Glycobiology vol. 16 no. 2 pp. 29R–37R, 2006
doi:10.1093/glycob/cwj016
Advance Access publication on July 21, 2005

REVIEW

Structures and mechanisms of glycosyltransferases

Christelle Breton[1,2], Lenka Šnajdrová[2,3], Charlotte Jeanneau[2], Jaroslav Koča[3], and Anne Imberty[1,2]

[2]CERMAV–CNRS, affiliated with Université Joseph Fourier, PO Box 53, 38041 Grenoble cedex 9, France; and [3]NCBR, Masaryk University, Kotlarska 2, CZ-611 37 Brno, Czech Republic

Accepted on July 13, 2005

Glycosyltransferases (GTs) catalyze the transfer of a sugar moiety from an activated donor sugar onto saccharide and nonsaccharide acceptors. A sequence-based classification spreads GTs in many families thus reflecting the variety of molecules that can be used as acceptors. In contrast, this enzyme family is characterized by a more conserved three-dimensional architecture. Until recently, only two different folds (GT-A and GT-B) have been identified for solved crystal structures. The recent report of a structure for a bacterial sialyltransferase allows the definition of a new fold family. Progress in the elucidation of the structures and mechanisms of GTs are discussed in this review. To accommodate the growing number of crystal structures, we created the 3D-Glycosyltransferase database to gather structural information concerning this class of enzymes.

Key words: fold recognition/glycosyltransferase/mechanism/superfamily/three-dimensional structure

Introduction

Glycosyltransferases (GTs; EC 2.4.x.y) constitute a large family of enzymes that are involved in the biosynthesis of oligosaccharides, polysaccharides, and glycoconjugates (Taniguchi *et al.*, 2002). These molecules of enormous diversity mediate a wide range of functions from structure and storage to signalling. Particularly abundant are the GTs that transfer a sugar residue from an activated nucleotide sugar donor, to specific acceptor molecules, forming glycosidic bonds. Transfer of the sugar residue occurs with either the retention or the inversion of the configuration of the anomeric carbon. These enzymes are present in both prokaryotes and eukaryotes, and they generally display exquisite specificity for both the glycosyl donor and the acceptor substrates. In eukaryotes, most of the glycosylation reactions that generate the diversity of oligosaccharide structures of eukaryotic cells occur in the Golgi apparatus. Golgi resident GTs are type-II transmembrane proteins with a large C-terminal globular catalytic domain facing the luminal side. Recent success in X-ray crystal structure determinations of Golgi GTs has provided a molecular basis accounting for donor and acceptor substrate specificities as well as catalysis. This review deals with the most recent structural data concerning sugar nucleotide-dependent GTs. It also includes fold recognition and homology-modelling studies that were shown to be of value for identifying the folds of many GT families or for rationalizing experimental data.

Sequence-based GT families

GTs have been classified into families by amino acid sequence similarities (Coutinho *et al.*, 2003) (available at http://afmb.cnrs-mrs.fr/CAZY). In addition to nucleotide sugar-dependent GTs, the CAZY database also integrates GTs that utilize dolichol-phospho-sugars, sugar-1-phosphates, and lipid diphospho-sugars as activated donors. At the time of writing (May 2005), the database comprises more than 12,000 known and putative GT sequences that have been divided into 78 families (denoted as GTx). Approximately 240 human sequences have already been listed in this database where they fall into 42 distinct families. However, not all GT sequences are present in the database, and the number of families will likely increase with the discovery of new GT genes. Large differences in the number and function of GTs are observed among families. A few families comprise a huge number of sequences from various sources with diverse functions. This is best illustrated with the family GT2 which contains more than 3500 sequences, originating from animal, plant, yeast, and bacterial species, and for which at least 12 distinct GT functions have already been characterized. The members of family 2 include cellulose synthase, chitin synthase, mannosyltransferase, glucosyltransferase, galactosyltransferase, rhamnosyltransferase, and so on. In contrast, other families are monofunctional and contain only a few sequences. In the case of "monospecific" families, sequence similarities are generally observed for the whole catalytic domain, whereas for the large "polyspecific" families, such as GT2 or GT4, they are mostly restricted to only a portion of the catalytic domain.

The prediction of the function of a putative GT of sequence homology can be problematic because there are many examples of closely related sequences having different catalytic activity. The best example comes from the study of blood group A and B transferases, which differ by only four amino acids. They use the same H-antigen acceptor but a different glycosyl donor (uridine diphospho N-acetyl galactosamine [UDP-GalNAc] for the A transferase and UDP-Gal

[1]To whom correspondence should be addressed; e-mail: breton@cermav.cnrs.fr and imberty@cermav.cnrs.fr

for the B transferase) (Yamamoto *et al.*, 1990). Therefore, if an open reading frame is related to a large polyspecific family, its precise function cannot be reliably predicted. This sequence-based classification is assumed to integrate both structural and mechanistic features within each family (Coutinho *et al.*, 2003). When applied to glycoside hydrolases, this classification system was shown to correlate well with enzyme mechanism (inversion or retention of the anomeric configuration), because once established for a member of a family, the mechanism can be safely extended to all other members of the family (Davies and Henrissat, 1995). This may not apply so safely to all GT families, because distant similarities between inverting and retaining CAZY families have been noted, suggesting that these families share common ancestors (Franco and Rigden, 2003; Liu and Mushegian, 2003). A similar observation has been made for the retaining GT27 family (corresponding to the animal polypeptide-αGalNAc transferases), which displays sequence similarities with inverting enzymes (Breton *et al.*, 1998), and that was initially grouped into GT2 family (Campbell *et al.*, 1997). Another example is given by the GT52 family, which may comprise both inverting (α2,3-sialyltransferase) and retaining (α2-glucosyltransferase) enzymes that use different nucleotide donors, cytidine monophospho (CMP)-β-N-acetyl neuraminic acid (NeuAc) and UDP-α-Glc, respectively. Very recently, the three-dimensional structure of a mannosylglycerate synthase (MGS) from *Rhodobacter marinus* has been published (Flint *et al.*, 2005). Although MGS, at the sequence level, displays similarity with protein members of family GT2, it is a retaining enzyme that has been classified in a new family (GT78). Altogether, these results demonstrate that the catalytic mechanism (inverting or retaining) cannot always be predicted with reliability from sequence comparison alone.

Crystal structures of GTs

Difficulties with high-level expression, purification, and crystallization hampered crystal structure determinations for GT enzymes. The first X-ray structure was reported in 1994 for bacteriophage T4-glucosyltransferase, an enzyme that transfers glucose from UDP-Glc to phage-modified DNA (Vrielink *et al.*, 1994). Since then, >100 crystal structures have been described for proteins corresponding to 23 different GTs, from prokaryotes and eukaryotes. Structural information is now available for 17 distinct GT families, including both retaining and inverting enzymes (Table I). In contrast to glycosylhydrolases that adopt a large variety of folds, including all α, all β, or mixed α/β structures, GT folds have been observed to consist primarily of α/β/α sandwiches (Figure 1), similar or very close to the Rossmann-type fold, a classical structural motif (six-stranded parallel β-sheet with 321456 topology) found in many nucleotide-binding proteins (Lesk, 1995). Until recently, only two structural superfamilies have been described for GTs, named GT-A and GT-B, and which were first observed in the original SpsA and β-glucosyltransferase (BGT) structures, respectively (Vrielink *et al.*, 1994; Charnock and Davies, 1999). A third family has recently emerged which comprises the bacterial sialyltransferase (CstII) belonging to family GT42 (Chiu *et al.*, 2004). This protein displays a similar type of fold than GT-A, but with some differences, so it can be considered as a new fold. The GT-A and GT-B folds are also shared by non-GT enzymes, such as nucleotidyltransferases and sugar epimerases, respectively (Brown *et al.*, 1999; Campbell *et al.*, 2000).

The GT-B fold consists of two separate Rossmann domains with a connecting linker region and a catalytic site located between the domains. There is an excellent structural conservation between protein members of the GT-B family, particularly in the C-terminal domain which corresponds to the nucleotide-binding domain. Variations are more pronounced in the N-terminal domains, in the loops and helices which point towards the active site, which have evolved to accommodate very different acceptors. Peptide motifs characteristic of the GT-B fold, notably a glutamate residue and glycine-rich loops interacting with the ribose and phosphate moieties of nucleotide donor, respectively, have been described (Wrabl and Grishin, 2001). But so far, GT-B enzymes do not appear to share any strictly conserved residue (Hu and Walker, 2002). Although divalent cations may be required for full activity of GT-B enzymes, there is no evidence of a bound metal ion associated with catalysis.

The GT-A fold consists of an α/β/α sandwich (a seven-stranded β-sheet with 3214657 topology in which strand 6 is antiparallel to the rest) that resembles a Rossmann fold. The central β-sheet is flanked by a smaller one, and the association of both creates the active site. A general feature of all enzymes of the GT-A family is the presence of a common motif, the DxD motif, and also their requirement for a divalent cation for activity (Breton *et al.*, 1998; Breton and Imberty, 1999). The DxD motif is shown in all crystal structures to interact primarily with the phosphate groups of nucleotide donor through the coordination of a divalent cation, typically Mn^{2+} (Figure 2). Depending on the GT, the two aspartate amino acids are not always conserved, but this particular motif, or its variants, can always be identified at the same location, in a short loop connecting one β-strand of the main α/β/α sandwich to a smaller one (Figure 1A). Comparison of the catalytic domains of enzymes of the GT-A family revealed the presence of two regions that are structurally well conserved in all members of the GT-A family, including inverting and retaining enzymes (Figure 2). This suggests that common structural elements are necessary for the glycosyl transfer reaction, irrespective of the stereochemistry of the reaction. The first region mostly corresponds to the Rossmann-type nucleotide-binding domain, encompassing the first 100–120 residues, and that is terminated by the DxD motif. The key amino acids that interact with UDP are mainly found at the C-term of strands β1 and β4. In some crystal structures, residues in the C-terminus of the catalytic domain were shown to make additional contacts with UDP. Differences are observed in the function of the residues of the DxD motif in retaining and inverting enzymes. In retaining enzymes, the two aspartate residues can interact with the Mn^{2+} ion, whereas only the last aspartate interacts with the metal cation in inverting enzymes (Persson *et al.*, 2001; Tarbouriech *et al.*, 2001). In both cases, the variable amino acid of the DxD motif (usually a polar or an aliphatic residue of moderate

Table I. Glycosyltransferases with available crystal structures

Organism	Glycosyltransferase	Name	GT family	Number of structure	Reference
Virus					
Phage T4	β-Glucosyltransferase	BGT	GT63	18	Vrielink *et al.* (1994)
Prokaryotes					
Agrobacterium tumefaciens	Glycogen synthase 1	AtGS	GT5	2	Buschiazzo *et al.* (2004)
Amycolatopsis orientalis	β-Epi-vancosaminyltransferase	GtfA	GT1	2	Mulichak *et al.* (2003)
	β-Glucosyltransferase	GtfB	GT1	1	Mulichak *et al.* (2001)
	β-Vancosaminyltransferase	GtfD	GT1	1	Mulichak *et al.* (2004)
Bacillus subtilis	Putative glycosyltransferase	SpsA	GT2	5	Charnock and Davies (1999)
Campylobacter jejuni	α-2,3/2,8-Sialyltransferase	CstII	GT42	2	Chiu *et al.* (2004)
Escherichia coli	β-1,4-GlcNAc transferase	MurG	GT28	2	Ha *et al.* (2000)
	Trehalose-6-phosphate synthase	OtsA	GT20	3	Gibson *et al.* (2002)
	Heptosyltransferase II	RfaF	GT9	1	PDB 1PSW
Neisseria meningitidis	α-1,4-Galactosyltransferase	LgtC	GT8	3	Persson *et al.* (2001)
Rhodothermus marinus	Mannosylglycerate synthase	MGS	GT78	4	Flint *et al.* (2005)
Eukaryotes					
Yeast	α-1,2-Mannosyltransferase	Kre2p/Mnt1P	GT15	3	Lobsanov *et al.* (2004)
Mouse	α-1,4-N-Acetylhexosaminyltransferase	Extl2	GT64	4	Pedersen *et al.* (2003)
	Polypeptide-α-GalNAc transferase	ppGalNAc-T1	GT27	1	Fritz *et al.* (2004)
Rabbit	α-Glucosyltransferase	Glycogenin	GT8	3	Gibbons *et al.* (2002)
	β-1,2-GlcNAc transferase I	GnTI	GT13	3	Unligil *et al.* (2000)
Bovine	α-1,3-Galactosyltransferase	α3GalT	GT6	13	Gastinel *et al.* (2001)
	β-1,4-Galactosyltransferase I	β4GalT1	GT7	16	Gastinel *et al.* (1999)
Human	β-1,3-Glucuronyltransferase	GlcAT-I	GT43	2	Pedersen *et al.* (2000)
	β-1,3-Glucuronyltransferase	GlcAT-P	GT43	3	Kakuda *et al.* (2004)
	α-1,3-GalNAc transferase A	GTA	GT6	12	Patenaude *et al.* (2002)
	α-1,3-Galactosyltransferase B	GTB	GT6	6	Patenaude *et al.* (2002)

All data are available from the 3D-Glycosyltransferase database accessible from the Glyco3D site (http://www.cermav.cnrs.fr/glyco3d/). Only reference to the original work is indicated. Glycogen and starch phosphorylases (GT35) are not included in the 3D-Glycosyltransferase database.

size) is involved in ribose binding. The C-terminal portion is highly variable and is mostly dedicated to the recognition of the acceptor. However, a common structural motif is seen that corresponds to the region β6–α4–α5 (as shown for α4-galactosyltransferase [LgtC] in Figure 2) which forms part of the active site. This corresponds to the second structurally conserved region which comprises residues that were shown in some crystal structures to interact with both the donor sugar and the sugar acceptor (Persson *et al.*, 2001; Boix *et al.*, 2002; Pedersen *et al.*, 2002; Ramakrishnan *et al.*, 2002). In inverting enzymes, the presumed catalytic base was proposed in this region (a Asp or Glu residue at the beginning of α5-helix) (Pedersen *et al.*, 2002; Kakuda *et al.*, 2004). Despite the similarity of their spatial folds, GT-A and GT-B enzymes appear to be unrelated. These two superfamilies have members in the three domains of life. In addition, some bifunctional enzymes, such as the exostosin family (GT47/GT64), are expected to contain both GT-A and GT-B folds.

The third-fold family has recently emerged with the crystal structure of a sialyltransferase (CstII) from *Campylobacter jejuni*, a highly prevalent foodborne pathogen (Chiu *et al.*, 2004). This is the first crystal structure of an enzyme that utilizes a nucleotide monophosphosugar (CMP-NeuAc). This enzyme which belongs to GT42, displays a different type of α/β/α sandwich (a seven-stranded parallel β-sheet with 8712456 topology) and has no DxD motif.

The ratio of loops to secondary elements is high in GTs, and many crystal structures do not describe the entire catalytic domain, because the polypeptide extremities and/or several loops are flexible and do not present clear electron density. In both the GT-A family and bacterial sialyltransferase, flexible loops appear to play an important role in substrate binding. Of paramount importance is the demonstration for some of these enzymes of an ordered binding of donor and acceptor substrates linked to a donor substrate induced conformational change (Boix *et al.*, 2001, 2002; Ramakrishnan *et al.*, 2002). A disordered loop (or C-term extremity) in the free enzyme becomes ordered upon nucleotide sugar binding and creates a lid over the donor substrate where additional residues make direct contacts with the diphosphate moiety (Figure 2). This new conformation, called the closed active conformation, creates a pocket that will serve as binding site for the acceptor (Figure 3). Affinity

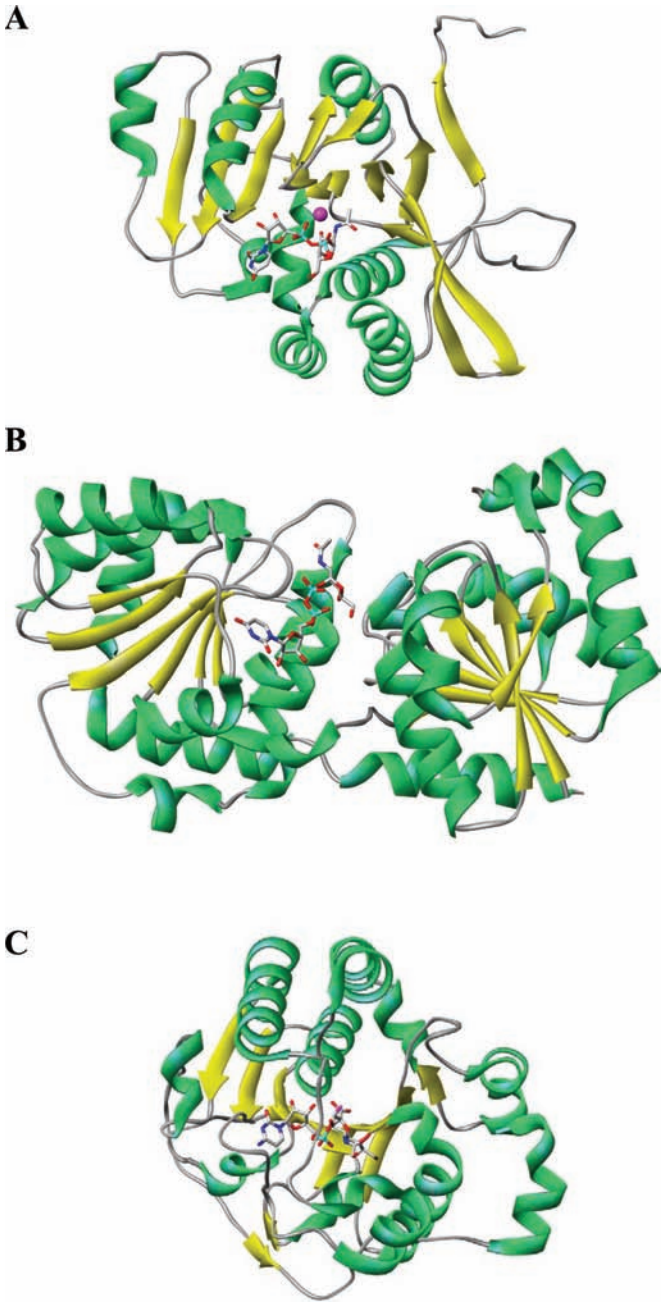

Fig. 1. Ribbon diagram of three glycosyltransferases (GTs) representative of the different folds. Bound nucleotide sugar are represented with stick model, and manganese, when present, by a ball. (**A**) GT-A fold, mouse α-1,4-*N*-acetylhexosaminyltransferase (EXTL2) complexed with UDP-GalNAc (PDB code 1OMZ) (Pedersen *et al*., 2003), (**B**) GT-B fold, *Escherichia coli* MurG complexed with UDP-GlcNAc (PDB code 1NLM) (Hu *et al*., 2003), and (**C**) *Campylobacter jejuni* sialyltransferase CstII complexed with cytidine monophospho 3-fluoro N-acetyl neuraminic acid (CMP-3FNeuAc) (PDB code 1RO7) (Chiu *et al*., 2004). Drawings were prepared with the Chimera program (Pettersen *et al*., 2004).

studies performed on LgtC (GT8), with the use of titration microcalorimetry confirmed that the open state (free enzyme) has no or little affinity for the oligosaccharidic acceptor (Boix *et al*., 2002). This complex mechanism is thought to prevent water molecules to act as acceptor for sugar transfer and, therefore, to limit the hydrolysis of energetically precious nucleotide sugar. Another significant feature is the observed distorted conformation of the bound nucleotide sugar in the active site that may be important in the catalytic mechanism (for a review, see Qasba *et al*., 2005).

A database is now available that provides an extensive list of the known three-dimensional structures of GTs (http://www.cermav.cnrs.fr/glyco3d). The GTs are classified not only according to CAZY systematics (Coutinho *et al*., 2003), but also according to the organism of origin, the linkage formed by the enzymatic reaction or the protein fold. Useful links permit retrieval of bibliographic information and also atomic coordinates at the Protein Data Bank and other protein databases with structural information. Images are provided to illustrate the details of protein/substrate interactions.

Molecular modelling of GTs

Molecular modelling of GTs is difficult. The number of available crystal structures is still limited. Only 17 of the 78 CAZY families include at least one GT for which an X-ray structure has been reported. The low degree of sequence similarity within some of the CAZY families and the absence of similarity between different families represent an unsurmountable barrier for classical sequence alignment procedure which is a prerequisite for homology building. Docking of substrates also appears to be a difficult task because of the flexibility of the nucleotide sugar and the presence of phosphate and divalent cation. Appropriate energy parameters have been recently developed for some modelling softwares (Petrova *et al*., 1999).

Fold recognition

Fold recognition is a theoretical approach which allows alignment of one sequence with one structure by a process called "threading" (Godzik, 2003). Although there is little or no sequence similarity between GTs, the limited number of observed folds facilitates the use of fold recognition methods to predict whether GT-A, GT-B, or "something else," will be the most probable for a given sequence. When performed on selected sequences representing all GT families present in CAZY database, such "threading" analyses predicted that many other GT families should adopt the GT-A or GT-B fold (Breton *et al*., 2002; Franco and Rigden, 2003). In 2002, of 56 GT families, we predicted with a high level of confidence the occurrence of either a GT-A or a GT-B fold for 23 families, five among them have later been confirmed by crystal structures: GT5, 9, 20, 27, 64 (formerly 47). Table II which gives the results of an updated fold-recognition study extended to include the new families that were performed using the three-dimensional position-specific scoring matrix (3D-PSSM) fold-recognition program (Kelley *et al*., 2000). With the growing number of crystal structures sharing a GT-A or GT-B fold and the existence of a new fold type (Cst II fold), the new picture that emerges from this study is the prediction of a GT-A fold for 22 families, a GT-B fold predicted for 13 families, and a CstII fold predicted for two families. Therefore, GT families, where a SpsA or a BGT fold has been experimentally

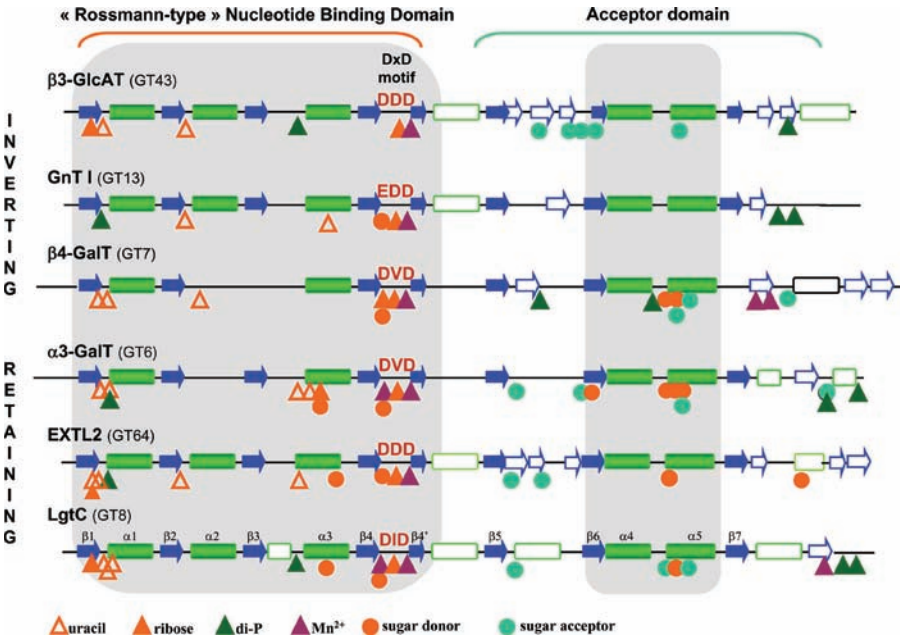

Fig. 2. Schematic representation of the catalytic domains of different members of the glycosyltransferase GT-A superfamily including inverting and retaining enzymes. The location of the two best structurally conserved regions are indicated by the gray shading. The first region encompasses ~100 residues at the N-terminus of the catalytic domain and corresponds roughly to the nucleotide-binding domain that is terminated by the DxD motif. The second region comprises the structural motif formed by β6-α4-α5 (α4-galactosyltransferase [LgtC] numbering) that is part of the acceptor domain. β-Strands and α-helices are represented by blue arrows and green cylinders, respectively. Plain arrows and cylinders correspond to the most conserved secondary structure elements in the GT-A superfamily and are denoted as indicated for the LgtC. Residues interacting with the nucleotide sugar and acceptor sugar are marked using the colour coding given at the bottom of the figure.

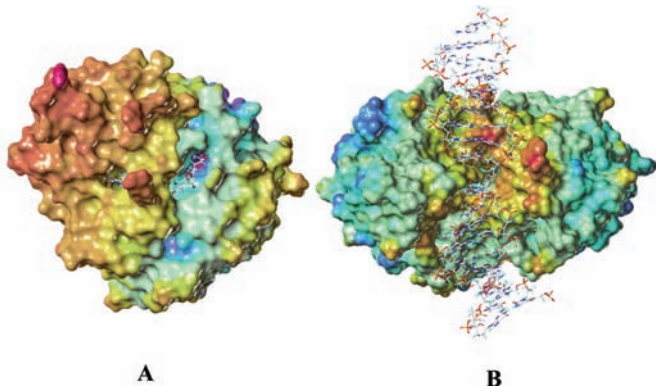

Fig. 3. Accessible surface representation of two glycosyltransferases (GTs). (**A**) *Neisseria meningitidis* α-1,4-galactosyltransferase LgtC complexed with donor and acceptor sugar analogs (PDB code 1GA8) (Persson *et al.*, 2001). (**B**) T4 phage β-glucosyltransferase (BGT) complexed with UDP and a 13-mer DNA duplex (PDB code 1IXY) (Lariviere and Morera, 2002). Connolly surfaces were calculated and represented using the MOLCAD program (Waldherr-Teschner *et al.*, 1992).

Table II. Classification of glycosyltransferase (GT) families in structural superfamilies

GT families	GT-A fold	GT-B fold	CstII fold
Known fold	2, 6, 7, 8, 13, 15, 27, 43, 64, 78	1, 5, 9, 20, 28, 35, 63	42
Predicted fold[a]	12, 14, 16, 21, 24, 25, 31, 32[b], 34, 40, 45, 46, 49, 54, 55, 60, 62, 67, 69, 71, 74, 77	3, 4, 10, 18, 19, 30, 33, 38, 41, 47, 56, 70, 72	29, 73

[a]Using 3D-PSSM program (Kelley *et al.*, 2000). GT families corresponding to integral membrane proteins with multitransmembrane domains were not included in this study (excluded GT families GT22, 39, 48, 50, 53, 57, 58, 59, 66, 76).
[b]GT-B fold predicted in Wrabl and Grishin (2001) and Liu and Mushegian (2003), using other methods (i.e., PSI-BLAST, HMMer).

demonstrated or predicted with a high probability level, now represent 51 (65%) of the 78 currently known families. Families corresponding to integral membrane proteins with multiple transmembrane domains were not considered in this study (named GT-C superfamily by some authors, Kikuchi *et al.*, 2003; Liu and Mushegian, 2003). It is striking to note that a CsT II fold can now be predicted with confidence for GT29, the enzyme family comprising the eukaryotic sialyltransferases. Some families gave weak or moderate scores in fold recognition (i.e., GT11, 17, 18, 23, 26, 37, 44, 52, 61, 65, 68, 75), and this probably means that novel folds can still be discovered in future structural studies. Probabilistic methods of database searching, such as PSI-BLAST (Altschul *et al.*, 1997) and HMMer (Eddy, 1998), can be used to detect homologs with lower sequence similarity. Using these methods, some authors were also able to predict a GT-A or GT-B fold for a number of GT families (Kikuchi *et al.*, 2003; Liu and Mushegian, 2003). No major conflicting results are observed between the different approaches

except for GT32 (Table II). Fold recognition has also been used in a more innovative manner to identify GTs in the fully sequenced genome of *Mycobacterium tuberculosis* (Wimmerova *et al.*, 2003). All of the 3995 translated peptide sequences were threaded on a small library of structures with GT-A, GT-B, and Rossmann folds. The resulting folding "scores" were combined with biochemical characteristics of the sequences (length, theoretical pI...) and treated with chemometric methods to discriminate and/or predict folds. Several sequences with unknown function could be proposed as putative GTs. The two-domain GT-B fold was easier to identify. Multivariate sequence analysis associated with fold recognition also prove to be a useful tool for predicting both folds and mechanisms (inverting or retaining) for the *Escherichia coli* and *Synechocystis* putative GTs present in the CAZY database (Rosen *et al.*, 2004).

Homology modelling and docking of substrates

Once a sequence has been attributed a fold, it is possible to predict the secondary structure elements, namely α-helices and β-strands, and to align it with sequences with known three-dimensional structure. When sequence identities are very low, comparison of hydrophobic clusters (HCA) has proven to be a useful method for aligning sequences with similar fold (Gaboriaud *et al.*, 1987). A new structure can then be modelled using homology-building methods. However, in general, models have a low confidence index for flexible loops and highly variable regions, and for GTs, this may be a major problem for modelling acceptor sites. This is illustrated by the modelling studies performed in GT-A family. The first attempt to model a GT was done on an α3-galactosyltransferase, using as a template the structure of BGT which was the only crystal structure available at that time. Although modelling was incorrect, the Rossmann-like topology of the nucleotide-binding domain and the role of the DxD motif in binding the divalent cation were later confirmed in the solved crystal structure of α3-galactosyltransferase (Gastinel *et al.*, 2001). In contrast, when the target and the template have sufficient identity, such as α3-GalT and the α3-GTs responsible for the synthesis of blood group A, blood group B, Forssman and iGb3 antigens, the models are accurate and allow for docking of nucleotide sugars and acceptors (Heissigerova *et al.*, 2003). Specificity towards the sugar donor and sugar acceptor was shown to be determined by a few critical residues in the binding site. In addition, a closed active conformation of the modelled enzymes was proposed that may complement data from the crystal structures of the blood group A and B transferases determined in an open conformation (Patenaude *et al.*, 2002).

Some homology modelling studies were performed in the GT-B family. The GT-B topology with a large cleft between the two domains appears to be well adapted to accommodate large acceptor molecules as shown in Figure 3B with the docking of DNA in the BGT structure. The configuration of the substrates in the proposed model (Moréra *et al.*, 1999) suggested the participation of an Asp residue (D100) as the catalytic base. A recent study supports a role in catalysis for this residue (Lariviere *et al.*, 2003). Edman and colleagues (Edman *et al.*, 2003), using the fully automated modelling server SWISS-MODEL (http://swissmodel.expasy.org//SWISSMODEL.html), proposed models for the monoglucosyl–diacylglycerol synthase and diglucosyl–diacylglycerol synthase of *Acholeplasma laidlawii* and *Streptococcus pneumonia* involved in the synthesis of membrane glucolipids. In plants, a large family of glucosyltransferases (belonging to GT1 family) are involved in the biosynthesis of natural products, such as alkaloids, terpenoids, and phenylpropanoids. Based on the crystal structure of a bacterial glucosyltransferase from *Amycolatopsis orientalis* that belongs to the same sequence family in CAZY, a model of the betanidin 5-*O*-glucosyltransferase from *Dorotheathus bellidiformis* has been proposed (Hans *et al.*, 2004) as well as of the cyanohydrin glycosyltransferase from *Sorghum bicolor* (Thorsøe *et al.*, 2005).

Molecular dynamics

Since both inverting and retaining glycosyltransferases undergo large conformational movements (Figure 4), the flexible loops that are proposed in GT mechanisms have been the subject of pioneering studies by molecular dynamic simulation. Two well-characterized systems, that is, the inverting bovine β4-galactosyltransferase (Ramakrishnan and Qasba,

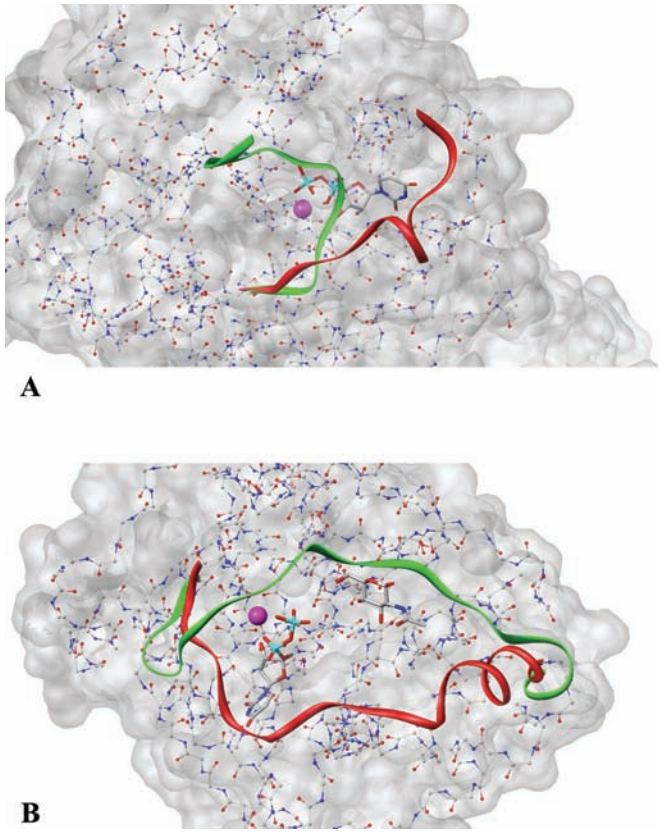

Fig. 4. Superimposition of open (green) and close (red) forms of glycosyltransferases (GTs). (**A**) Bovine α-1,3-galactosyltransferase: open form (PDB code 1G8O) (Gastinel *et al.*, 2001) and close form (PDB code 1K4V) (Boix *et al.*, 2001). (**B**) Bovine β-1,4-galactosyltransferase form I (PDB code 1PZT) and form II (PDB code 1PZY) (Ramasamy *et al.*, 2003). Drawings were prepared with the Chimera program (Pettersen *et al.*, 2004).

2001) and the retaining LgtC from *Neisseria meningitidis* (Persson *et al*., 2001), have been crystallized in their closed conformations. In both cases, when substrates are removed from the protein, several ns simulations with explicit water molecules allowed to see the first step of the loops opening (Gunasekaran and Nussinov, 2004; Snajdrová *et al*., 2004). Longer simulations would be required to visualize the complete loop movements (>20 Å change for one loop of β4-GalT), but these preliminary results demonstrate the correlated motions of several loops as well as the importance of contacts between loops in the mechanism.

Mechanisms

A specific feature of GTs is the use of an activated donor that can be a nucleoside diphosphosugar, nucleoside monophosphosugar, or lipid phosphosugar. Mechanistic analogies between GTs and glycosylhydrolases have been reviewed recently (Lairson and Withers, 2004). The mechanism of inverting GTs is believed to be similar to the one of inverting glycosylhydrolases with the requirement of one acidic amino acid that activates the acceptor hydroxyl group by deprotonation. All structural evidence to date supports the *SN2* mechanism originally proposed by Wong's group for α3-FucT (Murray *et al*., 1996). Mechanism of GT-A inverting GTs has the particularities to involve a Mn^{++} ion that plays the role of acid catalyst and also initiates a sequential ordered mechanism in which nucleotide sugar binding is followed by loops closing and acceptor binding (Ramakrishnan *et al*., 2004). GT-B fold-inverting GTs have completely different active sites with no ion involved. T4 phage BGT that transfers glucose to modified DNA also uses an *SN2* mechanism with the participation of one acidic amino acid from the protein (Lariviere *et al*., 2003).

The mechanism of retaining GTs has not been elucidated yet. Direct comparison with glycosylhydrolases would suggest a double displacement mechanism with a short-lived glycosyl–enzyme intermediate. Despite numerous efforts, the only intermediate that could be trapped involved an amino acid rather far from the active site (Lairson *et al*., 2004). An alternative mechanism suggests a direct attack by the acceptor, as previously proposed for glycogen phosphorylase (Klein *et al*., 1986). This so-called *SNi*-like mechanism was originally proposed by Persson *et al*. (2001) for bacterial LgtC. Quantum mechanical calculations using Density Functional Theory applied to a model site of LgtC (136 atoms included) confirmed that the one-step mechanism is energetically favoured (Tvaroska, 2004). These prediction were used for designing new putative inhibitors based on the scaffold of the transition state (Raab *et al*., 2005). Retaining GTs with GT-B fold, that have distant structural similarities with glycogen phosphorylases, have been proposed to adopt the same mechanism, based on structural studies of complexes with nontransferable analogues (Gibson *et al*., 2004). In all retaining GTs, the nucleotide sugar is forced to adopt a folded shape that brings the sugar over the pyrophosphate. This special conformation facilitates the transfer by several means: the C-1 is spatially accessible for the reactions, the anomeric bond is elongated and weakened by the torsion around the Π torsion angle as calculated by ab initio methods (Petrova *et al*., 1999), and a hydrogen bond can be established between O-2 of sugar and phosphate, lowering the energy barrier.

Conclusions

A sequence-based classification places GTs in many families. In contrast, folds appear to be more conserved because there is convergence to only a few topologies for most of the GT families. The small variety of folds observed is compensated by a large structural variability in the acceptor-binding domain, thus conferring some functional plasticity which allows fine tuning with respect to the acceptor. The catalytic mechanisms are still poorly understood, but current data suggest the importance of movement of loops and domains in catalysis. Additional structural, modelling, and mutational studies are needed to further progress in the understanding of these enzymes.

Acknowledgments

CERMAV is part of GDR CNRS n 2590 and FR CNRS n 2607. The work was supported by Ministry of Education (JK, LS contract MSM0021622413) of the Czech Republic. Travel and visits between NCBR and CERMAV were supported by a BARRANDE exchange program. C.J. benefited from a grant from French Ministry of Research and Education and L.S. from French Ministry for Foreign Affairs.

Abbreviations

3D-PSSM, three-dimensional position-specific scoring matrix; BGT, β-glucosyltransferase; CMP, cytidine monophosphate; GalNAc, N-acetylgalactosamine; GT, glycosyltransferase; MGS, mannosylglycerate synthase; NeuAC, N-acetyl neuraminic acid; UDP, uridine diphosphate.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Boix, E., Swaminathan, G.J., Zhang, Y., Natesh, R., Brew, K., and Acharya, K.R. (2001) Structure of UDP complex of UDP-galactose: β-galactoside-α-1,3-galactosyltransferase at 1.53A resolution reveals a conformational change in the catalytically important C-terminus. *J. Biol. Chem.*, **276**, 48608–48614.

Boix, E., Zhang, Y., Swaminathan, G.J., Brew, K., and Acharya, K.R. (2002) Structural basis of ordered binding of donor and acceptor substrates to the retaining glycosyltransferase alpha-1,3-galactosyltransferase. *J. Biol. Chem.*, **277**, 28310–28318.

Breton, C. and Imberty, A. (1999) Structure-function studies of glycosyltransferases. *Curr. Opin. Struct. Biol.*, **9**, 563–571.

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

Breton, C., Heissigerova, H., Jeanneau, C., Moravcova, J., and Imberty, A. (2002) Comparative aspects of glycosyltransferases. *Biochem. Soc. Symp.*, **69**, 23–32.

Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreulx, D., Cambillau, C., and Bourne, Y. (1999) Crystal structure of the bifunctional N-acetylglucosamine 1-phosphate uridyltransferase from *Escherichia coli*: a paradigm for the related pyrophosphorylase superfamily. *EMBO J.*, **18**, 4096–4107.

Buschiazzo, A., Ugalde, J.E., Guerin, M.E., Shepard, W., Ugalde, R.A., and Alzari, P.M. (2004) Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. *EMBO J.*, **23**, 3196–3205.

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**, 929–939.

Campbell, R.E., Mosimann, S.C., Tanner, M.E., and Strynadka, N.C. (2000) The structure of UDP-N-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferases. *Biochemistry*, **39**, 14993–15001.

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

Chiu, C.P., Watts, A.G., Lairson, L.L., Gilbert, M., Lim, D., Wakarchuk, W.W., Withers, S.G., and Strynadka, N.C. (2004) Structural analysis of the sialyltransferase Cst II from *Campylobacter jejuni* in complex with a substrate analog. *Nat. Struct. Mol. Biol.*, **11**, 163–170.

Coutinho, P.M., Deleury, E., Davies, G.J., and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.*, **328**, 307–317.

Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**, 853–859.

Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.

Edman, M., Berg, S., Storm, P., Wikstrom, M., Vikstrom, S., Ohman, A., and Wieslander, A. (2003) Structural features of glycosyltransferases synthesizing major bilayer and nonbilayer-prone membrane lipids in *Acholeplasma laidlawii* and *Streptococcus pneumoniae*. *J. Biol. Chem.*, **278**, 8420–8428.

Flint, J., Taylor, E., Yang, M., Bolam, D.N., Tailford, L.E., Martinez-Fleites, C., Dodson, E.J., Davis, B.G., Gilbert, H.J., and Davies, G.J. (2005) Structural dissection and high-throughput screening of mannosylglycerate synthase. *Nat. Struct. Mol. Biol.*, **12**, 608–614.

Franco, O.L. and Rigden, D.J. (2003) Fold recognition analysis of glycosyltransferase families: further members of structural superfamilies. *Glycobiology*, **13**, 707–712.

Fritz, T.A., Hurley, J.H., Trinh, L.B., Shiloach, J., and Tabak, L.A. (2004) The beginnings of mucin biosynthesis: the crystal structure of UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-T1. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 15307–15312.

Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J.P. (1987) Hydrophobic cluster analysis: an efficient way to compare and analyse amino acid sequences. *FEBS Lett.*, **224**, 149–155.

Gastinel, L.N., Cambillau, C., and Bourne, Y. (1999) Crystal structures of the bovine beta4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *EMBO J.*, **18**, 3546–3557.

Gastinel, L.N., Bignon, C., Misra, A.K., Hindsgaul, O., Shaper, J.H., and Joziasse, D.H. (2001) Bovine α1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases. *EMBO J.*, **20**, 638–649.

Gibbons, B.J., Roach, P.J., and Hurley, T.D. (2002) Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.*, **319**, 463–477.

Gibson, R.P., Turkenburg, J.P., Charnock, S.J., Lloyd, R., and Davies, G.J. (2002) Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chem. Biol.*, **9**, 1337–1346.

Gibson, R.P., Tarling, C.A., Roberts, S., Withers, S.G., and Davies, G.J. (2004) The donor subsite of trehalose-6-phosphate synthase: binary complexes with UDP-glucose and UDP2-deoxy-2-fluoro-glucose at 2 A resolution. *J. Biol. Chem.*, **279**, 1950–1955.

Godzik, A. (2003) Fold recognition methods. *Methods Biochem. Anal.*, **44**, 525–546.

Gunasekaran, K. and Nussinov, R. (2004) Modulating functional loop movements: the role of highly conserved residues in the correlated loop motions. *Chembiochem.*, **5**, 224–230.

Ha, S., Walker, D., Shi, Y., and Walker, S. (2000) The 1.9 A crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.*, **9**, 1045–1052.

Hans, J., Brandt, W., and Vogt, T. (2004) Site-directed mutagenesis and protein 3D-homology modelling suggest a catalytic mechanism for UDP-glucose-dependent betanidin 5-Oglucosyltransferase from *Dorotheanthus bellidiformis*. *Plant J.*, **39**, 319–333.

Heissigerova, H., Breton, C., Moravcova, J., and Imberty, A. (2003) Molecular modeling of glycosyltransferases involved in the biosynthesis of blood group A, blood group B, Forssman and iGb3 antigens and their interaction with substrates. *Glycobiology*, **13**, 377–386.

Hu, Y. and Walker, S. (2002) Remarkable structural similarities between diverse glycosyltransferases. *Chem. Biol.*, **9**, 1287–1296.

Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M., and Walker, S. (2003) Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 845–849.

Kakuda, S., Shiba, T., Ishiguro, M., Tagawa, H., Oka, S., Kajihara, Y., Kawasaki, T., Wakatsuki, S., and Kato, R. (2004) Structural basis for acceptor substrate recognition of a human glucuronyltransferase, GlcAT-P, an enzyme critical in the biosynthesis of the carbohydrate epitope HNK-1. *J. Biol. Chem.*, **279**, 22693–22703.

Kelley, L.A., MacCallum, R.M., and Sternberg, M.J.E. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.*, **299**, 499–520.

Kikuchi, N., Kwon, Y.D., Gotoh, M., and Narimatsu, H. (2003) Comparison of glycosyltransferase families using the profile hidden Markov model. *Biochem. Biophys. Res. Commun.*, **310**, 574–579.

Klein, H.W., Im, M.J., and Palm, D. (1986) Mechanism of the phosphorylase reaction. Utilization of D-gluco-hept-1-enitol in the absence of primer. *Eur. J. Biochem.*, **157**, 107–114.

Lairson, L.L. and Withers, S.G. (2004) Mechanistic analogies amongst carbohydrate modifying enzymes. *Chem. Commun. (Camb)*, **20**, 2243–2248.

Lairson, L.L., Chiu, C.P., Ly, H.D., He, S., Wakarchuk, W.W., Strynadka, N.C., and Withers, S.G. (2004) Intermediate trapping on a mutant retaining alpha-galactosyltransferase identifies an unexpected aspartate residue. *J. Biol. Chem.*, **279**, 28339–22844.

Lariviere, L. and Morera, S. (2002) A base-flipping mechanism for the T4 phage betaglucosyltransferase and identification of a transition-state analog. *J. Mol. Biol.*, **324**, 483–490.

Lariviere, L., Gueguen-Chaignon, V., and Morera, S. (2003) Crystal structures of the T4 phage beta-glucosyltransferase and the D100A mutant in complex with UDP-glucose: glucose binding and identification of the catalytic base for a direct displacement mechanism. *J. Mol. Biol.*, **330**, 1077–1086.

Lesk, A.M. (1995) NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.*, **5**, 775–783.

Liu, J. and Mushegian, A. (2003) Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.*, **12**, 1418–1431.

Lobsanov, Y.D., Romero, P.A., Sleno, B., Yu, B., Yip, P., Herscovics, A., and Howell, P.L. (2004) Structure of Kre2p/Mnt1p: a yeast alpha1,2-mannosyltransferase involved in mannoprotein biosynthesis. *J. Biol. Chem.*, **279**, 17921–17931.

Moréra, S., Imberty, A., Aschke-Sonnenbron, U., Rüger, W., and Freemont, P.S. (1999) T4 phage beta-glucosyltransferase: substrate binding an proposed catalytic mechanism. *J. Mol. Biol.*, **292**, 717–730.

Mulichak, A.M., Losey, H.C., Walsh, C.T., and Garavito, R.M. (2001) Structure of the UDPglucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure*, **9**, 547–557.

Mulichak, A.M., Losey, H.C., Lu, W., Wawrzak, Z., Walsh, C.T., and Garavito, R.M. (2003) Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 9238–9243.

Mulichak, A.M., Lu, W., Losey, H.C., Walsh, C.T., and Garavito, R.M. (2004) Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. *Biochemistry*, **43**, 5170–5180.

Murray, B.W., Takayama, S., Schultz, J., and Wong, C.H. (1996) Mechanism and specificity of human alpha-1,3-fucosyltransferase V. *Biochemistry*, **35**, 11183–11195.

Patenaude, S.I., Seto, N.O., Borisova, S.N., Szpacenko, A., Marcus, S.L., Palcic, M.M., and Evans, S.V. (2002) The structural basis for specificity in human ABO (H) blood group biosynthesis. *Nat. Struct. Biol.*, **9**, 685–690.

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

Pedersen, L.C., Darden, T.A., and Negishi, M. (2002) Crystal structure of beta 1,3glucuronyltransferase I in complex with active donor substrate UDP-GlcUA. *J. Biol. Chem.*, **277**, 21869–21873.

Pedersen, L.C., Dong, J., Taniguchi, F., Kitagawa, H., Krahn, J.M., Pedersen, L.G., Sugahara, K., and Negishi, M. (2003) Crystal structure of an alpha 1,4-Nacetylhexosaminyltransferase (EXTL2), a member of the exostosin gene family involved in heparan sulfate biosynthesis. *J. Biol. Chem.*, **278**, 14420–14428.

Persson, K., Ly, H.D., Dieckelmann, M., Wakarchuk, W.W., Withers, S.G., and Strynadka, N.C.J. (2001) Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat. Struct. Biol.*, **8**, 166–175.

Petrova, P., Koca, J., and Imberty, A. (1999) Potential energy hypersurfaces of nucleotide-sugars: *Ab initio* calculations, force-field parametrization, and exploration of the flexibility. *J. Am. Chem. Soc.*, **121**, 5535–5547.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004) UCSF chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

Qasba, P.K., Ramakrishnan, B., and Boeggeman, E. (2005) Substrate-induced conformational changes in glycosyltransferases. *Trends Biochem. Sci.*, **30**, 53–62.

Raab, M., Kozmon, S., and Tvaroska, I. (2005) Potential transition-state analogs for glycosyltransferases. Design and DFT calculations of conformational behavior. *Carbohydr. Res.*, **340**, 1051–1057.

Ramakrishnan, B. and 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.

Ramakrishnan, B., Balaji, P.V., and Qasba, P.K. (2002) Crystal structure of beta-1,4galactosyltransferase complex with UDP-Gal reveals an oligosaccharide binding site. *J. Mol. Biol.*, **318**, 491–502.

Ramakrishnan, B., Boeggeman, E., Ramasamy, V., and Qasba, P.K. (2004) Structure and catalytic cycle of beta-1,4-galactosyltransferase. *Curr. Opin. Struct. Biol.*, **14**, 593–600.

Ramasamy, V., Ramakrishnan, B., Boeggeman, E., and Qasba, P.K. (2003) The role of tryptophan 314 in the conformational changes of beta1,4-galactosyltransferase-I. *J. Mol. Biol.*, **331**, 1065–1076.

Rosen, M.L., Edman, M., Sjostrom, M., and Wieslander, A. (2004) Recognition of fold and sugar linkage for glycosyltransferases by multivariate sequence analysis. *J. Biol. Chem.*, **279**, 38683–38692.

Snajdrová, L., Kulhánek, P., Imberty, A., and Koca, J. (2004) Molecular dynamics simulations of glycosyltransferase LgtC. *Carbohydr. Res.*, **339**, 995–1006.

Taniguchi, N., Honke, K., and Fukuda, M. (2002) *Handbook of Glycosyltransferase and Related Genes*. Springer, Tokyo.

Tarbouriech, N., Charnock, S.J., and Davies, G.J. (2001) Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDP-sugar glycosyltransferases. *J. Mol. Biol.*, **314**, 655–661.

Thorsøe, K.S., Bak, S., Olsen, C.E., Imberty, A., Breton, C. and Møller, B.L. (2005) Determination of catalytic key amino acids and UDP-sugar donor specificity of the cyanohydrin glycosyltransferase UGT85B1 from *Sorghum bicolor*: molecular modeling substantiated by site specific mutagenesis and biochemical analyses. *Plant Physiol*, forthcoming.

Tvaroska, I. (2004) Molecular modeling insights into the catalytic mechanism of the retaining galactosyltransferase LgtC. *Carbohydr. Res.*, **339**, 1007–1014.

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

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

Waldherr-Teschner, M., Goetze, T., Heiden, W., Knoblauch, M., Vollhardt, H., and Brickmann, J. (1992) MOLCAD – computer aided visualization and manipulation of models in molecular science. In Post, F.H. and Hin, A.J.S. (eds), *Advances in Scientific Visualization*. Springer, Heidelberg, Germany, pp. 58–67.

Wimmerova, M., Engelsen, S.B., Bettler, E., Breton, C., and Imberty, A. (2003) Combining fold recognition and exploratory data analysis for searching for glycosyltransferases in the genome of *Mycobacterium tuberculosis*. *Biochimie*, **85**, 691–700.

Wrabl, J. and Grishin, N. (2001) Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. *J. Mol. Biol.*, **314**, 365–374.

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