The EMBO Journal Vol.18 No.13 pp.3546–3557, 1999

Crystal structures of the bovine β4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose

Louis Noël Gastinel[1], Christian Cambillau and Yves Bourne

Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS, UPR 9039, 31 Chemin Joseph Aiguier 13402 Marseille Cedex 20, France

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

β1,4-galactosyltransferase T1 (β4Gal-T1, EC 2.4.1.90/38), a Golgi resident membrane-bound enzyme, transfers galactose from uridine diphosphogalactose to the terminal β-*N*-acetylglucosamine residues forming the poly-*N*-acetyllactosamine core structures present in glycoproteins and glycosphingolipids. In mammals, β4Gal-T1 binds to α-lactalbumin, a protein that is structurally homologous to lyzozyme, to produce lactose. β4Gal-T1 is a member of a large family of homologous β4galactosyltransferases that use different types of glycoproteins and glycolipids as substrates. Here we solved and refined the crystal structures of recombinant bovine β4Gal-T1 to 2.4 Å resolution in the presence and absence of the substrate uridine diphosphogalactose. The crystal structure of the bovine substrate-free β4Gal-T1 catalytic domain showed a new fold consisting of a single conical domain with a large open pocket at its base. In the substrate-bound complex, the pocket encompassed residues interacting with uridine diphosphogalactose. The structure of the complex contained clear regions of electron density for the uridine diphosphate portion of the substrate, where its β-phosphate group was stabilized by hydrogen-bonding contacts with conserved residues including the Asp252ValAsp254 motif. These results help the interpretation of engineered β4Gal-T1 point mutations. They suggest a mechanism possibly involved in galactose transfer and enable identification of the critical amino acids involved in α-lactalbumin interactions.

Keywords: crystallography/enzyme/glycosylation/glycosyltransferase/nucleotide-binding protein

Introduction

Almost every organism contains a broad range of glycans, which are linked to lipids and proteins via either an N-type linkage with the amide group of the Asn residue contained in the Asn-X-Ser/Thr consensus sequence or an O-type linkage with the hydroxyl group of Ser or Thr residues. These oligosaccharides influence the biological activity of the glycoproteins that carry them and their clearance. They are synthesized in eukaryotic cells by several different glycosyltransferases and glycosidases that are clustered in an assembly line located in the endoplasmic reticulum and the Golgi apparatus. As many as a few hundred different glycosyltransferases and their corresponding genes account for the synthesis of the large variety of oligosaccharide structures observed at the surface of glycoproteins and glycolipids in mammalian species. The first glycosyltransferase to be cloned was the UDP-Gal:GlcNAc-R β1,4-galactosyltransferase (β4galactosyltransferase T1 or β4Gal-T1) (Narimatsu *et al*., 1986), and this enzyme is one of those that have been most thoroughly characterized to date. Meanwhile, more than 30 different mammalian glycosyltransferase genes have been cloned and classified (Campbell *et al*., 1997; Breton *et al*., 1998). Although they have very different primary amino acid sequences, these type II (i.e. cytoplasmic N-terminus) endoplasmic reticulum and Golgi resident membrane-bound glycoprotein enzymes have a similar pattern of domain organization (Paulson and Colley, 1989). All the eukaryotic glycosyltransferases cloned to date consist of a lumen-oriented C-terminus containing the catalytic domain, followed by an extended region called 'the stem region', a single membrane-spanning region and a short cytoplasmic N-terminus. Secreted forms of glycosyltransferases are produced by proteolysis in the Golgi apparatus at multiple protease-sensitive positions within the stem region of the protein. The resulting soluble enzymes are responsible for the glycosyltransferase activities detected in milk, serum and saliva (Joziasse, 1992).

The bovine β4Gal-T1 is a *trans*-Golgi resident, type II membrane-bound glycoprotein which, in the presence of manganese acting as the cofactor, catalyses the transfer of galactose (Gal) from uridine diphosphogalactose (UDP-Gal) to terminal β-*N*-acetylglucosamine (GlcNAc) residues forming the poly-β4-*N*-acetyllactosamine core structures present in glycoproteins and glycosphingolipids (Stous, 1986; Van den Eijnden and Joziasse, 1993). This 'housekeeping' activity of β4galactosyltransferases occurs widely among both mammalian and non-mammalian vertebrates, including avians (Shaper *et al*., 1997), and has also been found to occur in some plants (Powell and Brew, 1974). Murine β4Gal-T1 exists in two isoforms with similar types of catalytic activities, a long β4Gal-T1 (LGT) having an additional 13 amino acids in the N-terminal portion and a short β4Gal-T1 (SGT) isoform lacking this portion (Shaper *et al*., 1988).

One unique feature of the β4Gal-T1 in comparison with the other glycosyltransferases is the fact that a modulation of its activity contributes to the synthesis of lactose (Brodbeck *et al*., 1967). In mammals, the synthesis of lactose (Galβ1,4Glc) involves the modulation of the activity of β4Gal-T1 as a result of its tissue-specific association with α-lactalbumin, a non-catalytic protein which shows structural homologies with lysozyme (Morrison and Ebner, 1971; Pike *et al*., 1996). The association between β4Gal-T1

and α-lactalbumin has a low affinity and results in a decrease of the K_m for glucose from 1 M to 1 mM.

Bovine β4Gal-T1 was recently found to be a member of a large homologous gene family containing at least six human members, in which β4Gal-T1 (Narimatsu *et al*., 1986) -T2, -T3 and -T4 transfer Gal to GlcNAc in various glycoconjugates (Almeida *et al*., 1997; Schwientck *et al*., 1998), while -T5 (Lo *et al*., 1998) and -T6 (Nomura *et al*., 1998) transfer Gal to glucose (Glc) in glucosylceramide. Several short conserved sequence motifs, possibly involved in the donor and acceptor binding sites, have been identified among the human enzymes and among β4GalT orthologues in various animal enzymes (Van Die *et al*., 1997; Breton *et al*., 1998; Wiggins and Munro, 1998). These motifs were also conserved in the UDP-GlcNAc:βGlcNAc β1-4-*N*-acetylglucosaminyltransferase isolated from a snail (Bakker *et al*., 1994, 1997). Furthermore, four cysteine residues are conserved between the six human β4galactosyltransferase members and the bovine β4Gal-T1.

Moreover, in mice, in addition to the 'housekeeping' Golgi resident activity, β4galactosyltransferase activity also occurs on the cell surface of spermatozoids and is thought to be involved in diverse cellular functions including fertilization, embryonic cell migration and cell adhesion (Shur *et al*., 1998). Mice lacking β4Gal-T1 activity were engineered using the gene knockout technique and showed severe dermatosis, immune and neurological deficiencies and a lack of lactose in their milk (Asano *et al*., 1997; Lu *et al*., 1997).

Recombinant bovine β4Gal-T1 has been expressed in *Escherichia coli* in the form of an inclusion body, and an active enzyme has been recovered after *in vitro* renaturation (Boeggeman *et al*., 1993, 1995). In these studies, both the transfer to GlcNAc and the lactose synthase activities remained intact in the truncated enzyme lacking the 129 N-terminal amino acid residues, but the two activities were lost when the deletion extended to residue 142 and when the cysteine 134 was mutated into alanine or serine. The results of site-directed mutagenesis performed on the human β4Gal-T1, expressed in low amounts as an active enzyme exported into the periplasm of *E.coli*, suggested that Tyr284, Tyr309 and Trp310 might be important residues for GlcNAc binding (Aoki *et al*., 1990). A region spanning from amino acids Phe305 to Tyr309 was also identified as being necessary for catalytic activity, along with the Asn307 and Asn308 side-chains possibly constituting part of the UDP-galactose binding site (Zu *et al*., 1995).

Altered galactosyltransferase enzyme activity may be associated with some diseases: (i) an increase in β4Gal-T1 activity has been reported to occur in NIH 3T3 cell lines transfected with N-*ras*-proto-oncogene under the control of a glucocorticoid-inducible mouse mammary tumor virus promoter (Easton *et al*., 1991); (ii) anarchic galactosylation has been found to be associated with some forms of cancer (Kobata, 1989; Delves, 1998; Udagawa *et al*., 1998); (iii) a decrease in the galactosylation of the oligosaccharide structures at the Asn297 N-linked consensus sequence in the CH_2 domain of IgG has been found to occur in rheumatoid arthritis (RA) (Parekh *et al*., 1988; Jeddi *et al*., 1996); (iv) a decrease in β4Gal-T1 activity has been reported to occur in some RA patients (Axford *et al*., 1987; Furukawa *et al*., 1990), although no disease-associated polymorphisms of β4Gal-T1 have been described; and (v) the absence of galactose in the oligosaccharide present in immunoglobulins decreases the FcγR1 and C1q binding and may also play a pathogenic role in autoimmune diseases (Tsuchiya *et al*., 1989; Rademacher *et al*., 1994).

Like other recombinant glycosyltransferases, β4Gal-T1 has been used in the enzymatically assisted chemical synthesis of complex polysaccharides required for preparing massive quantities of the new oligosaccharides now being used as drugs, which have great biomedical potential (Guo and Wang, 1997).

Since it is very difficult to prepare large amounts of homogeneous, stable and soluble glycosyltransferases, no structural data are yet available on any eukaryotic glycosyltransferases, particularly the β4Gal-T1 enzyme protein family. The only glycosyltransferase structure solved to date is that of the β-glucosyltransferase of T4 bacteriophage (Vrielink *et al*., 1994). This enzyme catalyses the glucosylation of DNA by transferring the glucose moiety from UDP-glucose to DNA hydroxymethylated cytosines, thus protecting its own genome from the host nucleases. Various attempts have been made to use the T4 β-glucosyltransferase structure as a template for predicting the structure of other mammalian glycosyltransferases (Breton *et al*., 1996; Imberty *et al*., 1999). No biochemical and biological data confirming these fold predictions have yet been published, however.

Here we have determined the crystal structures of the bovine β4Gal-T1 catalytic domain alone (substrate-free) and in the presence of its substrate UDP-Gal (substrate-bound) at 2.4 Å resolution, in order to provide a scaffold for describing its glycosyltransfer mechanism as well as for designing structure-based drugs specifically modulating the *N*-glycosylation biosynthesis of glycoproteins.

Results and discussion

Quality of the models

Crystals of substrate-free β4Gal-T1 belong to space group $C222_1$, have unit cell dimensions a = 108.5 Å, b = 161 Å, c = 107.4 Å and contain two molecules of β4Gal-T1 in the asymmetric unit. The final substrate-free model comprises β4Gal-T1 residues Leu131 to Ser402. High temperature factors and weak electron density maps were obtained for a surface loop containing residues His347 to Glu354. The N-terminal portion Pro114 to Ser130 was not visible in the electron density maps. The final electron density maps unambiguously feature valine, proline, isoleucine and arginine residues at the respective positions 158, 187, 256 and 265, which were marked as conflicting in the Swiss-Prot entry for bovine β4Gal-T1, P08037. The substrate-free structure which was refined with R_{cryst} = 22.2% and with R_{free} = 27% in the resolution range 30–2.4 Å, includes 126 water molecules and the electron density unambiguously showed the residual density of uridine monophosphate (UMP). The substrate-bound structure comprises β4Gal-T1 residues Ala133 to Ser402 and includes 77 water molecules as well as the uridine diphosphate portion of UDP-Gal and has R_{cryst} = 25% with R_{free} = 28.6% in the resolution range 30–2.4 Å. The galactose residue was only very weak in the final electron density map and was therefore not included in the last

Table I. Data collection and refinement statistics

	Substrate-free	Mercury(II) acetate	Substrate-free cryo	Substrate-bound cryo
Resolution (Å)[a]	13.0–3.0	13.0–2.9	30.0–2.4	30.0–2.4
No. of unique reflections	19024	21016	32979	32664
Completeness (%)	99	99	93 (95)	93 (94)
$(I/\sigma_{(I)})$	5.7 (1.8)	5.6 (2.0)	7.2 (1.7)	6.7 (2.1)
R_{sym} (%)[b]	12.6 (41)	10 (38)	7.5 (36)	6.5 (35)
Resolution for the refinement (Å)			30.0–2.4	30.0–2.4
R_{cryst} (%)/R_{free} (%)			22.2/27	25/28.6
R.m.s.d. (bonds) (Å)			0.013	0.014
R.m.s.d. (angles) (°)			1.7	1.9
No. of atoms:				
protein/water/cofactor[c]			4441/126/21	4418/77/25
Average *B*-factor (Å^2):				
protein/water/cofactor[c]			54/46/95	57/53/66
No. of ϕ/ψ angles (%):				
most favoured and allowed			98.9	99.1

The values in parentheses refer to data in the high resolution shell.
[a]Wavelength 1.5418 Å for native and Mercury data; 0.932 Å for substrate-free cryo and substrate-bound cryo (crystals soaked with UDP-Gal).
[b]$R_{sym} = \Sigma_{hkl} \Sigma_i | I_i - (I) | / \Sigma (I)$
$R_{cryst} = \Sigma (|| Fp(obs) | - | Fp(calc) ||) / \Sigma |Fp(obs)|$ and R_{free} = *R*-factor for a randomly selected subset (3%) of data that were not used to minimize the crystallographic residual.
[c]Cofactor: UMP in the case of substrate-free structure and UDP-Gal in the case of substrate-bound structure.

refinement steps. Table I summarizes the data collection and refinement statistics on each of the final models.

Description of the overall structure

Figure 1A and B shows a ribbon diagram of the overall structure of substrate-free β4Gal-T1 catalytic domain containing 20% β-strands, 23% α-helices and 2% 3_{10} helices based on the PROMOTIF program (Hutchinson and Thornton, 1996). The β4Gal-T1 catalytic domain, residues Leu131 to Ser402, is composed of 11 β-strands (β1–β11), six α-helices (α1–α6) and two short 3_{10} helices with overall dimensions of 52×50×42 Å^3. The central part of the structure consists of an eight-stranded mixed twisted β-sheet with strand order 7(81)32465 surrounded by two α-helices on the one side and by four α-helices on the other (Figure 1A and C). Starting at the N-terminus, there is a long unstructured region followed by an exposed α-helical hairpin (α1β1) that enters the central β-sheet at strand β2 located in the middle of the twisted β-sheet. A large C-terminal region, which starts after the β-strand β8, consists of an α-helical region (α4,α5) and a short β-strand (β9) that completes a small three-stranded anti-parallel β-sheet above the central β-sheet, an α-helix α6 with a 3_{10} helix and a long β-hairpin (β10β11). A structural homology search using DALI databases showed that the overall fold of β4Gal-T1 is different from any fold previously determined (Holm and Sander, 1983), indicating that no structural homology was apparent with the bacteriophage T4 DNA modifying enzyme β-glucosyltransferase structure (Vrielink *et al*., 1994). The overall molecular surface of β4Gal-T1 can be compared to a cone containing a large deep open pocket 13 Å in diameter at its base.

The N-terminus part of the protein from Ala133 to Cys176, containing the α-helix α1 and the β-strand β1, is stabilized by a disulfide bridge between Cys134 and Cys176, whereas the second disulfide bridge between Cys247 and Cys266 is located in the middle of the central twisted eight-stranded β-sheet, and maintains the rigidity of the core between the β-strand β4 and the β5–β6 loop (Figure 1A and B). The C-terminus portion of the molecule containing β-strands β10 and β11 is stabilized by the hydrophobic stacking interactions between Trp312 and Pro401 side-chains. In previous studies, in which protein chemical modifications and site-directed mutagenesis experiments were performed, a single disulfide bridge was identified between Cys134 and Cys247 (Yadav and Brew, 1990, 1991; Wang *et al*., 1994). This is inconsistent with the distance of 16 Å existing between these two residues in our structure. The only free cysteine, Cys342, which was previously described as a critical amino acid for UDP-Gal binding, is not conserved in any of the members of the β4galactosyltransferase family (Almeida *et al*., 1997). Its side-chain is buried in the structure, 20 Å from the UDP-Gal binding site, which suggests that Cys342 is unlikely to play a direct role in the catalysis, as was previously thought to occur (Yadav and Brew, 1990, 1991).

UDP-galactose binding and catalytic pocket of β4Gal-T1

The substrate-bound complex structure was obtained using β4Gal-T1 crystals soaked overnight in mother liquor containing fresh UDP-galactose in the presence of manganese. Difference electron density maps showed clear regions of electronic density corresponding to the uridine diphosphate portion of the substrate, and showed weak density patterns of the galactose residue. However, it was not possible to determine from the electron density maps whether galactose is present in a disordered conformation, or whether the galatose may have been cleaved off during the soaking period. The uridine diphosphate portion of UDP-Gal binds in an extended position along the diameter of the surface pocket (Figures 2 and 3). The uracyl ring is held in position by a stacking interaction between the Phe226 side-chain and the Arg191 side-chain and by hydrogen-bonding contacts between O_2 and N_3 atoms of the uracyl ring and the main chain nitrogen and oxygen atoms of Arg189 residue, respectively (Figure 2A and B). The O_3 oxygen atom of the ribose ring is hydrogen-bonded to the main chain nitrogen atom of Val253 residue

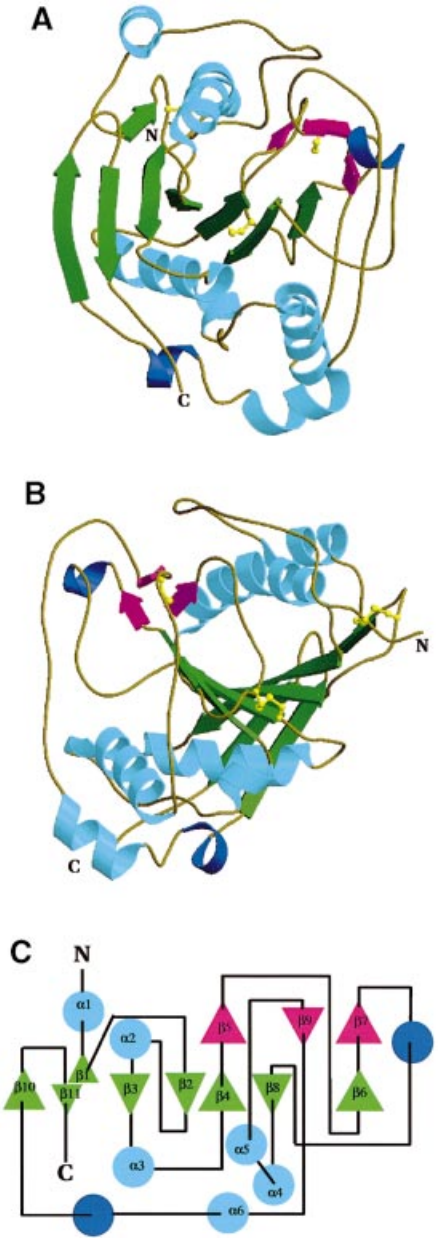

Fig. 1. Overall view of the substrate-free bovine β4Gal-T1 catalytic domain structure. (**A**) Ribbon diagram of the molecule viewed down the open pocket. The cysteine residues engaged in disulfide bridges, Cys134–Cys176, Cys247–Cys266 and the free Cys342 are shown in ball-and-stick form in yellow. (**B**) Ribbon diagram of the molecule viewed 90° from (A). (**C**) Topological representation of the β4Gal-T1 fold. N and C indicate the N- and C-termini of the molecule. Secondary structure elements are colour-coded as α-helices in light blue, 3_{10} helices in dark blue, the β-strands of the central twisted eight-stranded β-sheet in green, and the β-strands of the small three-stranded antiparallel β-sheet in cyan.

and located within hydrogen-bond distance from the carboxyl group side-chain of Asp252 and from the main chain oxygen atom of Pro187 (Figure 2C). The ribose ring O_2 oxygen atom is hydrogen-bonded to the main chain oxygen atom of Pro187 residue. Out of the two phosphate groups of the uridine diphosphate, only the β-phosphate group is apparently stabilized by a hydrogen-bonding contact with the ε-amino group of Lys279 residue and by a strong interaction mediated by a water molecule with the carboxyl group belonging to the Asp254 side-chain. The α-phosphate group of uridine diphosphate is solvent accessible and protrudes towards the external side of the pocket (Figure 2A and B).

The 29 residues lining the pocket include five polar, 10 ionizable, nine hydrophobic side-chains and five glycines (Figure 3). The majority of these residues are included in short conserved sequence motifs which were previously identified on the basis of sequence alignments (Barton, 1993) of the different members of the β4galactosyltransferase family (Bakker *et al*., 1994, 1997; Van Die *et al*., 1997; Breton *et al*., 1998; Wiggins *et al*., 1998) (Figure 4). At the bottom of the pocket, the side-chain of Arg228 (FNR228A) forms salt bridges with Asp252 (D252VD), Glu317 and Asp318 (WGWGGE317DD) side-chains, and forms hydrogen bonds with the Gly292 carbonyl oxygen, thus creating a large network of polar interactions in this region of the protein. The substrate-bound β4Gal-T1 structure brought to light a new motif PF(H)R189XR, lying close to the F226NRA motif; these two motifs together line the side of the pocket and are both involved in UDP-Gal binding (Figure 3A). The second motif, D252VD254, is located at the bottom of the pocket along with the side-chains of two of its Asp residues, Asp252 and Asp254, which are directed toward the surface and 5 Å apart. In the third signature sequence, WGWGGE317DDD, the WGWGG portion is located at the periphery of the pocket, while the cluster of negatively charged residues E317DDD is located at the bottom (Figure 3A). The C_α atoms of two glycine residues in this conserved sequence, Gly313 and Gly315, are exposed to the pocket surface, suggesting that any side-chain here would cause steric interference with the substrate binding. The DK351KN motif, which was identified as the unique common motif between β4Gal-T1 sequence and the bovine α1,3-galactosyltransferase sequence (Joziasse *et al*., 1989), is located in the only flexible and disordered surface loop in our structure, His347–Glu354. This motif, which lies above the pocket ~16 Å away from Arg228, increases the pocket surface area (Figure 3A, B and C). The residues of this loop are not strictly conserved in the β4galactosyltransferase family. The positive electrostatic potential of this surface loop may counteract the negative electrostatic potential produced by the DVD and EDDD motifs (Figure 3D, E and F), but no particular role for this surface loop can be deduced from the present structure.

Comparisons between the substrate-free and the substrate-bound β4Gal-T1 structures

Comparisons between the substrate-free and the substrate-bound structures did not show the existence of significant differences with a deviation value between the main chain C_α atoms of 0.25 Å. However, the substrate-free β4Gal-T1 structure was unambiguously found to contain a residual UMP molecule. UMP was used to elute specifically β4Gal-T1 protein bound to UDP-hexanolamine–Sepharose during its purification. Surprisingly, UMP interacts with β4Gal-T1 in a completely different way from UDP-Gal in the substrate-bound structure (Figure 5). The uracyl ring of UMP is held in position by the hydrophobic stacking interaction occurring between the Phe226 side-chain and Arg191 side-chain, but rotates almost 180° in comparison with the position taken by the uracyl ring of uridine diphosphate bound to β4Gal-T1. The N_3 nitrogen

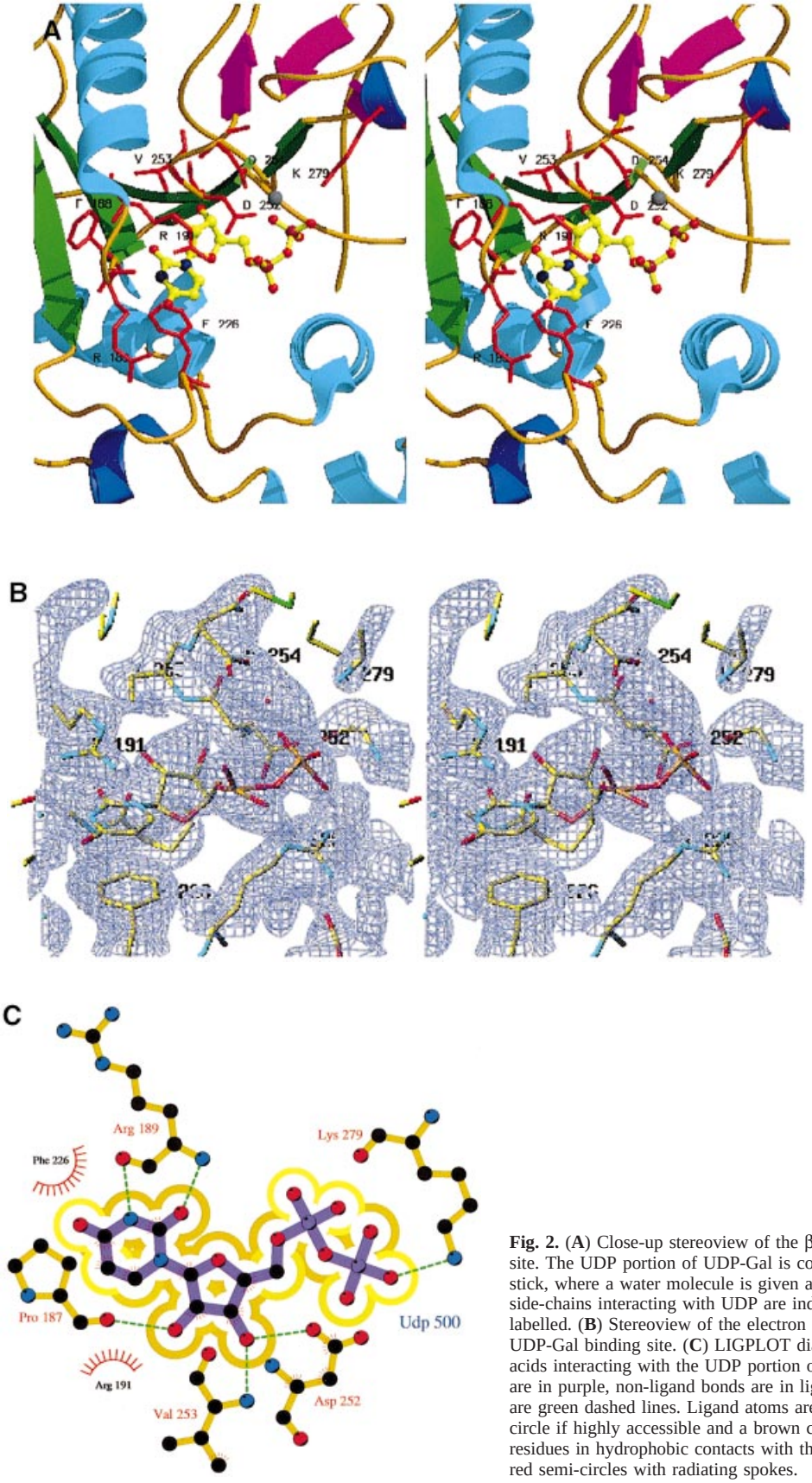

Fig. 2. (**A**) Close-up stereoview of the β4Gal-T1 substrate binding site. The UDP portion of UDP-Gal is coloured yellow in ball-and-stick, where a water molecule is given as a grey sphere. Amino acid side-chains interacting with UDP are indicated by red sticks and labelled. (**B**) Stereoview of the electron density omit map (2σ) of the UDP-Gal binding site. (**C**) LIGPLOT diagram of β4Gal-T1 amino acids interacting with the UDP portion of UDP-Gal. Ligand bonds are in purple, non-ligand bonds are in light brown, hydrogen bonds are green dashed lines. Ligand atoms are surrounded by a yellow circle if highly accessible and a brown circle if buried. Non-ligand residues in hydrophobic contacts with the ligand are presented by red semi-circles with radiating spokes.

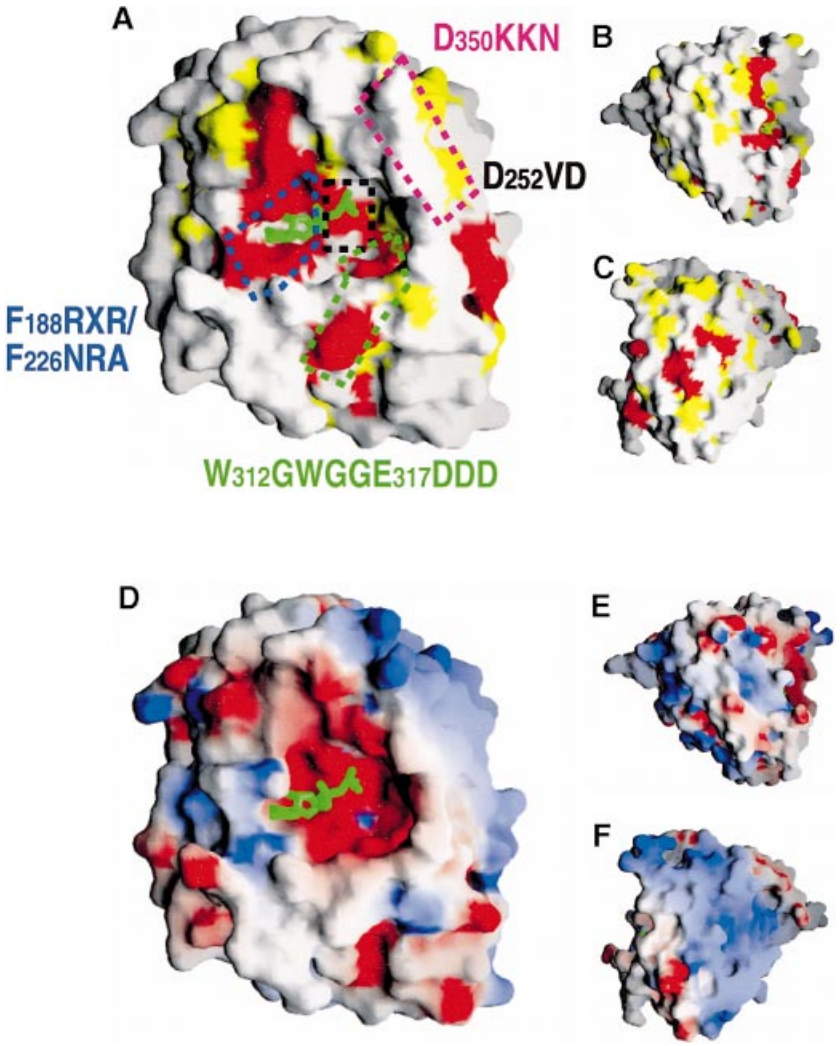

Fig. 3. β4Gal-T1 biologically important regions, catalytic pocket and surface electrostatic potential. (**A**) The β4Gal-T1 molecular surface is oriented to match Figure 1A, with the invariant residues coloured red, the conserved residues yellow, and the non-conserved residues white, as defined in Figure 4. A large patch of sequence invariant surface residues lie inside and around the pocket. The UDP portion of UDP-Gal is shown in ball-and-stick form in green. The topography of the conserved motifs is colour-coded black, blue, green and magenta for the DVD, F(H)RXR/FNRA, WGWGGEDDD and DKKN motifs, respectively (dashed boxes). (**B**) and (**C**) are β4Gal-T1 molecular surfaces oriented +90° and –90° from (A). (**D**), (**E**) and (**F**) are electrostatic potential mapped onto the molecular surface from –6 kT to +6 kT in the orientation of (A), (B) and (C), respectively. Red is negative, blue positive and white uncharged or hydrophobic. The bottom of the pocket shows the most electronegative surface potential (center), with discrete surrounding patches of electropositive potential. UDP is coloured green in a ball-and-stick form.

atom of the uracil ring is hydrogen-bonded to the main chain oxygen atom of Arg189 residue. As a result, the ribose ring and the α-phosphate group of UMP point outside the catalytic pocket with the ribose ring O_4 oxygen atom at a hydrogen-bonded distance from the nitrogen atom of the Arg189 side-chain (Figure 5A). The α-phosphate group no longer interacts with the Asp253VAsp motif, but is located 15 Å away and is hydrogen-bonded with nitrogen atoms of the Arg191 and the Asn190 side-chains, respectively (Figure 5B). An electrostatic repulsion may occur between the negative charge of the α-phosphate group and the two Asp side-chains of the Asp253VAsp254 motif, resulting in the reorientation of the uracyl ring of UMP, the biological significance of which still remains to be determined.

Galactose transfer mechanism suggested by the structure

The present substrate-bound β4Gal-T1 structure does not provide any accurate information about the exact location of the galactose residue, nor the presence of any bound manganese ion. The approximate location of the galactose residue can be inferred from the position of the UDP portion of the UDP-Gal substrate and from the presence of some unconnected visible difference electron density regions extending from the β-phosphate group. The galactose moiety could be accommodated in the very bottom part of the catalytic pocket with its O_6 oxygen atom at a hydrogen-bonded distance from the hydroxyl group of the Tyr289 side-chain and with hydrophobic interactions between Gly291 and Pro357 residues.

The reaction catalysed by the β4Gal-T1 involves the transfer of galactose from the substrate UDP-Gal to the O_4 atom of the acceptor GlcNAc as well as the anomeric inversion of the galactose O_1–C_1 bond, from α to β, and the release of UDP. β4Gal-T1 is known to require the presence of Mn^{2+} for catalysis to occur. The participation of a Mn^{2+} ion bound to the phosphate groups of GDP–fucose has been modelled as an important step in the anomeric inversion of the glycosidic bond catalysed by

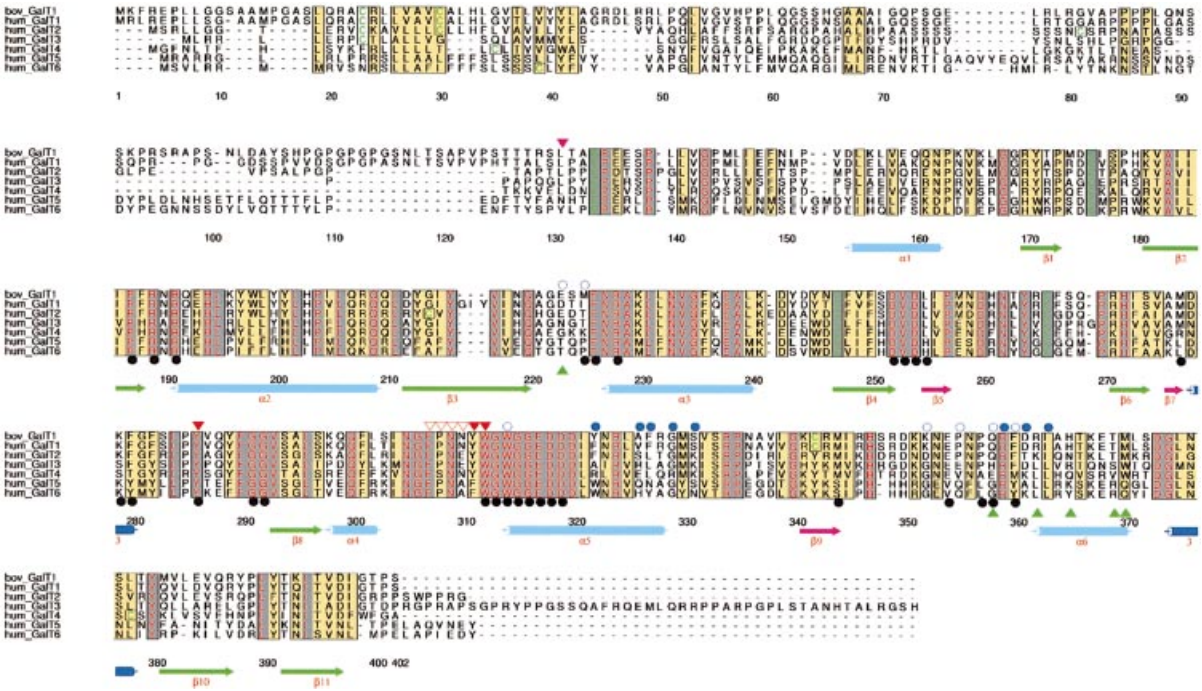

Fig. 4. Sequence alignment of bovine β4Gal-T1 and homologous proteins from the six known β4galactosyltransferase isoforms in the human genome (β4Gal-T1 to β4Gal-T6). The invariant residues are highlighted in red with a grey background, whereas conserved residues have an orange background. Cysteine residues are displayed in green. Secondary structure elements of bovine β4Gal-T1 are indicated beneath the sequences colour-coded as in Figure 1. Mutated residues are indicated by red triangles, which are closed in the case of those affecting UDP-Gal binding, and open with those affecting GlcNAc binding. Residues buried at the dimer interface are indicated by blue circles, which are open in the case of monomer A and closed in that of monomer B, whereas those that seem to play an important role in α-lactalbumin binding are indicated by closed green triangles. Residues exposed at the surface of the pocket are indicated by closed black circles. The bovine sequence numbering of every tenth residue is also given. The first well-defined residue in the electron density map of substrate-free structure, Leu131, is indicated (closed magenta triangle).

the human α1,3-fucosyltransferase V (Murray *et al*., 1996). The divalent cation might stabilize the oxygen phosphate negative charges and fragilize the GDP–sugar linkage, and even accelerate the hydrolysis reaction via the electrophilic catalysis occurring in the case of non-enzymatic hydrolysis of GDP–fucose (Murray *et al*., 1996). Under the crystallization conditions used here, β4Gal-T1 contained some manganese ions (2 mM), and some manganese ions were added during the protein purification procedure. Manganese ions were also added during the soaking of β4Gal-T1 crystals with UDP-Gal. Although the protein has been reported to bind to divalent metal ions, none were clearly visible in either our substrate-free or sub-strate-bound structures. However, in the case of the substrate-bound β4Gal-T1 structure, an electron density was detected between the carboxyl group Asp254 side-chain and the β-phosphate group of the substrate, where a water molecule was modelled at hydrogen-bond distance between the carboxyl group Asp254 side-chain and the β-phosphate group of the substrate.

Based on the substrate-bound structure, we can speculate about the glycosyl transfer mechanisms possibly involved. The anomeric inversion was expected to be favoured by the nucleophilic attack made on the 4-hydroxyl group of the acceptor substrate GlcNAc ring, by the carboxylic group of the Glu317 (or the Asp318) residue; the acceptor substrate GlcNAc was therefore likely to be positioned so that its C_4 atom was close to the βP-O-C_1 bond of UDP-Gal. The exact location of the sugar acceptor is not yet known, but on the basis of manual docking models for free GlcNAc and for a hepta-oligosaccharide chain obtained from the Fc neonatal receptor, FcRn glycoprotein (Protein Data Bank code: 3fru.pdb), the GlcNAc may be positioned deep inside the pocket with its C_4 atom close to the C_1 atom of galactose, 4.5 Å from the carboxyl group of the Glu317 residue, with its amino and acetyl group nested in a sub-site of the pocket interacting with Asp318 carboxyl group side-chain. The hepta-oligosaccharide chain of FcRn could be fitted on the side of the pocket opposite to the UDP-Gal binding site, where the solvent-exposed Trp314 may interact with the mannose residue of the sugar chain, as observed frequently in lectin structures (Figure 6C) (Elgavish and Shaanan, 1997).

Interpretation of β4Gal-T1 engineered point mutations

Extensive chemical protein modifications (Yadav *et al*., 1990, 1991) and site-directed mutagenesis experiments (Aoki *et al*., 1990; Zu *et al*., 1995) on the acceptor- and donor-binding sites of human β4Gal-T1 have been carried out to identify the most functionally relevant residues. These data suggested that the C-terminal region 276–328 of human β4Gal-T1 plays an important part in both UDP-Gal and GlcNAc binding and may involve residues

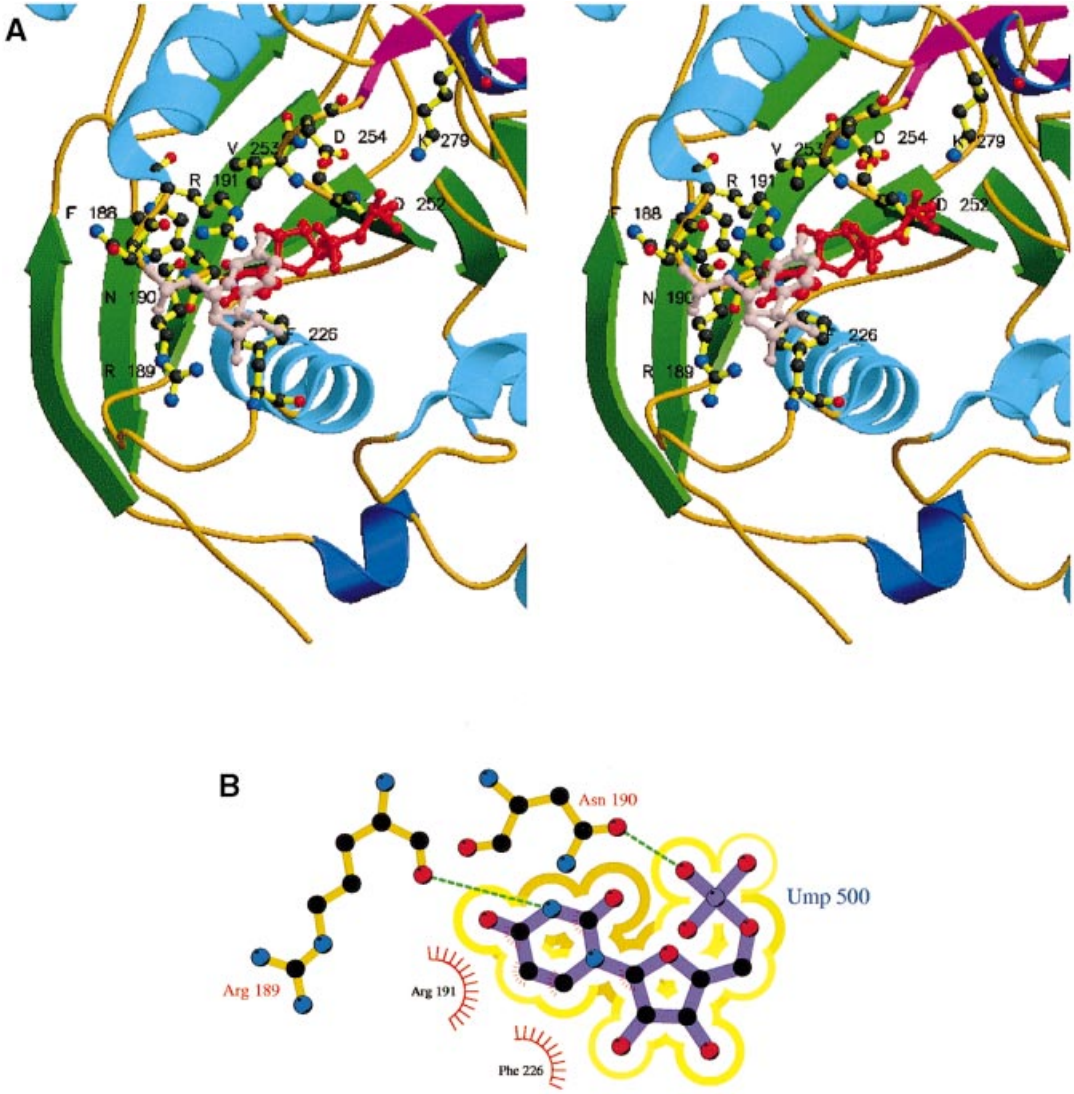

Fig. 5. Close-up stereo view of the superimposition of UMP from the substrate-free structure and UDP from the substrate-bound complex structure. (**A**) UMP and UDP are given in ball-and-stick forms where pink stands for UMP and red for UDP-Gal. β4Gal-T1 structure from the β4Gal-T1 structure complexed with UDP-Gal is shown as a ribbon diagram with amino acid side-chains interacting with UMP or UDP-Gal shown in a ball-and-stick form in yellow. (**B**) LIGPLOT representation of β4Gal-T1 amino acids interacting with UMP. The LIGPLOT legend is the same as in Figure 2C.

Phe305, Pro306, Asn307 and Asn308, and to a lesser extent, residues Tyr309 in UDP-Gal binding and Tyr284, Tyr309, Trp310 in GlcNAc binding (Zu *et al*., 1995). Among these residues, only the three bovine β4Gal-T1 residues, Tyr286, Asn309 and Trp312, are strictly conserved in the β4galactosyltransferase protein family (Figure 4). Our structural data suggest that most of the residues modified by mutagenesis do not lie exactly in the proposed UDP-Gal binding site and therefore that, when mutated, they may indirectly affect the UDP-Gal binding. The bovine β4Gal-T1 Asn309 side-chain located near Trp312 does not face the pocket and seems to stabilize the conformation of the W312GWGG sequence by establishing extensive hydrogen bonds with Asn233, Gly374 and Ser377 residues. Only the Tyr286Phe and Trp312Gly substitutions, which abolish the UDP-Gal binding without altering the manganese binding, are consistent with the proposed UDP-Gal binding site. Tyr286 which is located at the bottom of the pocket, forms hydrogen bonds between its side-chain hydroxyl group, the Asp319 side-chain and the Gln358 carbonyl oxygen and may possibly participate in the recognition of the O_4 oxygen atom of the galactose moiety. The Trp312 side-chain is buried at the pocket's periphery and establishes van der Waals contacts with the C-terminal residue Pro402, which suggests that any Trp312 substitutions (e.g. Trp312Gly) would destabilize the structure of the C-terminal region.

Amino acids thought to be critically involved in the α-lactalbumin binding

Among the members of the β4galactosyltransferase family, β4Gal-T1 and β4Gal-T2 synthesize lactose after being modulated by α-lactalbumin, while the other members are generally insensitive, or as in the case of β4Gal-T4, exhibit

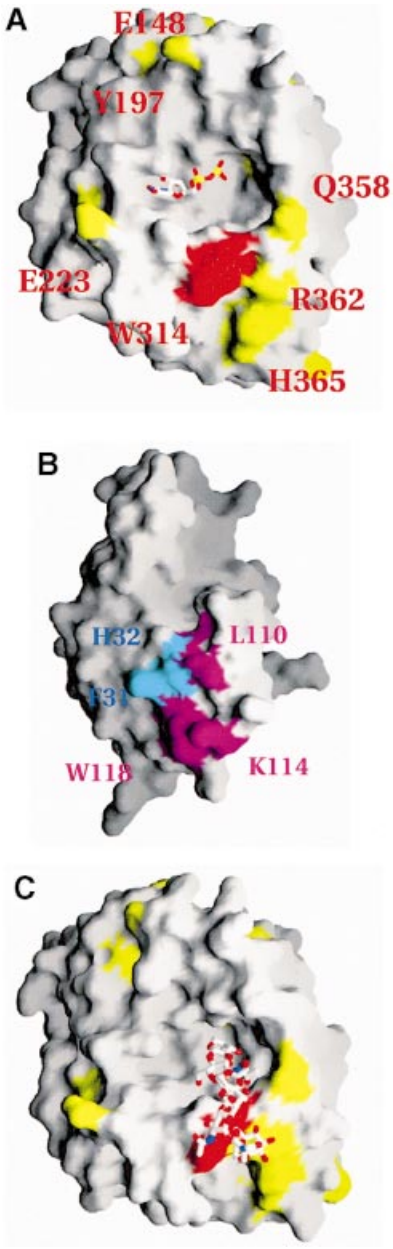

Fig. 6. α-lactalbumin binding site and oligosaccharide acceptor docking model. (**A**) Molecular surface of β4Gal-T1 is shown down to the catalytic pocket; the amino acid residues found to have the greatest functional importance for α-lactalbumin binding are coloured yellow and labelled, and the less important residues, white. UDP is given in ball-and-stick form. Solvent-accessible residues within the single secondary element α-helix α6 and Trp314 side-chain (red) are candidates for α-lactalbumin binding. (**B**) The molecular surface of bovine α-lactalbumin (Protein Data Bank code: 1alc.pdb) is given on the same scale as β4Gal-T1 in (A). Residues playing an important role in glucose binding are coloured blue, and those involved in the β4Gal-T1 interaction are coloured magenta. (**C**) The molecular surface of β4Gal-T1 is shown in the same orientation as in (A) with the hepta-saccharide of FcRn (PDB code: 3fru.pdb) docked manually into the acceptor site of the enzyme. The hepta-saccharide, Gal-GlcNAc-Man-Man-GlcNAc(Fuc)-GlcNAc, mannose (Man) and fucose (Fuc) is given in ball-and-stick form.

activation with GlcNAc as the substrate but only a poor lactose synthesis in the presence of α-lactalbumin (Schwientek *et al*., 1998). These data suggest that some of the residues involved in binding to α-lactalbumin might be conserved at the molecular surface, especially in the case of β4Gal-T1 and β4Gal-T2 family members. A single patch of non-conserved residues in the β4galactosyl-transferase protein family located in the α6 α-helix involves residues Gln358, Arg362, His365 and Met370 (Figures 4 and 6A). Since all these residues are fully exposed to solvent, with the exception of Met370, and are located close to Trp314, Trp314 may be involved in the acceptor binding (Figure 6A). The fact that most of these residues are buried at the β4Gal-T1 dimer interface (data not shown) suggests the existence of similarities between the α-lactalbumin–β4Gal-T1 interaction and the dimer (Figure 4). α-lactalbumin residues involved in the β4Gal-T1 interaction and glucose binding have been identified by performing chemical modifications and site-directed mutagenesis (O'Keeffe *et al*., 1980; Grobler *et al*., 1994). α-lactalbumin residues Phe31 and His32 have been found to play an important role in glucose binding and residues Leu110, Lys114 and Trp118 have been found to be critical for β4Gal-T1 interactions (Figure 6B) (Malinovskii *et al*., 1996). α-lactalbumin in complex with β4Gal-T1 might prevent access to some portions of its catalytic pocket, particularly the β4Gal-T1 acceptor site, due to a steric hindrance process involving a possible interaction between the Trp314 side-chain and α-lactalbumin Trp118 (Figure 6A and B).

Comparisons between the β4Gal-T1 structure and that of other glycosyltransferases

All the mammalian glycosyltransferases with various specificities that have been cloned so far have the same domain organization and are type II membrane-bound proteins, although their amino acid sequences do not show any evident homologies. However, upon careful inspection of numerous glycosyltransferase amino acid sequences, some common short sequence motifs were identified, such as the DXD motif (Busch *et al*., 1998; Wiggins and Munro, 1998), which is involved in the UDP-Gal binding site of the substrate-bound structure. A consensus sequence containing the short amino acid stretch WGGE, which contains the important catalytic β4Gal-T1 Glu317 residue, is present in human GalNAc-T4 transferase (Bennett *et al*., 1998) and in *Pasteurella multocida* hyaluronan synthase (DeAngelis *et al*., 1998). These particular glycosyltransferases might therefore have some enzymatic mechanism in common with β4Gal-T1.

In conclusion, the catalytic domain of a mammalian glycosyltransferase, β4Gal-T1, shows a novel topology with a large open pocket lined with mostly invariant residues forming the donor binding site; the acceptor binding region might be located at the periphery of the pocket in the close proximity of the α-lactalbumin binding site. This structure might provide a useful template for predicting the structures of other related glycosyltransferases.

Materials and methods

Protein expression and purification

Bovine β4Gal-T1, like other mammalian glycosyltransferases, consists of type II membrane-bound glycoproteins constitutively expressed in the *trans*-Golgi subcellular compartment. Various unsuccessful attempts to obtain β4Gal-T1 crystals have been made in the past, using truncated forms of enzyme containing quite a large portion of the Pro/Gly-rich segment of the N-terminal part of the enzyme (stem region). In order to

obtain β4Gal-T1 crystals with satisfactory diffraction qualities, a shorter truncated form of β4Gal-T1 completely devoid of the Pro/Gly-rich region (starting at residue Pro114) was produced using a transfected stable mammalian cell line (NSO cells). The catalytic properties of this truncated soluble β4Gal-T1 enzyme were found to be identical to those of the full-length enzyme, and it was able to synthesize lactose in the presence of α-lactalbumin and glucose as acceptor (data not shown). The β4Gal-T1 protein secreted in the culture medium was purified by performing UDP-hexanolamine–Sepharose affinity chromatography and eluted with 10 mM UMP. The protein was further purified by performing size-exclusion chromatography. The enzyme was characterized by performing standard assays with GlcNAc as the substrate.

Crystallization

Purified bovine β4Gal-T1 catalytic domain was crystallized at 20°C in a 2–4 μl hanging drop containing a 1:1 mixture of protein solution (5–15 mg/ml of β4Gal-T1 in 20 mM Tris Maleate pH 6.5, 2 mM $MnCl_2$ and 50 mM NaCl) and reservoir solution (25–30% PEG 5K, 100–150 mM Li_2SO_4, 100 mM Tris pH 8.5). Crystals ~0.2×0.2×0.3 mm^3 in size appeared within 2 weeks. Mercury(II) acetate derivatives were obtained by soaking β4Gal-T1 crystals for 48 h in a solution containing 2–5 mM mercury(II) acetate and 30% PEG 5K, 125 mM Li_2SO_4, 100 mM Tris pH 8.5. β4Gal-T1 crystals soaked with UDP-Gal were obtained by incubating β4Gal-T1 crystals with crystallization buffer containing 5 mM UDP-Gal (freshly prepared) and 5 mM $MnCl_2$ for 12 h. Before data collection at 100K, crystals were soaked briefly in a solution containing 12% ethylene glycol, 30% PEG 5K, 125 mM Li_2SO_4 and 100 mM Tris pH 8.5, and flash frozen in the gaseous stream of liquid nitrogen.

Data collection and structure determination

The diffraction intensities of the native and mercury derivative crystals were collected on a MarResearch/Rigaku RU-200 rotating anode X-ray generator with graphite monochromator at 40 kV×80 mA. Oscillation images were integrated with the DENZO program (Otwinoski and Minor, 1997), scaled and merged with SCALA (CCP4 programs, 1994) (Table I). The intensities were converted into structure factor amplitudes with TRUNCATE (CCP4 programs, 1994). The crystals belonged to the orthorhombic space group $C222_1$ and the unit cell dimensions were a = 108.5 Å, b = 161 Å, c = 107.4 Å. Two molecules of β4Gal-T1 are present in the asymmetric unit of the crystal. The sites of the mercury derivative were found by difference Patterson techniques using RSPS (CCP4 programs, 1994). Heavy atom sites were refined and phases calculated to 3.0 Å resolution with MLPHARE (CCP4 programs, 1994), giving an overall figure of merit of 0.36. High resolution data were obtained on the ID14-EH3 ESRF synchrotron beam line (Grenoble, France) at cryogenic temperatures.

Model refinement: substrate-free and substrate-bound β4Gal-T1 structures

All the main secondary structures were located in the solvent-flattened 3.0 Å SIRAS maps (DM, R_{free}=30%). A partial model for a β4Gal-T1 molecule was built using the TPPR option (Trigonal Plane Pseudo Residue) in the TURBO-FRODO program (A.Roussell and C.Cambillau, personal communication) and was used as a search model to locate the second β4Gal-T1 molecule using the AMoRe program package (Navaza, 1994). An averaged 3.0 Å resolution SIRAS map was calculated with a mask built around one monomer and its two-fold NCS-related partner, and used for subsequent model building. The assignment of sequences was aided by an accurate secondary structure prediction (Rost and Sander, 1994). The models were refined using X-PLOR (Brünger *et al*., 1987) and in later stages using CNS (Brünger *et al*., 1998). Successive rebuilding and simulated annealing refinement rounds provided electron density maps with which the β4Gal-T1 structure could be completely interpreted. With most of the residues in the structure, tight NCS-restraints (100 kcal/mol.$Å^2$) were applied, with the exception of those that did not obey the NCS (residues 133–156, 223–227, 311–317 and 344–345), which were refined with loose NCS-restraints (10 kcal/mol.$Å^2$). Maximum-likelihood refinement, bulk solvent correction and anisotropic *B*-factor correction of the data were applied in the final refinement. The final models comprise β4Gal-T1 residues Leu131 to Ser402 for the substrate-free structure and β4Gal-T1 residues Ala133 to Ser402 for the substrate-bound structure. The two molecules of the dimer are practically identical, with an r.m.s.d. of 0.23 Å in the case of the C_α atoms, excluding residues Trp312 and the loop region 346–353. The dimer hides ~700 $Å^2$ of the total 12000 $Å^2$ surface area of a single protein molecule and the dimer interface involves seven residues on each molecule, with the exception of Trp314, which are not conserved in the β4galactosyltransferase protein family, which suggests that dimerization is unlikely to occur in solution (Figure 4).

The substrate-free β4Gal-T1 structure, after solvent and UMP removal, was used as the starting model to determine the structure of the complex. Rigid body refinement of the substrate-bound complex decreased the *R*-factor to 28% in the 10–3.5 Å resolution range. Fourier difference maps clearly showed the location of bound UDP-Gal in the two molecules of the asymmetric unit. Refinement was subsequently carried out using CNS (Brünger *et al*., 1998). The addition of 77 solvent molecules and two UDP resulted in an *R*-factor of 25% using data from 30 to 2.4 Å (Table I). The stereochemistry model (Table I) was analysed with PROCHECK (Laskowski *et al*., 1993) and WHATIF (Hooft *et al*., 1996). Figures 1A and B, 2A and 5A were generated using MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt and Murphy, 1994), Figure 2B with TURBO-FRODO (A.Roussel and C.Cambillau, personal communication) and Figures 3 and 6 with GRASP (Nicholls *et al*., 1991). Figure 4 was prepared with ALSCRIPT (Barton, 1993). Figures 2C and 5B were prepared with LIGPLOT (Wallace *et al*., 1995).

Acknowledgements

We thank V.Zanboni and S.Dixon for their excellent technical assistance, and S.Munro and B.Henrissat for their helpful comments. We thank J.Paulson (Cytelcorp, La Jolla, CA) and J.Schultz for their encouragment in the initial phase of this work and for the gift of bovine β4Gal-T1 cDNA. We thank P.Bjorkman, A.Chirino, P.Marchot and H.Clausen for their critical reading of the manuscript and for providing unpublished information.

References

Almeida,R. *et al.* (1997) A family of human β4-galactosyltransferases: cloning and expression of two novel UDP-galactose-β-*N*-acetylglucosamine β1,4-galactosyltransferases, β4Gal-T2 and β4Gal-T3. *J. Biol. Chem.*, **272**, 31979–31991.

Aoki,D., Appert,H.E., Johnson,D., Wong,S.S. and Fukuda,M.N. (1990) Analysis of the substrate binding sites of human galactosyltransferase by protein engineering. *EMBO J.*, **9**, 3171–3178.

Asano,M., Furukawa,K., Kido,M., Matsumoto,S., Umesaki,Y., Kochibe,N. and Iwakura,Y. (1997) Growth retardation and early death of β1,4-galactosyltransferase knockout mice with augmented proliferation and abnormal differentiation of epithelial cells. *EMBO J.*, **16**, 1850–1857.

Axford,J.S., Mackensie,L., Lydyard,P.M., Hay,F.C., Isenberg,D.A. and Roitt,I.M. (1987) Reduced B-cell galactosyltransferase activity in rheumatoid arthritis. *Lancet*, **ii**, 1486–1488.

Bakker,H., Agterberg,M., Van Tetering,A., Koeleman,C.A.M., Van den Eijnden,D.H. and Van Die,I. (1994) A *Lymnaea stagnalis* gene with sequence similarity to that of mammalian β1,4galactosyltransferases, encodes a novel UDP-GlcNAc: GlcNAcβ-R β1,4-*N*-Acetylglucosaminyltransferase. *J. Biol. Chem.*, **269**, 30326–30333.

Bakker,H., Van Tetering,A., Agterberg,M., Smit,A.B., Van den Eijnden,D.H. and Van Die,I. (1997) Deletion of two exons from *Lymnaea stagnalis* β1,4-*N*-Acetylglucosaminyltransferase gene elevates the kinetic efficiency of the encoded enzyme for both UDP-sugar donor and acceptor substrates. *J. Biol. Chem.*, **272**, 18580–18585.

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

Bennett,E.P. *et al.* (1998) Cloning of a human UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete *O*-glycosylation of the MUC1 tandem repeat. *J. Biol. Chem.*, **273**, 30472–30481.

Boeggeman,E.E., Balaji,P.V., Sethi,N., Masibay,A.S. and Qasba,P.K. (1993) Expression of deletion constructs of bovine β1,4-galactosyltransferase in *E.coli*: importance of Cys134 for its activity. *Protein Eng.*, **6**, 779–785.

Boeggeman,E.E., Balaji,P.V. and Qasba,P.K. (1995) Functional domains of bovine β1,4galactosyltransferase. *Glycoconj. J.*, **12**, 865–878.

Breton,C., Oriol,R. and Imberty,A. (1996) Sequence alignment and fold recognition of fucosyltransferases. *Glycobiology*, **6**, VII–XII.

Breton,C., Bettler,E., Joziasse,D.H., Geremia,R. and Imberty,A. (1998) Sequence–function relationships of prokaryotic and eukaryotic galactosyltransferases. *J. Biochem.*, **123**, 1000–1009.

Brodbeck,U., Denton,W.L., Tanahashi,N. and Ebner,K.E. (1967) The isolation and identification of the B protein of the lactose synthase as α-lactalbumin. *J. Biol. Chem.*, **242**, 1391–1397.

Brünger,A.T., Kuriyan,J. and Karplus,M. (1997) Crystallographic *R*-factor refinement by molecular dynamics. *Science*, **235**, 1118–1135.

Brünger,A.T. *et al.* (1998) Crystallography and NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr.*, **D54**, 905–921.

Busch,C., Hofmann,F., Selzer,J., Munro,S., Jeckel,D. and Aktories,K. (1998) A common motif of eukaryotic glycosyltransferases is essential for the enzyme activity of large clostridial cytotoxins. *J. Biol. Chem.*, **273**, 19566–19572.

Campbell,J.A., Davies,G.J., Bulone,V. and Henrissat,B. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.*, **326**, 929–942.

CCP4, Collaborative Computer Projet N°4. (1994) The CCP4 suite : programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

DeAngelis,P.L., Jing,W., Drake,R.R. and Achyuthan,A.M. (1998) Identification and molecular cloning of a unique hyaluronan synthase from *Pasteurella multocida*. *J. Biol. Chem.*, **273**, 8454–8458.

Delves,P.J. (1998) The role of glycosylation in autoimmune disease. *Autoimmunity*, **27**, 239–253.

Easton,E.W., Bolscher,J.G. and Van den Eijnden,D.H. (1991) Enzymatic amplification involving glycosyltransferases forms the basis for the increased size of asparagine-linked glycans at the surface of NIH 3T3 cells expressing the N-ras proto-oncogene. *J. Biol. Chem.*, **266**, 21674–21680.

Elgavish,S. and Shaanan,B. (1997) Lectin–carbohydrate interactions: different folds, common recognition principles. *Trends Biochem. Sci.*, **22**, 462–467.

Furukawa,K., Matsuta,K., Takeuchi,F., Kosuge,E., Miyamoto,T. and Kobata,A. (1990) Kinetic study of a galactosyltransferase in the B cells of patients with rheumatoid arthritis. *Int. Immunol.*, **2**, 105–112.

Grobler,J.A., Wang,M., Pike,A.C.W. and Brew,K. (1994) Study by mutagenesis of the roles of two aromatic clusters of α-lactalbumin in aspects of its action in the lactose synthase system. *J. Biol. Chem.*, **269**, 5106–5114.

Guo,Z. and Wang,P.G. (1997) Utilization of glycosyltransferases to change oligosaccharide structures. *Appl. Biochem. Biotech.*, **68**, 1–20.

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

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., Flöckner,H., Rüger,W. and Breton,C. (1999) Fold recognition study of α3-Galactosyltransferase and molecular modeling of the nucleotide sugar-binding domain. *Glycobiology*, in press.

Jeddi,P.A., Bodman-Smith,K.B., Lund,T., Lydyard,P.M., Mengle-Gaw,L., Isenberg,D.A., Youinou,P. and Delves,P.J. (1996) Agalactosyl IgG and β1,4galactosyltransferase gene expression in rheumatoid arthritis patients and in the arthritis-prone MRL lpr/lpr mouse. *Immunology*, **87**, 654–659.

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

Joziasse,D., Shaper,J.H., Van den Eijnden,D.H., Van Tunen,H.J. and Shaper,N.L. (1989) Bovine α1,3-galactosyltransferase: isolation and characterization of a cDNA clone. Identification of homologous sequences in human genomic DNA. *J. Biol. Chem.*, **264**, 14290–14297.

Kobata,A. (1989) Altered glycosylation of surface glycoproteins in tumor cells and its clinical application. *Pigment Cell Res.*, **2**, 304–308.

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

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.

Lo,N.-W., Shaper,J.H., Pevsner,J. and Shaper,N.L. (1998) The expanding β4-galactosyltransferase gene family: messages from the databanks. *Glycobiology*, **8**, 517–526.

Lu,Q., Hasty,P. and Shur,B.D. (1997) Targeted mutation in β1,4galactosyltransferase leads to pituitary insufficiency and neonatal lethality. *Dev. Biol.*, **181**, 257–267.

Malinovskii,V.A., Tian,J., Grobler,J.A. and Brew,K. (1996) Functional site in α-lactalbumin encompasses a region corresponding to a subsite in Lysozyme and parts of two adjacent flexible substructures. *Biochemistry*, **35**, 9710–9715.

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

Morrison,J.F. and Ebner,K.E. (1971) Studies on galactosyltransferase: kinetic effects of α-lactalbumin with *N*-acetylglucosamine and glucose as galactosyl group acceptors. *J. Biol. Chem.*, **246**, 3992–3998.

Murray,B.W., Takayama,S., Schultz,J. and Wong,C-H. (1996) Mechanism and specificity of human α1,3-fucosyltransferase V. *Biochemistry*, **35**, 11183–11195.

Narimatsu,K., Sinha,S., Brew,K., Okayama,H. and Qasba,P.K. (1986) Cloning and sequencing of cDNA of bovine *N*-acetylglucosamine (β1-4)galactosyltransferase. *Proc. Natl Acad. Sci. USA*, **83**, 4720–4724.

Navaza,J. (1994) AMoRe: an automate package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Protein Struct. Funct. Genet.*, **11**, 281–293.

Nomura,T. *et al*. (1998) Purification, cDNA cloning and expression of UDP-Gal: glucosylceramide β1,4galactosyltransferase from rat brain. *J. Biol. Chem.*, **273**, 13570–13577.

O'Keeffe,E.T., Hill,R.L. and Bell,J.E. (1980) Active site of bovine galactosyltransferase: kinetic and fluorescence studies. *Biochemistry*, **19**, 4954–4962.

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

Parekh,R.B., Dwek,R.A. and Rademacher,T.W. (1988) Rheumatoid arthritis as a glycosylation disorder. *Br. J. Rheumatol.*, **27**, 162–169.

Paulson,J.C. and Colley,K.J. (1989) Glycosyltransferases: structure, localization and control of cell type-specific glycosylation. *J. Biol. Chem.*, **264**, 17615–17617.

Pike,A.C.W., Brew,K. and Acharya,K.R. (1996) Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure*, **4**, 691–703.

Powell,J.T. and Brew,K. (1974) Glycosyltransferases in the Golgi membranes of onion stem. *Biochem. J.*, **142**, 203–209.

Rademacher,T.W., Williams,P. and Dwek,R.A. (1994) Agalactosyl glycoforms of IgG autoantibodies are pathogenic. *Proc. Natl Acad. Sci. USA*, **91**, 6123–6127.

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

Schwientek,T., Almeida,R., Levery,S.B., Holmes,E.R., Bennett,E. and Clausen,E. (1998) Cloning of a novel member of the UDP-galactose: β-*N*-Acetylglucosamine β1,4-Galactosyltransferase family, β4Gal-T4, involved in glycosphingolipid biosynthesis. *J. Biol. Chem.*, **273**, 29331–29340.

Shaper,N.L., Hollis,G.F., Douglas,J.G., Kirsch,I.R. and Shaper,J.H. (1988) Characterization of the full length cDNA for murine β-1,4-galactosyltransferase. *J. Biol. Chem.*, **263**, 10420–10428.

Shaper,N.L., Meurer,J.A., Joziasse,D.H., Chou,T-D.D., Smith,E.J., Schnaar,R.L. and Shaper,J.H. (1997) The chicken genome contains two functional non allelic β1,4-galactosyltransferase genes. *J. Biol. Chem.*, **272**, 31389–31399.

Shur,B.D., Evans,S. and Lu,Q. (1998) Cell surface galactosyltransferase: current issues. *Glycoconj. J.*, **15**, 537–548.

Stous,G. (1986) Golgi and secreted galactosyltransferases. *CRC Crit. Rev. Biochem.*, **21**, 119–151.

Tsuchiya,N., Endo,T., Matsuta,K., Yoshinoya,S., Aikawa,T., Kusug,E., Takeuchi,F., Miyamoto,T. and Kobata,A. (1989) Effects of galactose depletion from oligosaccharide chains on immunological activities of human IgG. *J. Rheumatol.*, **16**, 285–290.

Udagawa,Y., Aoki,D., Ito,K., Uejima,T., Uemura,M. and Nozawa,S. (1998) Clinical characteristics of newly developed ovarian tumor marker, galactosyltransferase associated with tumor (GAT). *Eur. J. Cancer*, **34**, 489–495.

Van den Eijnden,D.H. and Joziasse,D.H. (1993) Enzymes associated with glycosylation. *Curr. Opin. Struct. Biol.*, **3**, 711–721.

Van Die,I., Bakker,H. and Van den Eijnden,D.H. (1997) Identification of conserved amino acid motifs in members of the β1,4galactosyltransferase gene family. *Glycobiology*, **7**, V–VI.

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.

Wallace,A.C., Laskowski,R.A. and Thornton,J.M. (1995) LIGPLOT: A program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.*, **8**, 127–134.

Wang,Y., Wong,S.S., Fukuda,M.N., Zu,H., Liu,Z., Tang,Q. and Appert,H.E. (1994) Identification of functional cysteine residues in human galactosyltransferase. *Biochem. Biophys. Res. Commun.*, **204**, 701–709.

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.

Yadav,S. and Brew,K. (1990) Identification of a region of UDP-galactose: *N*-acetylglucosamine β4-galactosyltransferase involved in UDP-galactose binding by differential labeling. *J. Biol. Chem.*, **265**, 14163–14169.

Yadav,S.P. and Brew,K. (1991) Structure and function in galactosyltransferase: sequence locations of α-lactalbumin binding site, thiol groups and disulfide bond. *J. Biol. Chem.*, **266**, 698–703.

Zu,H., Fukuda,M.N., Wong,S.S., Wang,Y., Liu,Z., Tang,Q. and Appert,H.E. (1995) Use of site-directed mutagenesis to identify the galactosyltransferase binding sites for UDP-galactose. *Biochem. Biophys. Res. Commun.*, **206**, 362–369.

Received February 19, 1999; revised April 30, 1999; accepted May 17, 1999