*seminars in* **DEVELOPMENTAL BIOLOGY**, Vol 6, 1995: pp 127–138

# Oligosaccharides in vertebrate development

*Ajit Varki and Jamey Marth**

*The outer surfaces of eukaryotic cells are covered with a dense and complex array of sugar chains (oligosaccharides). Most are found attached to other macromolecules, yielding glycoconjugates such as glycoproteins and glycolipids. Given their location and structural complexity, it is natural to predict their involvement in cell–cell interactions. Several examples of such interactions have been defined in animal systems. Since the expression of many oligosaccharides is tissue-specific and developmentally regulated, they may also be involved in embryonic development. Genetic evidence in favor of this notion has recently been obtained. This article provides a perspective on glycosylation in vertebrate development and experimental approaches towards elucidating oligosaccharide function.*

**Key words:** glycosylation / oligosaccharides / lectins / transgenic animals / gene-targeting

## Overview of the types of vertebrate glycoconjugates and their shared outer chains

Glycoconjugates consist of sugar residues attached to proteins and lipids and are classified into distinctive groups based upon unique 'core' structures (Figure 1), (for review see ref 1). Many proteins contain asparagine, serine or threonine residues in sequence contexts that target them for saccharide addition. With the exception of asparagine-linked glycosylation, the signals that determine whether a certain amino acid is to be glycosylated are not completely defined.[1] The most common types of oligosaccharides on cell surface and secreted proteins are defined by their linkage regions as 'N-linked glycans' (N-acetylglucosamine-asparagine), 'O-linked glycans' (N-acetylgalactosamine-serine/threonine), and 'proteoglycans' (Xylose-serine) (see Figure 1). Glycolipids are a distinct class that have glucose or galactose residues directly attached to a lipid (ceramide) core. 'GPI-anchors' have a characteristic core oligosaccharide that bridges between protein and membrane-anchored phosphatidyl inositol. In all these classes of molecules, the core structures are generally modified by glycosyltransferase-mediated monosaccharide addition to generate unique and complex oligosaccharide variations.[1] The discovery that nuclear and cytoplasmic glycosylation[2] is also widespread has greatly expanded the range of molecules that are glycoconjugates. In the commonest type of intracellular glycosylation, a single monosaccharide (N-acetylglucosamine) is added on to either serine or threonine (Figure 1). In addition to these major classes, there are other types of 'non-classical' glycosylation, e.g. O-linked fucose and mannose (see Table 1, and ref 1).

Although the proximal or 'core' oligosaccharide structures are unique among the different classes of glycoconjugates, continued biosynthesis by glycosyltransferases can produce structures that are identically modified at more distal biosynthetic points (See Table 1 and refs 1,3). Thus, there are both unique and shared attributes among the different classes of glycoconjugates.

## Oligosaccharide biosynthesis and glycosyltransferase specificity

Oligosaccharides are products of multiple enzymatic steps during which specific saccharides are either added or deleted. Glycosylation is thus a stepwise biosynthetic process that, in the case of secreted and cell surface glycoconjugates, occurs predominantly in the Golgi apparatus. While the substrate specificities for initial saccharide addition may reside in the protein, those for further addition are mostly (though not exclusively) based upon the oligosaccharide structures themselves. Hence, biological information can reside within oligosaccharide structures, potentially regardless of the peptide backbone, which may, in some cases serve as 'templates' for oligosaccharide production and variation. While glycosyltransferases act to link saccharides in generating oligosaccharide

*From the Division of Cellular and Molecular Medicine, Glycobiology Program, UCSD Cancer Center, and the *Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA 90293, USA*



*1044-5781/95/020127+12$8.00/0*

structures, glycosidases can remove specific saccharides. Both can be required for the final generation of a particular oligosaccharide structure.

Each monosaccharide can have multiple hydroxyl groups that are potential attachment sites for others, and can exist in two different anomeric states (α or β). Thus, the attachment (glycosidic linkage) of a given monosaccharide to an underlying oligosaccharide can be distinctive, both in terms of being α or β, and in terms of the attachment site on the ring (e.g. the disaccharide Galβ1-4GlcNAcβ1-, commonly called 'lactosamine' has a galactose residue beta-linked from its 1-position to the 4-hydroxyl group of the underlying N-acetylglucosamine). The sugar donors for these transferase reactions are made in the cytosol, and then translocated into the lumen of the endoplasmic reticulum-Golgi pathway.

While much evidence favors a 'one linkage: one enzyme' rule during biosynthesis, some exceptions do occur, mostly in the direction of 'one linkage: multiple enzymes' (see refs 4-6) for examples, and Table 1). Additionally, different glycosyltransferases may compete for a particular oligosaccharide on a particular glycoprotein. For example, certain fucosyltransferases, sialyltransferases and galactosyltransferases may all compete to add additional sugars

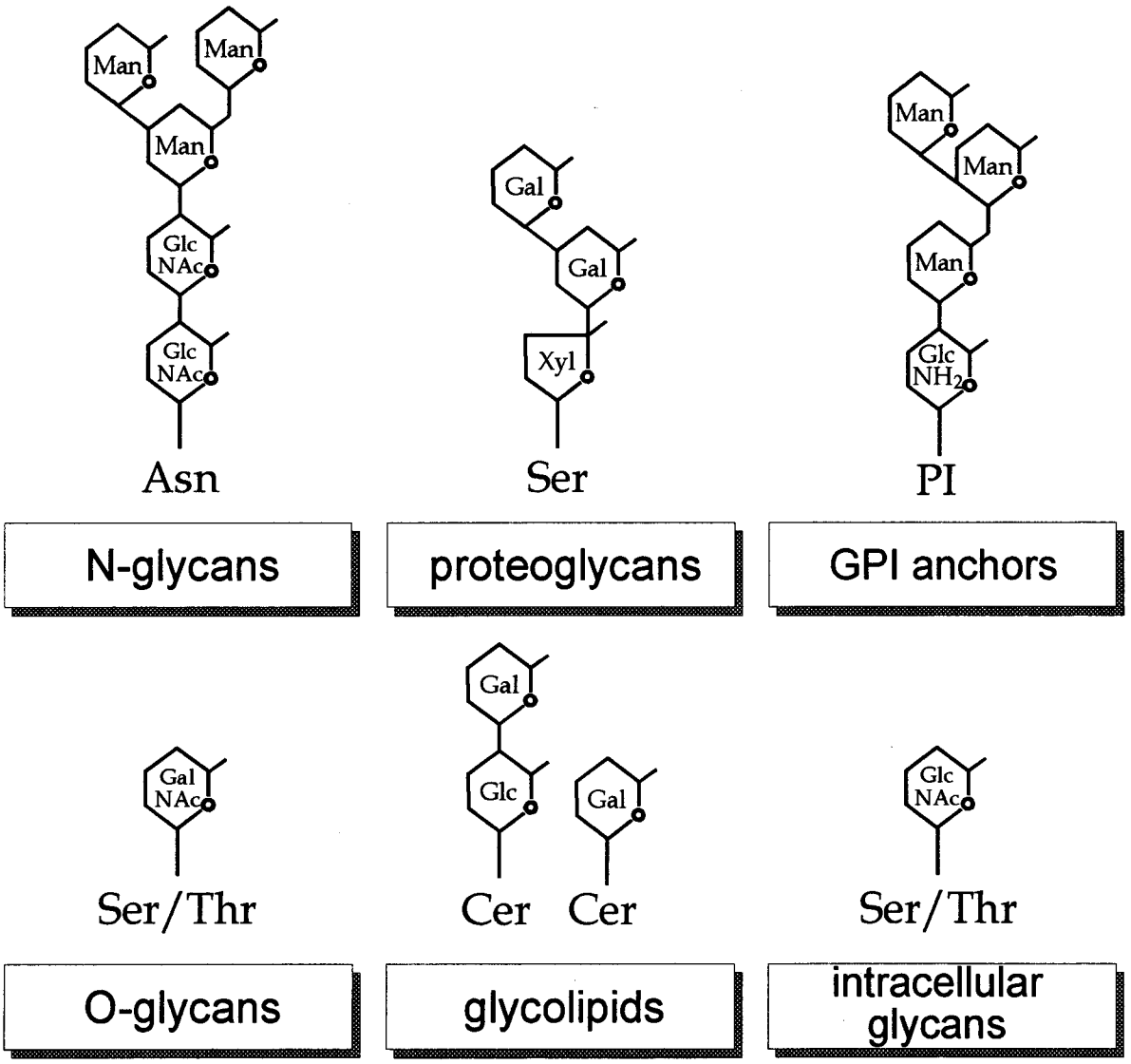


**Figure 1.** Invariant core structures of the major classes of vertebrate glycoconjugates are shown in stylized fashion. Commonly used terms for these classes are indicated in the boxes.

**Table 1.** Gene products involved in the biosynthesis of oligosaccharides in vertebrates: minimum estimates, genes cloned and occurrence of defects

| Gene product(s)* involved in: | Minimum estimate** | Genes cloned† | Defects§ |
|---|---|---|---|
| **Donor production pathways¶** | | | |
| UDP-Glc synthesis | 3 | | |
| UDP-Gal synthesis | 4 | | 3 |
| GDP-Man synthesis | 2 | | 1? |
| GDP-Fuc synthesis | 2 | | 1? |
| UDP-GlcNAc synthesis | 6 | | |
| UDP-GalNAc synthesis | 1 | | |
| UDP-Xyl synthesis | 1 | | |
| UDP-GlcA synthesis | 1 | | |
| CMP-Neu5Ac synthesis | 5 | | |
| CMP-Neu5Gc synthesis | 3 | 1 | |
| **Sugar nucleotide transporters¶** | | | |
| UDP-Glc transporter | 1 | | |
| UDP-Gal transporter | 1 | | |
| GDP-Man transporter | 1 | | |
| GDP-Fuc transporter | 1 | | |
| UDP-GlcNAc transporter | 1 | | |
| UDP-GalNAc transporter | 1 | | |
| UDP-Xyl transporter | 1 | | |
| UDP-GlA transporter | 1 | | |
| CMP-Sia transporter | 1 | | |
| **Nucleotide degrading enzymes¶** | | | |
| Dinucleotide pyrophosphatase(s) | 1 | | |
| CMP-Sia hydrolase | 1 | | |
| **N-GlcNAc-Asn linked pathway** | | | |
| Dolichol synthesis | 5 | 2 | |
| Dolichol monosaccharide synthesis | 3 | 2 | |
| Chitobiose core synthesis | 2 | | |
| GDP-Man dependent transferases | 5 | 1 | |
| Dol-P-GlcNAc dependent transferases | 2 | 1 | |
| Dol-P-Man dependent transferases | 4 | 1 | |
| Dol-P-Glc dependent transferases | 3 | 1 | |
| Oligosaccharyltransferase | 2 | 1 | |
| Glucosidase I | 1 | | |
| Glucosidase II | 1 | | |
| Reglucosylating glucosyltransferase | 1 | | |
| ER Man9->Man8α–Mannosidase(s) | 1 | 1 | |
| Endo-α-mannosidase | 1 | | |
| GlcNAc phosphotransferase | 1 | | 1 |
| ER/Golgi α–1,2 mannosidase(s) | 2 | 2 | |
| Phosphodiester glycosidase | 1 | | 1? |
| GlcNAc transferase I | 1 | 1 | 1‡ |
| α–Mannosidase II | 2 | 2 | 1 |
| GlcNAc transferase II | 1 | 1 | 1 |
| Core fucosyltransferase | 1 | | |
| Bissecting GlcNAc transferase III | 1 | 1 | |
| Branching GlcNAc transferase IV | 1 | | |
| Branching GlcNAc transferase V | 1 | 1 | |
| Bissecting GlcNAc transferase VI | 1 | | |
| **O-GalNAc-Ser linked Pathway** | | | |
| Ser: α–GalNAc transferase(s) | 2 | 2 | 1‡ |
| Core 1 β1-3 galactosyltransferase | 1 | | 1(s) |
| Core 2 β1-6 GlcNAc transferase | 1 | 1 | |
| Core 3 β1-3 GlcNAc transferase | 1 | | |
| Core 4 β1-6 GlcNAc transferase | 1 | 1 | |
| Core 1 β1-3 GlcNAc transferase | 1 | | |
| Core 2 β1-3 GlcNAc transferase | 1 | | |
| Core 5 α1-3 GalNAc transferase | 1 | | |
| Core 6 β1-6 GlcNAc transferase | 1 | | |
| Core 7 β1-6 GalNAc transferase | 1 | | |
| O-glycan α2–6 sialyl transferases | 2 | 2 | |
| O-glycan α2-3 sialyl transferases | 1 | 1 | |
| **Ceramide-linked Pathways** | | | |
| Ceramide synthesis | 3 | | |
| Glucosylceramide synthase | 1 | | |
| Glucosylceramide 'flippase' | 1 | | |
| Lactosylceramide synthetase | 1 | | |
| α2-3 Sialyltransferase ($G_{M3}$ synthase) | 1 | 1 | |
| GalNAcβ1-4 transferase ($G_{M2}$ synthase) | 1 | 1 | |
| Galβ1-3 transferase ($G_{M1}$ synthase) | 1 | | |
| Lacto β1-3 GlcNAc transferase | 1 | | |
| Lacto β1-3 galactosyltransferase | 1 | | |
| Neolacto β1-4 galatosyltransferase | 1 | | |
| Globo α1-4 galactosyltransferase | 1 | | |
| Globo β1-3 GalNAc transferase | 1 | | |
| Isoglobo α1-3 galactosyltransferase | 1 | | |
| Isoglobo β1-3 GalNAc transferase | 1 | | |
| Muco β1-4 galactosyltransferase | 1 | | |
| Forsmann α1-3 GalNAc transferase | 1 | | |
| Extending α1-4 galactosyltransferase | 1 | | |
| α2-3 sialyltransferase | 1 | | |
| α2-8 sialyltransferases | 3 | 1 | |
| Galactosylceramide synthase | 1 | 1 | |
| Galactosylceramide O-acyltransferases | 3 | | |
| α2-6 Sialyltransferase ($G_{M4}$ synthase) | 1 | | |
| Gal-Cer α1-4 galactosyltransferase | 1 | | |
| Glycolipid sulfotransferases | 3 | | |
| **Common Pathways of chain elongation** | | | |
| β1-4 galactosyltransferase(s) | 2 | 1 | |
| β1-3 galactosyltransferase(s) | 1 | | |
| β1-4 GalNAc transferase(s) | 2 | 1 | 1?(v) |
| β1-3 GlcNAc transferases | 2 | | |
| β1-6 GlcNAc transferase(s) | 1 | | |
| LacNAc: α2-3sialyltransferases | 2 | 2 | |
| LacNAc: α2-6sialyltransferase(s) | 1 | 1 | |
| α2-4 Sialyltransferase(s) | 1 | | |
| α2-8 Sialyltransferases | 2 | 1 | |
| β1-4 Glucuronosyltransferase | 1 | | |
| α1-4 Galactosyltransferase | 1 | | |
| α1-3(4) Fucosyltransferases | 5 | 5 | 2‡ |
| Branching β1-4 galactosyltransferase | 1 | | |
| Sialate: O-acetyltransferases | 3 | | |
| O-Sulfotransferases | 4 | | 1? |
| α1-2 Fucosyltransferases | 2 | 2 | 2(v) |
| A-blood group transferase | 1 | 1 | 1(v) |
| B-blood group transferase | 1 | 1 | 1(v) |
| α1-3 Galactosyltransferase | 1 | 1 | 1(v)‡ |
| **O-Xyl linked pathways** | | | |
| Ser: β-Xylosyltransferase | 1 | | |
| β-Galactosyltransferase I | 1 | | 1? |
| β-Galactosyltransferase II | 1 | | |
| Core xylosyl-phosphotransferase | 1 | | |
| Core galactosyl-sulfotransferases | 2 | | |
| β-Glucuronosyltransferase I | 1 | | |
| CS: GalNAcβ1-4 transferases | 2 | | |
| CS: β-Glucuronosyltransferase | 1 | | |
| Chondroitin 4/6 O-sulfotransferases | 2 | | |
| DS: Glucuronosyl epimerase | 1 | | |
| DS: Uronate 2-O-sulfotransferase | 1 | | |
| HS: GlcNAcα1-4 transferases | 2 | | |
| HS: β-Glucuronosyltransferase | | | |
| HS: de-N-acetylase | 2 | 2 | |

**Table 1. cont.**

| Gene product(s)* involved in: | Minimum estimate** | Genes cloned† | Defects§ |
|---|---|---|---|
| HS: N-sulfotransferase | | | |
| HS: Glucuronosylepimerase | 1 | | |
| HS: Uronate 2-O-sulfotransferase | 1 | | |
| HS: GlcN-O-sulfotransferase(s) | 2 | | |
| HS: Endoglucurononidase(s) | 1 | | |
| **Hyaluronan pathway** | | | |
| β1-4 GlcNAc transferase | 1 | | |
| β-Glucuronosyltransferase | 1 | | |
| **GPI-linked pathways** | | | |
| PI: αGlcNAc transferase complex? | 3 | 2 | 1(s) |
| GPI: GlcNAc de-N-acetylase | 1 | | |
| α1-4 Mannosyltransferase | 1 | | |
| α1-6 Mannosyltransferase | 1 | | |
| α1-2 Mannosyltransferase | 1 | | |
| Core ethanolamine phosphotransferase | 1 | | |
| GPI Palmitoyltransferase | 1 | | |
| GPI Palmitoyl deacylase | 1 | | |
| Transamidase(s) | 1 | | |
| α-Galactosyltransferase | 1 | | |
| Ethanolamine phosphotransferases | 2 | | |
| GalNAcβ1-4 transferase | 1 | | |
| GlcNAc transferase(s) | 1 | | |
| α-mannosyltransferase(s) | 2 | | |
| GPI-specific phospholipase(s) | 2 | | |
| **N-Glc-Asn linked pathway** | | | |
| Asn: β-Glucosyltransferase? | 1 | | |
| Elongating glucosyltransferase(s)? | 2 | | |
| **Xyl-Glc-Ser pathway** | | | |
| Ser: Glucosyltransferase | 1 | | |
| β-Xylosyltransferase(s) | 2 | | |
| **O-Linked Man Pathway** | | | |
| Ser: α-mannosyltransferase(s) | 1 | | |
| **O-Linked Fuc Pathway** | | | |
| Ser: α-Fucosyltransferase | 1 | | |
| β1-3GlcNAc transferase | 1 | | |
| β1-4Galactosyltransferase | 1 | | |
| α2-6Sialyltransferase | 1 | | |
| **Glc-hydroxylysine Pathway** | | | |
| Hydroxylysine: glucosyltransferase | 1 | | |
| β-Galactosyltransferase | 1 | | |
| **GalNAc-phospho-X pathway** | | | |
| GalNAc phosphotransferase(s) | 1 | | |
| **Cytosolic O-GlcNAc pathway** | | | |
| GlcNAc transferase(s) | 1 | | |
| GlcNAcase(s) | 1 | | |
| **Cytosolic O-Linked Man pathway** | | | |
| Ser: α-mannosyltransferase(s) | 1 | | |
| Man: Glc Phosphotransferase | 1 | | |
| **Cytosolic O-linked Fuc pathway** | | | |
| Thr: α-Fucosyltransferase | 1 | | |
| β-Glucosyltransferase | 1 | | |
| **Cytosolic glycogen pathway** | | | |
| Tyr: Glucosyltransferase (glycogenin) | 1 | 1 | 1 |
| α1-4Glucosyltransferase(s) | 2 | | |
| Branching α1-6 glucosyltransferase | 1 | | 1 |
| Totals: | 232 | 52 | 24 |

* Because of space limitations, incomplete names are used for most of the gene products.

** The minimum estimate is based on enzymology and other available information. It is conservatively assumed that when identical outer chain sequences are found on different types of oligosaccharides (e.g. on N-linked oligosaccharides and ceramide-linked oligosaccharides) the same gene product is involved in synthesizing the structure in question. In fact, some studies indicate that multiple gene products can encode enzymes that synthesize the same linkage. In those instances in which this has been clarified, the table indicates so.

† cDNA cloning and/or genomic cloning from vertebrates.

§ Reported in intact vertebrate organisms. The question marks (?) indicate that the defect has not been molecularly defined. Some defects indicated by (s) are somatic, and are thus not inheritable. In other cases (v) the defects represent intra- or inter-species variants not associated with any overt pathology.

¶ Excluding donors for modifications e.g. acetyl-Coenzyme A, since these act as donors in other unrelated pathways as well.

‡ Experimentally derived by gene targetting.

to the lactosamine unit indicated above. Also, some glycosyltransferases can synthesize linkages with differing efficiency, depending upon the substrate. Oligosaccharide variation is therefore dictated by multiple factors, which include the nature of the underlying glycoconjugate, glycosyltransferase donor and substrate availability, transit time through the Golgi apparatus and the levels and compartmentalized localization of competing glycosyltransferases. Further complexity is also generated by modifications (e.g. epimerization, sulfation, phosphorylation, or acetylation) of certain monosaccharide residues on selected oligosaccharide chains. Finally, many glycoproteins can be represented by multiple glycoforms, with oligosaccharide variation found upon the same peptide backbone, even as products of the same clonal cell population. Taken together, these attributes provide an explanation for the origin of oligosaccharide complexity as found in nature.

## How many gene products are involved in vertebrate oligosaccharide biosynthesis?

From the above, it follows that oligosaccharide production by vertebrate cells requires sugar nucleotide donors, sugar nucleotide transporters, specific glycosyltransferases and glycosidases, as well as enzymes

dictating subsequent modifications. Based upon these considerations, and assuming the most conservative 'one linkage:one enzyme' scenario, one can estimate that at least 232 genes in the vertebrate genome have evolved to function in oligosaccharide synthesis and structural variation (see Table 1). To date, less than 25% of these predicted gene products have been genetically cloned. When all of them are ultimately isolated, the figure of 232 may well prove to be a gross underestimate. This number also does not take into account the many cognate lectin receptors that exist primarily or exclusively to utilize this enormous amount of biological information (for review see refs. 7–9). Hence, at least 500 genes (i.e. ~ 0.5–1.0% of the translated genome) probably participate in oligosaccharide production and function.

## Phylogenetic and evolutionary considerations

Among eukaryotes, the biosynthesis and expression of mammalian oligosaccharides is best characterized at present. Although data from other organisms is sparse, available information indicates that increasing complexity correlates with phylogenic time. However, complex-type-N-linked oligosaccharides have not been found in invertebrates studied.[10] They are not common among, and may be absent from organisms within the largest phylum — *Arthropoda*.[11] Moreover, while all vertebrates display these oligosaccharides, species-specific variations are found. These variations include the presence of unique glycosyltransferases, as well as examples of null alleles.[12] As might be expected, the stepwise biosynthesis of oligosaccharides in vertebrates is somewhat recapitulated in phylogeny.

## Differentially regulated expression of glycosyltransferases and oligosaccharides

Oligosaccharide production and turnover have been shown to occur with a high degree of spatial and temporal regulation in vertebrates. Many glycosyltransferases, especially those acting more distally in the biosynthetic pathway, exhibit a significant amount of tissue-specific and developmentally-regulated, expression.[13-15] This is clear from work at several levels including gene expression analyses, lectin binding patterns, and oligosaccharide-specific antibody surveys (for some examples, see refs 16–25).

The biological roles of this multitude of glycosyltransferases and oligosaccharides have not been easily perceived. In part this has been due to technical limitations and the paucity of cloned genes (all genes isolated were cloned within the last five years). Additionally, while numerous inactivating mutations of glycosylation have been characterized in cell lines, most have not yielded remarkable phenotypes, even when the oligosaccharide repertoire in such cells has been severely restricted.[26,27] Hence, many aspects of oligosaccharide function are perhaps not manifested in such in-vitro studies. Taken together with the increasing complexity of oligosaccharide biosynthesis during vertebrate phylogeny, the implication is that these genes play important roles in ontogeny and tissue morphogenesis.

## Defects in glycosylation in intact animals can have serious pathologic consequences

Naturally occurring genetic disorders often point the way to biological function and relevant biochemical pathways. Glycosyltransferase action in disease etiology could theoretically occur either by recessive lesions, or following dominant mutations that act to antagonize and compete with the action of other glycosyltransferases. Several human diseases have been reported to result from defects in glycosylation, e.g. I-cell (inclusion cell) disease, hereditary erythrocyte multinuclearity with positive acidified serum test (HEMPAS), leucocyte adhesion deficiency (LAD) type II, carbohydrate deficient glycoprotein syndromes (CDGS) types I and II, paroxysmal nocturnal hemoglobinuria (PNH), and Tn polyagglutinin syndrome.[12] In some cases (see Table 1), the defects appear to be somatic and hence not inheritable (e.g. PNH and Tn polyagglutinin syndrome affect only hematopoeietic cells). Also, other genetic 'defects' in glycosylation exist (e.g. the ABO blood groups, see Table 1) that appear to be intra- or inter-species variations in glycosylation of somewhat uncertain biological significance.

Of the 232 predicted gene products listed in Table 1, about 10% are identified with vertebrate mutations *in vivo*, with differing biological consequences. As the genes that control carbohydrate production have only recently begun to be isolated, the number of associated diseases may increase significantly in the next several years. Nevertheless, it is interesting to speculate as to why more glycosylation disorders have not been detected earlier by non-genetic studies, as a large percentage of human inborn errors remain

biochemically undefined. One possibility includes the fact that phenotyping for oligosaccharide variation is presently beyond the capabilities of most clinical and research laboratories. In fact, CGDS was serendipitously defined as a glycosylation defect following a test normally used to screen for alcoholism.[28] Failure of detection may also occur due to the general requirement for many oligosaccharides in ontogeny (i.e. lethal mutations), although gene-targeting studies are required for strong support of this view. Perhaps mutations that are somatic or chimeric in nature, and/or $K_m$ alterations, are necessary for some glycosyltransferase deficiencies to result in viable animals. An alternative is that some glycosyltransferases are functionally redundant. However, evidence for this assertion is presently not available, and most linkages are in fact formed by unique glycosyltransferase gene products. Moreover, any redundancy found *in vitro* may not recapitulate the in-vivo situation, involving non-overlapping tissue-specific expression patterns of glycosyltransferases.

Dominant overexpression of glycosyltransferases has not yet been reported in genetic disease. However, the glycosyltransferase gene superfamily may contain members that can become pathogenic. The gene encoding GlcNAc transferase-V has long been known to be greatly induced during *src*- and *ras*-mediated transformation, and is implicated in enhancing metastastic cell behaviour.[29] Studies currently in progress may determine whether the gene encoding this glycosyltransferase can be defined as a proto-oncogene in mediating the tumorigenic conversion of immortalized cell lines.

## Clues to specific roles of oligosaccharides in development

With the advent of transgenic and gene-targeting methodologies, decisive in-vivo studies regarding oligosaccharide function in mammalian ontogeny, physiology and disease etiology can now be performed. While species-specific variations may restrict some comparisons with other vertebrates, the mouse remains at the forefront of genetic malleability and modeling.[30,31] Moreover, in the majority of gene-targeting studies to date, the mouse has either provided expected physiologic models of human conditions, or has revealed aspects of mammalian gene expression and physiology not previously appreciated.

Ontogeny in the mouse is relatively similar in comparison with other vertebrates, including humans. Following fertilization in mice, itself requiring O-linked oligosaccharides on oocyte zp3 for binding to the sperm receptor,[32,33] pre-implantation development occurs for approximately 4 days prior to blastocyst implantation in the uterus. While the first differentiated cell lineage is formed by the blastocyst stage (the trophoblast), little increase in total embryo volume occurs, mostly partitioning of previous mitotic products. Thus far, one endogenous gene has been shown to be required for pre-implantation development, while the embryonic genome itself is only significantly activated around the 8-cell stage. Hence, parental, and especially maternal, glycosylation factors could be involved in pre-implantation development. In this regard, transient expression of a virally-encoded sialic acid-specific O-acetylesterase from injected DNA blocks the first mitotic division,[34] and morulae compaction seems to require the $Le^x$ oligosaccharide structure also known as SSEA-1.[13,35]

Following these early stages, changes in the embryo include lectin- and antibody-visualized oligosaccharide variation,[13-15,17-23] implying that endogenous glycosyltransferases are operative. Studies to define a role of glycosylation in subsequent events have been few and far between. Inhibitors that block or alter the early steps of N-linked oligosaccharide production have been shown to affect development in the sea urchin,[36,37] and Swainsonine, an alkaloid that causes a moderate alteration in N-linked glycosylation,[38] is known to cause abortions in cattle. While these results are of interest, the possible non-specific toxic effects of glycosylation inhibitors may obscure the mechanisms behind such results.

With the induction of mesoderm from ectodermal precursors, the beginnings of significant cellular migration occur. By embryonic day 7 in the mouse, changes in oligosaccharides include increased branching of complex N-glycans[17] as L- and E-phytohemagglutinnin (L- and E-PHA) lectin binding has increased significantly (R. Campbell and J. Marth, submitted). L-PHA binding requires the action of the GlcNAc transferase V (GnT-V), the gene product mentioned above as affecting metastatic cell behaviour. By day 9, embryo vascularization is occurring while a major re-orientation of embryonic tissue begins. This includes organogenesis and the establishment of left–right body plan asymmetry. In the later stages of development, a role for cell surface β1-4 galactosyltransferase has also been proposed.[39]

The first gene-targeting studies involving a glycosyltransferase have been recently reported. Inactivation

of the mouse *mgat-I* locus, encoding GlcNAc transferase-I (GnT-I), was obtained and bred to homozygosity.[40,41] Without GnT-I, only high mannose forms of N-glycans are produced, structures much more prevalent in 'lower' eukaryotes and less 'complex' multicellular organisms. The results demonstrated that complex N-linked oligosaccharides are required for embryogenesis past embryonic day 9. Thus, while previous GnT-I inactivation in cultured CHO cells occurred without detectable phenotypes,[27] mouse embryos lacking this gene died by embryonic day 10. Morphogenic processes, and not cell differentiation *per se*, were affected while the determination of left–right body plan asymmetry was also impaired. These results have revealed the essential requirement for complex N-linked oligosaccharides during vertebrate ontogeny in the mouse.

There are presently no other published reports of gene-targeting studies involving glycosyltransferases, although many are underway. However, a role for polysialic acid in neural development may be indirectly inferred from studies of neural cell adhesion molecule (NCAM)-deficient mice, which displayed olfactory bulb defects.[42,43] NCAM is the major carrier of polysialic acid, generated by one or more α2,8 sialyltransferase(s). These oligosaccharide polymers may play an important role in neural axon branching by virtue of their enormous size and high negative charge (also see chapter by Rutishauser in this issue).[44] Also, overexpression of sialate: O-acetylesterase[34] or an α-galactosyltransferase[45] in mice caused variable abnormalities in specific tissues. In a very recent study, over expression of β1–4 galactosyltransferase in mice was reported to alter sperm function.[46] In conclusion, evidence for specific roles for glycosyltransferases and oligosaccharides in ontogeny is sparse at present, although the data from human mutations and *mgat-I*-deficient mouse embryos clearly indicates their importance and supports the view that further investigations are warranted.

## The need for in-vivo tissue-specific gene inactivation in complex multicellular systems

Oligosaccharides represent a large repertoire of highly regulated structures that exhibit tissue-specific expression patterns in development and in adult life. Since many are found coating the outer surface of cells, the specific information they contain likely contributes to cell–cell recognition events rather than directly modifying metabolic functions of the single cell. Thus, further genetic studies in complex multicellular systems are needed. Since conventional 'gene knockouts' may yield embryonic lethality as well as phenotypes involving multiple tissues and physiologic systems, it would be desirable to ablate specific oligosaccharide structures *in vivo*, in spatial and temporal patterns that are experimentally-determined. Recent developments with the Cre-lox transgene-recombination system provide the required technological advance.[47-49] Tissue-specific gene-targeting is done by first producing mice homozygous for lox-flanked ('floxed') genes by homologous recombination in embryonic stem cells. The resulting genetically altered and conditionally-mutant mice are mated with those expressing the Cre-recombinase enzyme under control of desired promoter sequences, yielding tissue-specific and developmentally-regulated gene deletions.[49] Such gene-targeting studies with 'floxed' glycosyltransferase genes should enable a systematic determination of tissue-specific functions of particular glycosyltransferases as well as the specific oligosaccharide products they generate.[31]

## The need to identify and ablate cognate receptors for oligosaccharides (endogenous lectins)

The more specific biological roles of oligosaccharides in development are likely to be mediated by their recognition and binding to lectins, the general term for oligosaccharide binding molecules. While it could be argued that each oligosaccharide structure will be recognized by a specific lectin, the actual correspondence may be less stoichiometric. Nevertheless, an understanding of the function of oligosaccharides will require more research into the identity of molecules that bind, or are functionally influenced by, carbohydrate structures. While many such lectins have already been identified,[7-9] recent studies indicate that a new family exists,[50] and it can be anticipated that many more will be discovered. Two naturally occurring defects in vertebrate lectins have been reported — deficiency of the serum mannose-binding protein in humans resulting in increased susceptibility to infection,[12] and deficiency of the cation-*independent* mannose-6 phosphate receptor in mice is associated with embryonic lethality.[8,51] Interpretation of the latter is complicated, because this receptor also binds insulin-like growth factor II, and by the nature of the genetic defect (imprinting of the *Tme* locus, that could

possibly contain other genes). In the few lectin gene-inactivation studies published to date, elimination of the cation-*dependent* mannose-6 phosphate receptor,[52,53] the selectins,[54-56] the asialoglycoprotein[57] or the L-14 galectin[58] did not give any grossly obvious abnormalities in development. However, in all but the two instances, further studies in adult animals revealed definite abnormalities in the appropriate systems.[52-56] Of course, in all these cases, related molecules exist that may have provided for functional redundancy.[7-9] Additional studies involving other vertebrate lectins are needed, especially with tissue-specific gene inactivation at defined stages of development.

## Genetic approaches to exploring oligosaccharide function: which experiments to do first?

In a biological situation involving a typical protein-:protein interaction, one can genetically explore the system by ablating either of the two components. In the case of glycosylation, the inherent complexities and interdependence of the biosynthetic steps increase the range of possibilities (see Figure 2). Given the large number of gene products involved in glycosylation, their extensive regulation in many tissues and cell types, and the paucity of information regarding their specific functions, multiple genetic approaches to determining their function may be contemplated. Each of these are mentioned below with some discussion of their merits (see also Figure 2).

One approach is to systemically inactivate crucial 'core' glycosyltransferases that are required for the production of entire chains and classes of oligosaccharides. The *mgat-I* gene-targeting studies represented an example of this strategy. Potential drawbacks include early lethality (preventing an understanding of downstream functions) and the enormous number of proteins and systems affected, making the likelihood of finding specific phenotypic effectors rather low. Instead, the *raison dêtre* for such a study is to provide general information on the physiologic processes that the particular class of oligosaccharides are involved in. This allows one to define the maximum developmental potential *in vivo* as well as any physiologic systems that are most highly affected, thereby aiding in identifying cell lineages for subsequent conditional gene-inactivation studies via the Cre-lox approach.

It can be noted that enzyme mutations, in general, can veil the identification of phenotypic effectors. For example, in-vivo studies with kinase and phosphatase deficiencies (results of natural mutations or gene-targeting experiments) often reveal a large number of affected phosphoproteins,[59,60] all of which are thus potential phenotypic effectors. Moreover, it remains unclear as to whether all kinase-mediated (or glycosyltransferase-generated) modifications are functionally significant. Any perceived pleiotropism resulting from such experiments does not present a valid argument against performing such studies. Rather such results highlight our current superficial understanding of molecular physiology, with the goal of understanding the etiology and function of complex physiologic systems *in vivo*.

A second approach involves eliminating the production of more distal and terminal events in oligosaccharide biosynthesis. Some glycosyltransferases, such as certain sialyltransferases and fucosyltransferases, provide terminal saccharide additions to oligosaccharides, while other enzymes can add specific modifications (sulfation or acetylation) to such structures. Of such genes thus far cloned and characterized, most exhibit tissue-specific and/or developmentally-regulated expression. Moreover, the number of substrates for such enzymes may be limited, since they require the step-wise action of many previous glycosyltransferases. Inactivating such terminally acting enzymes may thus evoke more subtle or specific phenotypes. Examples include the mechanism of selectin function in leukocyte trafficking, which can be governed by the action of specific fucosyltransferases and sialyltransferases,[61] and the potential role of sialylation and O-acetylation in regulating intercellular recognition in the immune system.[50]

Other approaches (depicted in Figure 2) involve the overexpression of specific glycosyltransferases that can result in the induction of certain oligosaccharide structures, the masking of oligosaccharide structures, or their elimination by biosynthetic competition. Alternatively, one can use the fact that many microbial glycosidases (unlike most proteases) are highly specific in their action on oligosaccharides. Thus, if a particular glycosyltransferase gene has not been cloned, the biology of the relevant oligosaccharide may be explored by expressing a microbial carbohydrate-modifying gene in a tissue-specific and temporally-regulated manner (Figure 2). In this case, the enzyme should be expressed as a cell-surface protein, to avoid non-intrinsic inductive effects associated with

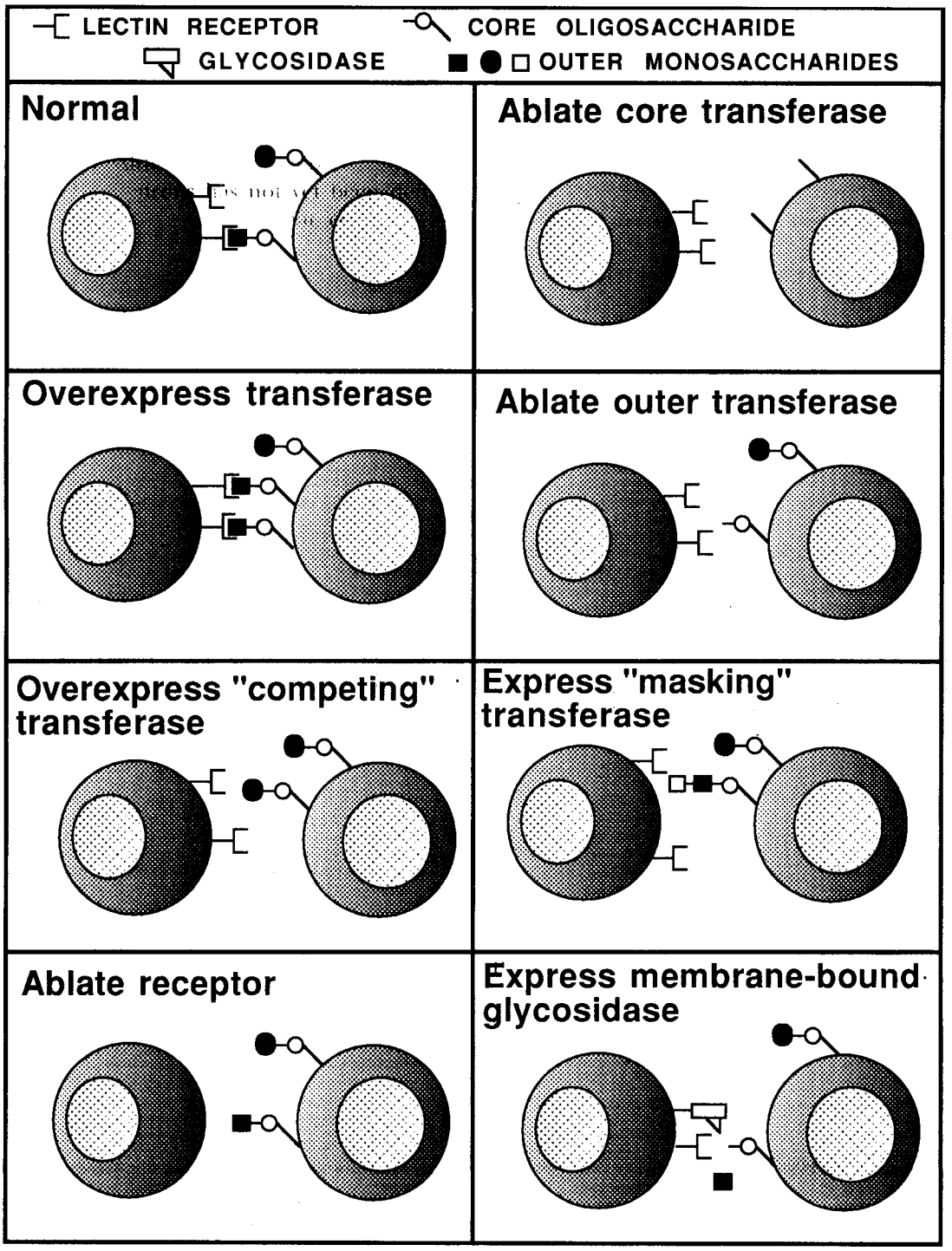


**Figure 2.** Assuming that many important biological roles of oligosaccharides in development involve cell–cell interactions, the various ways in which such interactions could be genetically interrupted in the intact animal are depicted. In each case, the change could be modulated in a spatial and temporal manner using specific promoters for targeting of gene expression or gene inactivation.

a soluble diffusable molecule. Finally, in cases where the primary or sole function of a protein is known to be carbohydrate recognition (i.e. a lectin), gene ablation is expected to yield very useful information regarding the role of the oligosaccharide substrate(s) (Figure 2).

It can also be pointed out that in some instances, certain highly-expressed glycoproteins (e.g. CD43 and CD45 on T cells, or NCAM on neural cells) appear to act as major templates for the changing oligosaccharide repertoire during ontogeny and activation. Hence, terminally-acting transferases may functionally modify one or very few glycoproteins on a particular cell type. In such cases, inactivation of the gene encoding the protein in question may also yield information about the role of glycosylation (e.g. the NCAM gene targeting studies mentioned above appear to result in loss of polysialic acid from the developing nervous system).

A combination of these approaches will aid in exploring the biological roles of specific oligosaccharides in development. As with all experimentation, one would hope to design studies to ask a specific question based upon a hypothesis. However, where no specific clues are available, it may be necessary to first carry out less-directed experiments. Based upon the outcome, further studies could be judiciously applied to specific cell types and tissues, and at specific times in development.

## Perspectives for the future

A large number of genes are involved in generating and maintaining the tissue-specific and developmentally-regulated expression of the diverse repertoire of oligosaccharides structures in vertebrates. Comparisons following mutations in cultured cells with those in intact organisms suggest that the major functions of oligosaccharides may be found in vertebrate development, tissue morphogenesis and cell trafficking. With the cloning and characterization of glycosyltransferase genes, and the advent of new in-vivo genetic approaches, the future appears bright for the elucidation of oligosaccharide function in vertebrates. The difficulty may initially lie in designing experiments that yield the least ambigious results.

## Acknowledgements

The authors are grateful to Andreas Conzelman, Jeff Esko, Jerry Hart, Carlos Hirschberg, John Lowe, Kelly Moremen, Karl-Anders Karlsson and Pamela Stanley for advice regarding the table, to Thierry Hennet for drawing Figure 1, and to Steve Hedrick and Nissi Varki for their critical review of the manuscript.

## Abbreviations used

The following standard abbreviations are used for the indicated monosaccharides: glucose, Glc; mannose, Man; galactose, Gal; N-acetylglucosamine, GlcNAc; N-acetylgalactosamine, GalNAc; fucose, Fuc; xylose, Xyl; glucuronic acid, GlcA; and iduronic acid, IdA. Sia denotes sialic acid, a common name encompassing a family of nine-carbon acidic monosaccharides (the term NeuNAc or NeuAc refers to the commonest type of Sia, but is often used to refer to any unknown Sia). All monosaccharides are in the D-configuration except for fucose and iduronic acid, which are in the L-configuration. All glycosidically linked monosaccharides are assumed to be in the pyranose form.

## References

1. Varki A, Freeze HH (1994) The major glycosylation pathways of mammalian membranes; a summary, in Subcellular Biochemistry (Maddy AH, Harris JR, eds), pp71-100. Plenum Press. New York
2. Hart GW, Haltiwanger RS, Holt GD, Kelly WG (1989) Glycosylation in the nucleus and cytoplasm. Annu Rev Biochem 58:841-874
3. Schachter H, Brockhausen I (1992) The biosynthesis of serine (threonine)-N-acetylgalactosamine-linked carbohydrate moieties, in Glycoconjugates: Composition, Structure, and Function (Allen HJ, Kisailus EC, eds) pp263-332. Marcel Dekker. Inc., New York
4. Weston BW, Nair RP, Larsen RD, Lowe JB (1992) Isolation of a novel human α(1,3)fucosyltransferase gene and molecular comparison to the human Lewis blood group α(1,3/1,4)fucosyltransferase gene. Syntenic, homologous, nonallelic genes encoding enzymes with distinct acceptor substrate specificities. J Biol Chem 267:4152-4160
5. Kitagawa H, Paulson JC (1994) Cloning of a novel α2,3-sialyltransferase that sialylates glycoprotein and glycolipid carbohydrate groups. J Biol Chem 269:1394-1401
6. Lee Y-C, Kojima N, Wada E, Kurosawa N, Nakaoka T, Hamamoto T, Tsuji S (1994) Cloning and expression of cDNA for a new type of Galβ1,3GalNAc α2,3-sialyltransferase. J Biol Chem 269:10028-10033
7. Drickamer K, Taylor ME (1993) Biology of animal lectins. Annu Rev Cell Biol 9:237-264
8. Kornfeld S (1993) Structure and function of the mannose 6-phosphate/insulinlike growth factor II receptors. Annu Rev Biochem 61:307-330
9. Barondes SH, Cooper DNW, Gitt MA, Leffler H (1994) Galectins. Structure and function of a large family of animal lectins. J Biol Chem 269:20807-20810

10. Freeze HH (1992) Developmental glycobiology of dictyostelium discoideum, Cell Surface Carbohydrates and Cell Development (Fukuda M, eds), pp285-317. CRC Press Inc. Boca Raton
11. Williams PJ, Wormald MR, Dwek RA, Rademacher TW, Parker GF, Roberts DR (1991) Characterisation of oligosaccharides from *Drosophila melanogaster* glycoproteins. Biochim Biophys Acta Gen Subj 1075:146-153
12. Varki A (1993) Biological roles of oligosaccharides: All of the theories are correct. Glycobiology 3:97-130
13. Feizi T (1985) Demonstration by monoclonal antibodies that carbohydrate structures of glycoproteins and glycolipids are onco-developmental antigens. Nature 314:53-57
14. Fenderson BA, Eddy EM, Hakomori S (1990) Glycoconjugate expression during embryogenesis and its biological significance. BioEssays 12:173-179
15. Muramatsu T (1988) Alterations of cell-surface carbohydrates during differentiation and development. Biochimie 70:1587-1596
16. Armant DR, Kaplan HA, Lennarz WJ (1986) N-linked glycoprotein biosynthesis in the developing mouse embryo. Dev Biol 113:228-237
17. Codogno P, Botti J, Font J, Aubery M (1985) Modification of the N-linked oligosaccharides in cell surface glycoproteins during chick embryo development. A using lectin affinity and a high resolution chromatography study. Eur J Biochem 149:453-460
18. Yoshinaga K, Muramatsu H, Muramatsu T (1991) Immunohistochemical localization of the carbohydrate antigen 4C9 in the mouse embryo: a reliable marker of mouse primordial germ cells. Differentiation 48:75-82
19. Griffith CM, Sanders EJ (1991) Changes in glycoconjugate expression during early chick embryo development: a lectin-binding study. Anat Rec 231:238-250
20. Hatten ME, Schachner M, Sidman RL (1979) Histochemical characterization of lectin binding in mouse cerebellum. Neuroscience 4:921-935
21. Cooper NG, Steindler DA (1986) Lectins demarcate the barrel subfield in the somatosensory cortex of the early postnatal mouse. J Comp Neurol 249:157-169
22. Sato M, Muramatsu T (1985) Reactivity of five N-acetylgalactosamine-recognizing lectins with preimplantation embryos, early postimplantation embryos, and teratocarcinoma cells of the mouse. Differentiation 29:29-38
23. Falk P, Roth KA, Gordon JI (1994) Lectins are sensitive tools for defining the differentiation programs of mouse gut epithelial cell lineages. Am J Physiol Gastrointest Liver Physiol 266:G987-G1003
24. Kitagawa H, Paulson JC (1994) Differential expression of five sialyltransferase genes in human tissues. J Biol Chem 269:17872-17878
25. Klein A, Krishna M, Varki NM, Varki A (1994) 9-O-acetylated sialic acids have widespread but selective expression: analysis using a chimeric dual-function probe derived from influenza C hemagglutinin-esterase. Proc Natl Acad Sci USA 91:7782-7786
26. Esko JD (1991) Genetic analysis of proteoglycan structure, function and metabolism. Curr Opin Cell Biol 3:805-816
27. Stanley P (1992) Glycosylation engineering. Glycobiology 2:99-107
28. Jaeken J, Carchon H, Stibler H (1993) The carbohydrate-deficient glycoprotein syndromes: Pre-Golgi and Golgi disorders? Glycobiology 3:423-428
29. Dennis JW (1992) Changes in glycosylation associated with malignant transformation and tumor progression, in Cell Surface Carbohydrates and Cell Development (Fukuda M, eds), pp161-194. CRC Press Inc., Boca Raton
30. Capecchi MR (1989) Altering the genome by homologous recombination. Science 244:1288-1292
31. Marth JD (1994) Will the transgenic mouse serve as a Rosetta Stone to glycoconjugate function? Glycoconjugate J 11:3-8
32. Wassarman PM (1990) Profile of a mammalian sperm receptor. Development 108:1-17
33. Lennarz WJ, Ruiz-Bravo N (1993) Identification of molecules involved in sperm-egg fusion. Methods Enzymol 221:261-272
34. Varki A, Hooshmand F, Diaz S, Varki NM, Hedrick SM (1991) Developmental abnormalities in transgenic mice expressing a sialic acid-specific 9-0 acetylesterase. Cell 65:65-74
35. Eggens I, Fenderson B, Toyokuni T, Dean B, Stroud M, Hakomori S-I (1989) Specific interaction between $Le^x$ and $Le^x$ determinants. A possible basis for cell recognition in preimplantation embryos and in embryonal carcinoma cells. J Biol Chem 264:9476-9484
36. Schneider EG, Nguyen HT, Lennarz WJ (1978) The effect of tunicamycin, an inhibitor of protein glycosylation, on embryonic development in the sea urchin. J Biol Chem 253:2348-2355
37. Kabakoff B, Lennarz WJ (1990) Inhibition of glycoprotein processing blocks assembly of spicules during development of the sea urchin embryo. J Cell Biol 111:391-400
38. Elbein AD (1991) Glycosidase inhibitors: Inhibitors of N-linked oligosaccharide processing. FASEB J 5:3055-3063
39. Hathaway HJ, Shur BD (1992) Cell surface β1,4-galactosyltransferase functions during neural crest cell migration and neurulation in vivo. J Cell Biol 117:369-382
40. Ioffe E, Stanley P (1994) Mice lacking *N*-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates. Proc Natl Acad Sci USA 91:728-732
41. Metzler M, Gertz A, Sarkar M, Schachter H, Schrader JW, Marth JD (1994) Complex asparagine-linked oligosaccharides are required for morphogenic events during post-implantation development. EMBO J 13:2056-2065
42. Cremer H, Lange R, Christoph A, Plomann M, Vopper G, Roes J, Brown R, Baldwin S, Kraemer P, Scheff S, Barthels D, Rajewsky K, Wille W (1994) Inactivation of the N-CAM gene in mice results in size reduction of the olfactory bulb and deficits in spatial learning. Nature 367:455-459
43. Tomasiewicz H, Ono K, Yee D, Thompson C, Goridis C, Rutishauser U, Magnuson T (1993) Genetic deletion of a neural cell adhesion molecule variant (N-CAM-180) produces distinct defects in the central nervous system. Neuron 11:1163-1174
44. Rutishauser U, Acheson A, Hall AK, Mann DM, Sunshine J (1988) The neural cell adhesion molecule (NCAM) as a regulator of cell–cell interaction. Science 240:53-57
45. Ikematsu S, Kaname T, Ozawa M, Yonezawa S, Sato E, Uehara F, Obama H, Yamamura K, Muramatsu T (1993) Transgenic mouse lines with ectopic expression of α-1,3-galacosyltransferase: Production and characteristics. Glycobiology 3:575-580
46. Youakim A, Hathaway HJ, Miller DJ, Gong X, Shur BD (1994) Overexpressing sperm surface β1,4-galactosyltransferase in transgenic mice affects multiple aspects of sperm-egg interactions. J Cell Biol 126:1573-1583
47. Orban PC, Chui D, Marth JD (1992) Tissue- and site-specific DNA recombination in transgenic mice. Proc Natl Acad Sci USA 89:6861-6865
48. Lakso M, Sauer B, Mosinger BJ, Lee EJ, Manning RW, Yu SH, Mulder KL, Westphal H (1992) Targeted oncogene activation by site-specific recombination in transgenic mice. Proc Natl Acad Sci USA 89:6232-6236
49. Gu H, Marth JD, Orban PC, Mossmann H, Rajewsky K (1994) Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. Science 265:103-106
50. Varki A, Powell LD (1995) I-Type Lectins. CD22 as a prototypic member of a new family of carbohydrate binding proteins belonging to the immunoglobulin superfamily. J Biol Chem, in press
51. Barlow DP, Stoger R, Herrmann BG, Saito K, Schweifer N (1991) The mouse insulin-like growth factor type-2 receptor is

imprinted and closely linked to the *Tme* locus. Nature 349:84-87
52. Köster A, Saftig P, Matzner U, von Figura K, Peters C, Pohlmann R (1993) Targeted disruption of the $M_r$ 46000 mannose 6-phosphate receptor gene in mice results in misrouting of lysosomal proteins. EMBO J 12:5219-5223
53. Ludwig T, Ovitt CE, Bauer U, Hollinshead M, Remmler J, Lobel P, Rüther U, Hoflack B (1993) Targeted disruption of the mouse cation-dependent mannose 6-phosphate receptor results in partial missorting of multiple lysosomal enzymes. EMBO J 12:5225-5235
54. Mayadas TN, Johnson RC, Rayburn H, Hynes RO, Wagner DD (1993) Leukocyte rolling and extravasation are severely compromised in P selectin-deficient mice. Cell 74:541-554
55. Arbones ML, Ord DC, Ley K, Ratech H, Maynard-Curry C, Otten G, Capon DJ, Tedder TF (1994) Lymphocyte homing and leukocyte rolling and migration are impaired in L-selectin-deficient mice. Immunity 1:247-260
56. Labow MA, Norton CR, Rumberger JM, Lombard-Gillooly KM, Shuster DJ, Hubbard J, Bertko R, Knaack PA, Terry RW, Harbison ML, Kontgen F, Stewart CL, McIntyre KW, Will PC, Burns DK, Wolitzky BA (1995) Characterization of E-selectin-deficient mice: demonstration of overlapping function of the endothelial selectins. Immunity 1:709–720
57. Ishibashi S, Hammer RE, Herz J (1994) Asialoglycoprotein receptor deficiency in mice lacking the minor receptor subunit. J Biol Chem 269:27803–27806
58. Poirier F, Robertson EJ (1993) Normal development of mice carrying a null mutation in the gene encoding the L14 S-type lectin. Development 119:1229-1236
59. Shultz LD, Schweitzer PA, Rajan TV, Yi T, Ihle JN, Matthews RJ, Thomas ML, Beier DR (1993) Mutations at the murine motheaten locus are within the hematopoietic cell protein-tyrosine phosphatase (Hcph) gene. Cell 73:1445-1454
60. Imamoto A, Soriano P (1993) Disruption of the csk gene, encoding a negative regulator of Src family tyrosine kinases, leads to neural tube defects and embryonic lethality in mice. Cell 73:1117-1124
61. Varki A (1994) Selectin ligands. Proc Natl Acad Sci USA 91:7390-7397