Crystal Structure of β1,4-Galactosyltransferase Complex with UDP-Gal Reveals an Oligosaccharide Acceptor Binding Site

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · MAY 2002

Impact Factor: 4.33 · DOI: 10.1016/S0022-2836(02)00020-7 · Source: PubMed

3 AUTHORS, INCLUDING:

Boopathy Ramakrishnan
Visterra, Inc.
76 PUBLICATIONS **2,030** CITATIONS
SEE PROFILE

Pradman K Qasba
U.S. Department of Health and Human Services
123 PUBLICATIONS **2,765** CITATIONS
SEE PROFILE

doi: 10.1016/S0022-2836(02)00020-7 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 318, 491–502

JMB

Crystal Structure of β1,4-Galactosyltransferase Complex with UDP-Gal Reveals an Oligosaccharide Acceptor Binding Site

B. Ramakrishnan[1,2], P.V. Balaji[1,2] and Pradman K. Qasba[1]*

[1]*Structural Glycobiology Section, Laboratory of Experimental and Computational Biology CCR, NCI, Frederick MD 21702-1201, USA*

[2]*Intramural Research Support Program, SAIC-Frederick, Inc. Laboratory of Experimental and Computational Biology, CCR NCI, Frederick, MD 21702-1201, USA*

**Corresponding author*

The crystal structure of the catalytic domain of bovine β1,4-galactosyltransferase (Gal-T1) co-crystallized with UDP-Gal and $MnCl_2$ has been solved at 2.8 Å resolution. The structure not only identifies galactose, the donor sugar binding site in Gal-T1, but also reveals an oligosaccharide acceptor binding site. The galactose moiety of UDP-Gal is found deep inside the catalytic pocket, interacting with Asp252, Gly292, Gly315, Glu317 and Asp318 residues. Compared to the native crystal structure reported earlier, the present UDP-Gal bound structure exhibits a large conformational change in residues 345–365 and a change in the side-chain orientation of Trp314. Thus, the binding of UDP-Gal induces a conformational change in Gal-T1, which not only creates the acceptor binding pocket for *N*-acetylglucosamine (GlcNAc) but also establishes the binding site for an extended sugar acceptor. The presence of a binding site that accommodates an extended sugar offers an explanation for the observation that an oligosaccharide with GlcNAc at the non-reducing end serves as a better acceptor than the monosaccharide, GlcNAc. Modeling studies using oligosaccharide acceptors indicate that a pentasaccharide, such as *N*-glycans with GlcNAc at their non-reducing ends, fits the site best. A sequence comparison of the human Gal-T family members indicates that although the binding site for the GlcNAc residue is highly conserved, the site that binds the extended sugar exhibits large variations. This is an indication that different Gal-T family members prefer different types of glycan acceptors with GlcNAc at their non-reducing ends.

Keywords: β1,4-galactosyltransferase; crystal structure; UDP-Gal binding; conformational changes; oligosaccharide binding site

Introduction

β1,4-Galactosyltransferase-I (Gal-T1; EC 2.4.1.90/38) is a Golgi resident, type II, transmembrane protein which transfers galactose (Gal) from UDP-Gal to a *N*-acetylglucosamine (GlcNAc) residue, generating a β1-4-linkage between Gal and GlcNAc.[1] Although it efficiently transfers Gal to the monosaccharide GlcNAc, oligosaccharides with GlcNAc at their non-reducing end are preferred, indicating the presence of an extended oligosaccharide binding site on Gal-T1.[2] In humans, a family of β1,4-galactosyltransferase (Gal-T) enzymes has been identified, which to date consists of at least seven members, Gal-T1 to Gal-T7.[3–5] These family members, which have tissue-specific expression, exhibit significant sequence identity and are involved in a variety of cellular processes. For example, Gal-T6 is more abundantly expressed in adult brain tissues and involved in lactosylceramide synthesis, while Gal-T7 is widely expressed in many tissues and is involved in proteoglycan synthesis. These enzymes differ from each other in their sugar acceptor specificity, although they all transfer Gal from UDP-Gal.[4] Gal-T6 and Gal-T7 transfer Gal to Glc and Xyl, respectively, whilst the remaining family members transfer Gal to GlcNAc linked at the non-reducing end of different oligosaccharides. Hence, these family members, Gal-T2 to Gal-T5,

Permanent address: P. V. Balaji, Biotechnology Center, Indian Institute of Technology, Powai, Mumbai 400 076, India.

Abbreviations used: Gal-T1, β1,4-galactosyltransferase; GlcNAc, *N*-acetylglucosamine; LA, lactalbumin; Glc, glucose; Xyl, xylose.

E-mail address of the corresponding author: qasba@helix.nih.gov

Table 1. Statistics for data collection and refinement

Wavelength used during data collection (Å)	0.98
Cell dimensions (Å)	$a = 49.7$, $b = 88.7$, $c = 140.9$
Resolution range (Å)	20–2.8
Space group	$P2_12_12_1$
No. of observations	58,965
No. of unique reflections	15,985
R_{sym}(%) (outermost shell)[a]	17.6(65)
Completeness (%) (outermost shell)[a]	98.9 (93)
$I/\sigma(I)$ (outermost shell)[a]	8.2 (1.7)
Crystallographic R-factor (%) (R_{cryst})	21.6
Free R-factor (%) (R_{free})[b]	28
Deviations from ideality	
Bond lengths (Å)	0.009
Bond angles (deg.)	1.6
Dihedral angles (deg.)	24.6
Improper angles (deg.)	1.0
Number of protein atoms	2210 atoms per each molecule
Number of molecules per asymmetric unit	2 (related by a pseudo-2-fold)[c]
Average B-factor statistics (Å^2)	
Protein atoms	35.2
Solvent atoms (54 in total)	38.8
UDP-Gal atoms (72 in total, 1/monomer)	51.5
Mn^{2+} (2 in total, 1/monomer)	41.9

[a] 3.0–2.8 Å.
[b] R_{free} is equal to R_{cryst} for a randomly selected 5% of reflections, not used in refinement.
[c] During the entire refinement the two molecules in the asymmetric unit were restrained using this pseudo-2-fold symmetry.

are expected to differ from Gal-T1, mainly at their oligosaccharide binding sites. In order to understand the structure–function relationship of Gal-T family members, a structural investigation of the oligosaccharide binding site of Gal-T1 is important. In the earlier crystal structure of Gal-T1,[6] the oligosaccharide binding site on the enzyme (see below) could not be clearly identified.

An interesting feature of Gal-T1 is that its sugar acceptor specificity is altered by α-lactalbumin (LA) from GlcNAc to the monosaccharide glucose (Glc), resulting in the synthesis of lactose.[7] LA is a mammary gland-specific protein expressed only during lactation. It is also closely related to C-type lysozyme.[8] Our crystal structure analyses of Gal-T1·LA complexes have clearly identified the UDP·Mn^{2+}, GlcNAc, Glc and LA binding sites on Gal-T1.[9] These studies clearly identified the monosaccharide acceptor binding site on Gal-T1. Since LA was bound to Gal-T1 at the oligosaccharide acceptor binding site, an observation supported by kinetic studies,[10] the oligosaccharide binding site could not be identified clearly. These studies also could not identify the binding site for the donor sugar galactose on Gal-T1, since the Gal moiety in UDP-Gal was hydrolyzed during crystallization of the Gal-T1·LA·UDP-Gal complex.[9] It has also been shown that Gal-T1 can transfer a variety of sugars, besides Gal from their UDP derivatives, albeit at lower efficiency;[11] hence, to better understand the donor sugar specificity of Gal-T1, it was deemed important to acquire information about the binding of Gal moiety.

Even though spectroscopic studies have demonstrated that in the absence of LA, Gal-T1 undergoes a conformational change upon UDP-Gal binding,[12] it was not until the crystal structure of the complex of Gal-T1·LA was determined that the detailed conformational change could be described.[9] These structural investigations suggested that Gal-T1, upon substrate binding, undergoes a conformational change that facilitates LA binding to form the Gal-T1·LA complex. In this conformational change, the secondary structure for residues from 358 to 365 also undergoes a change from coil to helix, and only in the helix conformation do these residues interact with both the LA and the acceptor molecules. Since LA strongly interacts with these residues in the Gal-T1·LA complex, even in the absence of the acceptor,[9] it raises the possibility that this change in the secondary structure may have been due to LA binding. In order to determine the conformational changes that are caused by substrate binding, and to identify the donor galactose, as well as the extended sugar acceptor binding sites, it is necessary to determine the crystal structure of Gal-T1 in the absence of LA, but in the presence of its substrates. These objectives have been achieved in the present study in which we describe the crystal structure of Gal-T1 co-crystallized with UDP-Gal and $MnCl_2$.

Results and Discussion

Crystallization and diffraction of C342T-Gal-T1·UDP-Gal complex

Our efforts to crystallize the native Gal-T1 in the absence of substrates were unsuccessful. Although in the presence of UDP-Gal and $MnCl_2$ we were able to grow single crystals, they diffracted only to

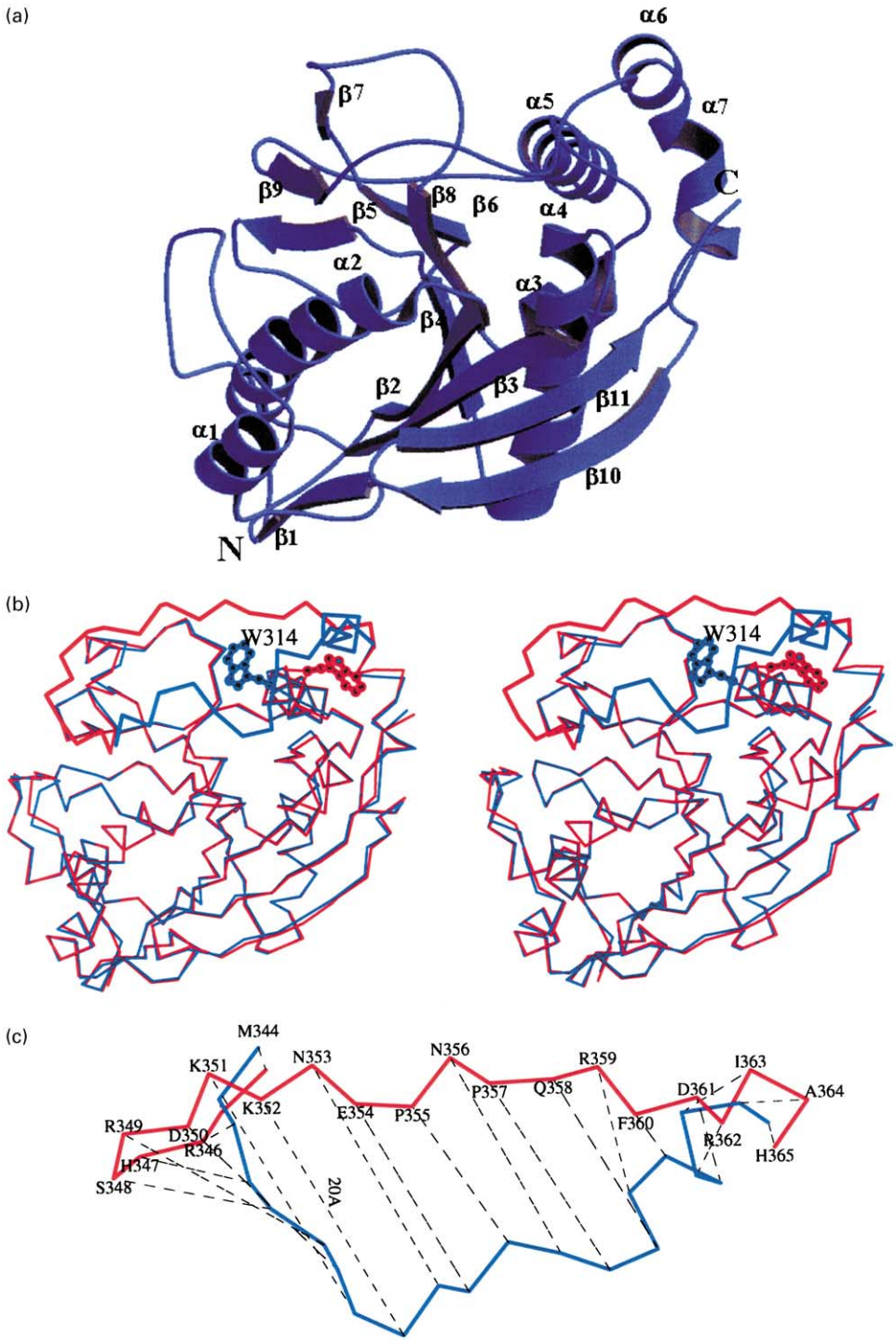

Figure 1. (a) Molecular structure of the Gal-T1 crystallized in the presence of UDP-Gal and Mn^{2+}. The molecule exists in conformation II where the side-chain orientation of Trp314 and the conformation for the residues 345–365 are different from the native crystal structure of Gal-T1[6] which is in conformation I (see the text). The secondary structural elements were determined by PROCHECK[36] and are very similar to that of conformation I. Since in conformation II the region 359–365, forms an α-helix, this helix is named as α6, and the α6 in conformation I is named as α7 in conformation II. The UDP-Gal molecule and Mn^{2+} are not shown. (b) Superposition of the C^{α} atoms of the present Gal-T1 molecule (blue) and the native crystal structure[6] (red). The superposition was carried out only for residues 132–345, and the mean rmsd value on C^{α} atoms for this region is 0.5 Å. Even though the residues from 345 to 402 were not used for the superposition, the residues 365–402 in both the structures are the same, while the residues 345–365 show large disagreement between these two structures. (c) The detailed deviation for the flexible region, residues 345–365, between conformation I (red) and conformation II (blue). The mean rmsd value of the C^{α} atoms for this region between the two structures is 9.6 Å. Further, residues 359–365 not only show a different fold but also a different secondary structure.

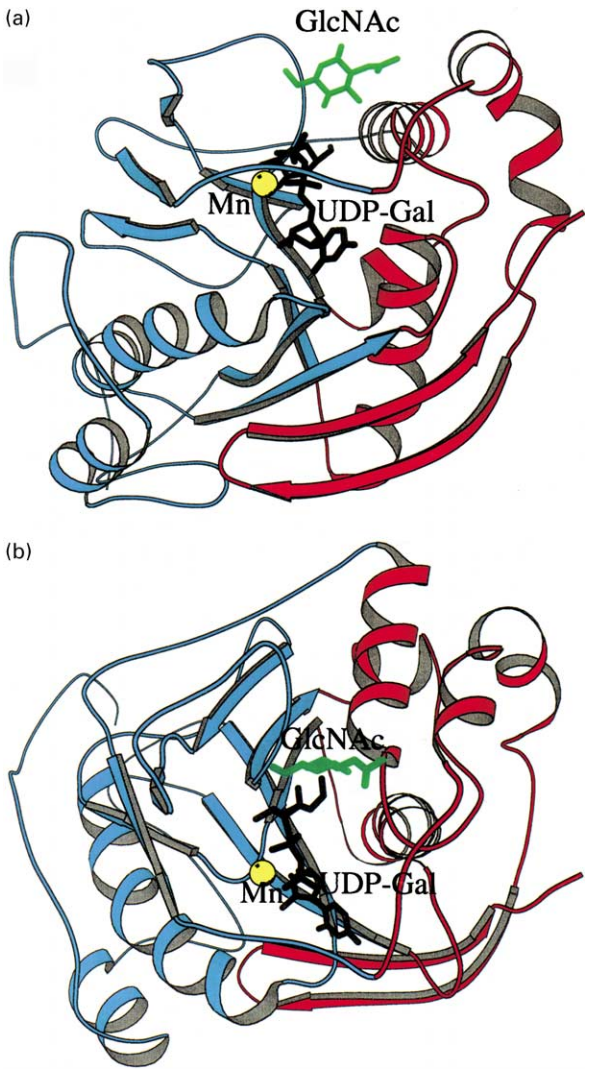

Figure 2. (a) Side and (b) top view of Gal-T1 molecule in conformation II, with the donor molecule, UDP-Gal (black), and the acceptor molecule, GlcNAc (green) assembled in a Gal-T1 molecule. The Mn^{2+} is shown as a yellow sphere. The protein regions are shown in two different colors to illustrate the domain structure. In conformation II, the catalytic pocket has been redefined, since the sugar acceptor pocket and the LA binding site in conformation I are only partially formed. A conformational change is essential to fully form these sites. In conformation I the protein was considered to be a single domain protein,[6] while in conformation II it can be considered as a two-domain protein with the substrate binding sites located between these domains from one end of the molecule to the other end. The substrate molecules bind in such a way that UDP-Gal binds at one end of the molecule with its Gal moiety facing the acceptor GlcNAc molecule, followed by the extended sugar binding site. The composite structure of UDP-Gal and GlcNAc bound substrates to Gal-T1 was constructed based on their individual bound structures. The GlcNAc binding was from the crystal structure of the Gal-T1·LA·GlcNAc complex,[9] and the UDP-Gal binding is from the present crystal structure.

4 Å resolution. In order to improve the diffraction quality of the crystals, numerous mutations on Gal-T1 were carried out and the mutants were tested for crystallization. A single mutation, Cys342 to Thr, resulted in a more stable protein compared to the recombinant wild-type Gal-T1, which crystallized readily in the presence of UDP-Gal and $MnCl_2$. Although these crystals diffracted to 2.3 Å resolution at room temperature, under cryo-conditions they diffracted poorly to a maximum resolution of only 2.8 Å (Table 1). Perhaps this was due to the fact that these crystals were grown in the presence of ammonium sulfate, which limited the use of different cryo-protectants during the data collection. In spite of the loss of resolution, in the crystal structure a bound molecule of UDP-Gal and Mn^{2+} could be located very clearly from the electron density maps (see below).

Conformation of Gal-T1 structure and comparison with native structure

The overall structure of Gal-T1 determined here is quite similar to its crystal structure of the apo-enzyme.[6] The superposition of these two structures reveals a large deviation in the region comprising residues 345–365 and the side-chain orientation of Trp314 (Figure 1(a) and (b)). The native crystal structure of Gal-T1[6] has been referred to as conformation I[9] and the present structure is called conformation II. The mean root-mean-square deviation (rmsd) on the C^{α} atoms between these two structures is 0.6 Å excluding the flexible region 345–365, whereas for this region it is 9.6 Å (Figure 1(c)). The conformation of the flexible region in the present crystal structure of Gal-T1 is quite similar to that found in the crystal structures of the Gal-T1·LA·substrate complexes,[9] with a mean rmsd value of only 0.6 Å for all residues. Between the two conformations, I and II, this flexible region not only exhibits a different fold but also differences in the secondary structure for several residues. In conformation II, residues 352–355 form a turn, and residues 359–365 an α-helix (Figure 1(c)). In contrast, in conformation I they are in random coil conformation. In conformation II these structural elements are important for the function of Gal-T1: the α-helix between the residues 359 and 365 for the binding of GlcNAc,[9] as well as for the binding of the extended sugar acceptors (see below). Residues Ile345 and His365, flanking this flexible region, appear to act like hinges which enable the region to undergo conformational changes between I and II. In fact, our preliminary molecular dynamics simulation studies suggest that the flexible region exhibits greater mobility and flexibility compared to the rest of the protein molecule (unpublished results).

It is highly unlikely that the observed conformational differences between the two Gal-T1 structures are due to the mutation of the Cys342 residue to Thr, since it has been shown that the crystal

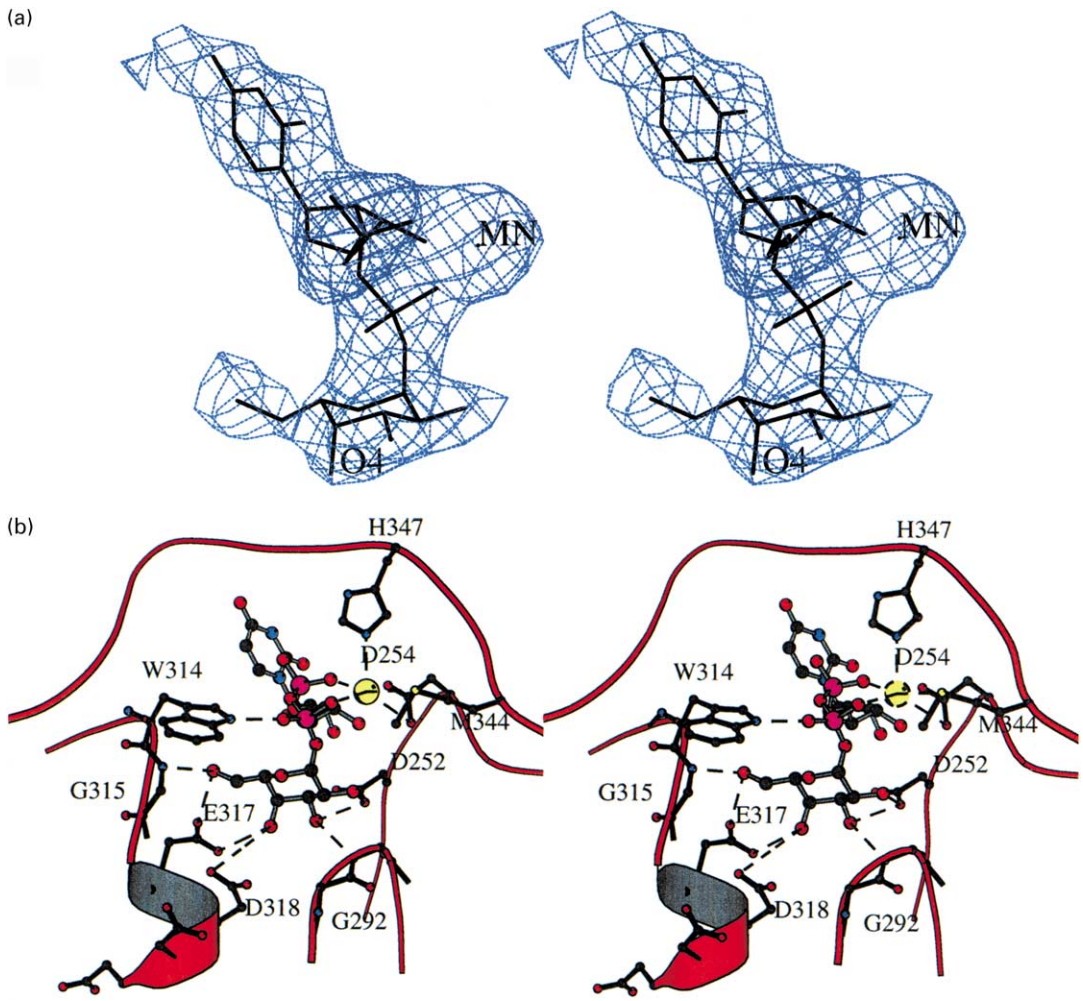

Figure 3. (a) A stereo view showing the difference electron density map for UDP-Gal and Mn^{2+}, contoured at 2σ level, to illustrate the ordered UDP-Gal molecule located in the crystal structure. (b) A stereo view showing the molecular interactions of the Gal moiety of UDP-Gal with the Gal-T1 molecule. All the residues interacting with the Gal moiety are highly conserved among the human Gal-T family members (data not shown). All the possible hydrogen-bonding interactions are shown in dotted lines. The axial O4 hydroxyl group of Gal does not strongly participate in hydrogen-bonding interactions with the Asp318 side-chain carboxylate group (see Table 2), and is pointed towards the acceptor binding site.

structure of the C342T-Gal-T1 in the C342T-Gal-T1·LA complex is the same as that of wild-type Gal-T1 in the Gal-T1·LA complex.[9] The present results provide direct evidence that in the absence of UDP-Gal, Gal-T1 exists in a conformation quite similar to its native crystal structure (conformation I), whereas in the presence of UDP-Gal it undergoes conformational changes for the region comprising residues 345–365 and in the side-chain orientation of Trp314 to conformation II. That these conformational differences are not due to different crystal packing is supported by the results of the spectroscopic[12] and photo-inactivation studies,[13] which suggested that Gal-T1 upon UDP-Gal and Mn^{2+} binding undergoes a conformational change. Photo-inactivation studies showed that UDP-Gal binding potentially protects a single Trp residue from hydrolysis by UV irradiation.[13] The present crystal structure of the Gal-T1·UDP-Gal complex supports this experimental observation. In the native crystal structure (conformation I), Trp314 is exposed to the solvent environment, while in the present crystal structure (conformation II), it is inside the catalytic pocket interacting with the UDP-Gal molecule, which most likely offers protection to Trp314 from UV irradiation. Therefore, it is clear from these observations that they represent two true conformational states, I and II, of Gal-T1 and are not due to the crystal environment or mutation.

Several protein regions in Gal-T1, which were buried in conformation I, are exposed in conformation II (see below), redefining the catalytic pocket of the Gal-T1. In the structure of the native Gal-T1,[6] this has been defined as a cone-shaped single domain protein containing a 13 Å wide catalytic pocket. However, in conformation II, the Gal-T1 molecule can be defined as a two-domain protein, with the substrate binding sites located between these two domains, which extend from

Table 2. Possible hydrogen bonding interactions of Gal moiety of UDP-Gal with Gal-T1

Atom in galactose	Residue (atom) in Gal-T1	Distance (Å) in molecule 1	Distance (Å) in molecule 2
O2	Asp252($O^{\delta 2}$)	2.78	3.02
O3	Asp252($O^{\delta 2}$)	2.75	3.41
	Asp252($O^{\delta 1}$)	3.27	2.88
	Gly292(N)	2.83	2.77
O4	Glu317($O^{\varepsilon 1}$)	3.45	2.76
	Asp318($O^{\delta 1}$)	3.64	3.43
O6	Glu317($O^{\varepsilon 2}$)	2.78	3.11
	Gly315(N)	2.66	3.10

one end to the other end of the molecule (Figure 2(a) and (b)). Earlier extensive site-directed mutagenesis studies on Gal-T1 have been carried out in several laboratories,[14–16] including ours[17]. Although the effect of several of the mutations on the enzyme kinetics could be explained assuming Gal-T1 to be in conformation I, not all the mutational data fit the explanation. With the redefinition of the catalytic pocket in conformation II, these mutational data can now be explained successfully. For example, the mutation of Trp314 was found to abolish the catalytic activity of Gal-T1.[14] In conformation I, Trp314 is found outside the catalytic pocket, and does not seem to participate in any direct interactions with the donor or acceptor molecule. In conformation II, Trp314 is found to reside inside the catalytic pocket and is involved in interactions with both the UDP-Gal and the acceptor molecule.[9] Similarly, Phe280 and Phe360 residues seem to affect the binding of GlcNAc binding to Gal-T1.[17] These two residues interact directly with the GlcNAc molecule in conformation II;[9] hence, mutation of these residues is expected to affect the binding of GlcNAc.[17] In contrast, in conformation I these residues are 14 Å apart and the acceptor binding site is only partially present (see below). Thus, the redefinition of the catalytic pocket seems to fit all the mutational data on Gal-T1.

UDP-Gal and Mn^{2+} binding to Gal-T1

Although the binding of UDP and the acceptor GlcNAc to Gal-T1 is well understood,[9] the exact nature of the binding of Gal moiety in UDP-Gal could not be determined in the earlier studies. Even though the present crystal structure is solved only at 2.8 Å resolution, the UDP-Gal molecule can be clearly located from the difference electron density maps (Figure 3(a)). A Mn^{2+} is located between the pyrophosphate group of UDP-Gal and the protein molecule, exhibiting five coordinations similar to that observed with UDP·Mn^{2+} bound in the Gal-T1·LA complex.[9] Of these, the coordination with His347 is only possible in conformation II, since in conformation I the $N^{\varepsilon 1}$ atom of His347 is nearly 12 Å away from the Mn^{2+}. Based on enzyme kinetic studies, two metal ion binding sites are present on Gal-T1, a primary site to which a Mn^{2+} binds and a second site that could bind either Mn^{2+} or Ca^{2+}.[18] In the present crystal structure the second metal ion binding site could not be unambiguously located from the electron density maps. A DXD amino acid motif, found in all the glycosyltransferases, is considered to be a metal ion binding motif.[19] In all the crystal structures of glycosyltransferases determined so far, the Asp residues in the DXD motif coordinate with the Mn^{2+}, either directly or through a water molecule.[20–24] In the present crystal structure, only Asp254 in the DXD motif forms a coordination bond with Mn^{2+}, whereas the side-chain carboxylate group of Asp252 is 6 Å away from the Mn^{2+} and it forms hydrogen bonds with the O2 and O3 hydroxyl groups of the donor sugar galactose (Figure 3(b) and Table 2).

As predicted earlier,[9] the Gal moiety is found deep inside the catalytic pocket, with all its hydroxyl groups interacting with protein residues (Figure 3(b) and Table 2). The O4 hydroxyl group of the galactose moiety in UDP-Gal forms a hydrogen bond with the Glu317 side-chain carboxylate oxygen atom, and this is similar to the binding of UDP-Glc to Gal-T1.[11] In addition to this hydrogen bond, there is a weak hydrogen bond observed between the O4 hydroxyl group and the Asp318 residue. This interaction is only possible in galactose because in glucose the O4 hydroxyl group is in an equatorial orientation, while it is axial in galactose. The lack of an O4 hydroxyl group in the axial orientation, as in the UDP-4-deoxy glucose or UDP-glucose, or the modification at the O4 hydroxyl group, results in a reduced catalytic efficiency (k_{cat}) of Gal-T1.[25,26] Thus, it appears that the axial orientation of the O4 hydroxyl group, which points towards the acceptor binding site, is important for the efficient catalysis (k_{cat}) of Gal-T1. Gal-T1 has also been shown to transfer galactose with O2 hydroxyl substitution such as *O*-methyl or even *N*-acetylgalactosamine (GalNAc),[26,27] but with very low catalytic efficiency (k_{cat}). Examination of the present crystal structure suggests that in the vicinity of the O2 hydroxyl group of galactose there is a small hydrophobic pocket formed by residues Leu255, Met277 and Met344, which can to some extent accommodate an O2-hydroxyl group substitution. Also, it is evident from the interactions between the side-chain carboxylate group of Asp252 and the O2 and O3 hydroxyl groups of galactose moiety (Table 2),

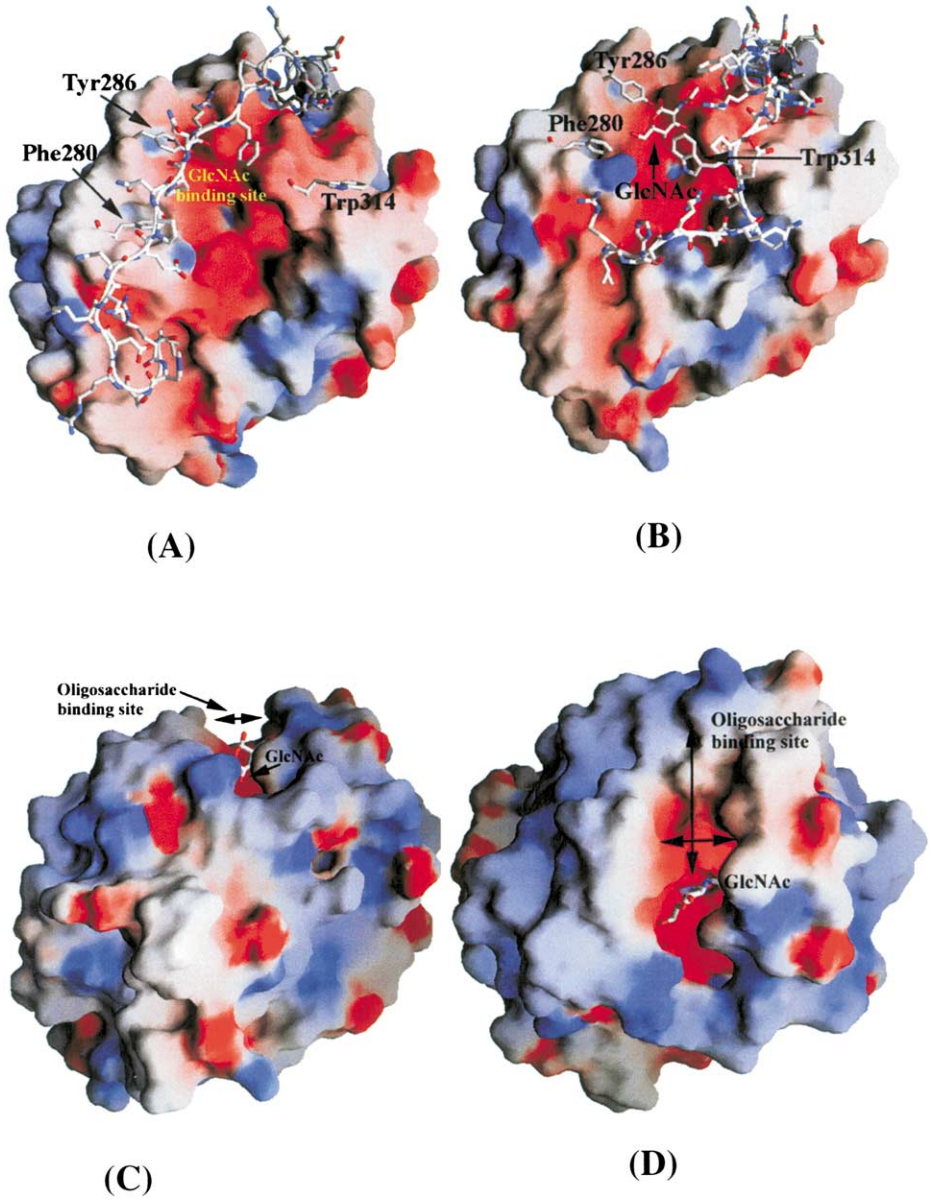

Figure 4. (a) In conformation I, the flexible region (white) blocks the GlcNAc binding site (shown with molecular surface diagram), residues 359–363 are in random coil conformation. (b) In conformation II, residues 359–363 are present in an α-helix conformation which exposes the hidden GlcNAc binding site and the *N*-acetyl group binding region. The GlcNAc binding to Gal-T1 is taken from the crystal structure of the Gal-T1·LA·GlcNAc complex.[9] (c) and (d) The oligosaccharide binding site edge view and the top view, respectively. In conformation II, not only the creation of the GlcNAc binding site but also that of the extended sugar binding site can be seen. The average width and length of this oligosaccharide binding site are 10 and 16 Å, respectively. This is the site where LA binds to Gal-T1 in order to modulate its sugar acceptor specificity.[9] The protein molecules are illustrated using the surface electrostatic potential and it is mapped onto molecular surface from −10 (red) to +10 kT (blue).

that minor variations in the side-chain conformation of Asp252 can facilitate the transfer of galactose with substitution or modification at the O2 hydroxyl group.

Conformational transition and creation of GlcNAc and extended sugar binding site

Comparing the GlcNAc or Glc bound crystal structures of the Gal-T1·LA complex[9] with the native Gal-T1 structure,[6] it is clear that in conformation I of Gal-T1, one part of the acceptor binding site is buried and the other part is absent. In conformation II the binding site is fully formed. Specifically, GlcNAc binding residues, Phe280 and Tyr286, are partially buried in the conformation I (Figure 4(a)), while in conformation II, these residues are well exposed (Figure 4(b)), their solvent accessibility increased two- to threefold. In the crystal structure of the Gal-T1·LA·GlcNAc complex,[9] the *N*-acetyl group of GlcNAc is seen to bind to residues Arg359, Phe360 and Ile363,

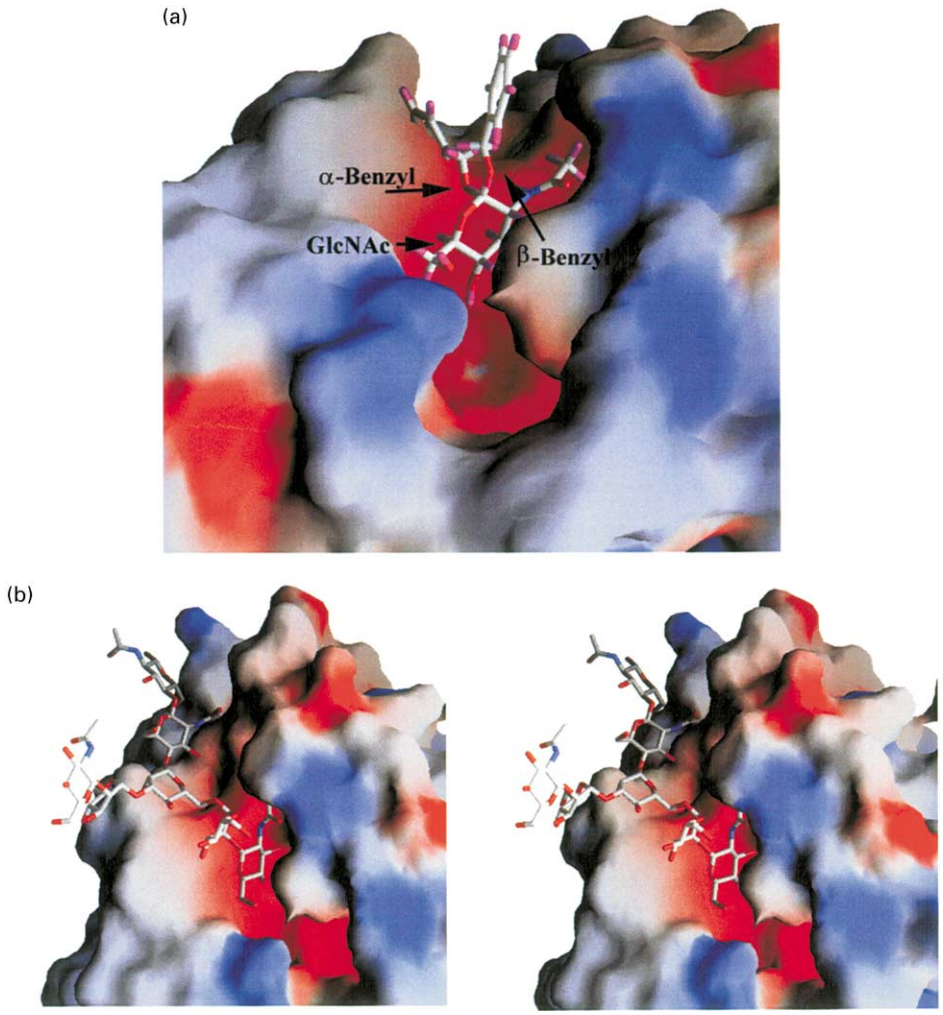

Figure 5. (a) Docking of both β-benzyl-GlcNAc and α-benzyl-GlcNAc to Gal-T1 in conformation II. The substrates are shown in a conformation with least contact with Gal-T1 molecule. As can be seen, the α-benzyl group exhibits steric hindrance with Gal-T1, while the β-benzyl group fits very well in the second sugar binding site; that is in agreement with the enzyme kinetic data, which show that α-benzyl-GlcNAc is less preferred, compared to β-benzyl-GlcNAc. (b) Binding of a biantennary *N*-glycan with GlcNAc at its non-reducing ends to Gal-T1. A stereo view showing the GlcNAc of the α-1-6 arm of GlcNAcβ1-2Manα1-6Manβ1-4GlcNAcβ1-4GlcNAcβ- is bound in the GlcNAc binding site on Gal-T1, while the α-1-3 arm is placed perpendicular to the *N*-glycan. In the α-1-6 arm bound conformation, the torsional values for the four disaccharide linkages are: (1) (φ, ψ) for the GlcNAcβ1-2Man linkage is (118°, 123°); (2) (φ, ψ, χ) for the Manα1-6Man linkage is (−124°, 166°, 114°); (3) (φ, ψ) for the Manβ1-4GlcNAc linkage is (−113°, 135°); and (4) (φ, ψ) for the GlcNAcβ-4GlcNAc linkage is (−98°, 120°). In the oligosaccharide binding site, the asparagine-linked GlcNAc of the chitobiose unit is pointed away from Gal-T1. The presence of the protein linked to this GlcNAc may not influence or interfere with the oligosaccharide binding to Gal-T1. The protein molecules are illustrated using the surface electrostatic potential.

present in an α-helix conformation. The presence of these residues in an α-helix structure is important as it is only in this conformation they can form a hydrophobic pocket that facilitates the binding of the *N*-acetyl group.[9] In conformation I, these residues are in random coil conformation and bury the GlcNAc binding site. In the present crystal structure, even though there is no GlcNAc, its binding site is fully present and is quite similar to that observed in the crystal structure of the Gal-T1·LA·GlcNAc complex.[9] Therefore, these observations indicate that the conformational change induced by UDP-Gal is essential for the creation of the acceptor binding site on Gal-T1.

Defining the oligosaccharide binding site by docking the substrates

Examination of the present crystal structure and the GlcNAc binding site in the Gal-T1·LA·GlcNAc crystal structure[9] revealed an "open canal" shaped extended sugar binding site, with an average width and length of 10 and 16 Å, respectively, that lies behind the GlcNAc binding site (Figure 4(c) and (d)). This site is formed by the residues from three regions: residues 280–289, residues 319–325, and residues 359–368. In conformation I, the region 359–368 is placed in such a way that there is no oligosaccharide binding site. Therefore, it is

```
                 286       319            360
                  |         |              |
B_β4GalT1      ..YVQY....DDIYNRLAF....FDRIAHTK..
H_β4GalT1      ..YVQY....DDIFNRLVF....FDRIAHTK..
H_β4GalT2      ..YAGY....DDIFNRISL....FTKIQNTK..
H_β4GalT3      ..YPQY....DDIATRVRL....FDLLVRTQ..
H_β4GalT4      ..YSGY....DDLRLRVEL....MKLLHQVS..
H_β4GalT5      ..YTEF....DDLWNRVQN....YALLRKSK..
H_β4GalT6      ..YKEF....DDLWNRVHY....YKLLRYSK..
```

Figure 6. Sequence comparison of the human Gal-T family members at the oligosaccharide binding site. Although the overall sequence identity between Gal-T5 and Gal-T6 is 92%, the sequences at the oligosaccharide binding site show differences between these family members. Residues shown in bold are those that constitute the oligosaccharide binding site.

clear that a change in Gal-T1 from conformation I to II is essential for the creation of an acceptor site that would bind the extended sugar acceptor. Since LA also binds to this region, as seen in the crystal structure of the Gal-T1·LA complex,[9] it can be expected to compete with the GlcNAc-containing oligosaccharides such as chitobiose. Enzyme kinetics show that when chitobiose is used as the acceptor molecule, LA inhibits the catalytic reaction. In fact, such an experiment is used to determine the LA affinity (K_i) for Gal-T1. Detailed enzyme kinetic studies have shown that disaccharides, such as GlcNAcβ,4-GlcNAc are preferred ($K_m = 1.9$ mM) over the monosaccharide GlcNAc ($K_m = 5.1$ mM).[2] These studies also show that of the different GlcNAc-containing disaccharides, the β-linked disaccharides such as GlcNAcβ1,4-GlcNAc or GlcNAcβ1,2-Man are preferred over α-linked disaccharides. Also, oligosaccharides such as *N*-glycan is preferred ($K_m = 0.4$ mM) over $(GlcNAc)_4$ ($K_m = 1.2$ mM).[2] The presence of a protein or peptide attached to such *N*-glycans does not influence the binding of an oligosaccharide.[2] Thus, all these experimental observations indicate that an extended oligosaccharide binding site exists on Gal-T1. Taking into consideration these experimental results, we have carried out a modeling study using various disaccharides and *N*-glycans to probe the size and the nature of the oligosaccharide binding site in Gal-T1. These modeling studies were carried out using the simple docking method (described in Materials and

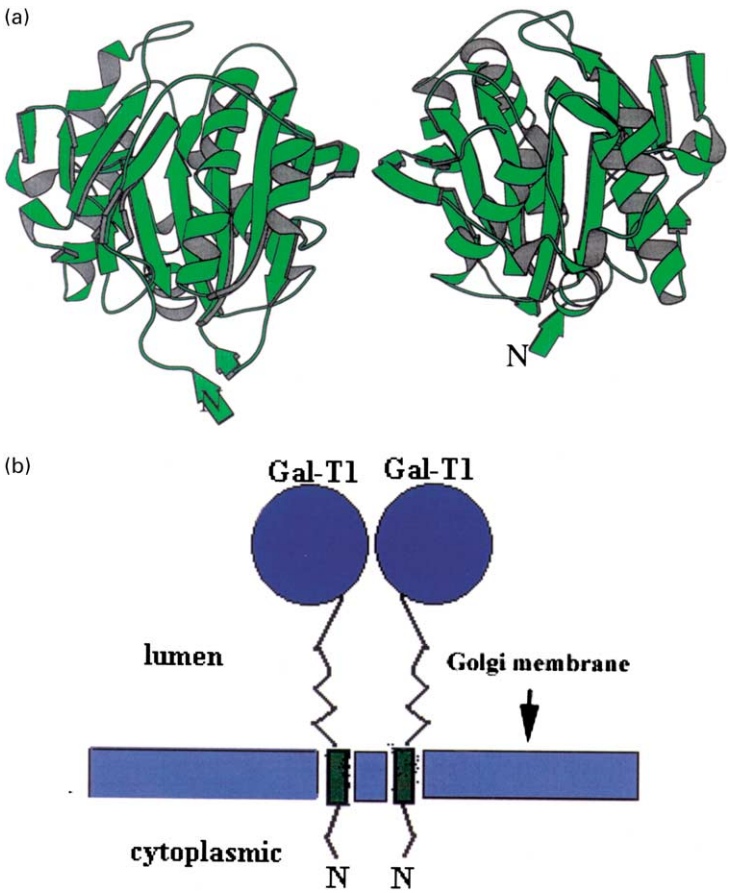

Figure 7. (a) The two Gal-T1 molecules in the asymmetric unit are related by a pseudo-2-fold symmetry, forming a homo-dimer. The dimerization interactions are through the C-terminal β10 sheet of each molecule, forming an intermolecular antiparallel β-sheet. The N termini of these two homodimers face in the same direction. It seems unlikely that the presence of the stem, transmembrane and the cytoplasmic domains would interfere with the dimerization. (b) Based on the crystal structures a schematic diagram showing the likelihood of the assembly of two Gal-T1 molecules in the Golgi (a).

Methods); no energy minimization was carried on either of the molecules. It was found that GlcNAc with an α-linked substitution, such as α-benzyl-GlcNAc, cannot bind to Gal-T1 because of severe steric contacts with the highly conserved Tyr286 residue, whereas GlcNAc with a β-linked substitution such as β-benzyl-GlcNAc can bind without any steric hindrance (Figure 5(a)). Docking of a biantennary-glycan with GlcNAc at its reducing ends in the extended sugar binding site reveals that the acceptor binding site in Gal-T1 can accommodate a linear pentasaccharide all the way from the GlcNAc moiety to the asparagine-linked GlcNAc. The binding site can also accommodate either the α-1-6 arm (GlcNAcβ1-2Manα1-6 Manβ1-4 GlcNAc β1-4 GlcNAcβ-N) (Figure 5(b)) or α-1-3 arm (GlcNAcβ1-2Manα1-3Manβ1-4 GlcNAc β1-4 GlcNAc β-N) (not shown) of the *N*-glycan without any steric hindrance.

In glycoproteins and glycolipids, the *N*-acetyl-lactosamine (Galβ1-4GlcNAc) moiety is found in a variety of oligosaccharides, such as *N*-glycan, type 1 or 2 core sugars. In humans, Gal-T family members are responsible for the synthesis of Gal moiety of a variety of the different oligosaccharides.[3] This implies that these enzymes also recognize the remaining monosaccharide units of the oligosaccharide moiety to which GlcNAc is attached. Therefore, the oligosaccharide binding site defined on bovine Gal-T1 will be important in elucidating the specificity of human Gal-T family members. The protein sequence comparison of the GlcNAc binding site of the Gal-T family members reveals little or no variation among the family members, whereas the extended oligosaccharide binding region showed significant variations (Figure 6), indicating that each enzyme may prefer different GlcNAc containing oligosaccharide as its preferred sugar acceptor. However, investigations into the exact nature of their preferences, based on the homology modeling and the crystal structure determination of human Gal-T family members, are underway.

Dimerization of Gal-T1 and its relevance to Golgi

In the crystal structure, two Gal-T1 molecules are present in the asymmetric unit. They are related by a pseudo-2-fold symmetry, forming homo-dimers (Figure 7(a)). The dimerization is through the formation of an intermolecular antiparallel β-sheet through the β10 sheet of each molecule, with eight hydrogen bonds between these two sheets. Such dimerization is also consistently observed in all the crystal structures of Gal-T1 complex with LA.[9] In the Gal-T1 homo-dimers, a total of 424 Å^2 solvent accessible surface is hidden between the two Gal-T1 molecules. This is quite significant when compared to that of Gal-T1 and LA bound heterodimer where 1310 Å^2 solvent accessible surface is buried between Gal-T1 and LA molecules. It has been suggested, based on the biochemical studies, that in the Golgi apparatus Gal-T1 exists as a dimer.[28] Although the Gal-T1 molecules in the native crystal structure[6] also form a dimer, the dimerization involves residues in the catalytic pocket, and the mode of dimerization is different from that observed in the present studies. In the present study and in all the Gal-T1·LA complex structures, the dimerization region is found to be away from the catalytic pocket, as well as from the LA binding site (Figure 7(a)). Furthermore, the N-terminal in these structures is also away from the homo-dimers, indicating that the presence of the stem, the transmembrane and the cytoplasmic domains would not interfere in the dimerization. Thus, it seems that the dimers described here are likely to be similar to the dimerization of Gal-T1 in the Golgi, where the catalytic pockets and LA binding sites of the two monomers face away from each other, thereby facilitating the binding of the substrates and LA without dissociating into monomers (Figure 7(b)).

Conclusion

The present study: (1) describes the detailed binding of Gal moiety of the donor molecule, UDP-Gal, to Gal-T1; (2) reveals the oligosaccharide binding site on Gal-T1; and (3) shows that the binding of UDP-Gal alone induces a conformational change in Gal-T1. The change in conformation creates not only the GlcNAc binding pocket on Gal-T1, but also a site that can bind the extended sugar acceptor, its natural substrate. The sugar binding site can accommodate an *N*-glycan, pentasaccharide in length. Sequence comparison of the human Gal-T family members shows sequence variations in their oligosaccharide binding site, which would account for their respective preferences for different oligosaccharides.

Materials and Methods

Crystallization and crystal data

C342T-Gal-T1 protein was prepared and purified as described.[10] Crystals of the catalytic domain of the recombinant bovine Gal-T1, residues 130–402 (d129-Gal-T1) (M_r~33 kDa) and the mutant C342T were grown at room temperature by hanging drop methods, using 20 mg ml^{-1} of Gal-T1 in the presence of 17 mM UDP-Gal and 17 mM $MnCl_2$ with the precipitant containing 10% (v/v) dioxane, Mes-NaOH buffer 100 mM (pH 6.5) and 2–2.5 M ammonium sulfate. Since the crystals of wild-type diffracted very poorly (only at 4–7 Å resolution), C342T mutant crystals were used. Although the crystals diffracted to 2.3 Å resolution at room temperature, they diffracted to a maximum resolution of 2.8 Å at 100 K. For room temperature measurements, typically a C342T mutant crystal of 0.15 mm × 0.15 mm × 0.4 mm dimension was mounted in a glass capillary with mother liquid on one side. The crystals were tested using CuKα radiation, on an in-house X-ray facility with a 30 cm MAR research imaging plate mounted on a rotating anode X-ray generator operating at 50 kV and 100 mA. Since the crystals exhibited severe

radiation damage in less than an hour, data could not be collected at room temperature. Trials using a number of cryo-protectants showed that the best data up to only 2.8 Å resolution could be collected at 100 K using an cryo-protectant containing the precipitant and 20% (w/v) sucrose. Complete three-dimensional X-ray diffraction data were collected at beam line X9B, NSLS, BNL, using a Quantum-4 ccd detector. The crystal data, data collection statistics are given in Table 1. The frames were processed using DENZO.[29]

The crystal structures were solved by molecular replacement methods using AmoRe.[30] The crystal structure of Gal-T1 from the lactose synthase[9] with neither the substrate nor LA, was used as a model for molecular replacement. The two protein molecules of a dimer in the asymmetric units are related by a pseudo-2-fold symmetry. Although initial refinement with and without NCS restraint did not show any significant difference in R_{free}, due to the low resolution, NCS restraint was used throughout the refinement. After initial refinement, the difference electron density maps revealed the presence of a UDP-Gal molecule and Mn^{2+} bound to the Gal-T1. The models were fitted into the electron density using O.[31] Further refinements were carried out using simulated annealing followed by B-group refinement. In the B-group refinement, the main and the side-chain atoms in each residue are considered as two individual groups. The model was corrected for any deviations based on the omit electron density maps. The coordination distances for a Mn^{2+} were not restrained. Due to the lack of an adequate number of reflections, only B-group refinement was carried out throughout the refinement stages. All the refinements were carried out initially by X-PLOR3.85, followed by CNS1.0.[32] The final refinement statistics are given in Table 1. All Figures were drawn using MOLSCRIPT[33] and GRASP.[34]

Docking studies

Docking of various di- and oligosaccharides into Gal-T1 were carried out using InsightII molecular modeling package (MSI, CA) on an Indigo2 workstation (SGI, CA). The coordinates for the ligands were generated using standard geometry and the bond angle at the glycosidic oxygen was set to 117.5°.[35] All the protein and ligand atoms were considered explicitly. In all the docking studies, the conformation and orientation of GlcNAc in the primary binding site were fixed as observed in the Gal-T1·LA·GlcNAc complex, since Gal-T1·GlcNAc interactions were found to be optimized for binding and catalysis in this complex.[9] The conformation of the ligand was varied in the Gal-T1 binding site by systematically scanning the entire stereochemically allowed region of the ligand. For each ligand conformation, inter-molecular interaction energy between Gal-T1 and the saccharide was calculated using the Discover module of InsightII. The CVFF force field was used for energy calculations and only residues within 9 Å from any of the ligand atoms were considered for energy calculations. The total energy, comprising inter-molecular protein–ligand interaction energy and intra-molecular ligand energy, was used as a guide in determining the possible allowed conformations of the ligand in the binding site.

Protein Data Bank accession numbers

The coordinates have been deposited under PDB ID code 1KYB.

Acknowledgments

We thank Dr Zbgniew Dauter at NSLS, BNL, for help with synchrotron data collection. We also thank Drs Soma Kumar and Elizabeth Boeggeman for helpful discussions and comments during preparation of the manuscript. The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organization imply endorsement by the US Government. This project has been funded in part with Federal funds from the NCI, NIH, under contract number NO1-CO-56000.

References

1. Brew, K., Vanaman, T. C. & Hill, R. L. (1968). The role of alpha-lactalbumin and the A protein in lactose synthetase: a unique mechanism for the control of a biological reaction. *Proc. Natl Acad. Sci. USA*, **59**, 491–497.
2. Prieels, J. P., Dolmans, M., Schindler, M. & Sharon, N. (1976). The binding of glycoconjugates to human-milk D-galactosyltransferase. *Eur. J. Biochem.* **66**, 579–582.
3. Lo, N. W., Shaper, J. H., Pevsner, J. & Shaper, N. L. (1998). The expanding beta 4-galactosyltransferase gene family: messages from the databanks. *Glycobiology*, **8**, 517–526.
4. Amado, M., Almeida, R., Schwientek, T. & Clausen, H. (1998). Identification and characterization of large galactosyltransferase gene families: galactosyltransferases for all functions. *Biochim. Biophys. Acta*, **1473**, 35–53.
5. Almeida, R., Levery, S. B., Mandel, U., Kresse, H., Schwientek, T., Bennett, E. P. & Clausen, H. (1999). Cloning and expression of a proteoglycan UDP-galactose: beta-xylose beta1,4-galactosyltransferase I. A seventh member of the human beta4-galactosyltransferase gene family. *J. Biol. Chem.* **274**, 26165–26171.
6. Gastinel, L. N., Cambillau, C. & Bourne, Y. (1999). Crystal structures of the bovine beta4-galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *EMBO J.* **18**, 3546–3557.
7. Brodbeck, U., Denton, W. L., Tanahashi, N. & Ebner, K. E. (1967). The isolation and identification of the B protein of lactose synthetase as alpha-lactalbumin. *J. Biol. Chem.* **242**, 1391–1397.
8. Qasba, P. K. & Kumar, S. (1997). Molecular divergence of lysozymes and alpha-lactalbumin. *Crit. Rev. Biochem. Mol. Biol.* **32**, 255–306.
9. Ramakrishnan, B. & Qasba, P. K. (2001). Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the beta1,4-galactosyltransferase-I. *J. Mol. Biol.* **310**, 205–218.
10. Powell, J. T. & Brew, K. (1976). A comparison of the interactions of galactosyltransferase with a glycoprotein substrate (Ovalbumin) and with alpha-lactalbumin. *J. Biol. Chem.* **251**, 3653–3663.
11. Ramakrishnan, B., Shah, P. S. & Qasba, P. K. (2001). Alpha-lactalbumin (LA) stimulates milk beta-1,4-galactosyltransferase-I to transfer glucose from UDP-glucose to *N*-acetylglucosamine: crystal structure of beta4Gal-T1·LA complex with UDP-Glc. *J. Biol. Chem.* **276**, 37665–37671.

12. Geren, C. R., Magee, S. C. & Ebner, K. E. (1975). Circular dichroism changes in galactosyltransferase upon substrate binding. *Biochemistry*, **14**, 1461–1463.
13. Clymer, D. J., Geren, R. & Ebner, K. E. (1976). Ultraviolet photoinactivation of galactosyltransferase. Protection by substrates. *Biochemistry*, **15**, 1093–1097.
14. Aoki, D., Appert, H. E., Johnson, D., Wong, S. S. & Fukuda, M. N. (1990). Analysis of the substrate binding sites of human galactosyltransferase by protein engineering. *EMBO J.* **9**, 3171–3178.
15. Zu, H., Fukuda, M. N., Wong, S. S., Wang, Y., Liu, Z., Tang, Q. & 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.
16. Zhang, Y., Malinovskii, V. A., Fiedler, T. J. & Brew, K. (1999). Role of a conserved acidic cluster in bovine beta1,4-galactosyltransferase-1 probed by mutagenesis of a bacterially expressed recombinant enzyme. *Glycobiology*, **8**, 815–822.
17. Ramakrishnan, B. & Qasba, P.K. (1999). Fourth Annual Conference of the Society for Glycobiology, 11–14 November 1999, San Francisco, CA (Poster 58)
18. Powell, J. T. & Brew, K. (1976). Metal ion activation of galactosyltransferase. *J. Biol. Chem.* **251**, 3645–3652.
19. Wiggins, C. A. & Munro, S. (1998). Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc. Natl Acad. Sci. USA*, **95**, 7945–7950.
20. Unligil, U. M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H. & Rini, J. M. (2000). X-ray crystal structure of rabbit *N*-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. *EMBO J.* **19**, 5269–5280.
21. Pedersen, L. C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T. A. & Negishi, M. (2000). Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I. *J. Biol. Chem.* **275**, 34580–34585.
22. Persson, K., Ly, H. D., Dieckelmann, M., Wakarchuk, W. W., Withers, S. G. & Strynadka, N. C. (2001). Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nature Struct. Biol.* **8**, 166–175.
23. Gastinel, L. N., Bignon, C., Misra, A. K., Hindsgaul, O., Shaper, J. H. & Joziasse, D. H. (2001). Bovine alpha1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases. *EMBO J.* **20**, 638–649.
24. Charnock, S. J. & 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.
25. Berliner, L. J. & Robinson, R. D. (1982). Structure–function relationships in lactose synthase. Structural requirements of the uridine 5′-diphosphate galactose binding site. *Biochemistry*, **21**, 6340–6343.
26. Endo, T., Kajihara, Y., Kodama, H. & Hashimoto, H. (1996). Novel aspects of interaction between UDP-gal and GlcNAc beta-1,4-galactosyltransferase: transferability and remarkable inhibitory activity of UDP-(mono-*O*-methylated gal), UDP-Fuc and UDP-man. *Bioorg. Med. Chem.* **4**, 1939–1948.
27. Do, K. Y., Do, S. I. & Cummings, R. D. (1995). Alpha-lactalbumin induces bovine milk beta 1,4-galactosyltransferase to utilize UDP-GalNAc. *J. Biol. Chem.* **270**, 18447–18451.
28. Yamaguchi, N. & Fukuda, M. N. (1995). Golgi retention mechanism of beta-1,4-galactosyltransferase. Membrane-spanning domain-dependent homodimerization and association with alpha- and beta-tubulins. *J. Biol. Chem.* **270**, 12170–12176.
29. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
30. Navaza, J. (2001). Implementation of molecular replacement in AmoRe. *Acta Crystallog. sect. D*, **57**, 1367–1372.
31. Jone, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these methods. *Acta Crystallog. sect. A*, **47**, 110–119.
32. Brunger, A. T. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
33. Karulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
34. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
35. Rao, V. S. R., Qasba, P. K., Balaji, P. V. & Chandrasekaran, R. (1998). *Conformation of Carbohydrates*, Harwood Academic Publishers, New York.
36. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760.

Edited by R. Huber

(Received 21 September 2001; received in revised form 18 January 2002; accepted 24 January 2002)