3 Glycosyltransferases

Glycosyltransferases are a group of enzymes that catalyze the transfer of a sugar moiety from an activated sugar onto carbohydrate or non-carbohydrate acceptors. With the exception of hyaluronan synthase, glycosyltransferases elongate glycans by adding monosaccharides to the non-reducing ends of acceptor substrates. The transfer reaction is linkage-specific and the resulting product is fixed in a given anomeric configuration. Glycosyltransferase genes make up a significant portion of prokaryote and eukaryote genomes. Adding all transporters and enzymes required for substrate biosynthesis and for glycan degradation, up to 5% of genomes encode genes involved in glycosylation. Globally, more than 30'000 glycosyltransferases are known to date.

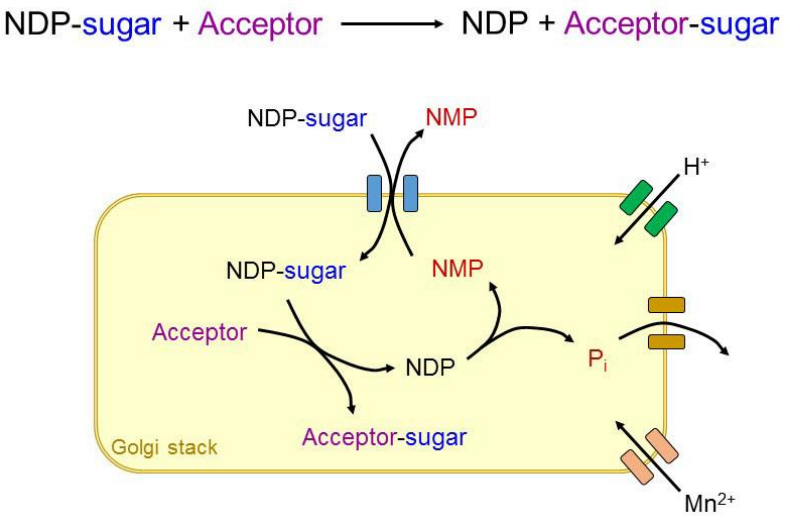


Figure 1 – Topology of the standard glycosyltransferase reaction in the Golgi apparatus. NDP dephosphorylation is critical to drive the reaction towards the formation of the glycosylated product (acceptor-sugar).

The standard glycosyltransferase reaction requires specific donor and acceptor substrates and yields a glycosylated product and an aglycone, which is typically NDP or a polyprenol-P. Additional parameters like pH, Mn^{2+} availability and

NDP levels affect many glycosyltransferase activities. Therefore, local concentrations of these parameters are strictly controlled in order to maintain proper glycosylation.

Glycosyltransferases have been first grouped by their substrate specificities. This classification system gave raise to families of galactosyltransferases, fucosyltransferases and sialyltransferases for example, which were further specified according to the type of linkage catalyzed. The broad availability of genomic data in the last decades led to the identification of multiple structurally related proteins for most glycosyltransferase families. As the enzymatic activity of new glycosyltransferases was characterized, it became soon evident that structurally related proteins may indeed catalyze the transfer of distinct donor substrates. Along this line, β1-3 galacatosyltransferases were found to be more similar to β1-3 N-acetylglucosaminyltransferases than to β1-4 galactosyltransferases.

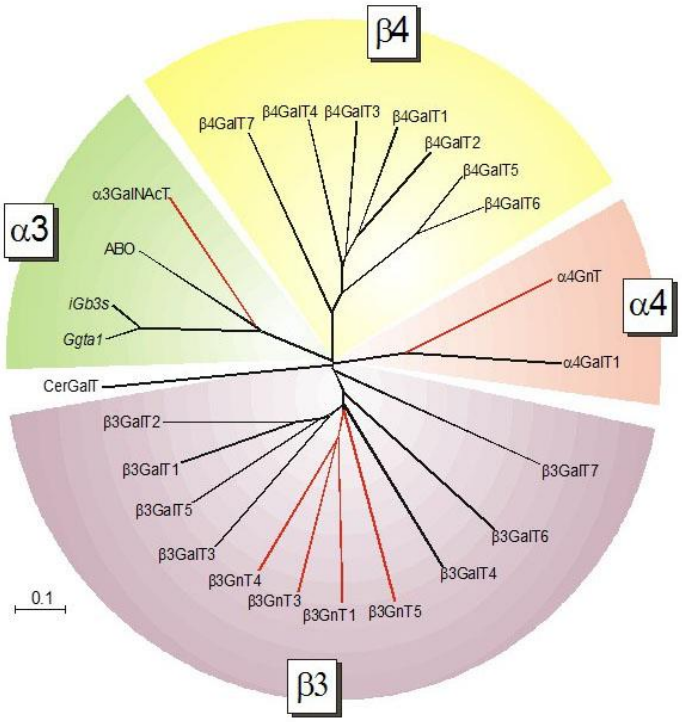
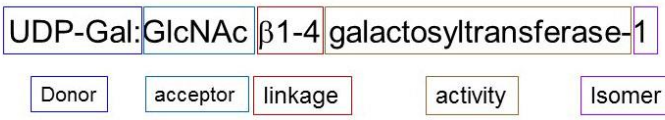


Figure 2 – Cladogram of Gal-transferases and structurally related GalNAc- and GlcNAc-transferases (marked with red lines) demonstrating the sequence conservation based on the type of linkage catalyzed.

The nomenclature of glycosyltransferases does not only consider the linkage and donor substrate, but also the acceptor substrate and the isomer number are included in the formal name. Nomenclature remains a difficult issue since many glycosyltransferases are usually called by trivial names. For example, a given sialyltransferase is described as SiaT9, GM3 synthase or ST3GAL5 depending on the publications considered. This lack of systematic nomenclature certainly does not simplify the access to glycobiology, which is already complex enough at the biological level.

As mentioned here above, the type of linkage dictates the structural conservation between glycosyltransferases. Despite the tremendous amount of combinations between donor, acceptor and linkage exists in theory, only few are really encountered in organisms. For example in animals, the donor GDP-Fuc is only transferred via α1-O linkage to Ser/Thr in the ER and to Gal via α1-2 linkage and to GlcNAc via α1-3, α1-4 and α1-6 linkages in the Golgi apparatus.



	Ser/Thr	Asn	hLys	Trp	Tyr	Cer	PI	Fuc	Gal	GalNAc	Glc	GlcNac	GlcA	Man	Sia	Xyl
GDP-Fuc	α1								α1-2			α1-3 α1-4 α1-6				
UDP-Gal			β1			β1			α1-3 α1-4 β1-3	β1-3	β1-4	β1-3 β1-4				β1-4
UDP-GalNAc	α1								α1-3 β1-3 β1-4	α1-3 α1-6		β1-4	β1-4			
UDP-Glc DolP-Glc	β1				α1	β1		β1-3	α1-2		α1-2 α1-3			α1-3		
UDP-GlcNAc	β1	β1					α1	β1-3	β1-3 β1-6	β1-6		α1-6 β1-4	α1-4 β1-4	β1-2		
UDP-GlcA									β1-3 β1-4	β1-3		β1-3 β1-4				
GDP-Man DolP-Man	α1			α1								α1-4 β1-4		α1-2 α1-3 α1-6		
CMP-Sia									α2-3 α2-6	α2-6					α2-8	
UDP-Xyl	β1										α1-3					α1-3

Figure 3 – Full name of glycosyltransferases include donor, acceptor, linkage, activity and isomer number. The table shows the linkages encountered in animal cells for each donor (left) and acceptor (top). Reactions taking place in the Golgi apparatus are in yellow.

Although most glycosyltransferases transfer monosaccharides to a multitude of glycoproteins and glycolipids, some glycosyltransferases recognize large protein domains and thereby are specific to selected types of proteins or even specific to a single protein. This is for example the case for the enzyme GlcNAc-1-phosphotransferase, which initiates the formation of Man-6-P on lysosomal enzymes, or for a GalNAc-transferase acting specifically on gonadotropin hormones.

Classification of glycosyltransferases

Whereas originally grouped by linkage and substrate specificity, glycosyltransferases are nowadays rather grouped by sequence similarity. This move is the logical consequence of the identification of thousands of potential glycosyltransferases across genomic databases. Because 95% of these potential glycosyltransferases have not been assigned any clear substrate specificities yet, they are best maintained in groups of structurally similar proteins. Nevertheless, several examples have shown that structural similarity does not imply similarity in substrate specificity.

All enzymes active on carbohydrates, including potential glycosylhydrolases and glycosyltransferases, have been annotated in a database called CAZY (Carbohydrate-Active enzymes, see www.cazy.org), which is curated by Bernard Henrissat and his team at the University of Marseille. CAZY classifies enzymes in families based on sequence similarity, protein fold prediction and reaction mechanisms. To date, 95 families of glycosyltransferases have been defined, which encompass more than 30'000 putative glycosyltransferases. All proteins of a same CAZY family have likely evolved from the same progenitor sequence, present conserved mechanisms of reaction, and share similar structural folds.

For example, the CAZY glycosyltransferase family #25 currently includes 1263 proteins, the majority coming from bacterial genomes (1145 proteins), but also features 85 eukaryotic proteins and even 33 virally encoded glycosyltransferases. The catalytically known glycosyltransferases from this family transfer Glc and Gal in β-linkage mainly in the context of bacterial LPS biosynthesis. By contrast, the eukaryotic glycosyltransferases in this family transfer Gal to collagen, thereby demonstrating that structural similarity does not relate to substrate specificity. Despite the limited usefulness of the CAZY database to predict glycosyltransferase activities, the direct comparison of protein sequences within a CAZY family outlines strongly conserved amino acid, which are likely to be involved in the catalytic activity or in maintaining protein stability.

Topology of glycosyltransferases

Glycosyltransferases show a great structural diversity. They can be as small as 200 amino acids or as large as 1500 amino acids long. They can be soluble or membrane-bound. As membrane proteins, glycosyltransferases can have a single transmembrane domain or be completely embedded in membranes through a dozen of transmembrane domains. In eukaryotes, most glycosyltransferases have a type-II transmembrane topology, with a short N-terminal cytosolic domain, a single transmembrane domain and a large luminal catalytic domain. The exceptions to the rule are ER-localized glycosyltransferases utilizing Dol-P based donor substrates, such as the enzymes involved in the assembly of lipid-linked oligosaccharides. The short cytosolic and single transmembrane domains include signal motifs for the proper localization of glycosyltransferases. Recent proteomic studies have revealed that several glycosyltransferases are phosphorylated either at their cytosolic or luminal domains. This discovery suggests that glycosyltransferase activities could be regulated by phosphorylation like signaling proteins.

Only few glycosyltransferase structures have been elucidated by crystallography to date. The three-dimensional structures determined show two major structural folds, which are often called GT-A and GT-B. Both folds include two Rossmann domains, which are often found in nucleotide-binding proteins. The typical Rossmann fold represents a series of three parallel β-strands intervened by two α-helices. For example, the β1-2 N-acetylglucosaminyltransferase MGAT1 and the β1-4 galactosyltransferase B4GALT1 have a GT-A fold topology while

fucosyltransferases have a GT-B fold. Glycosyltransferases that utilize polyprenol-P or Dol-P linked substrates do not follow the GT-A or GT-B fold organization.

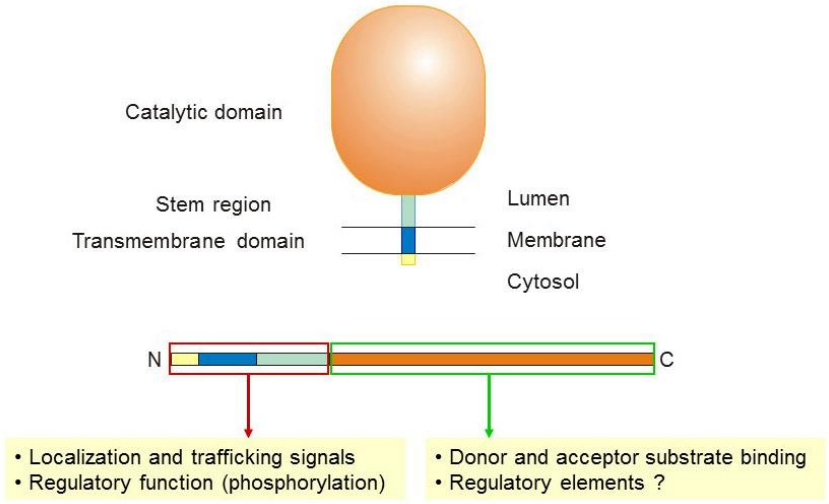


Figure 4 – Domain organization of typical Golgi glycosyltransferase (type-II tranmembrane topology).

Glycosyltransferase trafficking

Some Golgi glycosyltransferases are mainly found in *cis* cisternae whereas others are enriched in the *trans* cisternae. These gradients are likely the product of selective retrograde transport processes, such as those involving COG tethering factors. The localization of glycosyltransferase may be dictated by signal motifs found in the cytosolic domains, which are exposed to vesicular transport proteins. In fact, localization experiments performed on glycosyltransferases with mutated N-terminal sequences confirmed the presence of such localization signals. In some cases, the simple presence of a di-leucine motif appears to be sufficient for the retention of a glycosyltransferase in the Golgi apparatus.

Another mode of localization has been proposed, which is based on the size of the transmembrane domain of glycosyltransferases and the thickness of Golgi membranes. The cholesterol concentration increases from *cis* to *trans* in the Golgi, which increases correspondingly the thickness of the lipid bilayer in which glycosyltransferases are anchored. Some correlations have been documented between the length of the transmembrane domain and the localization of some glycosyltransferases in the Golgi apparatus. However, most glycosyltransferases do not show any correlation of such kind, meaning that the length of the transmembrane domain alone does not account for the localization of glycosyltransferase.

Another theory states that the retention of glycosyltransferases is mediated through the formation of heterodimers between enzymes that catalyze successive reactions in a glycosylation pathway. Along the line of this theory of "kin recognition", it has been shown that the β1-2 GlcNAc-transferase GnT1 interacts with the

mannosidase-II enzyme in the *medial* Golgi. However, experiments with cells from gene knockout mice did not support the kin recognition model, since the absence of a specific glycosyltransferase did not affect the localization of his presumed interacting partners.

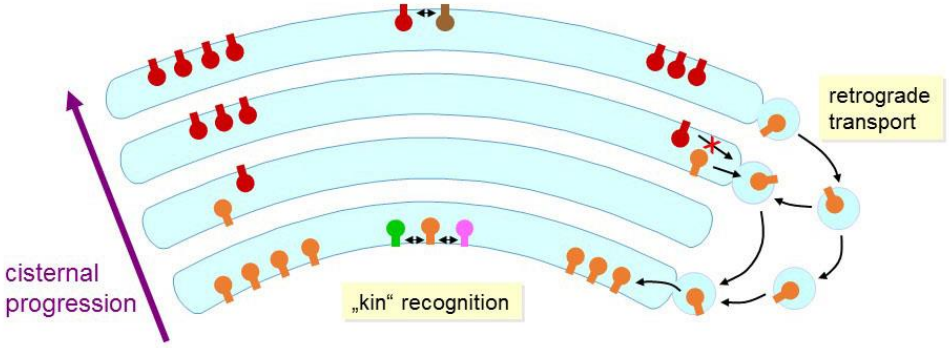


Figure 5 – Possible control mechanisms of glycosyltransferase localization in the Golgi apparatus.

Most core glycosyltransferases that initiate a class of glycosylation are localized in the ER. Only mucin-type glycosylation, GAG chain biosynthesis and glycosphingolipids formation are initiated in the Golgi apparatus based on the localization of their respective core glycosyltransferases. Some of the ER-located core glycosyltransferases transfer glycans to unfolded proteins, like collagen Gal-transferases and the oligosaccharyltransferase complex. Also, the protein O-fucosyltransferase OFUT1 even acts as a chaperone for the folding and trafficking of its substrate glycoprotein *Notch*.

Glycosylation Pathway	Glycosyltransferase	ER	Golgi
N-glycosylation	Oligosaccharyltransferase	✔	
O-GalNAc mucin-type	Polypeptide GalNAc-transferases		✔
O-Fucosylation	O-Fucosyltransferases	✔	
O-Glucosylation	Rumi O-glucosyltransferase	✔	
O-Mannosylation	Protein O-mannosyltransferases	✔	
O-Galactosyaltion collagen-type	β1-0 galactosyltransferase	✔	
C-Mannosylation	C-mannosyltransferase	✔	
GAG biosynthesis	Xylosyltransferases		✔
GPI anchor biosynthesis	PIG complex	✔	
Glycosphingolipid biosynthesis	Ceramide glucosyltransferase		✔

Figure 6 – Initiation of glycosylation pathways in the ER and Golgi apparatus based on the organelle localization of glycosyltransferases.

As shown recently (Gill *et al.*, 2010), the polypeptide GalNAc-transferase enzymes involved in mucin-type O-glycosylation can be relocated to the ER though through a Src kinase mediated COP1-retrograde transport. This redistribution of the glycosyltransferases even appeared to increase the level of O-glycosylation of marker O-glycoproteins. To date, this example of enzyme translocation between the Golgi apparatus and the ER is unique and additional work is required to further establish the biological relevance of the process. In any case, the study highlights an interesting concept with the dynamic regulation of mucin-type O-glycosylation through relocation of glycosyltransferases induced by signaling cascades.

Mechanisms of glycosyltransferase reaction

Most nucleotide-activated sugars are in the α-anomeric form but the products of many glycosyltransferase reactions are in the β-anomeric form. The enzymes catalyzing such reactions are called inverting glycosyltransferases, whereas those maintaining the anomerism in the product are called retaining glycosyltransferases. By the way, sialyltransferases and fucosyltransferases are inverting enzymes, because they catalyze the respective transfer of Sia and Fuc from CMP-β-Sia and GDP-β-Fuc to α-linked products. The catalytic mechanisms of the underlying transfer are totally different between inverting and retaining glycosyltransferases. The mechanism of inverting glycosyltransferases is a direct displacement S_N2-like reaction enabled by an enzymatic base catalyst and by Lewis acid activation of the departing phosphate leaving group. The catalytic mechanism of retaining glycosyltransferases is less clear and appears to involve a double-displacement mechanism with formation of a covalently bound glycosyl-enzyme intermediate. The type of glycosyltransferase reaction, *i.e.* inverting or retaining, does not correlate with the overall structural organization of the enzyme. Both GT-A and GT-B fold topologies have been found in inverting and retaining enzymes.

Most glycosyltransferases contain a single catalytic domain but few do comprise two catalytic domains. The latter enzymes are usually involved in the formation of heteropolymers such as glycosaminoglycan chains. At any case, each glycosyltransferase domain catalyzes a single enzymatic reaction. However, as always some exceptions exist to the rule. One of such exceptions deals with the type of linkage, namely the fucosyltransferase enzyme FUT3 catalyzes the transfer of Fuc in both α1-3 and α1-4 linkages. The second exception is the EXTL2 glycosyltransferase that transfers both the sugar donors GalNAc and GlcNAc to GlcA. Finally, the third exception is the β1-4 galactosyltransferase B4GALT1 that transfers Gal to either the acceptor GlcNAc or Glc. Note worthily, the latter reaction requires the association of the B4GALT1 enzyme with the protein α-lactalbumin to encompass the Glc acceptor.

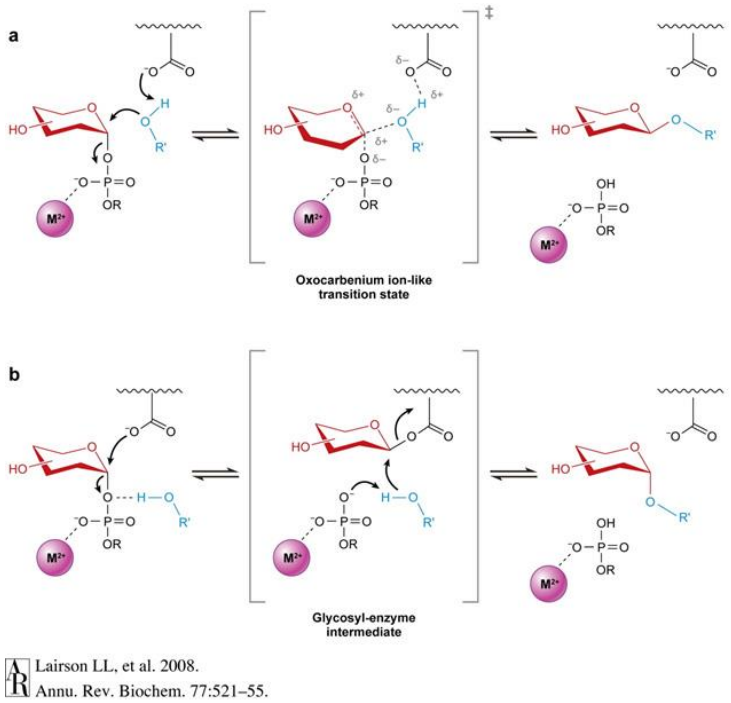


Figure 7 – Mechanisms of inverting (a) and retaining (b) glycosyltransferase reactions. Illustration taken from Lairson LL *et al.*, 2008, Annu Rev Biochem 77:521 ©.

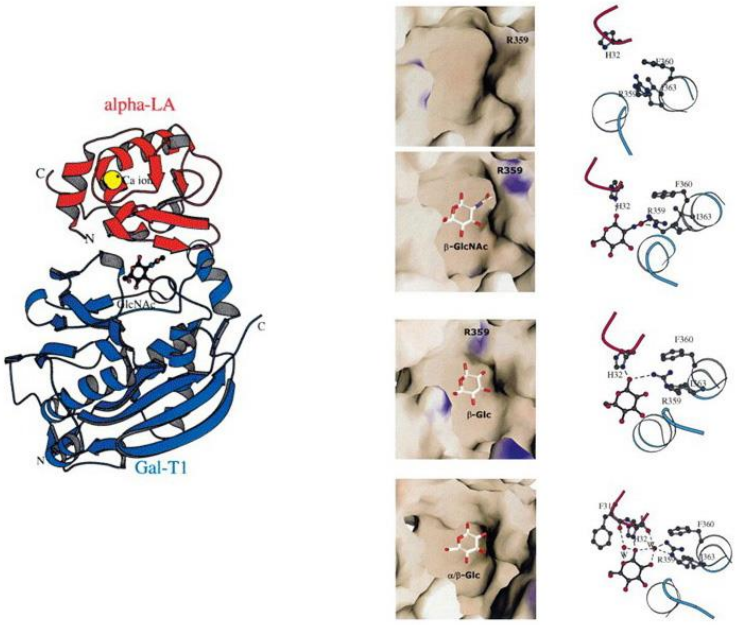


Figure 8 – Complex formation between B4GALT1 (blue) and α-lactalbumin (red). Illustration taken from Ramakrishnan and Qasba, 2001, J Mol Biol 310, 205 ©.

The complex of B4GALT1 with α-lactalbumin is essential for the production of lactose, i.e. Gal(β1-4)Glc, in the lactating mammary gland. The interaction with α-lactalbumin takes place within the catalytic domain B4GALT1, in a region

comprising the acceptor-binding site. The proximity of α-lactalbumin changes the orientation of side-chains in a loop of the acceptor-binding pocket of B4GALT1, thereby blocking the binding of GlcNAc and making Glc the preferred acceptor. This conformational change decreases the K_m of B4GALT1 for Glc by 1000-fold.