

# 4

## N-Linked glycosylation

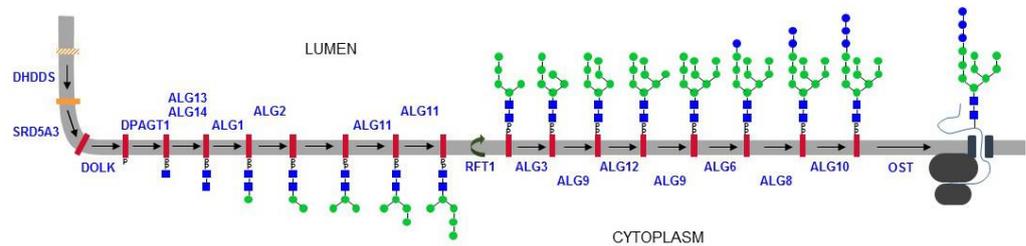
N-linked glycosylation is a common class of glycosylation encountered in all eukaryotes as well as in Archaea and some bacteria. In eukaryotes, the assembly of N-glycans follows a complex sequence of events spanning the ER and the Golgi apparatus. The ER pathway is strongly conserved within eukaryotes, but the Golgi pathway shows extensive differences between fungi, plants and animals. In addition to carrying epitopes involved in protein-protein interactions in the extracellular space, N-glycans also fulfill important roles in the maturation and intracellular trafficking of glycoproteins.

### ER N-linked glycosylation

N-linked glycosylation begins at the cytosolic leaflet of the ER membrane with the assembly of the oligosaccharide  $\text{GlcNAc}_2\text{Man}_9\text{Glc}_3$  on Dol. The pathway is initiated by the transfer of GlcNAc-P from UDP-GlcNAc to Dol-P, yielding Dol-PP-GlcNAc. The elongation proceeds sequentially through the action of glycosyltransferases anchored at the ER membrane. Once reaching Dol-PP-GlcNAc<sub>2</sub>Man<sub>5</sub>, the dolichol-linked oligosaccharide is translocated across the membrane and subsequently elongated by Man- and Glc-transferase enzymes using Dol-P-Man and Dol-P-Glc as donor substrates. The membrane-embedded protein RFT<sub>1</sub> is required for the translocation of Dol-PP-GlcNAc<sub>2</sub>Man<sub>5</sub>, although it is unclear if RFT<sub>1</sub> acts as a *bona fide* flippase. Yeast has been instrumental as a genetic model in the characterization of the ER pathway of oligosaccharide assembly. Several yeast N-linked glycosylation mutants (named alg for asparagine-linked glycosylation) led to the identification of the underlying glycosyltransferase genes, which were accordingly abbreviated alg<sub>1</sub> to alg<sub>14</sub>.

The family of alg yeast mutants has also greatly contributed to the elucidation of several congenital disorders of glycosylation (CDG). Because of the strong conservation of the dolichol-linked oligosaccharide pathway among eukaryotes, the impact of mutations identified in human CDG patients can be investigated by expressing human *ALG* genes in the corresponding alg yeast strains. Without the availability of alg yeasts, the detrimental effect of mutations can only be confirmed by assessing the enzymatic activity of the mutant protein. Such an

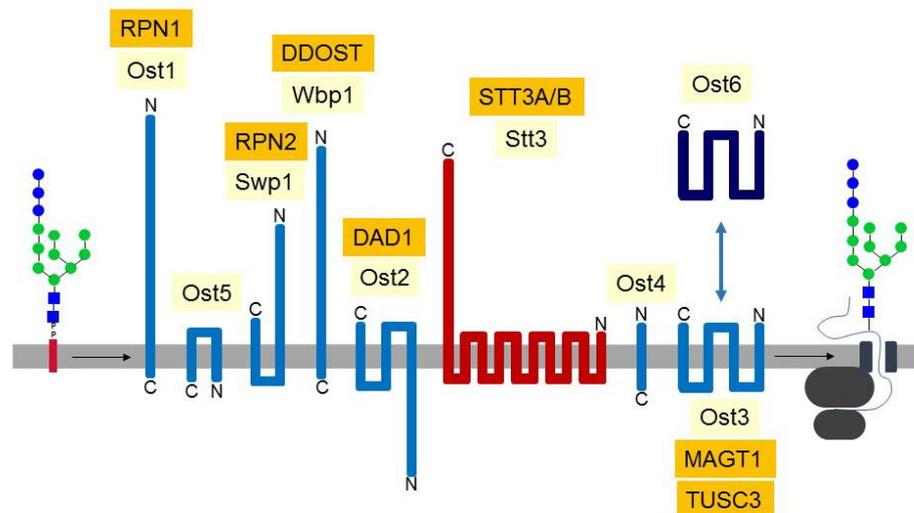
analysis is however extremely difficult considering the transmembrane nature of the alg enzymes and the requirement for complex Dol linked substrates for the assays.



**Figure 1** – Assembly of Dol-PP-oligosaccharides at the ER membrane and transfer of complete oligosaccharide to nascent protein.

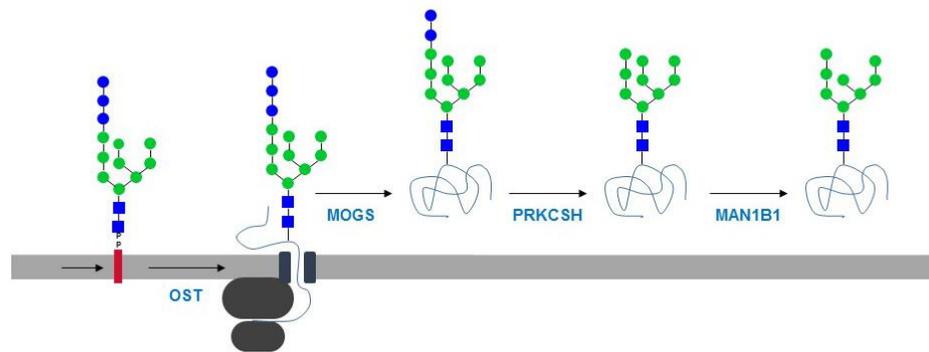
## Oligosaccharyltransferase

The complete Dol-PP linked  $\text{GlcNAc}_2\text{Man}_9\text{Glc}_3$  oligosaccharide is transferred to selected Asn residues on nascent glycoproteins by the oligosaccharyltransferase (OST) complex. In the vast majority of eukaryotes, OST consists of six to eight transmembrane proteins, which are clustered close to translocon complexes, through which nascent polypeptides enter the ER lumen. In contrast to alg glycosyltransferases that are functionally transferrable between yeast and animal cells, OST subunits are not functional when expressed across taxonomic classes.



**Figure 2** – Symbolic representation of OST subunits characterized in the yeast *Saccharomyces cerevisiae*. The Ost3 and Ost6 subunits do not occur together in the same OST complex. The abbreviations of the known mammalian orthologs are marked in orange. Yeasts express a single Stt3 subunit, whereas two variants (STT3A and STT3B) are found in animals. The [PglB protein](#) of *Campylobacter lari*, which is orthologous to eukaryotic STT3, has been crystallized and shown to include 13 transmembrane domains.

The STT<sub>3</sub> subunit is the catalytic center of OST, whereas the contribution of most of the other subunits remain unclear. OST recognizes on the one hand the terminal Glc residue and the proximal Dol-PP-GlcNAc of the oligosaccharide substrate, and on the other hand the side chain of an Asn residue in the context of the sequon N-X-S/T, where X can be any amino acid but Pro. After transfer to a nascent polypeptide, the GlcNAc<sub>2</sub>Man<sub>5</sub>Glc<sub>3</sub> oligosaccharide is readily trimmed by the glucosidase-I and -II enzymes, which cleave the 3 terminal Glc residues.



**Figure 3** – Transfer of complete oligosaccharide from the dolichol-PP carrier onto a nascent glycoprotein and subsequent trimming by glucosidase-I (MOGS), glucosidase-II (PRKCSH), and ER mannosidase-I (MAN1B1).

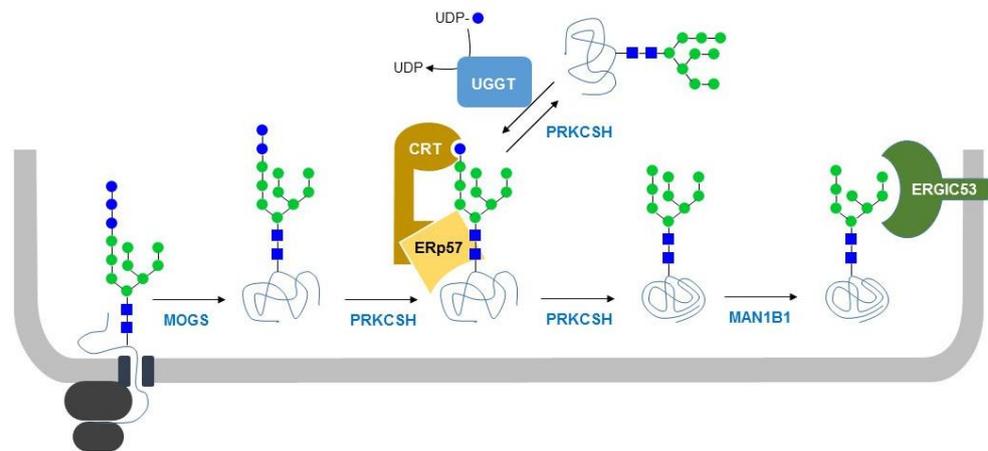
Bacterial N-linked glycosylation was first described in *Campylobacter jejuni*. N-linked glycosylation in *Campylobacter* is analog to the eukaryotic pathway in the sense that an oligosaccharide substrate is first assembled on a lipid carrier, which is then transferred by OST to Asn on acceptor polypeptides. The lipid carrier used in *Campylobacter* is bactoprenol-PP, which has 11 isoprenoid units (Dol has 17-24 isoprenoid units) and lacks the saturated residue carrying the primary alcohol group found in Dol. The oligosaccharide substrate in *Campylobacter* consists of the 7 monosaccharides GalNAc<sub>2</sub>[Glc]GalNAc<sub>3</sub>diNAcBac, where diNAcBac stands for 2,4-diacetamido-2,4,6-trideoxy-glucose. DiNAcBac is derived from GlcNAc and shares with GlcNAc the 2-acetamido group, which is essential for the recognition of the oligosaccharide substrate by OST. *Campylobacter* OST is encoded by the *pglB* gene and functions as a single protein localized in the plasma membrane, which mediates the glycosylation of acceptor proteins in the periplasm. The PglB protein is orthologous to the STT<sub>3</sub> subunit of eukaryotic OST. The acceptor sequon is similar to that of eukaryotes but requires Glu or Asp at -2 position, thus making: D/E-X-N-X-S/T. As in eukaryotes, X can be any amino acid but Pro.

In archaea, different oligosaccharide donors are assembled on Dol-P or Dol-PP and transferred by the OST AglB to variable acceptor sequons including N-X-S/T, N-X-N/L/V, and N-X-X-S. Recently, another type of N-linked glycosylation has been described in *Haemophilus influenzae* and *Actinobacillus pleuropneumoniae*. The acceptor sequon N-X-S/T is conserved, but the monosaccharides Glc

and Gal are transferred to acceptor proteins in the cytoplasm by a soluble OST enzyme.

## Intracellular functions of N-glycans

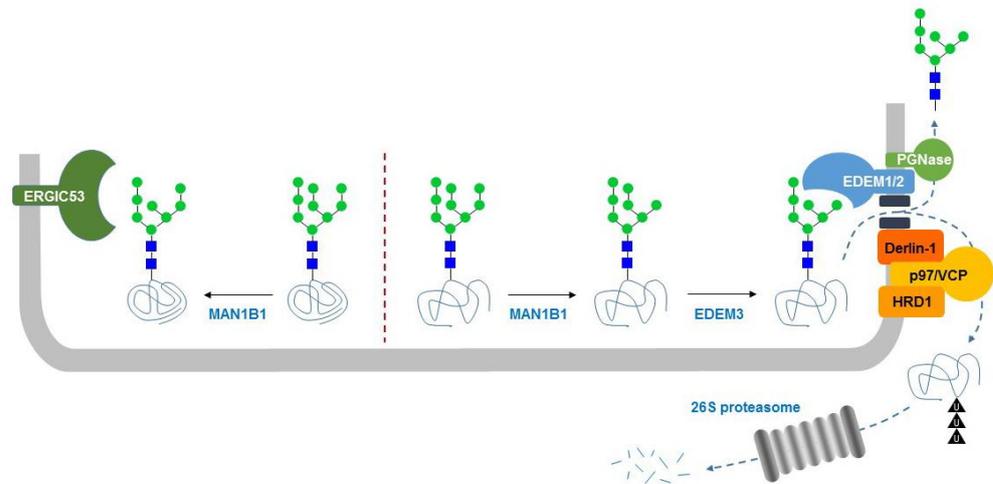
Immediately after transfer to Asn residues on glycoproteins, membrane anchored glucosidase-I trims oligosaccharides to  $\text{GlcNAc}_2\text{Man}_9\text{Glc}_2$ . During glycoprotein folding, the second and third Glc residues are cleaved by the soluble glucosidase-II enzyme whereas soluble UDP-Glc:glycoprotein Glc-transferase (UGGT) transfers back a Glc residue to  $\text{GlcNAc}_2\text{Man}_9$  on misfolded glycoproteins. This cycle of glucosylation-deglucosylation is part of the quality control of glycoprotein folding and enables interaction of misfolded glycoproteins with the Glc-specific lectins calnexin and calreticulin. These lectins are associated with the thiol-disulfide oxidoreductase ERp57, which assist in refolding glycoproteins.



**Figure 4** – Quality control of N-glycoprotein folding in the ER. Mono-glucosylated N-glycoproteins interact with the lectins calreticulin (CRT) and calnexin (CNX, not shown), which also bind to the oxidoreductase chaperone ERp57. Glucosidase-II (PPKCSH) cleaves the terminal Glc residue on folded and unfolded proteins. The latter are re-glucosylated by UDP-Glc:glycoprotein Glc-transferase (UGGT), which enables CRT/CNX binding and additional rounds of ERp57-mediated folding. After cleavage of the inner terminal Man residue, folded N-glycoproteins are bound by trafficking proteins such as ERGIC53 and transported out of the ER.

The transition from  $\text{GlcNAc}_2\text{Man}_9$  to  $\text{GlcNAc}_2\text{Man}_7$  represents another signal for the intracellular trafficking of glycoproteins. The ER mannosidase-I enzyme cleaves the terminal Man residue of the inner branch of N-glycans, yielding  $\text{GlcNAc}_2\text{Man}_8$ . Properly folded glycoproteins will leave the ER carrying this  $\text{GlcNAc}_2\text{Man}_8$  glycan. If the glycoprotein remains unfolded, a second mannosidase called HTM1 in yeasts (EDEM1 to EDEM3 in mammals) will cleave terminal Man from the outer  $\alpha$ 1-6 linked branch yielding  $\text{GlcNAc}_2\text{Man}_8$ , which is recognized by the Yos9p lectin (OS-9 in mammals). Yos9p/OS-9 is associated with the HRD ubiquitin ligase complex involved in the retro-translocation of misfolded

proteins out of the ER and of their degradation by the proteasome complex in the cytosol.



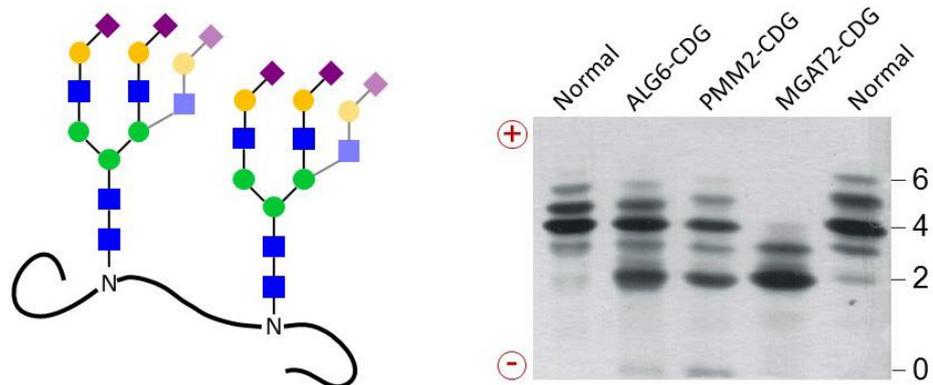
**Figure 5** – ER-associated protein degradation (ERAD) of misfolded N-glycoproteins.

After leaving the ER, glycoproteins are transferred to the ER-Golgi intermediary compartment (ERGIC). This process is in part mediated by the N-glycan lectin ERGIC<sub>53</sub>, which is also abbreviated LMAN<sub>1</sub> for *lectin, mannose-binding, 1*. Deficiency of ERGIC<sub>53</sub>/LMAN<sub>1</sub> causes a bleeding disorder (OMIM 227300) characterized by the defective secretion of coagulation factors V and VIII, which are N-linked glycoproteins.

N-glycans also carry the Man-6-P signal mediating the targeting of proteins to lysosomes. This lysosomal targeting signal is produced in two steps, consisting first of the addition of phospho-GlcNAc to Man residues on N-glycans by the Golgi-localized GlcNAc-1-phosphotransferase complex. This enzyme is a hexamer comprising two  $\alpha$ -, two  $\beta$ -, and 2  $\gamma$ -subunits, where the  $\alpha$ - and  $\beta$ -subunits are catalytically active and ensure the recognition of the lysosomal protein substrates. The contribution of the  $\gamma$ -subunits are unclear, but required for the full activity of the complex. Mutations in the *GNPTAB* gene encoding the polypeptide yielding the  $\alpha$ - and  $\beta$ -subunits cause I-cell disease (OMIM 252500) and pseudo-Hurler polydystrophy (OMIM 252600), which are also classified as mucopolysaccharidosis II and mucopolysaccharidosis III. Mutations in *GNPTG* encoding the  $\gamma$ -subunit cause a variant form of mucopolysaccharidosis III (OMIM 252605). After transfer of phospho-GlcNAc to Man, an uncovering enzyme localized in the trans Golgi network removes GlcNAc, thereby leaving the phosphate group at the 6'-position of Man. The Man-6-P signal is recognized by cognate Man-6-P receptors, which capture the glycoproteins bearing the signal and transports them to the lysosomes. The presence of Man-6-P receptors at the plasma membrane enables the treatment of defective lysosomal enzymes by intravenous administration of Man-6-P bearing recombinant enzymes.

## Diseases of ER N-glycosylation

Defects in the assembly of Dol-linked oligosaccharides decrease the pool of substrates for the N-linked glycosylation of nascent proteins. Such glycosylation defects can be diagnosed by isoelectric focusing (IEF) of blood serum transferrin, which carries two N-glycans. IEF separates the glycoforms of transferrin by negative charges, which are carried by terminal Sia on N-glycans. The simplicity of this test was instrumental in identifying more than 50 N-linked glycosylation defects in the last twenty years.

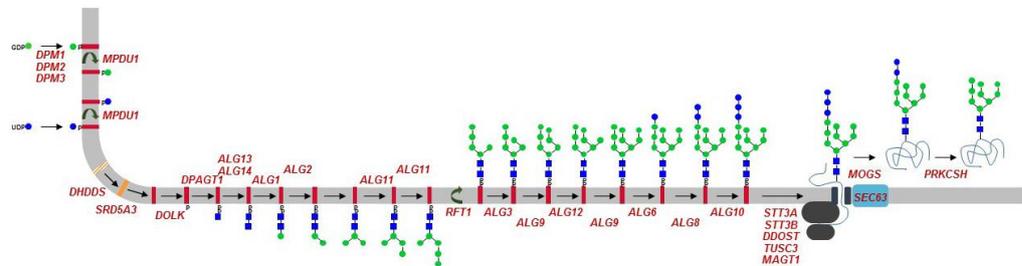


**Figure 6** – Serum transferrin carries two N-glycans normally terminated with four to six negatively charged Sia residues. Separation of transferrin glycoforms by IEF yields bands for each number of Sia. Transferrin from CDG patients lack complete N-glycans or only terminal carbohydrates, thereby decreasing the number of negative charges.

Nearly all genes involved in the biosynthesis of Dol, GDP-Man, Dol-P-Man and of Dol-linked oligosaccharides have been identified as cause of a glycosylation defect in humans. Mutations in some OST subunit genes and in the glucosidase-I and -II genes have also been described. Because of their similar yet broad clinical manifestations, defects of N-glycosylation have been grouped under the general term of congenital disorders of glycosylation (CDG). Soon after the description of the first cases, CDG have been divided into two groups, where CDG type-I, or CDG-I in short, encompassed glycosylation defects leading to the non-occupancy of N-glycosylation sites on glycoproteins, and CDG-II encompassing all other glycosylation defects. Accordingly, most ER N-glycosylation defects belonged to CDG-I. Individual gene defects were further specified by adding a distinct letter, thereby yielding CDG-Ia, CDG-Ib, CDG-Ic, etc. The identification of mutations affecting multiple glycosylation pathways prompted a revision of the nomenclature to a more open system. Therefore, CDG are currently designated by simply linking the [HGNC](#) approved symbol of the mutated gene to the suffix “CDG”, thus defining PMM2-CDG, ALG6-CDG, and DPM1-CDG.

Neurological alterations are prominent in defects in ER N-linked glycosylation. Typical symptoms include psychomotor retardation, ataxia, hypotonia,

epilepsy, and visual impairments. Endocrine disorders, coagulopathy, gastrointestinal and hepatic impairments, cardiomyopathy, are though also frequently observed. The severity of the diseases is variable and depends largely on the level of residual glycosylation enabled by individual gene defects. Most severe impairments lead to infant lethality and weak mutations lead to moderate intellectual delays.



**Figure 7** – ER N-glycosylation defects identified in CDG patients. Defective genes are marked in red.

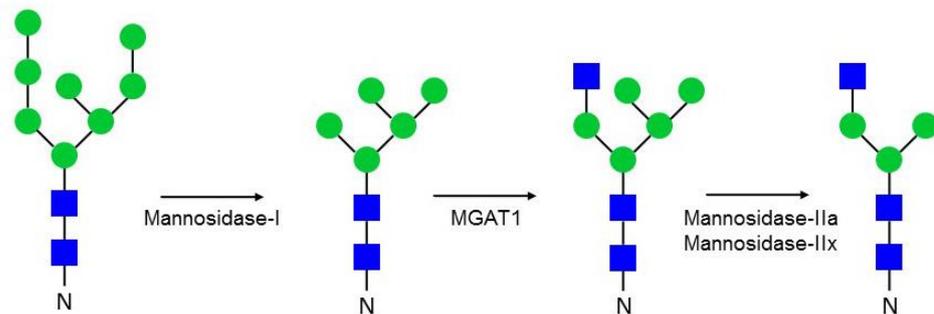
Defective phosphomannomutase activity is the most frequent form of CDG in respect to the number of mutations in the *PMM2* gene and number of patients identified to date. The *PMM2* enzyme catalyzes the conversion of Man-1-P from Man-6-P, which is an essential reaction in the biosynthesis of GDP-Man. Unfortunately, most CDGs remain without therapy, with the exception of mannose phosphate isomerase (MPI) defect, which can be circumvented by oral Man supplementation. MPI mediates the interconversion of fructose-6-P and Man-6-P, an important step in the maintenance of the cellular Man pool. MPI-CDG lacks the typical neurological symptoms of CDG and rather features gastrointestinal defects including diarrhea, protein-losing enteropathy, and intestinal bleeding.

## Golgi N-linked glycosylation

Whereas the pathway of N-linked glycosylation proceeds sequentially in the ER, the modifications of N-glycans in the Golgi follow parallel and alternative pathways, which contribute to the formation of variably branched and extended structures. The  $\text{GlcNAc}_2\text{Man}_8$  N-glycan entering Golgi can remain unchanged during transit and exit as high-mannose core structure. Alternatively, this core can be reduced down to three Man residues by mannosidases and elongated in various ways depending on the availability of glycosyltransferases and of donor substrates. Before trimming, N-glycans can be modified by addition of a phosphate group at the 6' position of selected mannose residues. Glycoproteins carrying N-glycans with such Man-6-P residues are recognized by the Man-6-P receptor in the Golgi and transported to lysosomes.

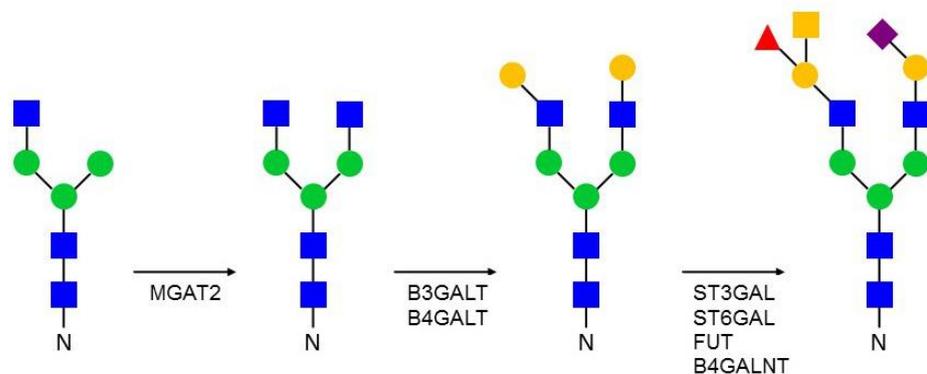
The first trimming step in the Golgi apparatus is performed by mannosidase-I enzymes, which cleave the two  $\alpha$ 1-2 linked Man sitting on the  $\alpha$ 1-3 linked Man branch. This step is necessary to enable further modifications, such as the addi-

tion of a  $\beta$ 1-2 linked GlcNAc to  $\alpha$ 1-3 linked Man. The branching GlcNAc-transferases are abbreviated MGAT for mannoside glycoprotein N-acetylglucosaminyl transferase. Accordingly, MGAT1 transfers the first branching  $\beta$ 1-2 GlcNAc to Man. This GlcNAc residue can be further elongated by Gal-, Fuc- and Sia-transferases, thereby giving rise to so-called hybrid N-glycans. The next Man trimming follows the addition of  $\beta$ 1-2 GlcNAc by MGAT1. Mannosidases-IIa and -IIx perform this trimming of the  $\alpha$ 1-3 and  $\alpha$ 1-6 linked Man residues. Both mannosidases-II share the same substrate specificity but are expressed in different tissues. The combined deficiency of mannosidases-IIa and -IIx is embryonic lethal in mice and precludes the formation of complex N-glycans.



**Figure 8** – Trimming of high-mannose N-glycans in the Golgi apparatus.

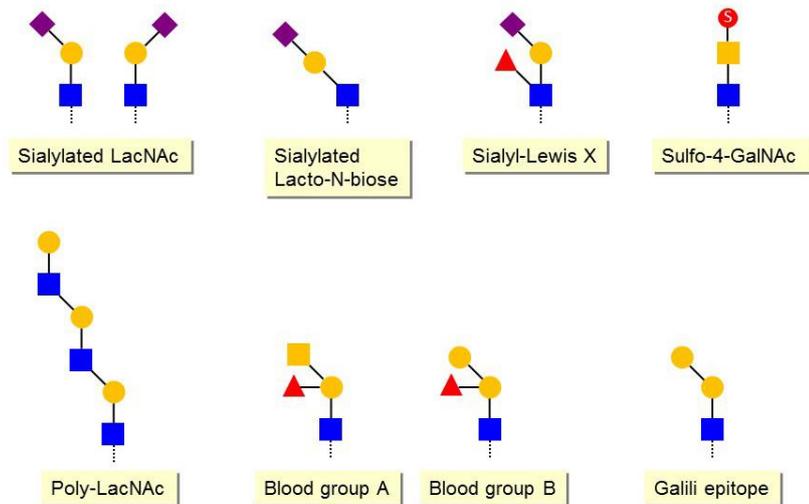
After mannosidase trimming, a second  $\beta$ 1-2 GlcNAc is added to the  $\alpha$ 1-6 linked Man by the MGAT2 enzyme. N-glycans comprising at least two elongated GlcNAc branches are called complex N-glycans. Also the first GlcNAc that is attached to Asn can be modified by addition of a  $\alpha$ 1-6 Fuc. The elongation of GlcNAc branches is variable since it depends on glycosyltransferases, which are expressed in developmental stage- and tissue specific patterns. Some of the glycosyltransferases elongating N-glycans can also act on O-glycans and on glycosphingolipids. The structural heterogeneity of N-glycans is such that a single glycoprotein can carry complex, hybrid and high-mannose N-glycans on the same polypeptide.



**Figure 9** – Elongation of biantennary N-glycans in the Golgi apparatus.



forming sialylated LacNAc. Gal can alternatively be  $\beta$ 1-3 linked, yielding sialylated lacto-N-biose. Other typical terminal structures of N-glycans are fucosylated epitopes, such as the blood group A and B determinants and the sialyl Lewis-X antigen, which acts as ligand for selectin proteins. N-glycans can also carry long extensions of poly-LacNAc, so-called polylactosamine chains, which are mainly found on the MGAT5-mediated  $\beta$ 1-6 branch. Terminal structures are not specific to N-glycans since most are also found on O-glycans and on glycosphingolipids. However, the sulfated-4-GalNAc epitope is indeed specific to the N-glycans of gonadotropin hormones and mediate essential signals to the function of these hormones (see below).



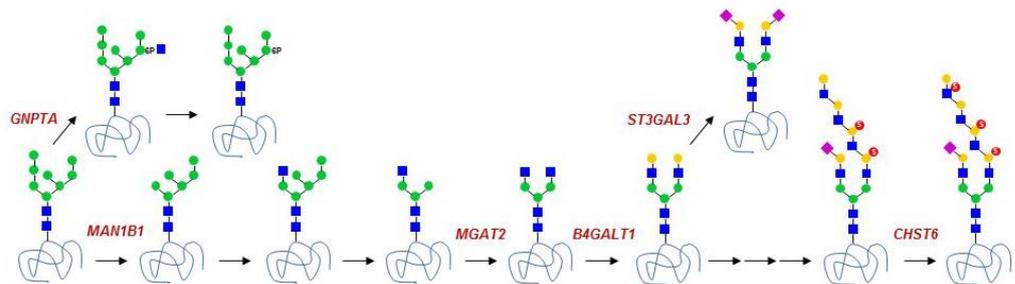
**Figure 11** – Typical terminal epitopes found on N-glycans. The Gal( $\alpha$ 1-3)Gal terminus (Galili epitope) does not occur on human cells.

Some terminal structures like the Gal( $\alpha$ 1-3)Gal Galili epitope are not found on human N-glycans, because the gene encoding the terminally acting  $\alpha$ 1-3 Gal-transferase is inactive in the human genome. Accordingly, humans have high titers of antibodies against the Galili epitope since it is commonly found on bacteria. The Galili epitope is also widespread on most mammalian cells and is therefore responsible for the hyperacute rejection of xenotransplanted mammalian organs in humans.

## Diseases of Golgi N-glycosylation

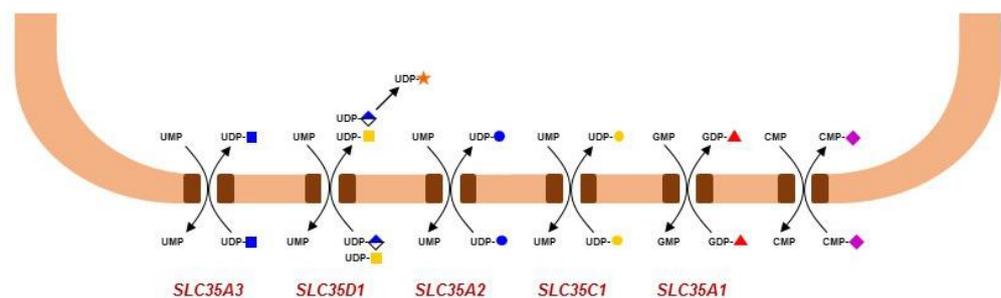
Only few defects of Golgi N-glycosylation are known in humans to date. Some are asymptomatic like the deficiencies of the terminal fucosyltransferases FUT3 and FUT6, which are found in about 9% of individuals on the island of Java in Indonesia. This lack of apparent phenotype is surprising considering that FUT6 is a major fucosyltransferase involved in the fucosylation of liver proteins.

A single defect of N-glycan branching has been identified to date. Mutations in the *MGAT2* gene lead to the loss of complex N-glycans. This defect causes a severe disease, characterized by multiple organ dysfunctions and elevated childhood lethality. A rare form of galactosylation deficiency is caused by mutations in the  $\beta$ 1-4 Gal-transferase *B4GALT1* gene. So far only two patients have been identified with *B4GALT1*-CDG. In the first case, *B4GALT1* deficiency was associated to severe clinical features such as psychomotor retardation, hypotonia, myotonia and macrocephaly. Paradoxically, the second case presented with mild symptoms of mental retardation and hepatic dysfunction. It is currently unclear whether the second case represented only a partial galactosylation defect.



**Figure 12** - Golgi glycosidase and glycosyltransferase defects known to date.

Glycosylation defects in the Golgi are usually not specific to N-glycans, because enzymes such as *B4GLAT1* also act on several types of O-glycans and glycosphingolipids. The same remark applies to defects of nucleotide-activated sugar transport into the Golgi, since these donor substrates are used to elongate multiple types of glycans. For example, mutations in the *SLC35C1* gene encoding the Golgi GDP-Fuc transporter leads to a generalized defect of fucosylation in multiple classes of glycans. By contrast, ER O-fucosylation is not affected by the *SLC35C1* defect, indicating that another GDP-Fuc transporter is active in the ER.



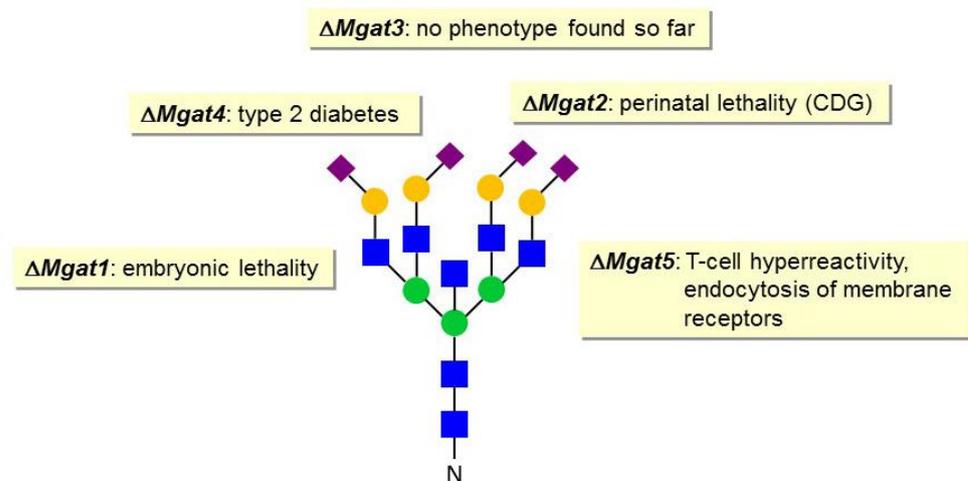
**Figure 13** - Golgi-localized nucleotide-activated sugar transporter defects known to date.

The loss of Golgi fucosylation leads to a form of immunodeficiency called Leukocyte Adhesion Deficiency type-II (LAD2) or *SLC35C1*-CDG. Fucosylated glycans, such as sialyl Lewis-X, are ligands of selectin proteins, which are important for the rolling of leukocytes on blood vessel walls and thereby for the extravasation of leukocytes into tissues. The treatment of LAD2 patients with

dietary Fuc has been shown to partially restore Golgi fucosylation and leukocyte rolling. Mutations in the CMP-Sia acid transporter gene *SLC35A1* have also been described in two infants presenting distinct pathologies. The first child was affected by with recurrent bleeding episodes, thrombocytopenia, neutropenia and multiple bacterial infections, whereas the second child presented with intellectual impairment, seizures, ataxia, renal and cardiac disorders. Additional cases are required to delineate the importance of sialylation in human development and physiology.

## Disruption of Golgi N-glycosylation genes in mice

Whereas MGAT2 deficiency is the only known N-glycan branching defects in humans, nearly all branching *Mgat* genes have been disrupted in mice. The inactivation of *Mgat1* gene leads to embryonic lethality, which is probably related to defective neural tube closure and vascularization defects. The disruption of the *Mgat2* gene is associated to perinatal lethality in a mouse strain (129/Sv background) and to survival to the adult stage but with multiple organ dysfunctions in another strain (C57Black/6 background). *Mgat2* knockout mice show similar phenotypes to those of MGAT2-CDG patients.

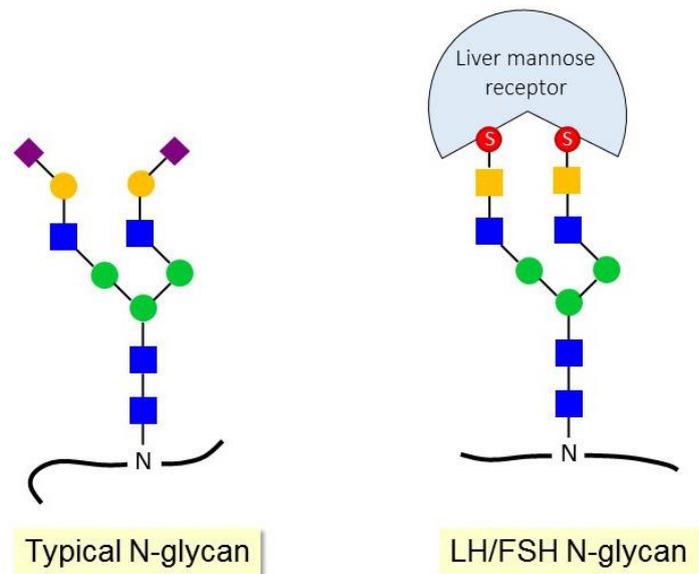


**Figure 14** – Phenotypes of branching defects in gene knockout mice.

The loss of the *Mgat3* gene does not impair mouse development and viability. In fact, no phenotype could be identified so far in this mouse knockout model. By contrast, the inactivation of the *Mgat4a* gene prevents the formation of tri-antennary N-glycans, thereby affecting the stability of glycoproteins like the Glc transporter GLUT2 at the surface of pancreatic  $\beta$ -cells. The decreased presence of GLUT2 at the cell membrane results in Glc insensitivity and to a form of type 2 diabetes. A similar decreased stability of membrane glycoproteins is observed in *Mgat5*-knockout mice. This instability is caused by increased endocytosis due to reduced interaction of glycoproteins with the galectin lattice at the cell membrane. The alteration of membrane glycoprotein expression results in various

alterations in signaling cascades such as exemplified by a T-cell hyperactivity phenotype.

The N-glycans of the gonadotropins follicle-stimulating hormone (FSH) and luteinizing hormone (LH) carry a unique sulfated GalNAc epitope instead of Sia. The sulfated groups are essential for the binding of these hormones to the mannose receptor in the liver. The liver mannose receptor is responsible for the rapid clearance of FSH and LH from the blood circulation, which is critical to maintain a pulsatile stimulation of the gonadotropin receptors on gonad cells.



**Figure 15** – Sulfated GalNAc termini on gonadotropin N-glycans.

In fact, the pulsatile release of gonadotropins by the pituitary gland mediates the hormonal signal to the target cells. The pulsatile release of gonadotropins is stimulated by the peptide hormone GnRH, which is itself secreted by hypothalamic neurons in a pulsatile manner. Alterations in the pulsatile pattern of GnRH and FSH/LH lead to various hormonal disorders ranging from the shut-down of gonadotropin production to abnormal gonadal and gamete development.

The inactivation of the GalNAc-4-sulfotransferase gene in mice impairs the clearance of FSH and LH through the liver mannose receptor, which leads to elevated levels of the gonadotropins in blood. The increased gonadotropin concentrations affect the production of the sexual hormones testosterone, estradiol and progesterone and thereby the development and function of gonads and accessory glands. For example, female mice lacking the GalNAc-4-sulfotransferase gene are constantly ovulating, which also yield litters larger than average.

Another typical feature of Golgi N-glycans is the fucosylation of the GlcNAc core. This step is catalyzed by the  $\alpha$ 1-6 fucosyltransferase *FUT8* gene product. Disruption of the *Fut8* gene in mice (see PNAS 102, 15791-15796, 2005) and the corresponding loss of core fucosylation affects the binding of TGF $\beta$  to its

receptors. This alteration of TGF $\beta$  signaling induces the expression of metalloproteinases and decreases the production of extracellular matrix proteins in lungs, which leads to pulmonary emphysema and increased lethality. The mechanisms underlying the regulation of TGF $\beta$  binding by core fucosylation are however unclear at this stage.