the E211DD213 sequence of GnT I and the D252VD254 sequence of β4GalT1 are all located between strands β4 and β5, and the D225VD227 sequence of α3GalT between the equivalent β-strands β5 and β6 (Figure 4G). Taken together, these observations show that both the inverting glycosyltransferases and a retaining enzyme contain a UDP-binding domain of very similar spatial structure, in which key acidic amino acids occupy equivalent positions.

In order to search for similar fold arrangements in proteins with known structures, a DALI search was performed with the same N-terminal subdomain of α3GalT (Thr128–Met224), which yielded numerous hits. One of the best scores ($Z$ = 5.0, r.m.s.d. = 2.5 Å, LALI/SEQ = 74/143) was obtained with the unrelated protein ATP-binding domain of Mj0577 from *Methanococcus jannaschii* (1mjh; Zarembinski *et al.*, 1998). This structure shows some resemblance to a shared topology of secondary structural elements within a central β-sheet with the similar parallel β-strand order 3, 2, 1, 4 (Figure 4F). In addition, upon superimposing secondary structural elements, the location of the ATP molecule bound in the Mj0577 structure was found to be highly similar to that of the UMP or the UDP portion of Hg-UDP-Gal bound to α3GalT (Figure 4G). However, no DXD motif is found in the Mj0577 structure. The triphosphate moiety of ATP is stabilized by an $Mn^{2+}$ ion and by hydrogen bonding interactions to protein main chain atoms, and is localized in the loop between the β-strand β4 and the α-helix α4.

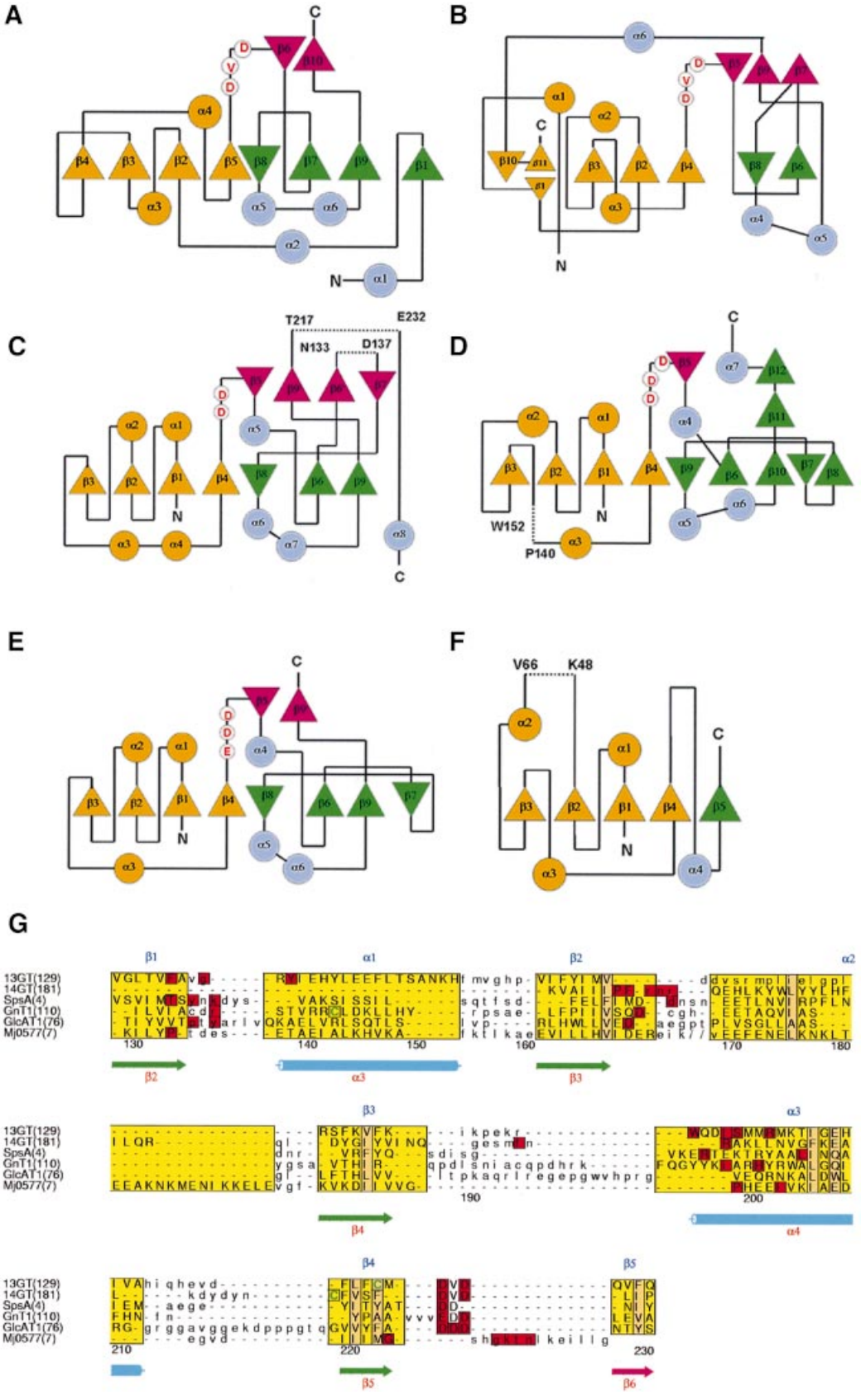
The subdomain as identified above including the DXD motif was termed SGC by Unligil *et al.* (2000). However, in view of its occurrence in other glycosyltransferases, the term UDP-binding domain (UBD) seems more appropriate. In glycosyltransferases, it encompasses virtually all those residues that bind $Mn^{2+}$ and the UDP portion of the donor substrate. Moreover, residues found interacting with the UDP moiety of the donor substrate are located in approximately the same structural secondary elements in all structures studied. These residues are located at the end of β-strands β1 and β2, in the middle of the α-helix α3, and in the DXD motif between β-strands β4 and β5 from the canonical UBD structure (Figure 4G).

The UBD may be an example of convergent evolution, or may point to a common ancestral origin, even though sequence identity between the enzymes studied here is minimal. As both structure and function are conserved, we favor the latter hypothesis. Alternatively, the glycosyltransferases may be the product of a gene fusion involving an ancestral UDP-binding protein. An analysis of intron–exon organization of the different glycosyltransferase genes demonstrates that the portion of the gene encoding the UBD has not been conserved as a single exon, which may argue against exon shuffling. We anticipate that the UBD will be common to all glycosyltransferases that use a UDP-sugar as donor substrate, which should facilitate the identification of additional glycosyltransferases present in the various genomic data banks.

**Fig. 4.** Topology diagrams of different glycosyltransferases and their respective UBD aligned sequences. Secondary structural elements of each protein are color coded as in Figure 1. (**A**) α3GalT (V129–Q231), (**B**) β4GalT1 (K181–P257), (**C**) SpsA (V4–Y101), (**D**) GlcAT-1 (T76–S200), (**E**) GnT I (I110–L217) and (**F**) Mj0577 (K7–G140). The portion of the molecules containing part of the twisted central β-sheet which has a similar organization in all five glycosyltransferases and Mj0577 is shown in yellow. N and C indicate the N- and C-termini of each molecule. The relative position of the DXD motif is indicated in the topology diagram. (**G**) Structure-based sequence alignments of the glycosyltransferase's UBDs and of the Mj0577 ATP-binding domain. Structures were first superimposed using TURBO (A.Roussel and C.Cambillau, personal communication). The individual secondary structure elements of superimposed structures were then aligned using Clustal X after removing the connecting loops. These loops (lower case) were then added back to the final alignment, which was optimized further by manual inspection. Secondary structural elements of the canonical glycosyltransferase UBD are shaded in yellow. Conserved residues have an orange background. Residues interacting with UDP are shaded in red. Secondary structure elements of α3GalT are color coded as in Figure 1, and those of the canonical UBD are indicated in blue over the alignment. The sequence numbering is from α3GalT beginning at residue Val129 and is numbered every tenth residue.

### *Structural implications for the reaction mechanism of α3GalT*

The α3GalT galactose transfer mechanism has not yet been well documented. It retains the anomeric configuration of the galactose C1 atom, whereas other glycosyltransferases such as SpsA, GlcAT-1, GnT I and β4GalT1 invert this configuration. By analogy with the glycosyl hydrolases, a double displacement mechanism of sugar



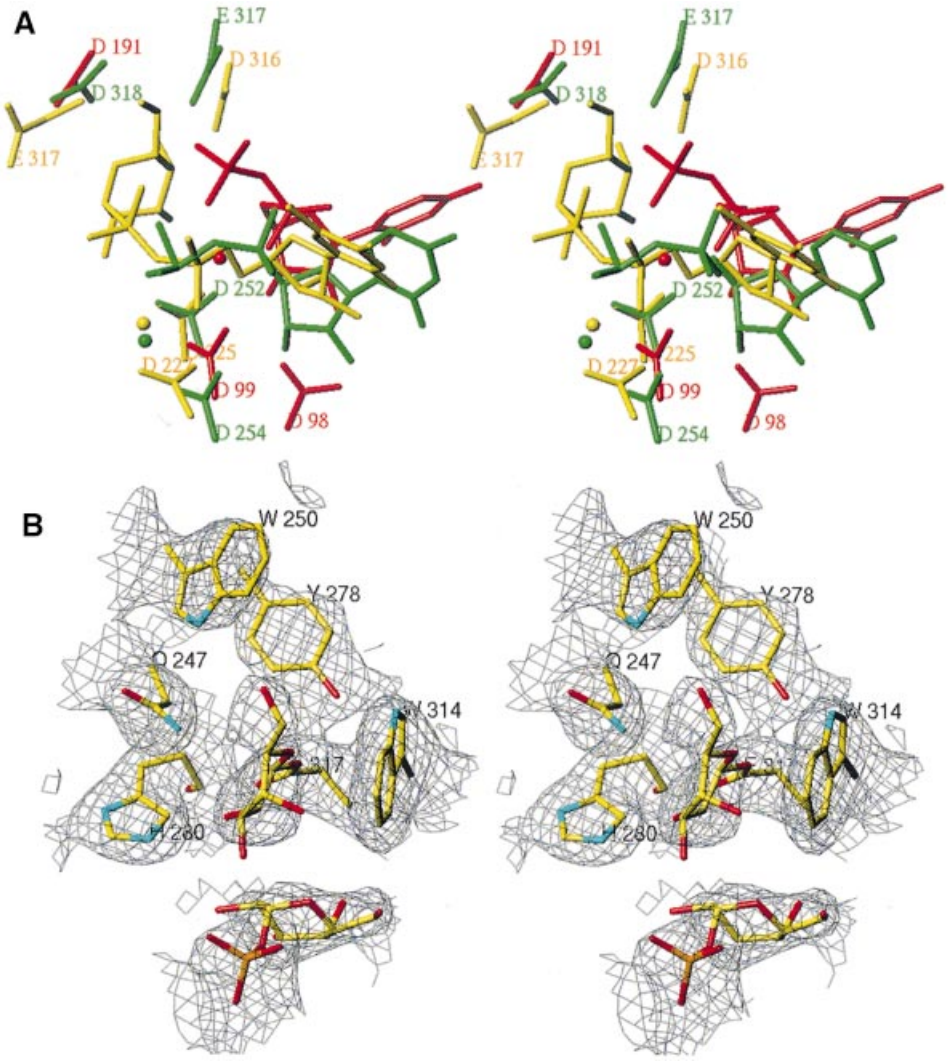



**Fig. 5.** Superimposition of residues important for catalysis in inverting and retaining glycosyltransferases and the α3GalT glycosyl-enzyme intermediate. (**A**) Stereoview of a superimposition of α3GalT (yellow), β4GalT1(green) and SpsA (red). UDP, Hg-UDP-Gal and the managenese ions found in the respective structures are displayed in the same color as the corresponding proteins. (**B**) Stereoview of the electron density map ($2F_o - F_c$, 1σ) centered around residue E317. α3GalT residues interacting with the galactose covalently bound to E317 are represented in stick form color coded according to the nature of the atom.

transfer by α3GalT has been proposed whereby the anomeric configuration is retained (Takayama *et al.*, 1999). This mechanism implies a first attack of a nucleophilic carboxyl group on the donor galactose C1 atom, resulting in the formation of a galactosyl-enzyme intermediate with a β-configuration of sugar or an oxo-carbenium sugar ion, and the release of the nucleotide diphosphate (first inversion). The acceptor substrate should subsequently attack the galactosyl-enzyme intermediate C1 atom with its deprotonated hydroxyl group linked to the C3 atom, inverting its configuration once more (second inversion).

In order to shed light on the differences and similarities between inverting and retaining catalytic mechanisms, we have superimposed the structures of two inverting glycosyltransferases, SpsA and β4GalT1, on that of α3GalT. Figure 5A shows the results of this superimposition based on 224 similar $C_\alpha$ atoms of both structures α3GalT and β4GalT1 and 210 $C_\alpha$ atoms of both α3GalT and SpsA with an r.m.s.d. of 2.9 Å. The α3GalT structure suggests that Glu317 (Figures 3 and 5), which is invariant in all the members of the α1,3-glycosyltransferase family, contains the carboxylic acid group, which reacts with the galactose (or GalNAc) C1 atom. A different amino acid, aspartic acid or glutamic acid, might play the role of nucleophile. In most inverting glycosyltransferases studied so far, the nucleophile is an aspartic acid residue: D191 for SpsA, D318 for β4GalT1 and D291 for GnT I. It is a glutamic acid residue in GlcAT1 (E281) and in α3GalT (E317).

Upon superimposition of the protein backbones, the uracil and ribose rings, the two α- and β-phosphate groups and the two equivalent DXD motifs, D254VD of β4GalT1 and D99 of SpsA, have almost the same relative position (Figure 5A). The positions of the D225VD, the $Mn^{2+}$ ion and the β-phosphate group of UDP-Gal in the substrate-bound α3GalT structure are slightly different, bringing the galactose C1 atom at a distance of 4.8 Å to the EO2 of

E317 (Figure 5A). The exact position of the sugar donor substrate cannot be derived from these superimpositions because no clear electron density was visible for it, neither in SpsA nor in the β4GalT1 structures. However, in the latter structures, the donor sugar C1 atom is expected to be sufficiently distant to be inaccessible to direct nucleophilic attack by D318 for β4GalT1 and by D191 for SpsA carboxyl groups.

The glycosyltransferases inverting the configuration of the anomeric carbon contain a general base (aspartic acid or glutamic acid), which abstracts a proton from the relevant acceptor sugar hydroxyl group. The acceptor sugar ring should be located 'above' the donor sugar plane or between the nucleophile and the donor sugar C1 atom. In the case of GlcAT-1 structure, the first structure solved with the substrate acceptor bound to it, the OE2 atom of the general base E281 is located 2.7 Å from the O3 atom of the 3-hydroxyl group of the Gal-Gal-Xyl acceptor sugar. In the case of α3GalT, E317 is the only residue capable of playing a similar nucleophile role. However, the distance (4.8 Å) between the OE2 of E317 and the galactose C1 atom is >3 Å.

Unexpectedly, the E317 residue of the substrate-bound α3GalT structure contains electron density extending its side chain carboxyl group (Figure 5B). We interpret this electronic density as being that of a galactose residue covalently bound to the OE1 atom of E317 upon its transfer from Hg-UDP-Gal. In this position, the galactose C1 atom has a β-configuration. The O2 atom of galactose is at hydrogen bonding distance from the NE2 atom of the H280 residue (3 Å) and its O6 atom is at 3.2 Å distance from the NE1 atom of Trp250. The galactose ring is in a stacking interaction with the Trp314 side chain.

From these observations, a schematic retaining mechanism can be proposed where the nucleophilic attack of E317 results in the first inversion of the C1 atom configuration (Figure 6A and B). The acceptor substrate should then be positioned in the catalytic pocket 'below' the donor sugar ring plane to present the O atom of its deprotonated 3-hydroxyl group at the correct distance, in order to attack the C1 donor sugar atom directly and invert its anomeric configuration once more (second inversion) (Figure 6C and D). The exact mechanism responsible for deprotonation of the 3-hydroxyl group of the acceptor substrate sugar is not known at present.

### *Conclusion*

The structure of the catalytic domain of α3GalT revealed a high degree of similarity with already known inverting glycosyltransferases. Its UBD is similar to those of SpsA, GnT I and GlcAT1, and to some extent to that of β4GalT1. This domain contains the residues necessary for binding the UDP moiety of UDP-Gal, as well as the DXD motif and the $Mn^{2+}$ ion. The substrate-bound α3GalT structure identifies the invariant residue E317 as an important catalytic nucleophile. Moreover, the substrate-bound α3GalT structure contains a galactose bound covalently to E317, forming a glycosyl-enzyme intermediate, suggesting that the first step in the retention mechanism (first inversion) is very similar to the first step common to retaining glycosyl-hydrolases (Lougheed *et al*., 1999). The position of the deprotonated hydroxyl group of the acceptor substrate sugar seems critical in order to specify the second inversion. A clearer view of the α3GalT mechanism might derive from knowledge of the exact atomic position of the acceptor substrate in the presence of UDP-Gal.

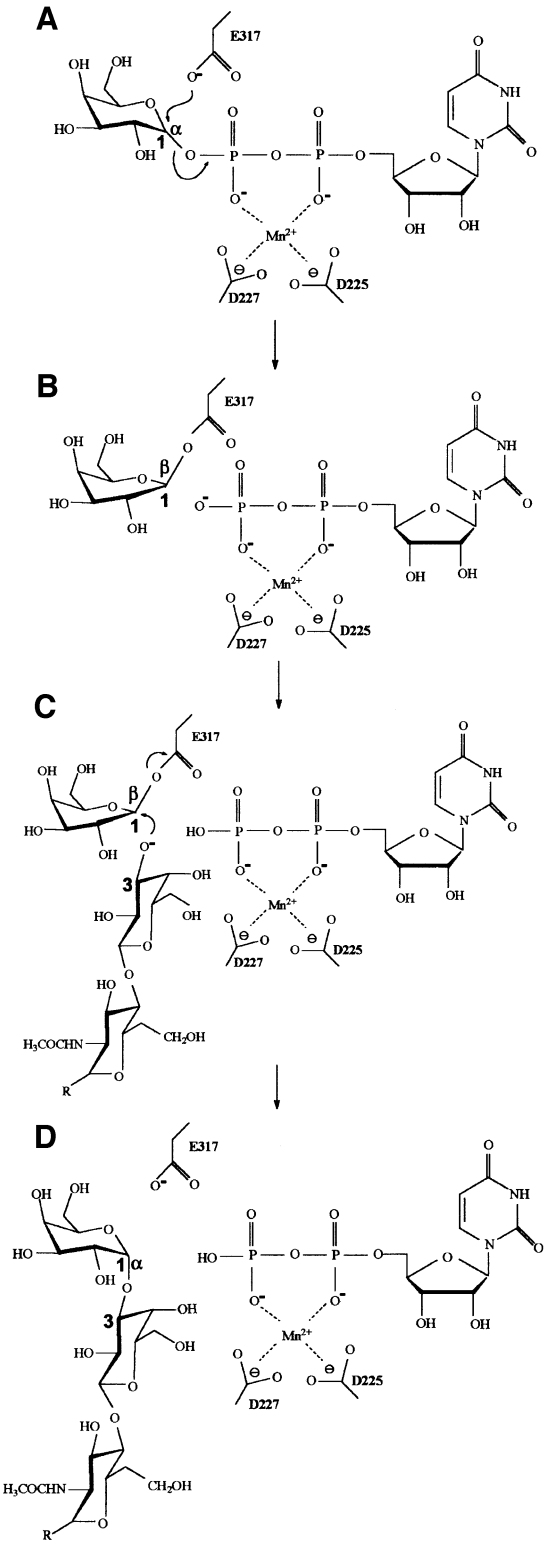



**Fig. 6.** Schematic representation of the α3GalT-retaining reaction mechanism. Steps (**A**) and (**B**) are derived from the substrate-bound α3GalT structure. The acceptor substrate schematized in steps (**C**) and (**D**) is a lactosamine-type glycan (Galβ1,4GlcNAc-R).

## Materials and methods

### *Protein expression and purification of native and seleniated α3GalT*

A truncated form of α3GalT (E80–V368), lacking the cytoplasmic domain, the transmembrane portion and part of the 'stem region', was prepared by PCR using two oligonucleotides containing an *Nde*I and a *Bam*HI restriction site at their 5′ ends for cloning purposes. The amplicon was subsequently subcloned into the expression plasmid pET15b (Novagen) at the *Nde*I and *Bam*HI sites, downstream from the plasmid polyHis tag. The N-terminal His tag α3GalT (E80–V368) fusion protein was expressed in *Escherichia coli* strain BL21($DE_3$) as described previously (Janczuk *et al*., 1999). The overexpressed protein accumulated mostly in inclusion bodies, but some soluble protein could be recovered and assayed as an active enzyme using lactose or lactosamine as the acceptor substrates and UDP-[$^3$H]galactose as the donor (data not shown). Expression of the soluble α3GalT was optimized by lowering the culture temperature to 20°C. Isopropyl-β-D-thiogalactopyranoside (IPTG) induction was omitted, because the expression was found to be constitutive in BL21($DE_3$). The soluble His-tagged α3GalT expressed by overnight, uninduced cell cultures was first purified by metal chelate Ni-NTA affinity chromatography. Elution using an imidazole gradient was followed by affinity chromatography on UDP-hexanolamine–Sepharose. The protein was eluted in the presence of 2 mM $MnCl_2$ and 10 mM UMP.

In order to solve the phase problem, we decided to label the α3GalT enzyme metabolically with selenomethionine (Hendrickson *et al*., 1990). The same plasmid construction described above was used with the *E.coli* strain B834($DE_3$) as a new host cell line and with LeMaster growth medium containing 100 μg/ml selenomethionine. The protein was expressed by induction with 0.1 mM IPTG in the host cell B834($DE_3$). Sel-α3GalT was purified as described above, but with the addition of the reducing agent TCEP (5 mM) in all buffers to prevent selenium oxidation. Biochemical integrity and biological activity of the α3GalT protein present in the crystals were assessed using the MALDI-TOF (MS) technique on protein obtained from single dissolved α3GalT crystals. A major MS peak of 35 666 Da was detected. A similar MS value was recorded upon testing the protein preparation before its crystallization. A galactose transfer activity equivalent to that of the native α3GalT enzyme was detected in dissolved crystals using lactose and lactosamine as the acceptor substrates (data not shown).

### *Crystallization, data collection, phase determination and refinement*

Purified bovine α3GalT catalytic domain was crystallized at 20°C in a 2–4 μl hanging drop containing a 1:1 mixture of protein solution (10–20 mg/ml α3GalT in 20 mM Tris pH 8.0, 2 mM $MnCl_2$, 10 mM UMP, 500 mM NaCl) and reservoir solution (1.3–1.6 M sodium acetate, 50 mM cacodylate pH 6.5). Tetragonal and bi-pyramidal shaped crystals ~0.2 × 0.2 × 0.3 mm$^3$ in size developed within 1 week.

α3GalT crystals soaked with Hg-UDP-Gal (the mercury atom is covalently bound to the uracil C5 atom; compound prepared by A.Misra and O.Hinsgaul, unpublished) were obtained by incubating α3GalT crystals with crystallization buffer containing 10 mM Hg-UDP-Gal and 10 mM $MnCl_2$ for 12 h. Before the data collection at 100 K, crystals were soaked briefly in the crystallization buffer containing 25% ethylene glycol and flash frozen in the gaseous stream of liquid nitrogen.

Seleniated α3GalT crystal data sets were recorded on the ESRF ID14-EH4 beam line (ESRF Grenoble, France). Three data sets at three different wavelengths were recorded on the same crystal flash frozen in liquid nitrogen in order to perform MAD analysis. Oscillation images were integrated with the DENZO program (Otwinowski and Minor, 1997), scaled and merged with SCALA (CCP4, 1994). The intensities were converted into structure factor amplitudes with TRUNCATE (CCP4, 1994). The crystals belonged to the tetragonal space group $P4_12_12$. One molecule of α3GalT was present in the asymmetric unit of the crystal (Table II).

High resolution data of the native crystals and α3GalT crystals soaked with Hg-UDP-Gal were recorded on the ESRF ID14-EH4 and ID14-EH2 synchrotron beam lines at cryogenic temperatures (ESRF Grenoble, France).

In the selenomethionyl α3GalT crystals, eight selenomethionines provide sufficient anomalous scattering to obtain experimental phasing information using the MAD technique (Hendrickson *et al*., 1990). The selenium atom sites were localized initially using SOLVE with $\langle m \rangle = 0.5$ and $Z$-score = 32 (Terwilliger *et al*., 1999). The initial overall figure of merit (FOM) value was 0.35–2.8 Å using 12 473 reflections. Electron density maps calculated with experimental MAD phases from SOLVE showed the presence of protein electron density signals, which were difficult to interpret at 2.8 Å resolution. Solvent flattening was performed with a 60% solvent content using DM (CCP4, 1994), which dramatically improved the quality of the electron density maps. The polypeptide chain of α3GalT was traced using the eight selenium sites as markers for the methionine side chains, which facilitated the amino acid assignment in the electron density map. An accurate secondary structure prediction using J-PRED also facilitated this assignment (Rost *et al*., 1994). A complete model from Lys82 to Thr358 was built using an experimental $2F_o - F_c$ phase combination with CNS (Brünger *et al*., 1998) and TURBO-FRODO to 2.8 Å resolution with a final $R_{cryst}$ = 23% and $R_{free}$ = 27% (Table II). The final seleniated α3GalT model includes 276 residues from Lys82 to Q358 and 43 water molecules.

The seleniated α3GalT structure solved at 2.8 Å resolution was used subsequently as a search model to solve the α3GalT non-seleniated structure, which diffracts to Bragg spacing of 2.0 Å resolution. A single molecular replacement solution was obtained using AMoRe with a $C$-factor = 76.2% and an $R$-factor = 34.7% (Navaza, 1994). Multiple model building steps were performed using TURBO-FRODO (A.Roussell and C.Cambillau, personal communication) and simulated annealing refinement using CNS was also necessary to refine the structure to 2.3 Å with a final $R_{cryst}$ = 21% and $R_{free}$ = 25% (Brünger *et al*., 1998). The final α3GalT model contained residues from Ser81 to Asn367, 130 water molecules, one UMP molecule and one $Mn^{2+}$ ion.

The α3GalT model (omitting the UMP molecule and manganese atom) solved to 2.3 Å resolution was used as a search model to solve the substrate-bound Hg-UDP-Gal α3GalT structure. A single molecular replacement solution was obtained using AMoRe and multiple model building steps were performed with simulated annealing refinement using CNS to refine the substrate-bound structure at 2.5 Å resolution with a final $R_{cryst}$ = 22% and $R_{free}$ = 27%.

The stereochemistry of each model (Table II) was analyzed with PROCHECK (Laskowski *et al*., 1993) and WHATIF (Hooft *et al*., 1996). Figures 1 and 2A were generated using MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt *et al*., 1994), and Figures 2B and 5 with TURBO-FRODO (A.Roussel and C.Cambillau, personal communication). Figures 3 and 4G were prepared with ALSCRIPT (Barton, 1993) using ClustalX as the alignment program. Figure 6 was prepared using Chemwindows.

## Acknowledgements

We would like to thank G.Davies for access to the SpsA coordinates prior to public release. Access to the synchrotron source at ESRF (ID14-EH4) is deeply appreciated. The authors thank R.Ravelli and the ID14-EH4 staff for expert technical assistance during the data collection, J.Bonicel for performing MALDI-TOF analysis, C.Cambillau for support during the initial phase of the work, B.Henrissat, A.Roussel and K.Brown for helpful discussions, and J.Blanc for correcting the English manuscript. This study was supported by the Mizutani Foundation for Glycosciences (L.N.G.) and the CNRS. The coordinates of the bovine α1,3GalT catalytic domain structure have been deposited in the Protein Data Bank (1fg5.pdb, 1G8O.pdb and 1G93.pdb).

## References

Barton,G.J. (1993) ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.*, **6**, 37–40.

Blanken,W.M. and Van den Eijnden,D.H. (1985) Biosynthesis of terminal Galα1→3Galβ1→4GlcNAc-R oligosaccharide sequences on glycoconjugates. Purification and acceptor specificity of a UDP-Gal:*N*-acetyllactosaminide α1,3-galactosyltransferase from calf thymus. *J. Biol. Chem.*, **260**, 12927–12934.

Brünger,A.T. *et al*. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.

CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.

Charnock,S.J. and Davies,G.J. (1999) Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry*, **38**, 6380–6385.

Cooper,D.K.C. (1998) Xenoantigens and xenoantibodies. *Xenotransplantation*, **5**, 6–17.

Galili,U., Clark,M.R., Shohet,S.B., Buehler,J. and Macher,B.A. (1987) Evolutionary relationship between the anti-Gal antibody and the Galα1,3Gal epitope in primates. *Proc. Natl Acad. Sci. USA*, **84**, 1369–1373.

Galili,U., Shohet,S.B., Kobrin,E., Stults,C.L. and Macher,B.A. (1988) Man, apes and Old World monkeys differ from other mammals in the expression of α-galactosyl epitopes on nucleated cells. *J. Biol. Chem.*, **263**, 17755–17762.

Gastinel,L., Cambillau,C. and Bourne,Y. (1999) Crystal structures of the bovine β4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *EMBO J.*, **18**, 3546–3557.

Hakomori,S. (1999) Antigen structure and genetic basis of histo-blood groups A, B and O: their changes associated with human cancer. *Biochim. Biophys. Acta*, **1473**, 247–266.

Haslam,D.B. and Baezinger,J.U. (1996) Expression cloning of Forssman glycolipid synthetase: a novel member of the histo-blood group ABO gene family. *Proc. Natl Acad. Sci. USA*, **93**, 10697–10702.

Hendrickson,W.A., Horton,J.R and LeMaster,D.M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.*, **9**, 1665–1672.

Henion,T.R., Macher,B.A., Anaraki,F. and Galili,U. (1994) Defining the minimal size of catalytically active primate α1,3galactosyltransferase: structure–function studies on the recombinant truncated enzyme. *Glycobiology*, **4**, 193–201.

Holm,L. and Sander,C. (1983) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Holm,L. and Sander,C. (1995) Evolutionary link between glycogen phosphorylase and a DNA modifying enzyme. *EMBO J.*, **14**, 1287–1293.

Hooft,R.W.W., Vriend,G., Sander,C. and Abola,E.E. (1996) Errors in protein structures. *Nature*, **381**, 272–276.

Hutchinson,E.G. and Thornton,J.M. (1996) PROMOTIF—a program to identify and analyse structural motifs in proteins. *Protein Sci.*, **5**, 212–220.

Imberty,A., Monier,C., Bettler,E., Morera,S., Freemont,P., Sippl,M., Flockner,H., Ruger,W. and Breton,C. (1999) Fold recognition study of α3-galactosyltransferase and molecular modeling of the nucleotide sugar-binding domain. *Glycobiology*, **9**, 713–722.

Janczuk,A., Li,J., Zhang,W., Chen,W., Chen,Y., Fang,J., Wang,J. and Wang,P.G. (1999) α-Gal oligosaccharides: chemistry and potential biomedical application. *Curr. Med. Chem.*, **6**, 155–164.

Joziasse,D.H. (1992) Mammalian glycosyltransferases: genomic organization and protein structure. *Glycobiology*, **2**, 271–277.

Joziasse,D.H. and Oriol,R. (1999) Xenotransplantation: the importance of the Galα1,3Gal epitope in hyperacute vascular rejection. *Biochim. Biophys. Acta*, **1455**, 403–418.

Joziasse,D.H., Shaper,J.H., Van den Eijnden,D.H., Van Tunen,A.J and Shaper,N.L. (1989) Bovine α1,3galactosyltransferase: isolation and characterization of a cDNA clone. *J. Biol. Chem.*, **264**, 14290–14297.

Joziasse,D.H., Shaper,J.H., Wang,J.E. and Shaper,N.L. (1991) Characterization of an α1,3galactosyltansferase homologue on human chromosome 12 that is organized as a processed pseudogene. *J. Biol. Chem.*, **266**, 6991–6998.

Joziasse,D.H., Shaper,N.L., Kim,D., van den Eijnden,D.H. and Shaper,J.H. (1992) Murine α1,3-galactosyltransferase. A single gene locus specifies four isoforms of the enzyme by alternative slicing. *J. Biol. Chem.*, **267**, 5534–5541.

Keusch,J., Manzella,S.M., Nyame,K.A., Cummings,R. and Baenziger, J.U. (2000) Expression cloning of a new member of the ABO blood group glycosyltransferases, $iGB_3$ synthase, that directs the synthesis of isoglobo-glycosphingolipids. *J. Biol. Chem.*, **275**, 25308–25314.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Larsen,R.D., Rajan,V.P., Ruff,M.M., Kukowska-Latallo,J., Cummings, R.D. and Lowe,J.B. (1989) Isolation of a cDNA encoding murine UDPgalactose:β-D-galactosyl-1,4-*N*-acetyl-D-glucosaminide α-1,3-galactosyltransferase: expression cloning by gene transfer. *Proc. Natl Acad. Sci. USA*, **86**, 8227–8231.

Larsen,R.D., Rivera-Marrero,C.A., Ernst,L.K., Cummings,R.D. and Lowe,J. (1990) Frameshift and nonsense mutations in a human genomic sequence homologous to a murine UDP-Gal:β-D-Gal(1,4)-D-GlcNAc α(1,3)-galactosyltransferase cDNA. *J. Biol. Chem.*, **265**, 7055–7061.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lougheed,B., Ly,H.D., Wakarchuk,W.W. and Withers,S.G. (1999) Glycosyl fluorides can function as substrates for nucleotide phosphosugar-dependent glycosyltransferases. *J. Biol. Chem.*, **274**, 37717–37722.

Merritt,E.A. and Murphy,M. (1994) Raster3d version 2.0—a program for photorealistic molecular graphics. *J. Appl. Crystallogr.*, **D50**, 869–873.

Morera,S., Imberty,A., Aschke-Sonnenborn,U., Ruger,W. and Freemont, P.S. (1999) T4 phage β-glucosyltransferase: substrate binding and proposed catalytic mechanism. *J. Mol. Biol.*, **292**, 717–730.

Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157–163.

Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

Pedersen,L.C., Tsuchida,K., Kitagawa,H., Sugahara,K., Darden,T.A. and Negishi,M. (2000) Heparan/chondroitin sulfate biosynthesis: structure and mechanism of human glucuronyltransferase I. *J. Biol. Chem.*, **275**, 34580–34585.

Rost,B. and Sander,C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*, **19**, 55–72.

Seto,N.O., Compston,.A., Evans,S.V., Bundle,D.R., Narang,S.A. and Palcic,M.M. (1999) Donor substrate specificity of recombinant human blood group A, B and hybrid A/B glycosyltransferases expressed in *Escherichia coli*. *Eur. J. Biochem.*, **259**, 770–775.

Strahan,K.M., Gu,F., Preece,A.F., Gustavsson,I., Andersson,L. and Gustafsson,K. (1995) cDNA sequence and chromosome localization of pig α1,3galactosyltransferase. *Immunogenetics*, **41**, 101–105.

Takayama,S., Chung,S.J., Igarashi,Y., Sepp,A., Lechler,R.I., Wu,J., Hayashi,T., Siuzdak,G. and Wong,C.H. (1999) Selective inhibition of β-1,4- and α-1,3-galactosyltransferases: donor sugar-nucleotide based approach. *Bioorg. Med. Chem.*, **7**, 401–409.

Terwilliger,T.C and Berendzen,J. (1999) Automated structure solution for MIR and MAD. *Acta Crystallogr. D*, **55**, 849–861.

Unligil,U.M., Zhou,S., Yuwaraj,S., Sarkar,M., Schachter,H. and Rini,J.M. (2000) X-ray crystal structure of rabbit *N*-acetylglucosaminyltransferase I: catalytic mechanism and a new glycosyltransferase superfamily. *EMBO J.*, **19**, 5269–5280.

Vrielink,A., Rüger,W., Dreissen,H.P.C. and Freemont,P.S. (1994) Crystal structure of the DNA modifying enzyme β-glucosyltransferase in the presence or in the absence of the substrate uridine diphosphoglucose. *EMBO J.*, **13**, 3413–3422.

Wiggins,C.A.R. and Munro,S. (1998) Activity of the yeast MNN1 α1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc. Natl Acad. Sci. USA*, **95**, 7945–7950.

Xu,H., Storch,T., Yu,M., Elliott,S.P. and Haslam,D.B. (1999) Characterization of the human Forssman synthetase gene. An evolving association between glycolipid synthesis and host–microbial interactions. *J. Biol. Chem.*, **274**, 29390–29398.

Yamamoto,F. and Hakomori,S. (1990) Sugar–nucleotide donor specificity of histo-blood group A and B transferases is based on amino acid substitutions. *J. Biol. Chem.*, **265**, 19257–19262.

Yamamoto,F.-I., Clausen,H., White,T., Marken,J. and Hakomori,S.-I. (1990) Molecular genetic basis of the histo-blood group ABO system. *Nature*, **345**, 229–233.

Yamamoto,F., McNeill,P.D. and Hakomori,S. (1995) Genomic organization of human histo-blood group ABO genes. *Glycobiology*, **5**, 51–58.

Yen,T.-Y., Joshi,R.K., Yan,H., Seto,N.O.L., Palcic,M.M. and Macher,B.A. (2000) Characterization of cysteine residues and disulfide bonds in proteins by liquid chromatography/electrospray ionization-tandem mass spectrometry. *J. Mass Spectrom.*, **35**, 990–1002.

Zarembinski,T.I., Hung,L.H., Mueller-Dieckmann,H.-J., Kim,K.-K., Yokota,H., Kim,R. and Kim,S.-H. (1998) Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc. Natl Acad. Sci. USA*, **95**, 15189–15193.

*Received August 21, 2000; revised and accepted December 19, 2000*