Chapter 3

Oligosaccharides and Glycoconjugates in Recognition Processes

Thisbe K. Lindhorst

Otto Diels Institute of Organic Chemistry,
Christiana Albertina University of Kiel,
Otto-Hahn-Platz 3/4, D-24098 Kiel, Germany
tklind@oc.uni-kiel.de

3.1 Introduction

All cells, eukaryotic cells in a special way, are covered with carbohydrates of enormous diversity. These are part of different glycoconjugates, which are embedded into the lipid bilayer of the plasma membrane or associated to the glycocalyx of the cell. The glycocalyx is a highly complex sugar coating, which is typical for every eukaryotic cell and indispensable for living organisms. It can be considered to form an interconnecting supramolecular entity between the extracellular matrix and the cytoskeleton and, apparently, is an indispensable "cell organelle."

Saccharides are major constituents of the glycocalyx, playing an essential role in cell biology. This is well known for the famous blood group antigens of red blood cells, all of them

Handbook of Carbohydrate-Modifying Biocatalysts
Edited by Peter Grunwald

ISBN 978-981-4745-08-6 (Hardcover), 978-981-4745-09-3 (eBook)
www.panstanford.com

being carbohydrates. The biological significance of cell surface carbohydrates in cell communication unfolds in a highly complex interplay with other molecules, both membrane-anchored receptors and soluble proteins. Such secreted or membrane-bound proteins, which can recognize carbohydrates to form carbohydrate–protein complexes, are called lectins. They occur ubiquitously in all organisms and are involved in cell development, the immune system, signal transduction, and also states of disease and malignancy. Intracellular lectins often recognize core structures from glycoconjugate oligosaccharides, while cell surface and extracellular lectins frequently bind to terminal carbohydrate residues. The diversity of lectins, the molecular details of their interaction with glycans, the predominant multivalency effects occurring in carbohydrate–lectin interactions and the biological significance of carbohydrate–protein complex formation, including bacterial adhesion, are subject of this chapter.

It might appear that the variety of possible lectin–carbohydrate interactions is large; however, it can be considered rather modest when compared to the amount of variations that is suggested theoretically. Still, the complexity of structures and the diversity of contexts in the carbohydrate regime present a significant challenge for glycobiological research. In an attempt to investigate the molecular details of carbohydrate–protein interactions, many different (multivalent) glycomimetics have been developed and intensively studied to deepen our understanding of glycobiology. Some of them are included here and discussed as tools of the glycosciences. This chapter is meant to let the reader acquire a taste of the fascinating field of glycobiology and its implications in human life.

3.2 The Lectins and Their Ligands

Lectins are carbohydrate-binding proteins, other than enzymes and antibodies. The word lectin is derived from the Latin word *lectus*, the past participle of *legere*, meaning to select or choose. Current studies, however, consider the role of lectins in *selection* and *assortment* of carbohydrates to a lesser extent; rather glycobiology is focused on their ability to *bind* carbohydrates. Lectins apparently recognize discrete glycans on cell surfaces as well as in solution, followed by the formation of complexes with certain carbohydrate epitopes that

are contained in the variously occurring, structurally highly diverse glycoconjugates.

Lectins were first discovered in plants, where they have been considered as defense agents against predators. In the meantime it has turned out that lectins are found in most (if not all) living organisms, ranging from viruses and bacteria to plants and animals, and also they are implicated in a lot of different fundamental biochemical processes. Many of the biological roles of carbohydrates, involved in physiological as well as in pathological cell biology, are effective through the recognition by lectins. For example, the glycoconjugate decoration of eukaryotic cell surfaces, called the glycocalyx, evolves its specific biological function in the interaction with lectins. Thus, carbohydrate–lectin interactions appear to be essential for live organisms. Many details of this molecular interplay within its complex environment, however, remain unknown so far.

3.2.1 *Classification of Lectins*

Glycan-binding proteins (GBPs), with a few exceptions, can be classified into two major groups, the lectins and the glycosaminoglycan-binding proteins, the latter binding to the highly negatively charged glycosaminoglycan polysaccharides. The lectins, on the other hand, can be classified into families with defined carbohydrate recognition domains (CRDs) that apparently evolved from shared ancestral genes, and often exhibit similar amino acid sequences or three-dimensional shapes.

Lectins reversibly interact with carbohydrates noncovalently through complexation of their ligands in a specific carbohydrate-binding site called CRD (Rini, 1995; Ambrosi et al., 2005). Lectins can contain one, two, or more CRDs and thus their interaction with their carbohydrate ligands is often multivalent. Multiple lectin–carbohydrate contacts can lead to precipitation by crosslinking interactions, called agglutination. Precipitation of the highly glycosylated erythrocyte cells is known as hemagglutination. Agglutinates can be dissolved by addition of the carbohydrate for which the lectin is specific. This inhibition of agglutination is routinely used to determine the carbohydrate specificity of a lectin and to obtain an estimate for its carbohydrate affinity, such as in the classical hemagglutination inhibition assay.

Lectins have been identified in all kinds of living organisms, e.g., plants, microbes, and vertebrates. The viral lectins are often referred to as hemagglutinins (Gamblin and Skehel, 2010), and bacterial lectins are described as adhesins and toxins. Lectins from plants were the first to be identified as carbohydrate-binding proteins. However, some proteins we now regard as animal lectins were discovered before plant lectins, though many were not recognized as carbohydrate-binding proteins for many years after first being reported (Kilpatrick, 2002). The first lectin activity was probably found in snakes. Agglutination of erythrocytes by rattlesnake venom, containing the rattlesnake venom lectin, was observed already around 1860. In 1888 the plant lectin ricin was described as the first agglutinin, in 1899 the first animal lectin from the albumin gland of the snail *Helix pomatia* was reported, and in 1919 Concanavalin A (ConA) from jack beans was the first lectin that was isolated in pure form (Sumner, 1919).

In an attempt to order the increasing number of discovered lectins, first classifications were introduced in the plant lectin field and were related to the carbohydrate specificity of lectins. Thus, according to the monosaccharide ligand toward which a lectin exhibits the highest affinity, lectins were classified into five groups: (i) mannose specific, (ii) galactose and *N*-acetylgalactosamine specific, (iii) *N*-acetylglucosamine specific, (iv) L-fucose specific, and (v) *N*-acetylneuraminic acid (sialic acid) specific (Lis and Sharon, 1998). Presently lectin classification relies on structural homology and evolutionary relatedness, which form a clearly laid out basis for classification of this large group of proteins. Lectin databases are available online, providing an overview of lectin structures (Peréz et al., 2013). Based on numerous crystallographic and NMR-spectroscopic studies, typical folds of lectin CRDs have been identified that can be affiliated to specific lectin classes. In Fig. 3.1 four typical CRD folds were selected as examples for the structural diversity exhibited by lectins. Probably the most famous lectin fold is what has been called the "jellyroll" motif or just the legume lectin fold (Fig. 3.1a). It is the most widely observed lectin fold, first seen in ConA. It comprises a nearly flat six-stranded "back" β-sheet, a curved seven-membered "front" β-sheet, a short five-membered "top" β-sheet, which has an important role in holding the two larger sheets together, and a number of loops interconnecting the sheets.

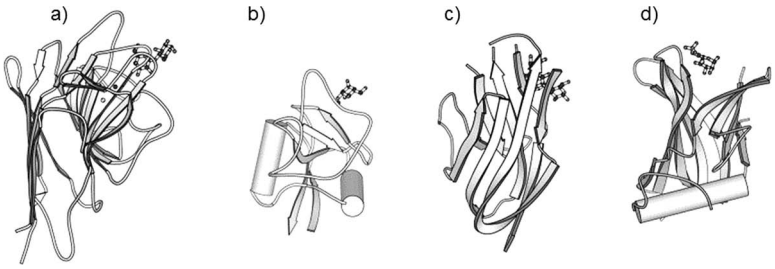


Figure 3.1 Typical folds of lectin CRDs: (a) The legume lectin fold represented by a subunit of the lectin PNA (peanut agglutinin) complexed with lactose (ball-and-stick model). The separate small spheres (filled and open) represent calcium and manganese ions, respectively. (b) A typical CRD in a C-type lectin, represented by mannose-binding protein A (MBP-A) from rat complexed with methyl mannoside (ball-and-stick model). (c, d) The I-lectin fold in a sialoadhesin domain and the P-lectin fold in a subunit of the cation-dependent mannose-6-phosphate receptor, respectively, both complexed with their ligands. Helices and β-strands are represented by cylinders and arrows, respectively, in all cases. According to Vijayan and Chandra (1999). Copyright Elsevier Science Ltd.

As lectins also play an important role as medicinal tools and diagnostic markers, the search for lectins with novel physical properties and ligand specificities is continuing. Recently, the crystal structure of a fungal lectin, the α-galactosyl binding *Lyophyllum decastes* lectin (LDL) was determined and found to adopt a unique fold, unprecedented among lectins (van Eerde et al., 2015).

Nowadays, especially animal lectins have gained interest throughout biochemistry (Drickamer, 2014). They are implicated in numerous biological functions including cell adhesion, cell recruitment, intracellular trafficking, and the immune system. Animal lectins are found in soluble, secreted form as well as associated with membranes, often representing transmembrane proteins. Transmembrane proteins span the entire phospholipid bilayer of a biological membrane, which is approximately 30 Å thick. That part of the protein, spanning the membrane is called the transmembrane domain; other parts are called extracellular and cytoplasmic. Many lectins are cell adhesion molecules (CAMs) which are classified into type I and type II membrane proteins. Type I proteins have a single transmembrane region of hydrophobic residues, with the N-terminus

Table 3.1 Overview about some important classes of animal lectins

Animal lectins →	Type	Subtypes	Main ligands
	R-type		Gal, GalNAc
	L-type		Gal
	P-type		Man-6-P
	C-type	endocytic lectins, collectins, selectins,	Ca^{2+} required, Man, Fuc, Gal
	I-type	siglecs	Sialic acid
	Galectins (formerly S-type)		Gal, LacNAc, Lac

exposed on the exterior of the cell and the COOH-terminal portion of the protein exposed on the cytoplasmic side; type II membrane proteins have their amino terminus on the cytoplasmic side of the cell and the carboxy terminus on the exterior.

Animal lectins are a structurally highly diverse multifunctional group of carbohydrate-recognizing proteins and are difficult to classify. Until today, at least 12 structural families have been defined based on structural homologies, and in addition there are further classes and examples of lectins that do not show any obvious sequence homologies or evolutionary relationships. Still it has remained difficult to get a good overview about lectin structure and function and after all there is no single universally accepted system for lectin classification at the present time. Table 3.1 represents a selection of lectin classes that is suited to explain principles of lectin structure and function.

3.2.2 *Characteristics of Important Lectin Classes*

3.2.2.1 The R- and L-type lectins

Lectins of the R- and L-type are mostly plant lectins (Loris et al., 1998). Proteins of the R-type contain CRDs that are structurally similar to the CRD of the plant lectin ricin, isolated from *Ricinus communis.* They show specificity for galactose and *N*-acetylgalactosamine and are found in plants, animals, and bacteria. Thus, the CRDs of R-type plant lectins have large structural similarity with many animal lectin CRDs. L-type lectins (legume lectins) are found in seeds of

leguminous plants and also in some eukaryotic organisms and represent the largest and most thoroughly studied family of simple lectins. Around 100 members have been characterized almost all isolated from plant seeds. Typically, legume lectins consist of two or four identical or nearly identical subunits of 25–30 kDa each. Each monomeric subunit contains one CRD, to which a Ca^{2+} and usually a Mn^{2+} ion are tightly coordinated. From the study of carbohydrate binding with legume lectins, many fundamental insights of how proteins bind carbohydrates could be obtained. Well-established concepts such as subsite multivalency were first formulated and the formation of crosslinked lattices was first observed during studies of legume lectins and proved to be valid outside the legume lectin family. An important additional reason for the interest in legume lectins is their structural similarity to lectins from other sources, such as the galectins of animals (Sharon and Lis, 2002).

Legume lectins have large similarity at the primary, secondary, and tertiary structural monomeric level. The tertiary structure is folded into what has been called a "jellyroll" motif, also known as the "lectin fold" (*vide supra*). On the other hand, legume lectins exhibit considerable variation in their carbohydrate specificities. They also differ in their quaternary structures, which mostly depends on the pH value. ConA, binding mannose and glucose or the mannotrioside Manα1,3-(Manα1,6)-Man (*vide infra*), is the prototype member of this family. At pH 5.0, ConA exists as a 55 kDa dimer of two identical monomeric units. At pH 7.0, ConA associates from this dimeric form to a tetrameric form with a molecular weight of 110 kDa (Fig. 3.2). The proportion of each species present depends on conditions of temperature and ionic strength. Although this association is rapid, the species do not appear to be in equilibrium and the reaction is largely irreversible.

3.2.2.2 The P-type lectins

P-type lectins do not show any sequence homology with other lectins. Only two examples are known: (i) the 46 kDa cation-dependent mannose-6-phosphate receptor (CD-MPR) and (ii) the 300 kDa insulin-like growth factor II/cation-independent MPR (IGF-II/CI-MPR). These receptors recognize mannose-6-phosphate (Man-6-P) found on N-linked oligosaccharides on lysosomal

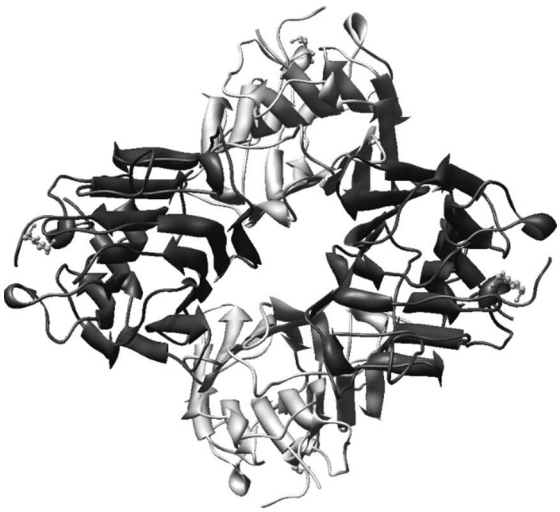


Figure 3.2 Structure of the ConA tetramer complexed with mannose based on PDB code PDB 5CNA. Graphic rendered with Sybyl6.8.

enzymes. Lysosomes require a repertoire of over 60 different acid hydrolases to carry out the degradative metabolism of proteins and other macromolecules. Delivery of newly synthesized soluble acid hydrolases from the Golgi apparatus to lysosomes is carried out by the MPRs (Dahms et al., 2008). When this targeting process in the generation of lysosomes containing a full complement of hydrolytic enzymes is defective, different lysosomal storage diseases can occur.

3.2.2.3 The C-type lectins

The C-type lectins recognize a variety of different carbohydrates, but they all require calcium ions for binding of their ligands (Drickamer, 1999). In all of the available CRD-ligand structures, the carbohydrate ligand is complexed to the protein by forming coordination bonds with a conserved Ca^{2+} ion that also coordinates to the CRD. In addition, ligand complexation is accomplished by hydrogen bonding with carboxylic acid and amide functional groups of amino acid side chains in the CRD. Besides the C-type lectins, part or all of the C-type CRD motif is found in other proteins, serving other functions than saccharide recognition. Hence, theses motifs were named C-type

lectin-like domains (CTLDs) to reflect their similarity to the CRDs of C-type lectins without necessarily implying common function (Drickamer, 1999).

The C-type lectins consist of three major classes, (i) endocytic lectins, (ii) collectins, and (iii) selectins. The prototype of the large group of C-type lectins and the first animal lectin discovered is the hepatic asialoglycoprotein receptor (ASGPR) originally isolated from rat liver. It is also called the "Ashwell receptor" after his discoverer (Ashwell and Morell, 1974). The prototype of avian lectins is the chicken hepatic lectin (CHL). Similar lectins as in rats are present in human liver and also in other mammals and are called mammalian hepatic lectins (MHLs). These receptors are present on hepatocyte cell surfaces as well as on their inner membranes. The MHLs are specific for galactose and GalNAc, the CHL is specific for GlcNAc, and the alligator hepatic lectin (AHL) is specific for mannose and L-fucose (Lee and Lee, 1995).

The C-terminal ends of hepatic lectins contain a CRD which faces the outside of the cell. To the CRD is attached a stalk (or neck) region, followed by a transmembrane segment and a short cytosolic tail at the N-terminus. The ASGPR is specific for Gal- and GalNAc-terminated oligosaccharides that appear after desialylation of complex type glycoconjugates (*vide infra*). Such glycoproteins are bound by ASGPR at the surface of hepatocytes and then internalized. All hepatic type II transmembrane proteins have a strong tendency to associate. The human ASGPR is a heterotetramer, whereas the CHL is a homotrimer. On rat or rabbit hepatocyte surface, MHL may further associate into tightly packed aggregates (Lee and Lee, 1995).

Oligomerization of hepatic lectins leads to high affinity for specific branched oligosaccharide ligands. Tri- and tetra-antennary N-glycans with appropriate branching, presenting nonreducing terminal galactose (Gal) or *N*-acetylgalactosamine (GalNAc) residues bind to rat ASGPR with greater than 10^5 times higher affinity than ligands with a single Gal or GalNAc residue. Recently it has been found that ASGPR can also bind to certain sialylated ligands.

The fucose-, mannose-, and galactose-specific receptor found on macrophages (MMR) and on Kupffer cells is another endocytic lectin. Whereas the hepatic lectins have a single C-type CRD, MMR has eight CRDs on one polypeptide chain. MMR plays a part in innate immunity by helping macrophages to bind and internalize pathogens.

3.2.2.4 The collectins: A subgroup of C-type lectins

The collectins belong to the superfamily of collagen-containing C-type lectins. Each collectin polypeptide contains an N-terminal cysteine-rich domain, a collagen-like domain, an α-helical coiled-coil neck region and a C-terminal carbohydrate-recognition domain (Drickamer and Taylor, 1993). To date, nine different collectins are known: mannose-binding protein (MBP; also called mannan-binding lectin, MBL), surfactant proteins A and D (SP-A, SP-D), conglutinin (from Bovine serum), and the collectins CL-L1 (liver), CL-P1 (placenta), CL-K1 (kidney), CL-43 (43 kDa), and CL-46 (46 kDa). Except CL-L1 and CL-P1, which are membrane proteins, all other collectins are soluble proteins. These molecules are major modulators of the innate immune system where they have a key role in the first line of defense against invading microorganisms. MBPs are known in serum (MBP-A) and liver (MBP-C). MBP-A is an important component in the mammalian innate immune system that binds carbohydrates on the surfaces of pathogenic microorganisms and activates complement in an antibody-independent manner (Hansen and Holmskov, 1998). The ability to distinguish exogenous structures (of bacteria) from endogenous mammalian glycans found on endogenous membranes, thus, provides a mechanism for identification of pathogens and their neutralization.

The mannose-binding site in MBP interacts only with the terminal residues in an oligosaccharide. MBP shows equal affinity for mannose, glucose, and GlcNAc residues in terminal positions. The single monosaccharide–CRD interaction in MBP is weak with K_d in the millimolar range, whereas multimerization of MBPs lead to specific and biologically significant interactions. Hence, MBPs form trimers (Weis and Drickamer, 1994) based on collagen triple helices (Fig. 3.3).

3.2.2.5 The selectins: A subgroup of C-type lectins

The selectins are a group of three different vascular adhesion receptors that participate in the recruitment of leukocytes into the sites of inflammation and their emigration into lymphatic tissues (McEver, 2002; Rosen, 2004). E-selectin and P-selectin appear on vascular endothelium in response to inflammation, while L-selectin

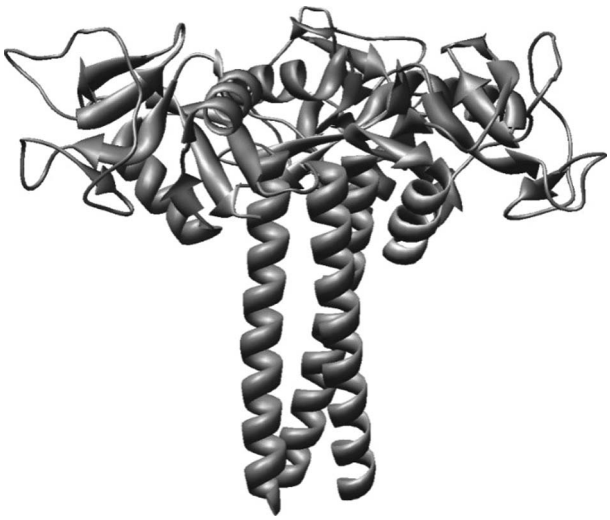


Figure 3.3 Trimeric structure of MBP-A based on PDB code PDB 1KWT. Graphic rendered with Sybyl6.8.

is expressed on leukocytes. All three selectins, L, P, and E, are relatively rigid, extended type I transmembrane proteins that share a common domain organization. At their N-terminal tips they carry a C-type CRD, which is followed by a consensus epidermal growth factor (EGF)-like domain and a series of short consensus repeats, called sushi modules, which project the lectin domain away from the cell surface. All lectins have a single transmembrane domain and a C-terminal cytoplasmic domain (Fig. 3.4).

The minimal saccharide ligand for selectins is the tetrasaccharide sialyl Lewis^X (SLe^X), Neu5Acα2,3Galβ1,4(Fucα1,3) GlcNAcβ1-R (R = glycoprotein residue). This tetrasaccharide terminates many N-glycoproteins and is contained in many mucins as well. In addition to SLe^X P- and L-selectin, but not E-selectin, bind to some forms of heparin sulfate and heparan sulfate. The affinities selectins show for SLe^X are relatively weak with K_d values in the millimolar range. The K_d of the P-selectin–SLe^X complex was determined as 7.8 mM, whereas E-selectin binds the same tetrasaccharide with 10-fold higher affinity (Somers et al., 2000). Certain glycoproteins, however, such as P-selectin glycoprotein ligand-1 (PSGL-1) interact with selectins with enhanced affinities.

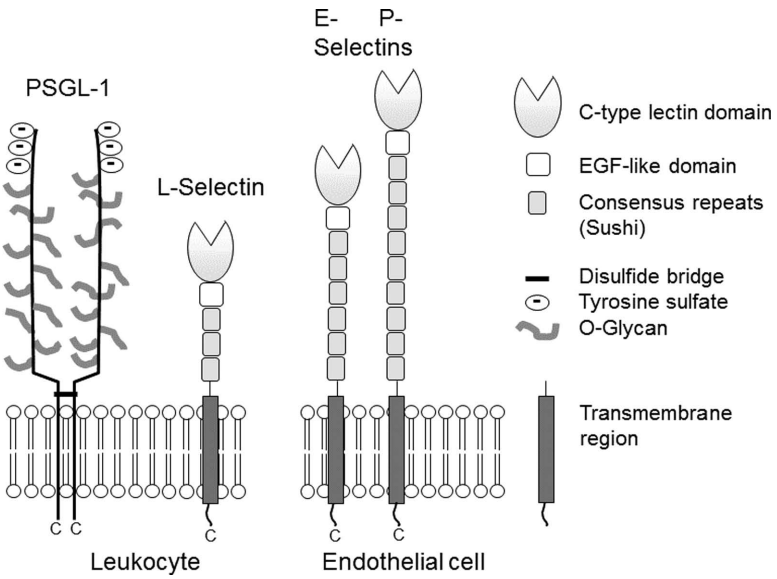


Figure 3.4 Overall domain structures of L-, E-, and P-selectin and the highly glycosylated mucin-like cell adhesion molecule PSGL-1, which can function as ligand for all three selectins.

PSGL-1 is a mucin-like CAM, which is expressed on leukocytes as 240 kDa homodimer (Fig. 3.4). If properly glycosylated, it can bind all three lectins, P-, E-, and L-selectin. Most O-linked glycans of PSGL-1 are short sialylated structures that keep the polypeptide backbone in an elongated conformation. Near the N-terminus there is an extended oligosaccharide, bearing the SLe^x motif, bound in close proximity to a sulfated tyrosine residue, and this combination is crucial for tight binding of selectins.

Beyond the so far described classes of C-type lectins, another lectin, having a C-type CRD, was recently found to be of importance in the immune system: DC-SIGN, originally named as dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin. DC-SIGN has a dual binding specificity for both high-mannose oligosaccharides and glycans bearing the $Lewis^x$ trisaccharide. It is a tetrameric transmembrane protein containing C-terminal C-type CRDs that are projecting from the cell surface of dendritic cells. Dendritic cells are mammalian immune cells and present in small

quantities in tissues that are in contact with the external environment, mainly the skin and the inner lining of the nose, lungs, stomach, and intestines. Once activated, dendritic cells migrate to the lymphoid node where they interact with T cells and B cells to initiate and shape the adaptive immune response. At certain development stages they grow branched projections, the dendrites, that give the cell its name. DC-SIGN promotes antigen-independent interaction between T cells and the surface of dendritic cells that may be presenting cognate antigens. Recently, a DC-SIGN-related molecule has been described and termed DC-SIGNR.

3.2.2.6 The I-type lectins

I-type lectins are GBPs of the immunoglobulin superfamily other than antibodies and T-cell receptors. So far, the siglec family of sialic acid binding lectins is the only well-characterized group of I-type lectins both structurally as well as functionally. Siglec is the abbreviation for sialic acid recognizing immunoglobulin superfamily lectins (Angata et al., 2004). Members of the immunoglobulin superfamily contain at least one immunoglobulin (Ig)-like fold that is classically known from antibodies. The first nonantibody immunoglobulin superfamily GBP was sialoadhesin (Sn), which is today called siglec-1. Siglec-1 molecules are found on the surface of macrophages, where they mediate interactions with cells and pathogens (such as sialylated bacteria or viruses).

There are 12 known functional human siglecs and one siglec-like molecule, named as siglec-12. Siglec-1 (sialoadhesin), siglec-2 (CD22), siglec-4 (myelin-associated glycoprotein, MAG), and siglec 15 form a separate distantly related evolutionary group. Siglecs-1, -2, and -4 were the first siglecs characterized and grouped together as the sialoadhesin family (Kelm et al., 1994; Crocker and Varki, 2001). Siglec-3, siglecs-5, -11, and siglec-15 share a high degree of sequence similarity in their extracellular and intracellular regions, and based on the structural features they have in common they are referred to as CD33-related siglecs. The genes encoding CD33-related siglecs map very close to each other and are clustered on chromosome 19 in humans. Siglec-12 is not found as a sialic-binding protein in humans and has been called a siglec-like molecule.

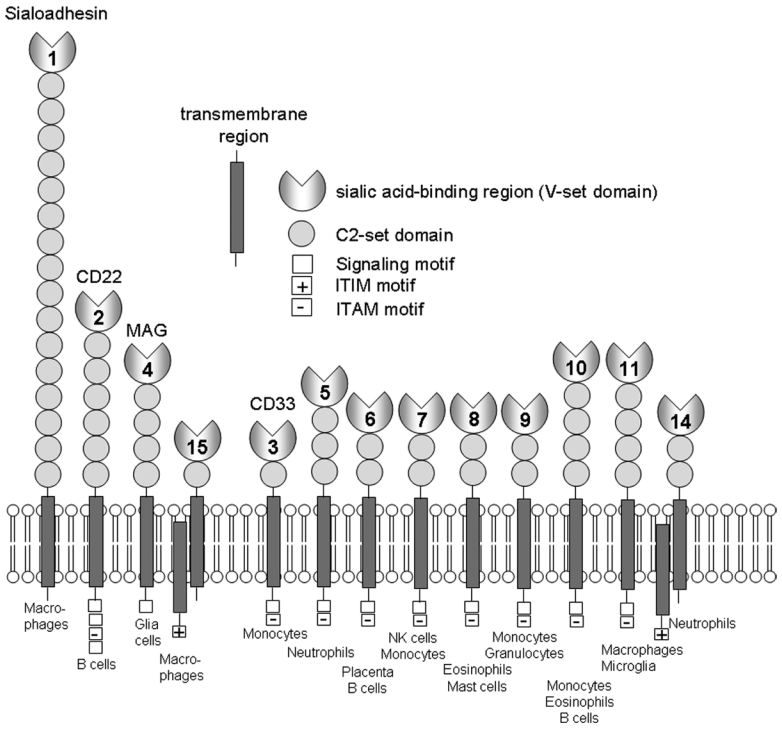


Figure 3.5 Domain structures of known siglecs in humans and where they can be found. One subgroup of siglecs contains siglec-1 (sialoadhesin), siglec-2 (CD22), siglec-4 (myelin-associated glycoprotein, MAG), and siglec-15. The other group of siglecs contains the so-called CD33 (siglec-3)-related siglecs, namely, siglec-3, siglecs-5 to -11, and siglec 14. Siglecs have one V-set domain (a domain similar to Ig's variable region) and 1–16 C2-set domains (domains similar to Ig's constant region), followed by transmembrane and cytoplasmic signaling domains (Varki, 2007).

All siglecs are type I membrane proteins, containing a sialic acid binding domain, an amino-terminal V-set domain and varying numbers of C2-set Ig domains (Fig. 3.5). The latter act as spacers, projecting the Sia-binding site away from the plasma membrane of the cell. Disulfide bonds are organized different from what is known from the typical intersheet disulfide organization of other molecules with immunoglobulin fold. Siglecs differ from most other Sia-binding

lectins, such as the selectins, in that they require an entire sialic acid residue for binding.

The siglecs are involved in cell adhesion and signaling (Varki, 2007) based on cytoplasmic tyrosine-based signaling motifs. Most prevalent is the ITIM motif (immunoreceptor tyrosine-based inhibition motif), comprising a conserved sequence of four amino acids, including tyrosine (ITAM is the antagonistic case). The tyrosine residues within the ITIM motifs become phosphorylated following interaction of the receptor molecules with their ligands and thus forming docking sites for other proteins involved in the signaling pathways of the cell. Siglecs also mediate regulation of immune cell function in disease (Macauley et al., 2014).

3.2.2.7 The S-type lectins

The S-type lectins have been named so when it was discovered that they require free thiols for stability. However, not all lectins of this type are thiol dependent, and hence today they are called galectins, as they all recognize β-galactosides. Galectins are the most widely expressed class of lectins sharing a primary structural homology in their CRDs. They can be found in species ranging from fungi to human.

The folding pattern of galectins is an antiparallel β-sandwich, known as the jellyroll-like fold. The structures of the galectins can be classified in three categories (Morris et al., 2004): (i) The prototypical galectins (galectins-1, -2, -5, -7, -10, and -11, -13, -14) with one CRD, which might exist as monomers or homodimers; (ii) the chimera-type galectins, which in addition to a C-terminal CRD contain a large (~130 amino acid) amino-terminal nonlectin domain; and (iii) tandem-repeat type galectins (galectins-4, -6, -8, -9, and -12) composed of two CRD domains in a single polypeptide chain connected by a linker peptide (Fig. 3.6). Galectin-3 is the only known chimeric galectin found in vertebrates. For galectins-5, -6, -14, and -15 no human counterparts have been discovered so far.

Galectins preferentially bind to glycans carrying units of the ligand *N*-acetyllactosamine (LacNAc: Galβ1,4GlcNAc), either as disaccharide units at the termini of tri- and tetra-antennary N-glycan chains, or as repeating units in a poly-*N*-acetyllactosamine chain

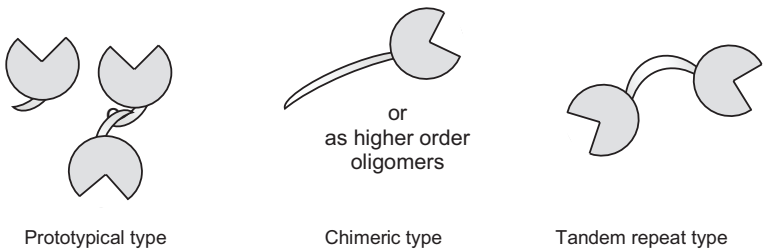


Figure 3.6 Galectins are found in three subclasses: (i) Prototype galectins are galectins-1, -2, -5, -7, -10, -13, and -14; (ii) galectin-3 is the only representative of the chimera type; (iii) tandem repeat galectins are galectins-4, -6, -8, -9, and -12. (Chimeric proteins are fusion proteins that are created through the joining of two or more genes that originally coded for separate proteins. Translation of this fusion gene results in a single polypeptide with functional properties derived from each of the original proteins.)

on N- or O-glycans. Galectins are active both intracellularly and extracellularly. They have diverse effects on many cellular functions including adhesion, migration, chemotaxis, proliferation, apoptosis, and differentiation (Gabius, 2006). A common function of galectins is to selectively crosslink glycoproteins on the cell surface to form a uniform lectin–carbohydrate lattice (*vide infra*). In galectin-3 both the C-terminal CRD domain and the N-terminal nonlectin domain are essential for its role in signal transduction, cellular adhesion, and lattice formation (Nieminen et al., 2007).

The galectin CRD is a β-sandwich of about 135 amino acids forming a groove in which the carbohydrate ligand is bound, and which is long enough to accommodate a linear tetrasaccharide. Schematically, the galectin carbohydrate binding site can be described as having four subsites, A–D and a fifth, less-defined site E (Leffler et al., 2002). In this model subsite, C is the defining *β*-galactoside binding site of the galectins, and subsite D contributes the second part of the conserved core disaccharide-binding site. The binding of a galactose residue in site C is the most conserved feature of galectin binding activities. Six of the seven motif amino acids interact with the galactose ring. The binding of a saccharide in site D is the second most conserved feature, but here the structure requirements for interaction can be fulfilled by different saccharides (Fig. 3.7). A source for variation in specificity between

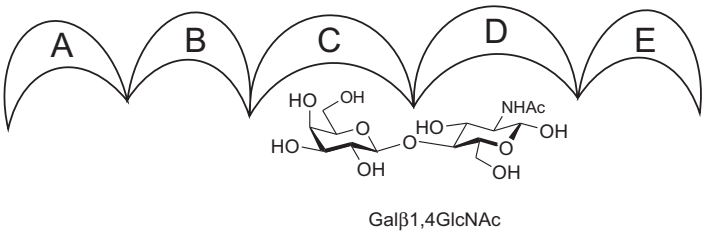


Figure 3.7 For the galectin CRD four subsites are established (A–D) and an additional subsite E has been proposed. The core binding sites are C and D, binding preferentially to *N*-acetyllactosamine (LacNAc: Galβ1,4GlcNAc). The 4-OH of the galactose ring in LacNAc or in lactose is hydrogen-bonded to the side chains of three invariant amino acids, histidine, asparagines, and arginine (not shown).

galectins is their different ability to accommodate saccharides (GlcNAc, Gal, GalNAc, sialic acids) or other groups (e.g., sulfate) in subsite B and further extensions into subsite A. The structurally less defined site E can interact with moieties linked at the reducing end of the saccharide in site D. In cells this would be another saccharide, protein, or lipid. Although all galectins are specific for the *N*-acetyllactosamine ligands, they can specify between different glycoproteins both in solution and on cell surfaces. Interactions in site E might help explain some of these distinct galectin binding activities that are not easily explained based on the properties of the well-defined sites A–D.

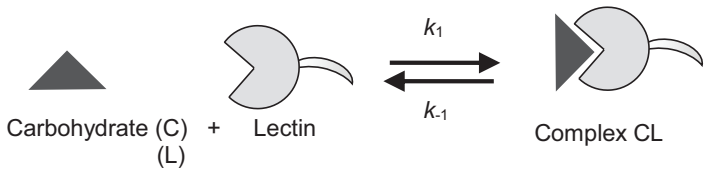
An important consequence of the galectin structures described above is that most are topologically multivalent, either by forming noncovalent dimers or higher oligomers, or by having two CRDs within one peptide chain. They are therefore able to crosslink β-galactoside-containing glycoconjugate ligands at physiological concentrations (Ahmad et al., 2004; Brewer, 2004). Apparently, multivalency of galectins enables their involvement in cell–cell and cell–pathogen interactions along with signal transduction events and lattice formation upon ligand crosslinking (Dennis et al., 2002; Lowe 2001; Rabinovich et al., 2002, 2004; Sato 2002; Sato and Nieminen 2004). It could also be shown that crosslinking of cell surface glycoproteins by galectins controls their spatiotemporal dynamics (Möckl et al., 2015).

3.3 Recognition of Carbohydrates

How do lectins precisely interact with carbohydrates, how can they recognize their ligands? How is the specificity of carbohydrate–lectin interactions governed and how can lectins select among the thousands of different glycans produced by a cell? How is the proximity of ligands and receptors ruled and how their geometrical arrangement controlled? Not all of these questions can be answered comprehensively; however, fundamental principles and some of the typical features of carbohydrate binding have been discovered, mainly by X-ray studies and NMR measurement of carbohydrate–lectin complexes.

3.3.1 *Formation of a Carbohydrate–Lectin Complex*

The formation of a carbohydrate–lectin complex "CL" is an equilibrium as shown in Eq. (3.1), which can be described thermodynamically and kinetically. The rate constant k_1 describes the forward reaction (association and formation of the CL complex) and k_{-1} is the backward rate constant (dissociation of CL). The rate constant k_1 is often referred to as k_{on} and k_{-1} as k_{off}. The free energy change of this binding process is described by the Gibbs–Helmholtz equation (Eq. (3.2)), with G the Gibbs free energy, H the enthalpy, T the absolute temperature, and S the entropy. The Gibbs energy is related to the equilibrium constant of complex formation K, which is used to quantify the affinity of binding (R is the gas constant). The association constant K_a is equal to k_1/k_{-1}, the dissociation constant K_d equals k_{-1}/k_1 (Eqs. (3.3) and (3.4)). The affinity of carbohydrates for their lectins is mostly discussed as dissociation process because the unit of K_d is a concentration such as millimolar, micromolar, or nanomolar, whereas the unit of K_a is 1/molarity (with $K_d = 1/K_a$).



(3.1)

$$\Delta G = \Delta H - T\Delta S = \mathrm{R}T \ln K \tag{3.2}$$

$$K_a = [\mathrm{CL}]/[\mathrm{C}][\mathrm{L}] = k_1/k_{-1} \tag{3.3}$$

$$K_d = [\mathrm{C}][\mathrm{L}]/[\mathrm{CL}] = k_{-1}/k_1 \tag{3.4}$$

Lectins interact with carbohydrates noncovalently in a reversible, often highly specific interaction. Much of the initial work important to our understanding of carbohydrate–lectin interactions was gathered in studies on the combining sites of plant lectins and antibodies toward specific blood group antigens. From numerous known structures of glycan–protein complexes that have been solved by X-ray crystallography and NMR spectroscopy, details about carbohydrate complexation are known. It has turned out that specific recognition of sugars by proteins occurs by multiple mechanisms. Typically one to four sugar residues are bound by a lectin's CRD. The interactions can sometimes extend over several binding sites and may include moieties of the glycoside aglycon portion. Lectins can be monovalent or multivalent, when they contain two or more CRDs; monovalent lectins are also found as multimers or clusters, respectively. Accordingly, lectin–carbohydrate interactions can be mono-, di-, or multivalent. Complexation of a mono- or oligosaccharide resembling a single ligand epitope is defined as monovalent binding. When multiple interactions between several CRDs and branched oligosaccharides or glycoconjugates possessing multiple carbohydrate epitopes occur, this results in multivalent binding.

Lectins bind carbohydrates through a network of hydrogen bonds and hydrophobic interactions. The physical forces that stabilize carbohydrate–lectin complexes are intermolecular hydrogen bonding (direct and water mediated), van der Waals interactions (hydrophobic forces between hydrophobic sites of the carbohydrate ligand and hydrophobic amino acid side chains in the lectin CRD), CH-π stacking interactions, and electrostatic interactions such as an ion bridge, typically the coordination of Ca^{2+} (Fig. 3.8).

The hydroxyl groups of a carbohydrate ring can serve as hydrogen bond acceptor as well as hydrogen bond donor. Acidic side chains such as in the sialic acids act as hydrogen bond acceptor under physiological conditions. The endocyclic ring oxygen of the sugar

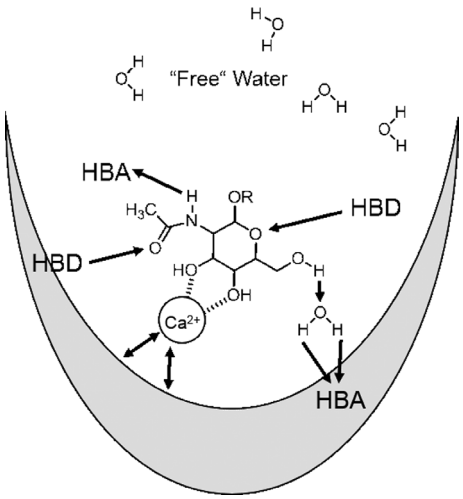


Figure 3.8 Some interactions operative in carbohydrate complexation. Upon formation of a carbohydrate–protein complex in aqueous phase, originally ordered water molecules are released to the bulk ("free" water). HBD: hydrogen bond donor; HBA: hydrogen bond acceptor.

can serve as hydrogen bond acceptor and usually shares a donor amino acid with another OH group of the same sugar. In GlcNAc and GalNAc the acetamido groups are also involved in the hydrogen bond network between ligand and receptor. A common type of hydrogen bond in lectin–sugar complexes are bidentate hydrogen bonds, involving two adjacent OH groups of a sugar ring (Fig. 3.8).

In addition, solvent reorganization of the hydrated surface of the lectin as well as the carbohydrate can add favorably to the entropy change during complex formation. Before complex formation, the polar groups of both the ligand and the receptor are extensively hydrogen-bonded to water molecules. Replacement of ordered water molecules from the CRD into the bulk results in an entropic gain. Conformational changes of the carbohydrate ligand can also contribute to stabilization of the carbohydrate–protein complex.

3.3.2 *Networks of Stabilizing Interactions*

The carbohydrate-binding sites of lectins are, apart from some exceptions, rather shallow structures on the surface of the protein

that apparently do not undergo significant conformational changes upon ligand binding. The carbohydrates, on the other hand, are notoriously flexible and in aqueous or physiological environments their conformational structures can be influenced by interaction with neighboring ions or molecules and particularly, by explicit hydration (Eriksson et al., 2008; Reynolds et al., 2008). This may in turn influence their selective molecular recognition at protein–carbohydrate receptor sites (Simons et al., 2009). Carbohydrate binding is thought to involve their preferred solution conformation(s); however, it is not always that the energy minimum conformation is found in a solution that is bound by a lectin. E-selectin, for example, selects one specific conformation among the populated solution conformations of SLe^X (Rinnbauer et al., 2003). Key factors controlling conformational preference and site selectivity in hydrated monosaccharides include the flexibility of their exocyclic hydroxymethyl groups (in glucose, galactose, and mannose), their anomeric configuration, and the relative orientations (axial vs. equatorial) of their hydroxyl groups. These factors can operate separately or collectively, adapting the carbohydrate conformation and configuration to optimize the sequence of intra- and intermolecular hydrogen-bonded interactions in the hydrated complex.

The pattern of recognition of hydroxyl groups of a sugar ring is lectin specific and lectin class specific. For example, mannose-specific legume lectins bind to the 3-, 4-, and 6-OH of the sugar ring and can therefore not distinguish between mannose and glucose. Galactose-specific legume lectins recognize the 3,4,6-trihydroxy pattern of galactose and often tolerate a rather bulky group at C-2. Thus, GalNAc is mostly equally recognized by these lectins.

In all legume lectins, irrespective of their specificity, four invariant amino acid residues participate in ligand binding: aspartic acid, asparagine, glycine, and an aromatic amino acid or leucine (Sharon and Lis, 2002). Although these key amino acids are highly conserved among legume lectins, different lectins from this class show rather different carbohydrate specificities. This apparently arises from a variability of amino acid residues in other regions of the lectin than the CRD. This parallels with the finding that fine-tuning of enzyme specificity as well as enzyme activity can be regulated by mutagenesis of amino acids far from the active site of the enzyme.

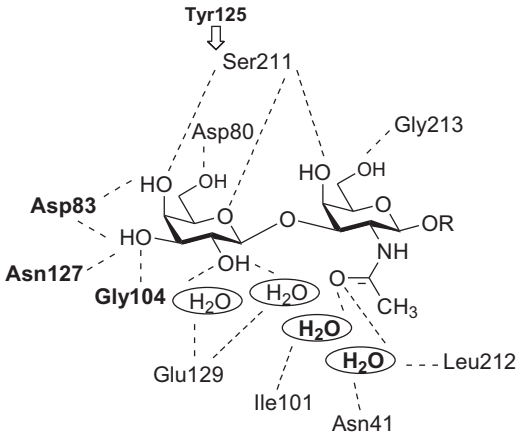


Figure 3.9 Schematic representation of the protein–carbohydrate interactions in the complex of PNA with the T-antigen Galβ1,3GalNAc (Ravishankar et al., 1997). The two water molecules mediating the especially high specificity of PNA for the T-antigen disaccharide are printed in bold.

For the well-known peanut agglutinin (PNA), which has a high specificity for galactose, the amino acid residues Asp83, Gly104, Asn127, and Tyr125 are essential for ligand binding (Banerjee et al., 1996), as it can be seen in its complex with the T-antigenic disaccharide Galβ1,3GalNAc. There are four invariant hydrogen bonds to Asp83 (establishing two hydrogen bonds), Gly104 and Asn127 (each establishing one hydrogen bond) and in addition the complex is stabilized by a stacking interaction between Tyr125 and the hydrophobic α-face of the galactose ring. Furthermore, about 60 van der Waals contacts are formed between the disaccharide ligand and the lectin's amino acid residues within a distance of 4 Å of the carbohydrate (Fig. 3.9).

PNA binds the T-antigen Galβ1,3GalNAc 20 times more strongly than lactose. Comparison of the two complexes reveals that the high specificity of PNA for the T-antigen Galβ1,3GalNAc is generated primarily by two specific water-mediated interactions (cf. Fig. 3.9), not present in the PNA-lactose complex (Ravishankar et al., 1997). Thus, water in the hydration shell of a lectin can be considered to be an extension of the protein surface and a strategy for generating

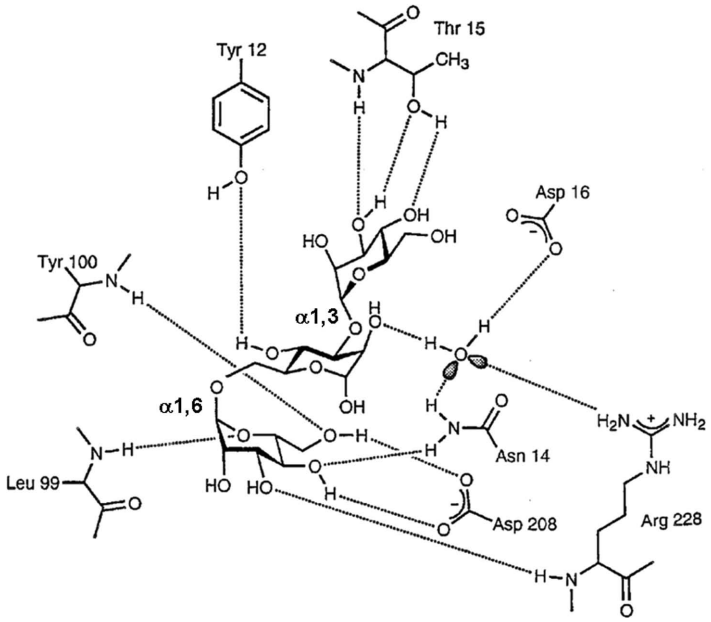


Figure 3.10 The extended trimannoside-binding site of Con A in complex with the mannotrioside Manα1,3-(Manα1,6)-Man. The hydrogen bonds between the sugar and protein are schematically represented. The distances for the hydrogen bonds depicted lie between 2.6 and 3.1 Å. Reproduced and adapted from Naismith and Field (1996). Copyright by the American Society for Biochemistry & Molecular Biology ASBMB.

carbohydrate specificity (Natchiar et al., 2006). The PNA-T-antigen complex (Ravishankar et al., 1997) was probably the first example, where it was shown that specific water bridges can generate carbohydrate specificity.

The legume lectin ConA, from *Canavalia ensiformis*, specifically recognizes the trimannoside core of many complex glycans. In the complex shown in Fig. 3.10, the 1,6-linked mannose residue is bound at the monosaccharide binding site of the lectin; the other two sugars bind in an extended cleft formed by residues Tyr12, Pro13, Asn14, Thr15, and Asp16. Hydrogen bonds are formed between the protein and all three sugar residues. In particular, the 1,3-linked mannose residue makes a strong hydrogen bond with the main chain of the protein. In addition, a water molecule, which is conserved in other

ConA structures, plays an important role in anchoring the reducing sugar unit to the protein. The complex is further stabilized by van der Waals interactions. The structure shows that, in addition to hydrogen bonds established in the monosaccharide binding region of a lectin such as ConA, ligand binding can be supplemented by additional hydrogen bonding in the extended binding sites of lectins. This explains, why the mannotrioside Manα1,3-(Manα1,6)-Man is bound with higher affinity to ConA than methyl α-D-mannoside (MeMan): K_a(MeMan) = 8.2 × 10^3 M^{-1} and ΔH = –8.2 kcal M^{-1}, whereas K_a (trisaccharide) = 4.9 × 10^5 M^{-1} and ΔH = –14.4 kcal M^{-1}.

Thus, lectins can achieve carbohydrate specificity by fine-tuned hydrogen bond networks, including water bridges (Vijayan and Chandra, 1999). However, homologous lectins with conserved CRDs, interacting with the same carbohydrate ligand through the same set of hydrogen bonds, can feature quite different thermodynamics of this interaction. This can be explained by small changes in the quaternary structure of different lectins (Dwek 1996; Weis and Drickamer, 1996). Especially the legume lectins exhibit large variability in quaternary association resulting from small alterations in essentially the same tertiary structure (Jiménez-Barbero et al., 1999; Fernández-Alonso et al., 2005).

3.3.3 *Complexation via Ca^{2+}*

Divalent cations such as Ca^{2+} and Mn^{2+} are also often important for the formation of lectin–carbohydrate complexes. In legume lectins, divalent cations such as Ca^{2+} and Mn^{2+} are not directly involved in sugar binding but are important for maintaining the integrity of the CRD. For sugar binding of C-type lectins, on the other hand, the calcium ion Ca^{2+} is essential. In this case the divalent cation is directly involved in carbohydrate complexation. A protein-bound calcium ion interacts with the 3-OH and 4-OH of either mannose or glucose or with the 2-OH and 3-OH of L-fucose. Four amino acids are conserved in the CRDs of all mannose-specific C-type lectins, two glutamic acid and two asparagine residues. This bonding pattern is common for all C-type lectins (Lis and Sharon, 1998).

The crystal structure of MBP-C complexed with mannose (Iobst et al., 1994) reveals that Glu185, Asn187, Glu193, and Asn205 are intimately involved in sugar binding. Each of these residues

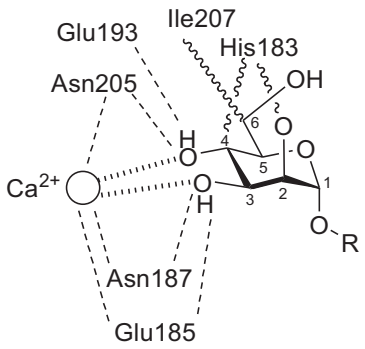


Figure 3.11 Principal interactions of an α-mannoside complexed in the CRD of the rat liver mannose-binding protein MBP-C (Iobst et al., 1994). Hydrogen bonds are denoted by dashed lines. Van der Waals contacts are indicated by wavy lines.

contributes to Ca^{2+} ligation and forms hydrogen bonds to hydroxyl groups 3-OH or 4-OH, respectively, of mannose or a terminal mannose residue of an oligomannose ligand (Fig. 3.11). Lone pairs of electrons on oxygen 3 and 4 of mannose accept protons in hydrogen bonds from Asn187 and Asn205, respectively, and the protons of 3-OH and 4-OH, on the other hand, donate hydrogen bonds to Glu185 and Glu193, respectively. The remaining lone pairs of electrons on oxygens 3 and 4 of mannose form coordination bonds with the Ca^{2+} ions (broad dashed lines in Fig. 3.11). In addition, three van der Waals contacts further stabilize the complex.

Direct involvement of a metal ion in ligand binding is typical, but not totally unique to C-type lectins. Direct involvement of calcium ions in ligand binding has for example been shown for the fucose-binding protein PA-IIL that was isolated from the Gram-negative bacterium *Pseudomonas aeruginosa*. (Mitchell et al., 2005). Here, the 2-OH, 3-OH, and 4-OH of the L-fucose ring participate in the coordination of the two calcium ions.

Often, only certain parts of a saccharide ligand, called the binding epitope, form direct interactions to the carbohydrate-binding protein within a CRD. This is the case in binding of SLe^X by the selectins. The tetrasaccharide binding sites of P- and E-selectin are very similar and they bind the tetrasaccharide ligand SLe^X in identical conformation. Interactions to the carbohydrate ligand are established to only one

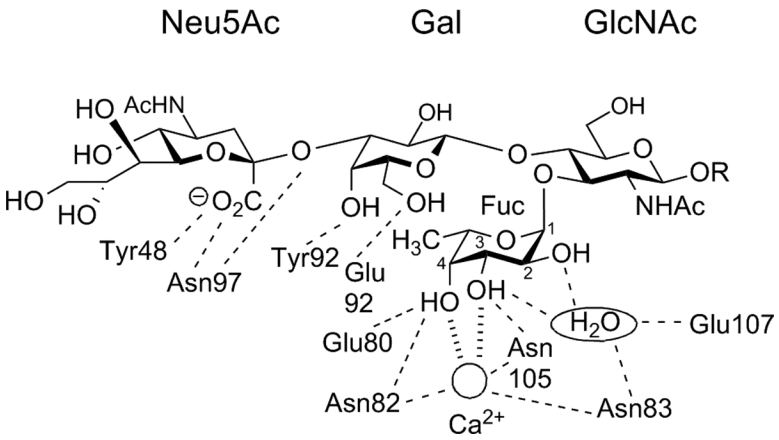


Figure 3.12 Interactions of the tetrasaccharide ligand SLe^X with the CRD of E-selectin (Somers et al., 2000).

specific site of the carbohydrate and are predominantly electrostatic in nature. The 3- and 4-hydroxyl groups of the L-fucose ring interact with the selectin-bound Ca^{2+} ion and the fucose-OH groups make hydrogen bonds with selectin residues which are also involved in coordination of the calcium ion. The 4- and 6-hydroxyl group of galactose and the carboxylate group of the neuraminic acid residue are involved in further hydrogen bonds with the carbohydrate binding site of the selectin. Figure 3.12 shows the complexation of SLe^X with E-selectin (Somers et al., 2000).

In fact, SLe^X is necessary but not sufficient to generate optimal binding by selectins (Varki, 2007). Placing SLe^X in the context of other structures is important for biologically relevant selectin recognition. Thus, certain leukocyte glycoproteins carrying SLe^X in specific arrangements seem to be optimal ligands for E-selectin, and functionally significant recognition by L- or P-selectin requires additional sulfate esters. A crucial high-affinity molecule for P-selectin recognition is PSGL-1 (*vide supra*). This mucin-like CAM is expressed on all leukocytes, but becomes specialized for selectin recognition on neutrophils and monocytes. PSGL-1 carries sulfate groups on tyrosine residues of its polypeptide chain. Optimal PSGL-1-binding sites for P-selectin comprise a short O-linked sugar chain bearing an SLe^X motif adjacent to two or three sulfated tyrosine

residues, all contained within a short sequence at the N-terminal region of PSGL-1.

3.3.4 *Weak Interactions and the Role of Water*

Typically, protein–carbohydrate interactions have been described as weak interactions with characteristic dissociation constants in the millimolar to high micromolar range. It remains a difficult question, how the rather weak complexes between carbohydrates and their lectin receptors can lead to effective biological signals. Multivalency of carbohydrate–lectin interactions is obviously important for their biological significance (*vide infra*), yet how multivalency effects control carbohydrate-protein binding and contribute to the regulation of essential cellular processes is not fully understood. Moreover, calculation of the free energy of the association of a carbohydrate and a lectin is difficult, especially in terms of the determination of the entropic balance of this process that occurs in water! Hence, water clusters as well as water layers can occur in ordered form in both the CRD and its proximity as well as a stable hydration shell surrounding a carbohydrate. What forces a specific glycan or a carbohydrate portion to leave the aqueous phase and enter the CRD of a lectin?

When the free energy gain of carbohydrate–lectin complex formation is considered, one always faces the "water problem." Saccharides in living systems are hydrated by water molecules including ions that reside in more or less proximity to the hydrated carbohydrate rings (Eriksson et al., 2008). Also, the carbohydrate binding site of a protein is hydrated until a carbohydrate ligand is approaching. For carbohydrate–protein complex formation water molecules have to be displaced from the carbohydrate as well as the carbohydrate binding site of the protein. Thus solvation/desolvation energies are very important for the affinity of carbohydrates for their protein receptors, but they are difficult to be reliably determined. In addition the occurring conformational changes within the ligand as well as the receptors render the overall calculation of binding energy problematic (Simons et al., 2009). Given that carbohydrate conformation is determined by interactions with water (Kirschner and Woods, 2001) and, vice versa, that specific saccharides have their specific water shell, the question then arises as to whether

the recognition motif is that of the carbohydrate or its solvation shell. The famous father of modern glycoscience, Ray Lemieux, has addressed this question, "how water provides an impetus on molecular recognition in aqueous solution" already in the last century (Lemieux, 1996).

3.3.5 *Hydrophobic Interactions with Carbohydrates*

Carbohydrates are not solely hydrophilic molecules, they also display sites that are relatively hydrophobic. Mannose is an especially striking example of a nearly amphiphilic monosaccharide (Fig. 3.13). Hydrophobic sites of monosaccharides can interact with hydrophobic patches of the protein. Thus, molecular recognition of carbohydrates by proteins often involves their selective binding at sites adjacent to aromatic amino acid residues, tryptophan, tyrosine, or phenylalanine, where they can adopt a stacking motif with the aromatic side group providing a platform for the bound ligand (Boraston et al., 2004).

This has been associated with "apolar CH-π bonding," which can supplement hydrogen bonded or electrostatic interactions in

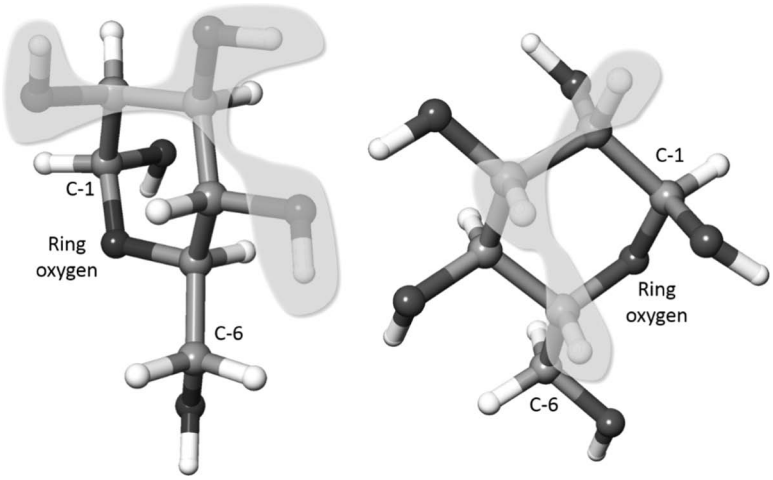


Figure 3.13 The monosaccharide α-D-mannose possesses a hydrophilic face (left), prone to hydration, and a rather hydrophobic face (right) that has been shown to remain "dry" in molecular recognition processes in aqueous systems. Hydrophilic OH groups and hydrophobic CH moieties are highlighted in the respective structures.

carbohydrate–lectin complex formation (Jiménez-Barbero et al., 2006; Ramírez-Gualito et al., 2009). Such van der Waals (London) forces are rather weak; however, they are frequently numerous and thus contribute significantly to overall binding.

3.3.6 *Binding to the Bacterial Lectin FimH*

Adhesion of bacteria to the surface of their host cells is, inter alia, mediated by interactions between carbohydrate ligands within the glycocalyx of the host cell and bacterial lectins, such as FimH. FimH is a lectin on the tips of long filamentous appendages projecting from the bacterial surface in hundreds of copies, called type 1 fimbriae (Knight and Bouckaert, 2009). Thus, fimbriae (also called pili) function as adhesive organelles. FimH contains a CRD specific for α-D-mannosides. Crystallographic studies have shown that the carbohydrate is buried within the binding site of the lectin, with the α-glycosidic aglycon moiety sticking out of the pocket (Bouckaert et al., 2005). At the entrance of the FimH CRD Tyr48 and Tyr137 form what has been called a "tyrosine gate" (Fig. 3.14). Hydrophobic

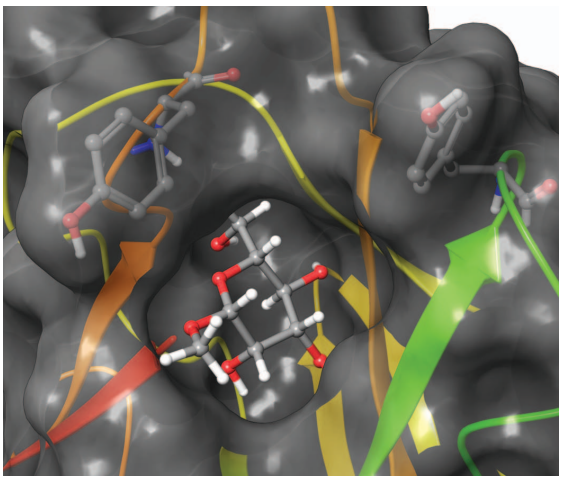


Figure 3.14 Connolly surface of the FimH carbohydrate-binding site with complexed methyl α-D-mannoside based on PDB code 1KLF. The α-glycosidic methyl moiety is sticking out of the carbohydrate binding site. Amino acid side chains of Tyr48 and Tyr137 at the entrance of the CRD are depicted as ball and stick models. They form a hydrophobic "gate." Graphic rendered with Maestro 9.5 as implemented in Schrödinger, LLC, New York.

CH-π or stacking interactions, respectively, that occur between a ligand and the tyrosine gate of the FimH CRD, can stabilize the resulting complex, both in case of artificial mannoside ligands such as p-nitrophenyl α-D-mannoside as well as when natural oligosaccharides are complexed by FimH (Wellens et al., 2008).

3.4 The Biological Role of Carbohydrate–Lectin Interactions

It is not possible to assign one precise role to the lectins especially in the plant regime. The C-type animal lectins on the other hand, represent an important recognition mechanism for oligosaccharides at cell surfaces, attached to circulating proteins and in the extracellular matrix. Binding of specific sugar structures by these lectins mediates biological events, such as cell–cell adhesion, serum glycoprotein turnover and innate immune responses to potential pathogens (Drickamer, 1999). Strikingly, amongst all the great diversity of potential and known glycan structures on cell surfaces, a very limited set of glycan epitopes is targeted for binding interactions involved in adhesion. Chief amongst these structures is the Lewis^{x} trisaccharide, along with its sulfated derivatives and SLe^{x}. It is unlikely that the diversity of cell surface glycans will be matched by a similar range of adhesion receptors. Nevertheless, specific glycosylation of mammalian cells provides a basis for selective adhesion. Changes in the glycosylation pattern such as in the case of tumor cells leads to altered cell adhesion, providing a route for tumor cell migration.

However, it remains a mystery what the biological orchestra behind these many interactions is. The simple paradigm of adhesion resulting from receptors on one cell binding to glycans on another cell applies in only a limited number of systems. Instead, glycans and receptor–glycan interactions often modulate adhesion in indirect ways, such as by changing the organization of cell surface glycoproteins and by antagonizing the effect of protein adhesion systems (Taylor and Drickamer, 2007).

Still, protein–carbohydrate interactions are fundamentally important in a wide array of biological organisms and many processes such as infection, fertilization, inflammation, and cellular

recognition, depend upon them. The numerous biological processes mediated by lectins can broadly be divided into those that involve recognition of endogenous ligands, such as for trafficking of glycoproteins and for cell adhesion and signaling events at the cell surface, and those that involve recognition of foreign (exogenous) cell surfaces (microbes) and mediate or modulate immune response to pathogens. There are a number of well-understood examples of cell adhesion based on interactions with cell surface glycans, such as ASGPR-mediated clearance of glycoproteins from the circulatory system and selectin-mediated leukocyte adhesion and recruitment in inflammation. Our current understanding of lectin-mediated biological processes is illustrated in the following on the basis of four examples, (i) the function of the ASGPR, (ii) selectin-mediated leukocyte trafficking, (iii) galectin-triggered signaling, and (iv) adhesion of type 1 fimbriated bacteria.

3.4.1 *Clearance of Glycoproteins by Interaction with ASGPR*

In the 1970s, Ashwell and colleagues discovered the hepatic ASGPR, a MHL and the first glycan-binding receptor ever discovered in animals (Ashwell and Morell, 1974). The liver controls the removal of proteins in the bloodstream and ASGPR is the involved hepatic receptor (van den Hamer et al., 1970). The C-type lectin ASGPR is specific for galactose- and GalNAc-terminated oligosaccharides, which appear after desialylation of senescent complex-type glycoconjugates. Such glycoproteins are bound by ASGPR at the surface of hepatocytes and then internalized through clathrin-coated pits (clathrin is a protein forming a vesicle coat). The cytoplasmic domain of ASGPR has a tyrosine-based internalization motif that promotes ligand delivery to early endosomes. There, the ligand–receptor complex dissociates and the receptor is recycled to the cell surface. The glycoproteins on the other hand are directed to lysosomes where they are degraded (Fig. 3.15). ASGP is a trimer in which each polypeptide chain contains a Gal/GalNAc-binding C-type CRD. Whereas the monomeric receptor has a rather weak affinity for its carbohydrate ligands with K_d values in the millimolar range, receptor clustering can lead to a logarithmic enhancement of receptor binding, a phenomenon that was termed

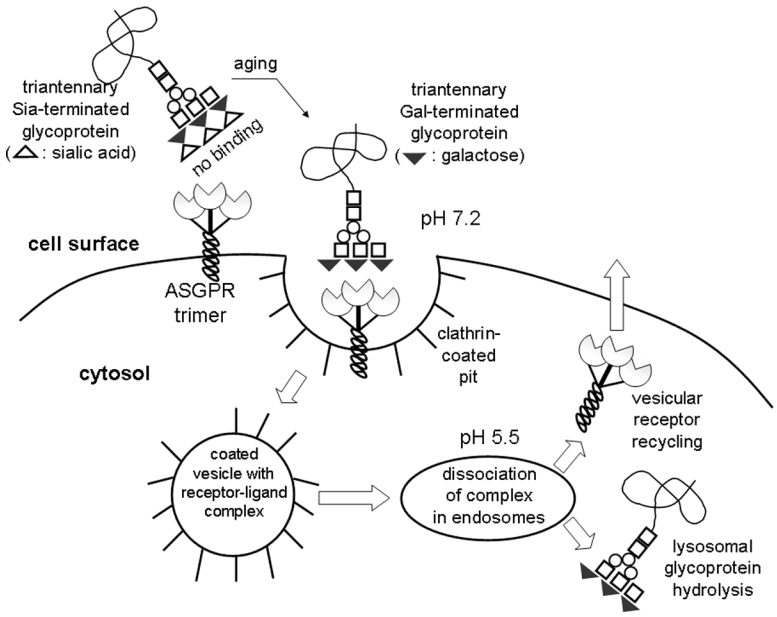


Figure 3.15 Schematic view of how the ASGPR trimer on liver cells binds to desialylated glycoprotein ligands, followed by endocytosis and thus clearance of senescent glycoproteins from the blood stream.

the "cluster effect" (Lee and Lee, 1995; *vide infra*). The ASGPR trimer binds with high affinity to triantennary N-linked glycans leading to their endocytosis. As in many other examples, here clustering of CRDs determines both the specificity and affinity of the lectin for its ligands.

ASGP seems to have a general role in clearance of serum glycoproteins. Yet the story of the Ashwell receptor has remained mysterious since endogenous ligands have not been seen. Only recently endogenous ligands of the Ashwell receptor were found and discovered to play a role as regulatory components in blood coagulation and thrombosis (Grewal et al., 2008).

3.4.2 *Leukocyte Trafficking*

Leukocytes (white blood cells) circulate continually in the blood and lymph and also migrate into the tissues at sites of infection or tissue

injury. This recirculation is critical to development of an inflammatory response in case of injury or other trauma. The vascular endothelium serves as a "gate keeper" regulating the movement of leukocytes into the tissues. Inflammation involves various immune system cells and numerous mediators.

In order for circulating leukocytes to enter the underlying tissue, the leukocyte cells must adhere to the endothelial cells and then pass between the cells that line the walls of blood vessels, a process called extravasation. To accomplish leukocyte trafficking, endothelial cells express leukocyte-specific CAMs, abbreviated as CAMs. Most CAMs belong to four families of proteins, (i) to the selectins, (ii) to the mucin-like family, (iii) the integrins, and (iv) to proteins from the immunoglobulin superfamily.

The selectins are a group of carbohydrate-binding CAMs that are involved in the first step of the leukocyte–endothelium interaction (Lowe, 2003). Three selectins mediate slow-down and so-called rolling of leukocytes, which finally triggers integrin-mediated firm adhesion and extravasation of leukocytes. Four principal steps can be distinguished in this process, as detailed in Fig. 3.16: (i) Rolling, (ii) chemoattractant activation, (iii) arrest and adhesion, and (iv) transendothelial migration (extravasation) (Rosen, 2004).

The rolling process requires a good balance between making and braking of contacts between the leukocytes and the endothelium. This balance is achieved through a variety of factors, some of which are intrinsic to the selectin molecules. For example, the rate constants for ligand binding (k_{on}) and release (k_{off}) are rapid. The extended shape of the selectins also allows them to act as a mechanical lever arm. In addition, the density and clustering of selectins as well as their carbohydrate ligands lead to multivalency effects that are critical to leukocyte trafficking (Ley and Kansas, 2004).

3.4.3 *Galectins in Signaling*

Many biological roles of galectins in cancer, immunity, inflammation, development, and signaling have been shown (Leffler et al., 2002; Liu and Rabinovich, 2010), but a unifying picture of their biological function is lacking. Instead galectins appear to have a particularly diverse but intriguing array of activities both inside and outside cells. Apparently, galectins are involved in the formation of clusters,

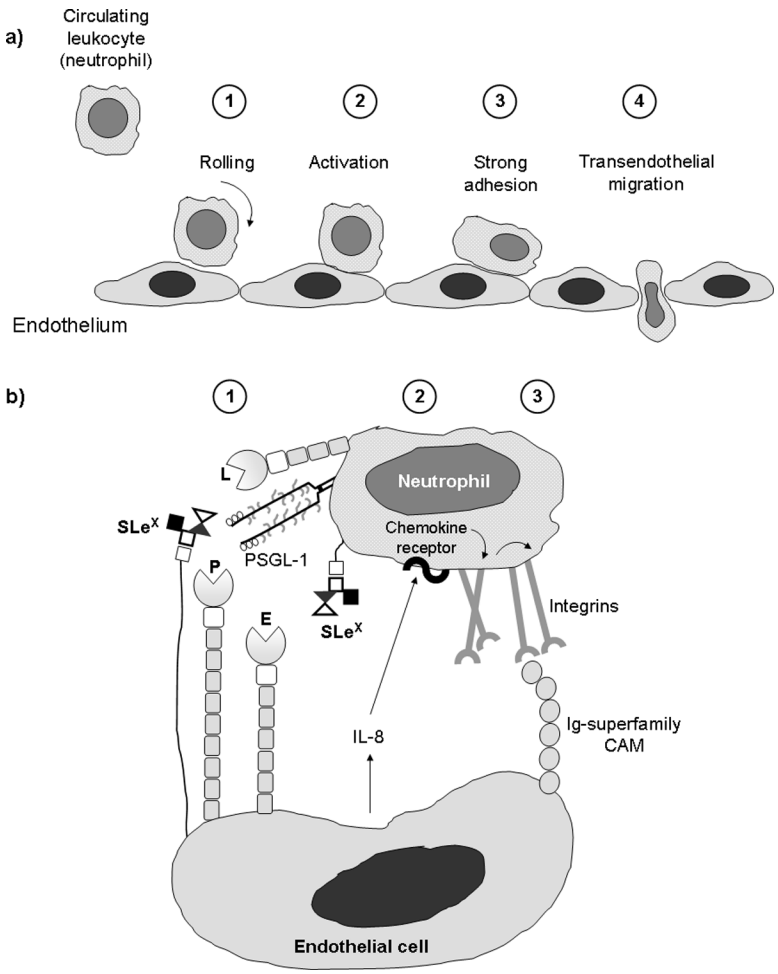


Figure 3.16 (a) Four sequential, but overlapping, steps lead to extravasation of leukocytes (neutrophils): (1) Slow-down of circulating leukocytes leading to "rolling"; (2) activation of neutrophils; (3) their firm adhesion; and (4) migration to the inflamed tissue. (b) The initial rolling (1) is mediated by interactions of L-selectin with sialyl Lewis^X (SLe^X) ligands on glycoproteins of cell surfaces of the vascular endothelium. Inversely, the endothelial P- and E-selectins interact with SLe^X epitopes on the leukocytes as well as PSGL-1 (P-selectin glycoprotein ligand-1, a mucin-like CAM, *vide supra*), in case of P-selectin. These initial interactions lead to secretion of a chemokine (interleukin-8, IL-8) that triggers an activation signal on the neutrophil (2). This signal induces a conformational change in the integrin molecules enabling them to adhere firmly to IG-superfamily molecules on the endothelial cells (3).

bundles, arrays, and lattices (Brewer et al., 2002), providing a special mechanism for lectin–saccharide-mediated cellular interactions.

Cells must be able to receive signals from the extracellular milieu and deliver those signals to the inside of the cell, a principle of cell communication. Protein–protein interactions have been described as the classical basis of signal transduction; however, today it has become clear that carbohydrate–protein interactions are also critical triggers in cell signaling. One mechanism of signal transduction starts with binding of a ligand to a transmembrane receptor, leading to a specific cellular response via a cascade of complex events. For example, signaling can occur when cell surface lectins bind to saccharide ligands or when lectins bind to cell surface glycans, respectively. Binding of soluble galectin-1 to its ligand *N*-acetyllactosamine on a T cell surface, for instance, triggers the T cell to die. However, signaling is not necessarily triggered by bimolecular receptor–ligand interactions. For many systems the clustering of protein receptors and ligands is required. Thus, signaling turns out to be a highly supramolecular process, involving wide-area molecular rearrangements such as the assembly of ordered domains on the cell surface (Dam and Brewer, 2010). Binding of galectin-3 to cell surface receptors was shown to lead to lectin–carbohydrate lattice formation at the cell surface. This process depends on carbohydrate ligand valency and density and has an impact on receptor internalization and cell signaling.

Carbohydrate ligand valency and density is connected to the level of glycosylation of cell surface glycoproteins and this, in turn, depends on the respective metabolism. In the biosynthesis of N-linked oligosaccharides a number of highly specific glycosyltransferases are active in a particular sequence. That is, a preceding glycosylation renders an oligosaccharide product that is a high-affinity substrate for a specific glycosyltransferase acting successively. The mannoside-*N*-acetylglucosaminyltransferases, abbreviated as Mgats, are a family of glycosyltransferases operative in the biosynthesis of N-glycans of the complex and hybrid type. These enzymes catalyze the addition of bisecting *N*-acetylglucosamine (GlcNAc) residues to a core mannoside residue on a glycoprotein glycan chain. For the synthesis of complex-type oligosaccharides, Mgat1 activity is essential, since other sugar-modifying enzymes only become active once Mgat1 (also called GlcNAc transferase I) has been in play. Mgat5, also called

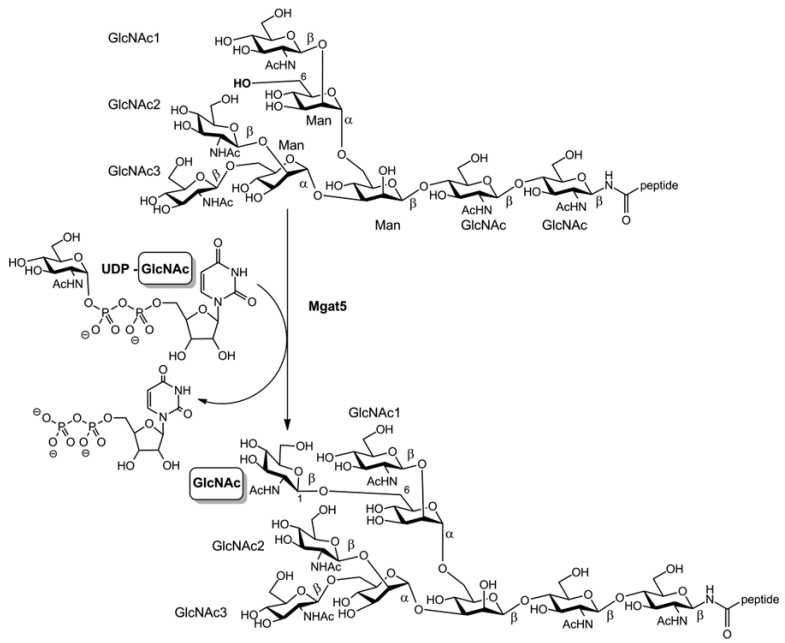
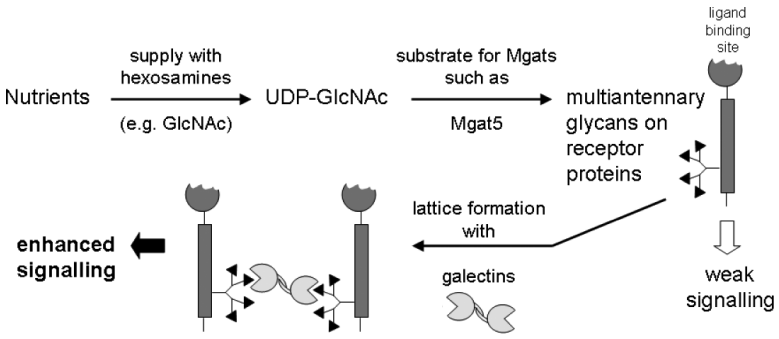


Figure 3.17 The glycosyltransferase Mgat5 is critical in the biosynthesis of complex-type oligosaccharides that are important ligands for galectins. It catalyzes the β1,6-glycosidic transfer of a GlcNAc moiety from the respective activated substrate (UDP-GlcNAc) on to triantennary glycans.

GlcNAc-transferase V (GlcNAc-TV), is one of the glycosyltransferases acting subsequently. Mgat5 catalyzes the β1,6-glycosidic transfer of a GlcNAc moiety from the respective activated substrate (GlcNAc-uridine diphosphate, UDP-GlcNAc) on to triantennary glycans, the key step in the sequential action of different monosaccharide transferases that yield complex-type oligosaccharide chains (Fig. 3.17).

The extent of glycosylation regulates receptor activity. Individual glycoprotein receptors may engage individual ligands, each interaction sending a weak signal. However, appropriately glycosylated receptors can be crosslinked by galectins and subsequent engagement by ligands sends signals that are temporarily and spatially concentrated to result in an overall increase in signal strength and/or duration. Thus, transmembrane receptor signaling is regulated by galectin-induced lattice formation (Brewer et al., 2002; Horst and Wagener,



AU: Please cite figure 3.18 in text.

Figure 3.18 Supply with hexosamines, such as GlcNAc, by nutrients has consequences on the biosynthesis of branched N-glycans of the complex type as part of glycoproteins which act as receptors on cell surfaces (Lau et al., 2007). Once exposed on the cell, galectins may bind to the glycosylated receptors. Whereas, individual receptors may engage individual ligands, each sending a weak signal, appropriately glycosylated receptors can be crosslinked by galectins to form lattices. Subsequent engagement by ligands leads to increased signal strength.

2009). As the extent of glycosylation of transmembrane receptor glycoproteins with oligoantennary complex-type glycans regulates their interaction with galectins, signaling can apparently depend upon the N-glycan biosynthesis (Rabinovich et al., 2007).

It was shown that N-glycan branching is highly sensitive to metabolic flux through the hexosamine pathway that regulates the UDP-GlcNAc level in the Golgi. N-glycan branching in turn regulates the strength of the glycoprotein association with cell-surface galectins and lattice formation with galectins. Thus, receptors with high N-glycan content, because they are involved in lattices, are subject to prolonged cell surface exposure and effective in signaling. On the other hand, receptors with few N-linked glycans are routed for internalization (endocytosis) and thus suppressed in their signaling activities (Lau et al., 2007). Cell arrest receptors are kept in check by their comparatively high rate of endocytosis, due to their weak association with the formed galectin lattice.

Cell signaling receptors that regulate cell growth or motility are not independent of one another, and they are coupled via long-distance feed-backloops. The long-distance loops are embedded in a complex regulatory network, involving different glycoprotein

glycoforms and molecular communication processes mediated by lectin–carbohydrate interactions (Horst and Wagener, 2009). Given that virtually every protein on the surface of a cell is glycosylated, lectin–carbohydrate interactions including the formation of lattices, may be ubiquitous participants in all types of cellular communication.

3.4.4 *Type 1 Fimbriae-Mediated Bacterial Adhesion*

Wheresoever cells are in contact with the outside environment, e.g., in the case of epithelial cells, their glycocalyx is utilized by bacteria to colonize the cell surface. Specific proteins are initiators of the adhesion process, which is in turn amplified, leading to the development of a well organized superstructure, a so-called biofilm, being highly advantageous for the colonizing microbes. Biofilm formation facilitates firm and irreversible adhesion to a surface, interlinking bacteria of different species, which produce a carbohydrate mucus to maintain the biofilm. Through this exopolysaccharide layer bacteria can achieve chemical communication and profit from favorable coordination, a process that is called quorum sensing. Often biofilm formation is the basis of a beneficial symbiosis between a microorganism and its host; however, as soon as microorganisms invade into another habitat or only slightly change their genes, disorders can arise leading to inflammatory diseases, or even apoptosis or uncontrolled cell growth (Sgouros and Bergele, 2006).

The cell surface forms an ideal site for multiplication and persistency for bacteria and thus adhesion and colonization of cell surfaces is most advantageous for bacteria. To facilitate adhesion, bacteria such as *Escherichia coli* have developed long hair-like organelles that project from the bacterial cell surface in several hundreds of copies. These so-called fimbriae mediate carbohydrate-specific adhesion to the highly glycosylated host. One class of bacterial fimbriae is type 1 fimbriae that consist of interlinking subunits of FimA proteins forming a coiled helix-shaped rod of some micrometers in length. The tip of this rod is capped by a lectin called FimH (*vide supra*). FimH mediates binding to α-mannosidic carbohydrate ligands on the host, which is important for uropathogenic *E. coli* (UPEC) bacteria (Fig. 3.19) as mannose is frequently found on epithelial cells.

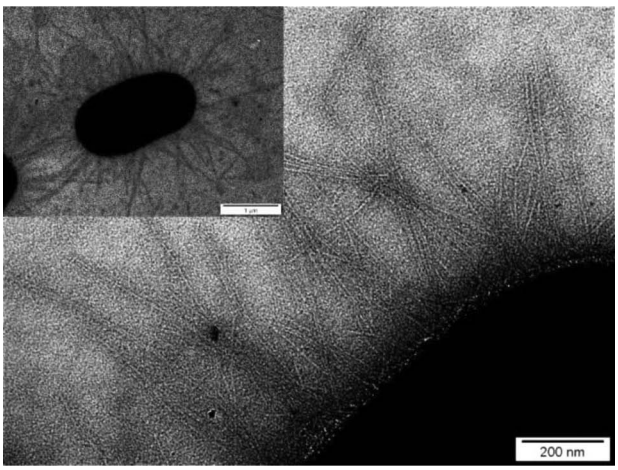


Figure 3.19 Electron micrograph of *E. coli* bacteria, covered with adhesive organelles called fimbriae (here type 1 fimbriae). Both, the bacteria as well as their fimbrial tentacles are some micrometers long.

Although the available knowledge allows a satisfactory visualization of carbohydrate recognition by FimH, biological assays have indicated that carbohydrate binding by the fimbrial lectin, the FimH–mannoside interaction, is rather weak and reversible, and therefore can hardly account for the firm attachment of *E. coli* to their host, nor can it explain progression into irreversible adhesion. At this stage it is worthwhile to consider additional molecular mechanisms and as yet unknown multivalency effects, which apparently play a vital role in glycocalyx biology.

3.5 Multivalency of Carbohydrate–Protein Interactions

Practically all known plant lectins possess one CRD per peptide chain. On the other hand, all mammalian lectins can act as multivalent lectins. While the carbohydrate ligands are presented as multiple copies both in solution as well as on cell surfaces, multivalency of lectin CRDs is established in different ways. Mammalian lectins can occur as tandem-repeat proteins such as in the case of many galectins, or they express two or more CRDs on one peptide chain. These CRDs can be homologous or heterologous. In

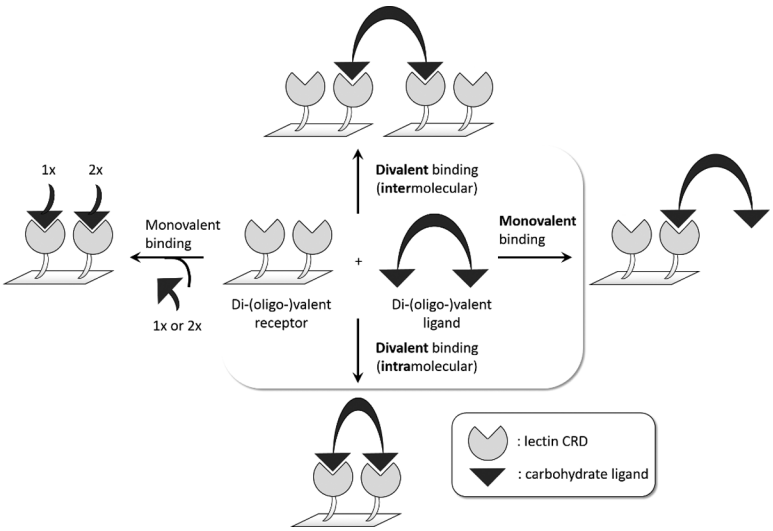


Figure 3.20 Multivalency in carbohydrate–protein interactions can occur once a lectin with clustered sugar binding sites and a multivalent ligand that can present sugar ligands with proper orientation are involved. Multivalent (shown: divalent) binding can occur intra- or intermolecularly. Similarly, monovalent interactions can happen according to different modes.

addition, multivalency of CRDs can be achieved by oligomerization of monovalent lectins or by lectin clustering within the lipid bilayer. Whereas lectin binding to a single carbohydrate ligand is usually a low-affinity interaction with K_d values in the millimolar range, multivalent lectin–carbohydrate interactions lead to tighter and often more specific binding. Thus, many low-affinity binding events result in a high overall avidity. This has been first observed by Y. C. Lee during studies on carbohydrate binding of ASGPR (Lee and Lee, 1995). He found that a linear increase of carbohydrate ligands on an appropriate scaffold molecule could result in a logarithmic gain in binding strength. From this result he concluded that multiple lectin CRDs must be involved in ligand binding accounting for the observed phenomenon, which he called "cluster effect." Thus, a significant cluster effect obviously requires two partners, a lectin with clustered sugar binding sites and a multivalent ligand that can present sugar ligands with proper orientation and spacing (Fig. 3.20). Enhancement in affinity on a concentration corrected basis can range from a few

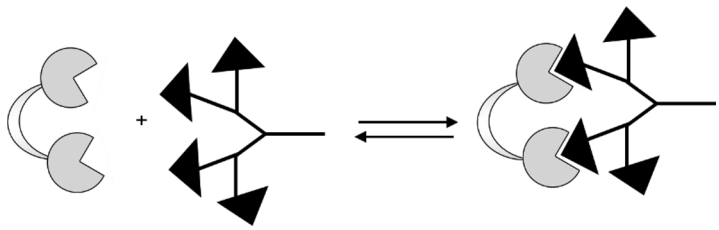


Figure 3.21 A bivalent receptor–ligand interaction forms the minimal case of multivalency. Complexation occurs with the free energy of association ΔG_2^{bi} and may involve some conformational changes within the binding partners.

orders of magnitude to nearly 10^9. Today it is generally accepted that multivalent interactions being much stronger than the monovalent ones, is an important glycobiological principle that is often necessary to regulate physiological processes.

How can binding of two (N) molecules occur with much higher affinity than the sum of two (N) corresponding monovalent interactions? There are a number of different ways in which N receptor sites can interact with N ligands, and the free energy of binding of an interaction depends strongly on its details. In a systematic approach, an interaction between N ligands and N receptors distributed on two entities can be regarded an Nth-order polyvalent interaction that occurs with free energy of association ΔG_N^{poly}. Figure 3.21 demonstrates the simple case of a di(bi)valent interaction, occurring with the free energy of association (not dissociation!) ΔG_2^{bi}. Here, as in analogous processes of another valency, the average free energy of interaction, ΔG_{avg}^{poly} for the association of a single carbohydrate ligand and a single lectin CRD in a polyvalent interaction is equal to $\Delta G_N^{poly}/N$:

$$\Delta G_{avg}^{poly} = \Delta G_N^{poly}/N;$$

and since $\Delta G = -RT \ln K$,

$$K_N^{poly} = (K_{avg}^{poly})^N.$$

Since a monovalent ligand–receptor interaction occurs with free energy change of ΔG^{mono}, N monovalent, independent receptors

interact with N monovalent, independent ligands with a free energy change of $N\Delta G^{mono}$. Accordingly, the term "avidity," which refers to the association constant of a polyvalent interaction (in contrast to "affinity," referring to the association constant of a monovalent interaction) is defined as the quantity given by K_N^{poly}.

The average free energy of interaction between a ligand moiety and receptor moiety in a polyvalent interaction G_{avg}^{poly} can be greater than, equal to, or less than the free energy in the analogous monovalent interaction ΔG^{mono}. Accordingly, these classes of polyvalent interactions were called positively cooperative (synergistic), noncooperative (additive), or negatively cooperative (interfering), respectively. In biochemistry, e.g., enzymology, positive cooperativity is assessed when binding of the first substrate molecule increases the affinity of the other active sites for the substrate (Badji et al., 2005). Negative cooperativity occurs when binding of the first substrate decreases the affinity of the enzyme for other substrate molecules. In a multivalent ligand–receptor interaction, the degree of cooperativity, α, can be defined as

$$\alpha = \lg(K_N^{poly})/\lg(K^{mono})^N, \text{ with } K_N^{poly} = (K_{avg}^{poly})^N = (K^{mono})\alpha^N.$$

Consequently, if the free energy of the second binding event in a bivalent complexation is more favorable than the first one, the cooperativity factor α is greater than 1. It is unitless. This is because in this case $(K_{avg}^{bi})^2 > (K^{mono\ (first)})^2$. This process is called a positively cooperative polyvalent interaction.

Interestingly, multivalent carbohydrate–protein interactions typically are processes with negative cooperativity, with $\alpha < 1$ (interfering) (Dam et al., 2002). A famous example for a negatively cooperative polyvalent interaction is binding of di- and trivalent galactose-containing ligands to the Ashwell receptor, a C-type lectin on the surface of hepatocytes, studied by Y. C. Lee and co-workers (Lee and Lee, 1995). They observed $K^{mono} = 7 \times 10^4\ M^{-1}$, $K_2^{bi} = 3 \times 10^7\ M^{-1}$, and $K_3^{tri} = 2 \times 10^8\ M^{-1}$. Thus, K_2^{bi} turned out to be equal to $420 \times K^{mono}$ and K_3^{tri} was determined as equal to $2800 \times K^{mono}$. Since $K_2^{bi} < (K^{mono})^2$ and $K_3^{tri} < (K^{mono})^3$ these di- and trivalent ligands bind with negative cooperativity. Thus, this example illustrates an important characteristic of polyvalent carbohydrate–protein interactions:

even though di- and trivalent binding in this case was negatively cooperative ($\alpha < 1$), the measured avidity for a di- and trivalent ligand was much higher than for the monovalent molecule (this observation was called the "cluster effect" by Y. C. Lee, *vide supra*). In other words, tight binding does not require positive cooperativity in the sense that this phrase is traditionally used in biochemistry.

Whitesides and co-workers (Mammen et al. 1998) have thus suggested the use of an enhancement factor β as a parameter that is more descriptive for polyvalent systems. The quantity β is the ratio of avidity and the component affinity of the monovalent equivalent to the interaction

$$\beta = K_N^{poly} / K^{mono}.$$

Factor β can also be used, when the valency of a process under investigation is not known. In this case, any experimentally obtained measure for the polyvalent interaction, using a multivalent carbohydrate ligand such as a glycopolymer, for example, K^{poly}, can be related to the obtained constant for the monovalent interaction by $K^{poly} = \beta K^{mono}$.

The reason why the classical definition of cooperativity is not sufficient to explain multivalency effects in interactions between polyvalent carbohydrate ligands with multiple CRDs can be found when changes of entropy are considered. There are many possible ways how multivalent molecules, ligand and receptor, can interact with each other (Fig. 3.22). The involved interactions might occur in two subsequent steps, first intermolecularly and then intramolecularly. However, binding does not necessarily have to follow a single pathway; combinations of binding modes are possible and might even contribute to specificity and regulation mechanisms. Both enthalpy (ΔH) and entropy (ΔS) of binding are important in this respect (Kitov and Bundle, 2003).

The traditional idea is that multivalency is mainly governed by entropy like in the case of a chelate effect (Lundquist and Toone, 2002). In case of the chelate effect the enthalpy term for each "monovalent portion" of the complexation reaction should be approximately the same for all steps, whereas the entropy term is different in each step of a multivalent complexation reaction. This means that less entropy

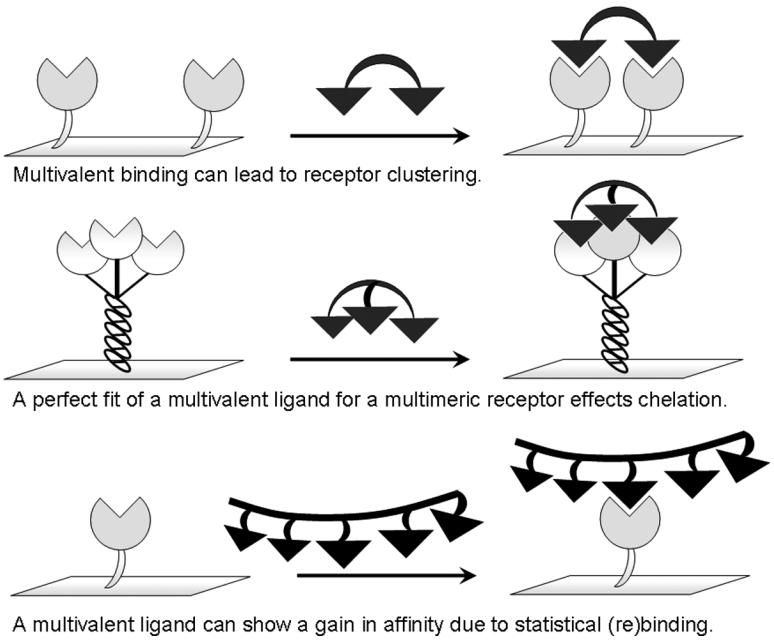


Figure 3.22 There are many different ways how multivalent ligands can interact with (multivalent) receptors leading to a gain in binding energy and avidity effects, respectively. Complex formation between ligands and receptors may influence the proximity of receptors on a surface (receptor clustering), or their relative orientation. Depending on the nature of a multivalent ligand, different effects can be seen, including agglutination and lattice formation (not shown). High receptor affinity of multivalent carbohydrate ligands can be just a statistical effect based on the high concentration of the ligand molecules in close proximity to the CRD of the receptor. In these cases it is important to report valency-corrected affinities.

of disorder is lost when the chelate complex is formed than when the complex with monovalent ligands is formed. This results in the well-known chelate effect that corresponds to those instances when K_N^{poly} is greater than K of any of the involved monovalent binding events. For example, K would refer to the association constant for the interaction of methylamine with iron, and K_2^{bi} is the association constant of ethylenediamine with iron. There is a chelate effect in

case that K_2^{bi} is greater than K. The chelate effect can be maximized by an optimal geometrical relationship of the oligovalent receptor and its oligovalent ligand.

Thus, in this approach the binding enthalpy is typically assumed to be proportional to the number of interactions and the mode of binding is dependent on the entropy of binding. Intramolecular multivalent binding would be entropically favorable as the multivalent complex is assumed to involve the same rotational and translational entropy loss as its corresponding monovalent interaction. Therefore, it is reasoned that the mode of binding is determined by the loss of conformational entropy upon intramolecular binding. However, it has been suggested that this is an idealized view and that this idealized entropy-based understanding of multivalent binding should not be taken too seriously (Mulder et al., 2004; Badji et al., 2005).

Because multivalent glycomimetics (*vide infra*) are often equipped with long flexible tethers, their binding should cost considerable entropic penalty. In spite of this, they often show considerable multivalency effects (Kiessling et al., 2006). Therefore, it seems to be appropriate to consider an effective concentration, C_{eff}, to understand the phenomenon of multivalent binding of tethered glycomimetics. Effective concentration represents a probability of interaction between two reactive or complementary entities. A plethora of multivalent ligands have been synthesized and tested in order to achieve the search for high affinity/avidity and improve our understanding of multivalency effects occurring in glycobiology and beyond (Badji et al., 2005). However, it has been difficult to design multivalent glycomimetics for appropriate multivalent binding and to predict the thermodynamics of the multivalent interactions aimed at. Defining geometries and binding motifs is problematic, even more so as the architecture of any multivalent entity has a strong influence on the mode of binding. Nevertheless, as the intrinsic complexity of biological molecules and systems imposes limitations on quantitative kinetic and thermodynamic analyzes of molecular recognition processes, small and well-defined artificial model systems, wherein quantification is quite feasible, can in principle contribute enormously to a better understanding of multivalent phenomena.

3.6 Model Systems for the Investigation of Carbohydrate–Protein Interactions

Because of the enormous structural diversity of carbohydrates and the manifold aspects of their biological effects, it is a major challenge to understand and investigate the mechanisms underlying their biological function. The molecular details of carbohydrate–protein interactions are almost impossible to elucidate in a complex scenario like a living system, or even a cell assay, for example. Therefore, individual molecules have been frequently applied in glycobiological studies and tested in simple, clearly laid-out setups. Due to the microheterogeneity of glycans (glycoconjugates frequently occur in different glycoforms), it is difficult to isolate complex oligosaccharides and glycoconjugates from natural material in sufficient quantity and purity. Thus, much research in the glycoscience utilizes synthetic molecules to address specific biochemical questions (Bertozzi and Kiessling, 2001). Either saccharides or glycoconjugates, respectively, are made according to the naturally occurring molecules (Herzner et al., 2000; Gamblin et al., 2009; Zhu and Schmidt, 2009; Piontek et al., 2009a,b), or structurally varied molecules are synthesized, which are often called "glycomimetics." A glycomimetic approach is advantageous in that specific molecular parameters can be systematically varied with comparatively limited time and work. In a solid phase supported glycopeptide synthesis, for example, the number of carbohydrate epitopes as well their spatial arrangement on the peptide backbone can be easily modified (Hilaire and Meldal, 2000; Barkley and Arya, 2001). Hundreds of examples of such man-made glycoconjugates have been published (Köhn et al., 2004; Gómez-García et al., 2005; André et al., 2008; Ernst and Magnani, 2009). Some of the invented molecules do not resemble much of the native glycoconjugates, whereas other architectures are more related to the natural prototypes.

Testing of monovalent interactions between carbohydrate ligands and their lectin receptors typically reveal low binding constants in the millimolar to micromolar range. Employing tailor-made glycomimetics has often led to significantly increased affinities. However, as it is known that affinity is especially increased by multivalent binding of carbohydrates to proteins, also multivalent glycomimetics have been gaining importance.

Multivalent glycoconjugates allow to study multivalency effects in relation to conformational availability, ligand presentation, or ligand density dependencies. Thus, to realize some of the complexity of glycoconjugates in a glycomimetic approach, an oligovalent scaffold molecule is regularly employed for assembly and ligation, respectively, of the carbohydrate epitopes of biological interest (Lindhorst, 2002; Ortiz Mellet et al., 2002). Many different scaffolds and ligation chemistries have been utilized in such an attempt. Some decades ago, TRIS (tris(hydroxymethyl)methylamine) was used in Lee's laboratory (Lee and Lee, 1995) for the synthesis of first so-called "cluster glycosides." In the meantime many other small but oligovalent molecules, including pentaerythritol, peptides, or carbohydrate cores, have been used for the same purpose. Figure 3.23 exemplifies some representative examples for small glycoclusters that were prepared using different synthetic methods.

3.6.1 *Multivalent Glycomimetics: Glycodendrimers and Successors*

The complexity of interactions as well as structures and the diversity of contexts in the carbohydrate regime present have prompted researchers to synthesize glycoconjugates in which multivalency is featured to an even larger extent than in case of the smaller glycoclusters. Many different architectures of multivalent glycoconjugates have been synthesized and tested and this research has been extensively reviewed (Chabre and Roy, 2010; Lindhorst, 2002; Lahmann, 2009; Kiessling et al., 2006). A typical example of how multivalency of carbohydrate ligands has been achieved in artificial conjugates are the glycodendrimers (Lindhorst, 1996). In classical glycodendrimers a (hyper)branched noncarbohydrate core, such as a PAMAM (polyamidoamine)-dendrimer is employed as scaffold molecule and its periphery is ligated to the principal carbohydrates of interest (often just simple monosaccharides). Thus, an oligoantennary oligosaccharide backbone, providing the multivalent presentation of carbohydrate ligands in the native molecules, is substituted by an unequally easy to synthesize dendritic core. The valency of a dendritic core can be varied with ease according to the general principal of generation-wise growth

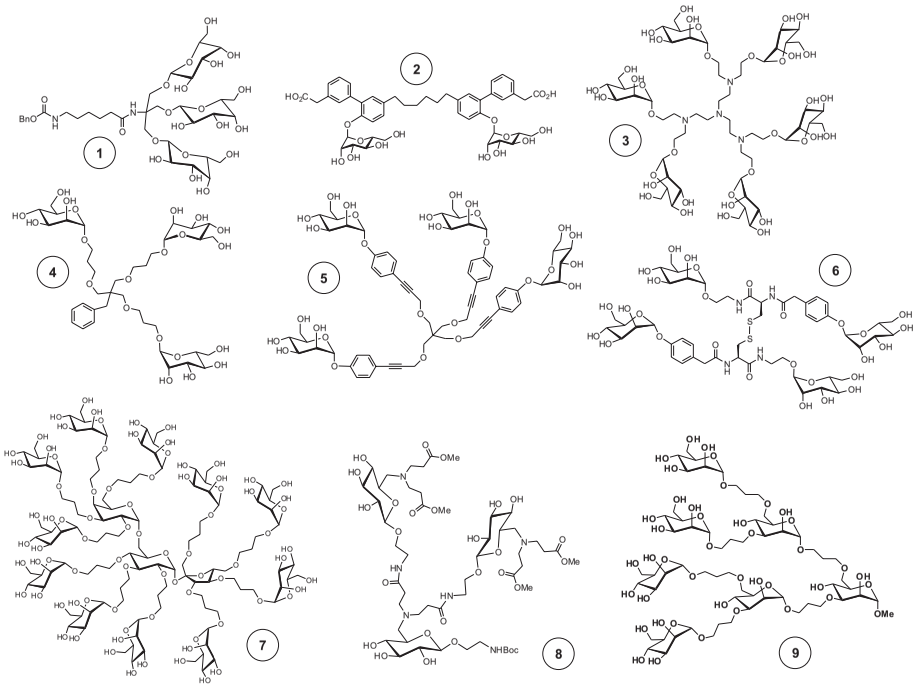


Figure 3.23 Different types of small glycoclusters, representing examples of glycomimetics, that are used in the glycosciences: 1 (Lee and Lee, 1995); 2 (Kogan et al., 1998); 3 (Dubber and Lindhorst, 2001); 4 (Röckendorf et al., 2002); 5 (Imberty et al., 2008); 6 (Schierholt et al., 2010); 7 (Dubber et al., 2006); 8 (Sadalapure and Lindhorst, 2000); 9 (Heidecke and Lindhorst, 2007).

of dendritic molecules (Röckendorf and Lindhorst, 2001; Touaibia and Roy, 2007; Turnbull and Stoddart, 2002). In addition, it is feasible to vary the ligation chemistry that is used to decorate a dendrimer or other scaffold molecule with carbohydrates. Besides glycosylation, many other chemistries have been employed to make glycodendrimers, such as coupling of isothiocyanato-functionalized carbohydrates with branched oligoamines to form as thiourea-bridged products.

When glycodendrimers are used in biological studies, it is assumed that the natural scaffolding of carbohydrate ligands is negligible, thus less important, and that the biological effect of a multivalent ligand can be reproduced by just the multivalent presentation of those carbohydrate moieties that are bound to the CRDs of their lectin receptors. This idea, however, has not always held true.

The first glycodendrimer reported (Roy et al., 1993) was in fact a "glycodendron." Dendrons are molecules branching out from just one functional group, called the "focal point." This can be an advantageous design, as the functional group at the focal point of the molecule can be used for eventual modification such as to attach a label, a reporter group or a linker moiety. This ligation step can be planned orthogonal to the chemistry that is used to equip the rest of the molecule with carbohydrate epitopes. Regular dendrimers on the other hand branch out at least from a di- or trifunctional molecule delivering hyperbranched multivalent molecules with uniform, indistinguishable functional groups.

Since the first glycodendron was published in 1993, the field of synthetic multivalent glycoconjugates has grown further (Chabre and Roy, 2013; Roy and Shiao, 2015). In Fig. 3.24, some of the principal architectures that have been used for the construction of multivalent glycoconjugates are summarized. Whereas glycoclusters and glycodendrimers can be considered to be monodisperse individual molecules, constructs like glycopolymers (Hartmann and Becer, 2015) and neoglycoproteins are more or less broad mixtures of multivalent molecules of different molecular weight, thus polydisperse. This has been kept in mind for the interpretation of results, obtained in biological assays with such molecules of one type or the other. To resemble the multivalent nature of cell surface glycans even more closely, larger, supramolecular models of multivalent

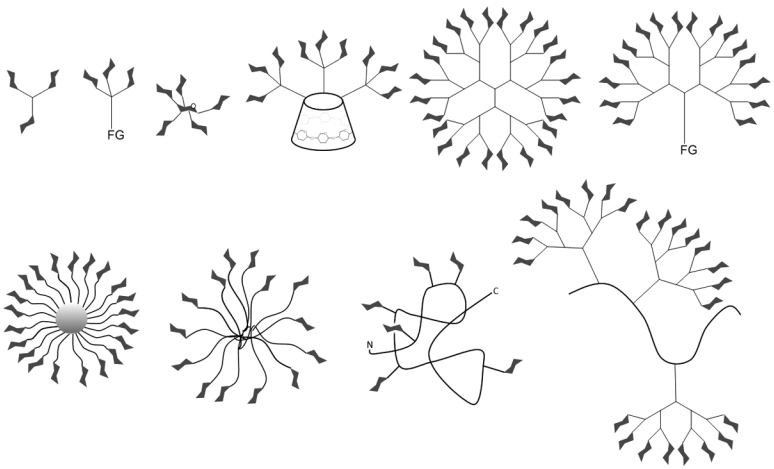


Figure 3.24 Many different classes of multivalent glycoconjugates and glycomimetics, respectively, are known. The depicted cartoons represent a selection of exemplary molecular architectures. First row from left to right: small glycoclusters and glycodendrons, carbohydrate-centered glycoclusters, multiglyco-cages based on calixarenes or cyclodextrins, respectively, glycodendrimers and glycondendrons; FG = functional group. Second row from left to right: supramolecular constructs such as glyconanoparticles, glycomicelles (glycoliposomes are not shown), neoglycoproteins, and glycodendronized polymers.

presentation are also used in glycobiology, such as micelles and liposomes (Liu et al., 2009) have been introduced. Using mercapto-functionalized carbohydrate derivatives gold glyconanoparticles (de la Fuente and Penadés, 2006) or carbohydrate-decorated self assembled monolayers (SAMs) (Houseman and Mrksich, 2002) could be achieved. These supramolecular constructs as well as carbohydrate surfaces (Jonathan et al., 2009) open up new prospects in the glycosciences (Timmer et al., 2007; Schmidt-Lassen and Lindhorst, 2014).

It has been notoriously difficult to draw the structures of large dendrimers and glycodendrimers and a great deal of effort has been spent on this "artwork" (Lee et al., 2005; Wang et al., 2008; Astruc et al., 2010). However, it cannot be assumed that glycodendrimers adopt a perfectly symmetric conformation as suggested by the

published chemical drawings. Rather many different conformations of a glycodendrimer are in equilibrium with one another, the perfectly symmetric conformations being less likely. This was shown in molecular dynamic (MD) simulations with four different mannose-decorated glycoclusters and PAMAM-based glycodendrimers (von der Lieth et al., 2002). The accessible conformational space of the molecules depicted in Fig. 3.24 is illustrated by an overlay of 1000 snapshots of 1 ns long MD simulations including explicit water molecules. Three atoms in the core region of each molecule were used to orient all of the 1000 archived conformations in the same way in space. The centre of the sugar termini as well as the centre of each glycodendrimer were defined as pseudo atoms and displayed as spheres in different shade of gray. In such a way, the occupied conformational space for each pseudo atom can be visualized (Fig. 3.25).

3.6.2 *Self-Assembled Monolayers: Glyco-SAMs.*

Self-assembled monolayers (SAMs) are an ideal system for a reliable as well as flexible fabrication of ordered glycoarrays (Love et al., 2005). SAMs can be functionalized with carbohydrate head groups to obtain "glyco-SAMs" (Houseman and Mrksich, 2002 Kleinert et al., 2004; Ban and Mrksich, 2008), which allow the study of carbohydrate–protein interactions using different biophysical methods (Kind and Wöll, 2009). For the formation of glyco-SAMs on gold, carbohydrate-terminated long-chain thiols, or thioacetates are needed, as well as their noncarbohydrate analogs allowing "dilution" of the monolayer to avoid steric hindrance at the surface, which continuously increases during assembly (Fig. 3.26). The alkane thiols that form the monolayer have to be attached to an OEG (oligoethylene glycol) chain, typically with EG_3 to EG_7, to resist unspecific adsorption of proteins (biorepulsive properties). Protein-repelling properties of the unfunctionalized SAM are a prerequisite for employment of glyco-SAMs in carbohydrate-protein binding studies. Finally, for glyco-SAM construction, molecules of the type $HO(CH_2CH_2O)_n(CH_2)_{11}SH$ can be functionalized with a biofunctional group such as a saccharide.

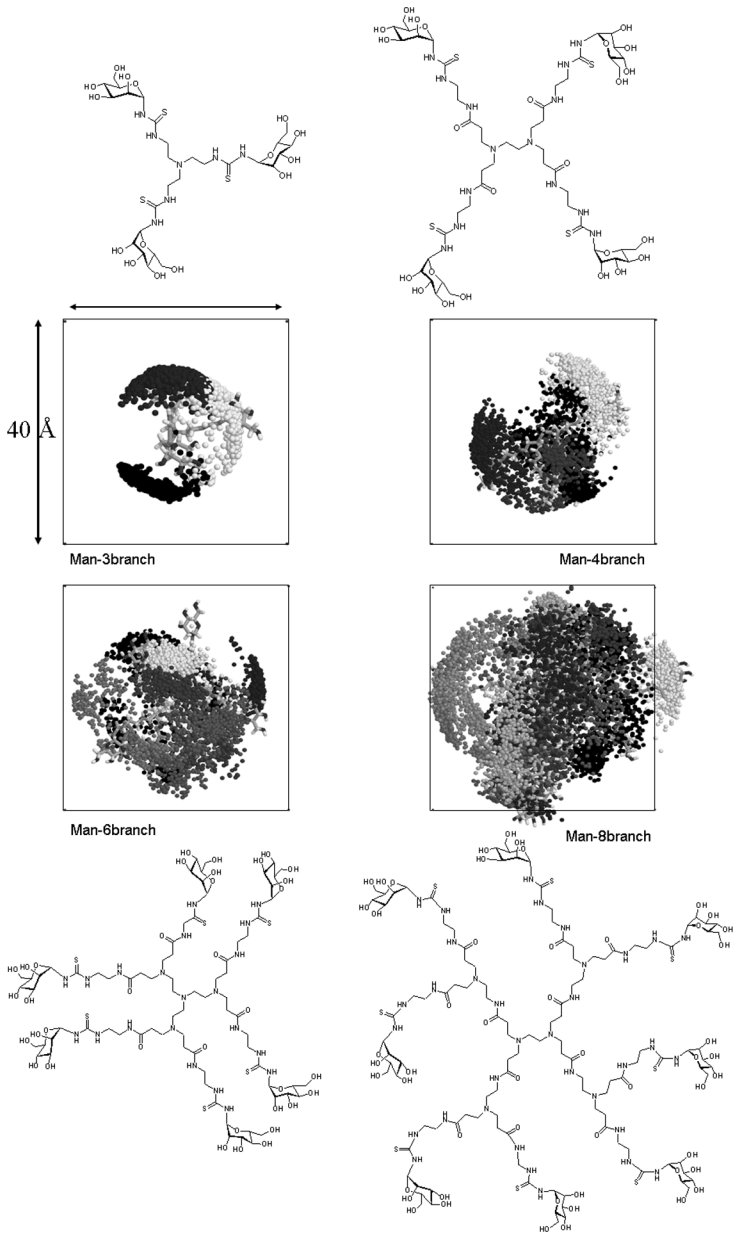


Figure 3.25 Accessible conformational space of thiourea-bridged mannose-decorated glycoclusters and glycodendrimers of different valency; as revealed by molecular dynamic simulations (von der Lieth et al., 2002).

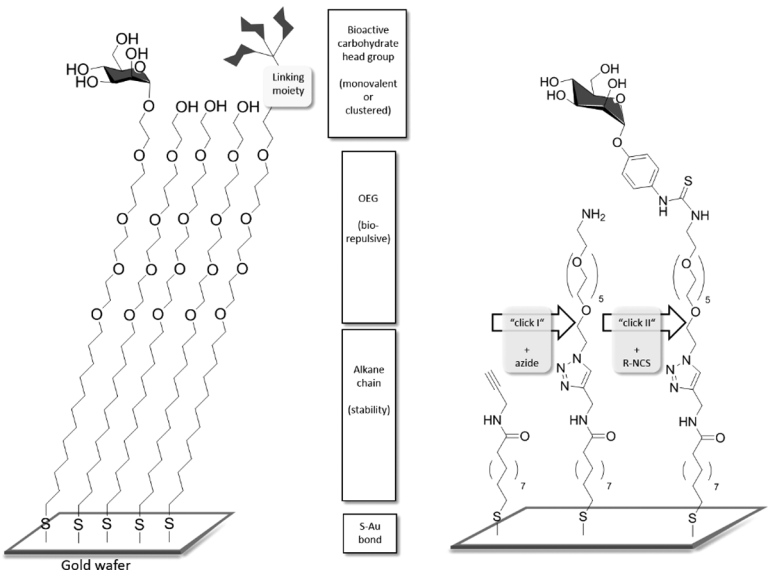


Figure 3.26 Left: Principal composition of carbohydrate-decorated SAMs (glyco-SAMs) in which oligoethylene glycol (OEG)-substituted alkane thiols are employed for chemisorption onto a gold surface. Right: Modular fabrication of glyco-SAMs by a "dual click" approach (Grabosch et al., 2013).

Glyco-SAMs can be constructed in two different ways: (i) either the completely functionalized thiol is synthesized first and subsequently assembled on the gold substrate; or (ii) a basic monolayer is formed first and this blank is then further refined "on SAM" (Fig. 3.26, right). The latter concept enables greater flexibility for biological testing, because a particular SAM can be investigated before and after further modification. Thus, "switching" the condition of a monolayer may allow to draw conclusions about the biological consequences of a specific change, which is not possible otherwise. Two chemistries suited to achieve ligation of carbohydrates "on SAM" are the 1,3-dipolar cycloaddition of azides to alkynes, which was named "click chemistry" by Sharpless (Tornøe et al., 2002; Rostovtsev et al., 2002) or thiourea-bridging between isothiocyanato-functionalized carbohydrates and amino-terminated SAMs (Grabosch et al., 2010; Grabosch et al., 2013; Kleinert et al., 2008). Figure 3.26 displays the principal modular approach to

glyco-SAMs of various kind employing simple monosaccharide head groups or cluster glycosides, respectively.

Glycoarrays such as glyco-SAMs allow to "scan" the carbohydrate specificity of new proteins and the evaluation of their lectin characteristics (Rillahan and Paulson, 2011). However, in addition to studies on the carbohydrate specificity of lectins, the investigation of the supramolecular context of carbohydrate recognition is important. This involves carbohydrate diversity and complexity, respectively, of a glycosylated surface, its carbohydrate density as well as the mode of carbohydrate exposition on a surface, in other words their orientation. In order to explore the influence of carbohydrate orientation on carbohydrate-specific cell adhesion azobenzene glycosides were utilized. In azobenzene glycosides an azobenzene substructure is glycosidically attached to a specific carbohydrate and serves as a photoresponsive hinge unit. Irradiation of the planar and more stable *E*-form of an azobenzene glycoside with UV light ($\lambda \sim 365$ nm) effects transition to the bent *Z*-isomer. By exposure to visible light ($\lambda > 440$ nm) or by thermal equilibration, the azobenzene *Z*-isomer relaxes back to the *E*-form according to a half life $\tau_{1/2}$ which is an individual parameter of a specific azobenzene derivative. Hence, E/Z isomerization of the N = N double bond in azobenzene glycosides can be employed to effect a considerable change in the spatial orientation of the conjugated sugar moiety. When an azobenzene glycoside is immobilized on a surface, E/Z isomerization allows control over carbohydrate orientation (Fig. 3.27).

Monitoring of the isomerizaion of azobenzene derivatives is easily accomplished by UV-vis spectrosopy. However, on metal surface such as on gold, UV-vis spectroscopy is less feasible. Recently, photoisomerizaion of azobenzene glyco-SAMs on gold was proven by IRRA (infrared reflection absorption) spectroscopy by the groups of Lindhorst and Tuczek (Chandrasekaran et al., 2014). In the next step a joint effort of the Lindhorst and the Terfort group it was shown that the mannose-specific adhesion of *E. coli* cells to an azobenzene glyco-SAM can be controlled by photochemically "switching" the surface between an adhesive (*E*-form) and a much less adhesive state (*Z*-form of the azobenzene hinge region) (Weber et al., 2014). This is an exciting finding as it shows that carbohydrate-lectin interactions are regulated by more parameters than just the nature of the carbohydrate. In a natural environment such as a cell surface, sugars

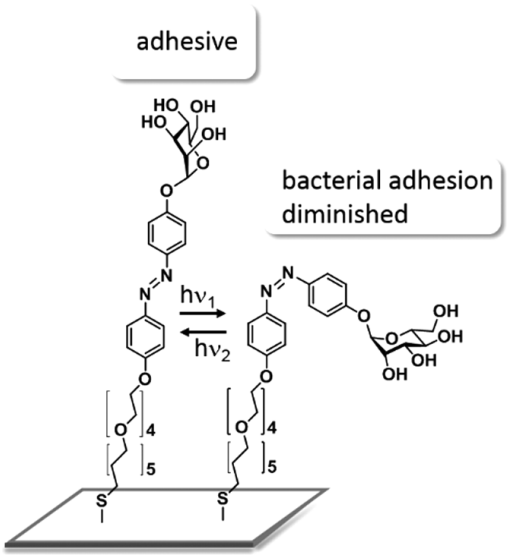


Figure 3.27 SAMs fabricated with azobenzene glycosides can be reversibly switched between an adhesive and a less adhesive state. The azobenzene substructure allows for a defined reorientation of the carbohydrate ligand under photochemical control (Weber et al., 2014).

may aid the orientational control of other sugars in their proximity. This is suggested by discoveries made with heteroglycoclusters. Here it was shown that the addition of different sugars into one glycocluster molecule, such as lactose together with mannose, e.g., can influence affinity to certain lectins more drastically than just increasing the valency of a respective homoglycocluster (Jiménez Blanco et al., 2013).

3.6.3 *Testing Bacterial Adhesion*

Bacterial infections constitute a major global health problem, especially threatening the health of young children (Mulholland and Adegbola, 2005). The most common serious neonatal infections involve bacteraemia, meningitis, and respiratory tract infections (Osrin et al., 2004). Key pathogens in these infections are *E. coli*, *Klebsiella* sp., *Staphylococcus aureus,* and *Streptococcus pyogenes*.

To cause infection, bacteria often need to adhere to target cells and to colonize the glycosylated cell surface. For the attachment to cells, most bacteria depend on the expression of specialized adhesive organelles, which are hair like, 1–2 μm long, and ~7 nm wide protein structures on the bacterial cell surface. They are referred to as fimbriae (or pili). Much studied examples include P fimbriae and type 1 fimbriae that provide UPEC with the ability to bind to carbohydrate receptors (Ohlsen et al., 2009; Klemm and Schembri, 2000). Type 1 fimbriae are common throughout the *Enterobacteriaceae* and mediate agglutination of guinea pig red blood cells in a mannose-inhibitable manner. *E. coli* type 1 fimbriae have been classified as type 1^{E}, which are mostly referred to simply as type 1 fimbriae (Boyd and Hartl, 1999). Type 1 fimbriae are widely expressed by *E. coli* and are used by uropathogenic strains to mediate attachment to specific niches in the urinary tract. Thus, type 1 fimbriae have a well-established role in urinary tract infections (Kau et al., 2005) and have also been implicated in neonatal meningitis, Crohn's disease and bovine mastitis (Barnich et al., 2003).

Even though many molecular details of mannose-specific bacterial adhesion have been elucidated (Knight et al., 2000; De Greve et al., 2007), the complex and highly dynamic interplay of different molecular interactions connected to type 1 fimbriae-mediated adhesion of *E. coli* is not yet completely understood. Bacterial adhesion to the highly diverse and nanodimensioned glycocalyx of target cells is a highly multivalent process that involves molecular recognition of α-D-mannosidic residues as well as additional molecular interactions with the fimbrial adhesin and moreover with the fimbrial rod, which mainly consists of FimA subunits (Kline et al., 2009; Virji, 2009). Type 1 fimbriae may even undergo conformational changes upon shear stress, leading to shear-enhanced adhesion (Thomas et al., 2002; Forero et al., 2006).

It depends on the tissue type on the one hand and the type of bacteria on the other, whether or not bacterial adhesion causes a problem for the host. It can be a synergistic advantage like *E. coli* colonization of the bowel or it might lead to inflammation, apoptosis, or even problems such as peptic ulcer (Blaser, 2005). Thus, investigation of the mechanisms of bacterial adhesion is an important research field to promote our understanding of its consequences in health and disease.

Again, as the molecular diversity of the carbohydrates that constitute the glycocalyx is extremely difficult to handle, the problem of complexity was tackled by narrowing it down to simplified but highly specified systems. A two-step methodology can be followed for testing bacterial adhesion to glycosylated surfaces.

(i) Reduction of the supramolecular complexity of the glycocalyx to distinct saccharide moieties, such as certain monosaccharides. This approach allows to study the effect of detailed structural variations of the glycoside aglycon and glycon, respectively.
(ii) Approximation of the surface scenario of the glycosylated cell by assembling the distinct glycoside constituents under investigation in the form of a glycoarray. This setup allows to study cellular adhesion to a glycosylated surface rather than just looking at receptor–ligand interactions in solution.

To allow the study of the molecular mechanisms of cellular adhesion to tailored glycosylated surfaces a powerful assay is an important prerequisite. A classical method is an ELISA (enzyme-linked immunosorbent assay), which has often been used in studies on bacterial adhesion with synthetic α-mannoside ligands. In this adhesion–inhibition assay mannose-specific bacteria together with the serially diluted mannoside are incubated in mannan-coated microtiter plates. The polysaccharide mannan and the synthetic mannosides in solution compete for binding to the type 1 fimbrial lectins. After washing, adhered bacteria are visualized by addition of a monoclonal antibody against a protein on the type 1 fimbrial rod. Treatment with a horseradish peroxidase-conjugated secondary antibody and subsequent staining with ABTS [2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)] facilitates the optical density (OD) readout. Higher OD-values thus parallel with better bacterial adhesion to the mannan surface and a low inhibitory potency of the added inhibitor.

The bottleneck of any ELISA is the primary monoclonal antibody, which has to be produced exclusively for one specific application. A second disadvantage is that the binding event is detected indirectly, as each single step is an individual error source. Next generation

bacterial adhesion assays were initiated by the idea to equip the bacteria directly with a detectable moiety. Therefore, prior to use, bacteria were biotinylated. For assaying bacterial adhesion, these biotin-labeled bacteria are again allowed to bind either the inhibitor or the mannosides on the surface. In a following incubation step, a streptavidin-HRP conjugate can practically be anchored to the bacteria through the very strong biotin-streptavidin interaction. Subsequent staining with ABTS allows readout of the OD, analogous to the ELISA. When fluorescing bacteria containing green fluorescent protein (GFP) (Hartmann et al., 2010) are applied with mannan-coated microtiter plates, competitive binding between the well surface and a specific inhibitor in solution was tested as before. However, after incubation and washing, readout of bacterial adhesion could be performed directly with a fluorescence intensity reader at 485 nm.

3.6.4 *Tailoring Carbohydrate Surfaces to Mimic Cellular Adhesion to the Glycocalyx*

The kit-like character of this GFP-based assay makes it a valuable tool for quick screening of potential new inhibitors of type 1 fimbriae-mediated bacterial adhesion and for testing adhesion to glycoarrays. Instead of using mannan coating, prefunctionalized 96-well microtiter plates can be modified according to the characteristics of a natural glycocalyx. In order to cope with the complexity of a glycosylated cell surface environment, specific structural elements of a glycoarray can be varied with rational planning. For example, the aglycon moieties of the exposed glycosides can be modified (Fig. 3.27), altering parameters such as scaffolding, flexibility of the linker, or additional affinity effects. However, the complexity of the glycoarray can be expanded by employing a mixture of carbohydrate ligands in a "multiplexing" approach (Fig. 3.27b,d), or increase the saccharide complexity by going from mono- to oligosaccharide ligands. This approach has the potential, to investigate subtle structural changes within the carbohydrate ligands and maintain the supramolecular character of a glycosylated surface, which forms the basis of the adhesion processes operating at cell surfaces.

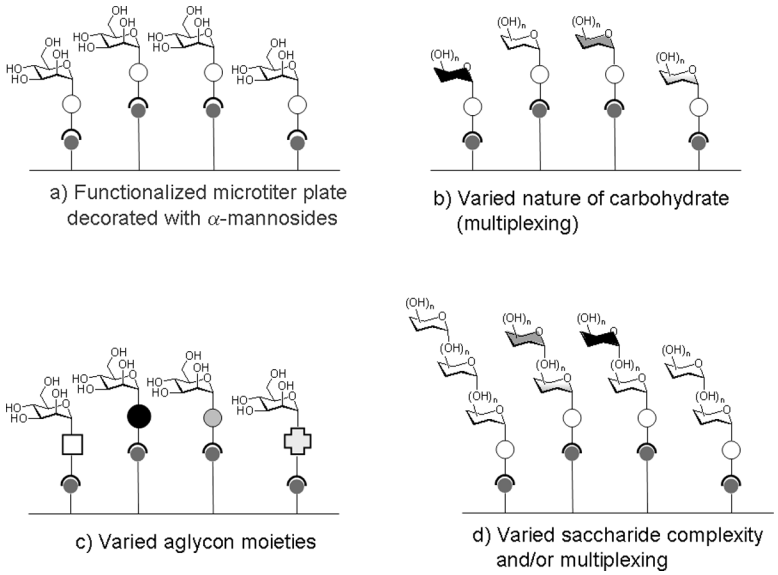


Figure 3.28 Tailor-made carbohydrate surfaces are composed of three variable building blocks, the carbohydrate glycon moiety (varying in size and complexity), the aglycon (structurally differing aglycon moieties are depicted as different forms), and the functional group to chemically link the molecule to the microtiter plate surface. (a) Mannoside ligands, the aglycon moiety and the linkage chemistry are kept constant; (b) the nature of the carbohydrate (glycon moieties) are varied (multiplexing); (c) aglycon moieties of the immobilized mannosides are varied, and (d) complexity and nature of saccharides (multiplexing) are varied within the respective glycoarray.

3.7 Conclusions and Outlook

Besides proteins and nucleic acids, the third class of essential biopolymers, the carbohydrates, serve as molecular key players in cell recognition, cell adhesion, cell communication, differentiation and immunity, as well as in pathogen adhesion. Carbohydrates occur on the cell surface as principal parts of the glycocalyx, literally the sugar coat of a cell, and in soluble form, both extracellularly and intracellularly. Their structural diversity exceeds that of nucleic acids and polypeptides by a large extent. From the millions of possible structural variations arising from configurational plurality, different

branching patterns and modifications of the sugar rings, respectively, apparently "only" some tens of thousands structural combinations are found in nature as part of the "mammalian glycospace" that is the conformational basis of glycan structures (Werz et al., 2007). Thus the mammalian oligosaccharide diversity could be much smaller than expected on the basis of theoretical considerations.

Nevertheless, carbohydrate sequences are a vast source of information and it remains to be elucidated how this information is released within the complex supramolecular networks in living systems. It has been argued that oligosaccharides could harbor a "glyco code" that is waiting to be deciphered in various contexts of biological and medical importance (Feizi and Chai, 2004). Such an oligosaccharide coding system (sugar code) could be "decoded" by specific interactions with lectin receptors to allow cells to communicate efficiently (Gabius et al., 2004); however, it is unassured that decoding of a sugar "alphabet," in analogy to reading the genetic code, works that way.

As a matter of fact, cell surface as well as secreted glycans mediate many receptor–lectin interactions by virtue of their recognizable chemical features. These are fine-tuned by a sophisticated set of enzymatic modifications. This "remodeling" of glycans can alter their binding properties and consequently their biological function (Parker and Kohler, 2010). To understand more about the multiple modes of carbohydrate–protein interactions, in addition to multivalent glycomimetics, high-throughput methodologies for their study have been created, such as the glycoarray technology (Krishnamoorthy and Mahal, 2009; Song and Pohl, 2009). This has supported the development of a systems-level study of carbohydrates, termed "glycomics," that is meant to increase our knowledge of biological systems that operate through oligosaccharide recognition.

The arsenal of synthetic and analytical tools that has been made available has generated an immense amount of information on carbohydrate structures and functions in a short period of time. However, from such glycomic database information, there is still a long way to go in understanding of the fundamental control mechanisms underpinning the glycome. Much remains to be discovered about this large class of molecules. Chemists and biologists have learned reasonably well to study even the most complex biological molecules when they are isolated, but how could

one get a molecular-level glimpse of the chemical processes as they unfold in living systems? After all, understanding the function of biomolecules is crucially dependent on dynamical changes during their biosynthesis, distribution, and recognition in their native environment at the cellular and organismal levels. It will be necessary to also investigate and understand aspects of conformational control, organization of multivalency (Ciuk and Lindhorst, 2015), and the importance of orientation in carbohydrate recognition, research that has recently been started (Weber et al., 2014). Given the importance of carbohydrates in various aspects of biology, studying oligosaccharides and glycoconjugates in recognition processes has the potential to usher a new understanding of biological recognition that will have a strong impact on biological chemistry.

Acknowledgment

The author gratefully acknowledges Dr. rer. nat. Femke-Isabell Beiroth for her help with some of the figures.

AU: Please provide the details of the reference citation Jiménez Blanco et al. (2013) so that it can be included in the reference list.

References

Ahmad N., Gabius H.-J., André S., Kaltner H., Sabesan S., Oscarson S., Brewer C. F., *Glycobiology*, **14** (2004), 817–825.

Ambrosi M., Cameron N. R., Davis B. G., *Org. Biomol. Chem*., **3** (2005), 1593–1608.

André S., Sansone F., Kaltner H., Casnati A., Kopitz J., Gabius H.-J., Ungaro R., *ChemBioChem.*, **9** (2008), 1649–1661.

Angata T., Margulies E. H., Green E. D., Varki A., *PNAS,* **101** (2004), 13251–13256.

Ashwell G., Morell A., *Adv. Enzymol*., **41** (1974), 99–128.

Astruc D., Boisselier E., Ornelas C., *Chem. Rev*., **110** (2010), 1857–1959.

Badji J. D., Nelson A., Cantrill S. J., Turnbull W. B., Stoddart J. F., *Acc. Chem. Res*., **38** (2005), 723–732.

Ban L., Mrksich M., *Angew. Chem*., 120 (2008), 3444–3447; *Angew. Chem. Int. Ed*., **47** (2008), 3396–3399.

Banerjee R., Das K., Ravishankar K., Suguna K., Surolia A., Vijayan M., *J. Mol. Biol*., **259** (1996), 281–296.

Barkley A., Arya P., *Chem. Eur. J*., **7** (2001), 555–563.

Barnich N., Boudeau J., Claret L., Darfeuille-Michaud A., *Mol. Microbiol*., **48** (2003), 781–794.

Bertozzi C. R., Kiessling L. L., *Science,* **291** (2001), 2357–2364.

Blaser M. J., *Sci. Am.,* **292** (2005), 38–45.

Boraston A. B., Bolam D. N., Gilbert H. J., Davies G. J., *Biochem. J.,* **382** (2004), 769–781.

Bouckaert J., Berglund J., Schembri M., De Genst E., Cools L., Wuhrer M., Hung C.-S., Pinkner J., Slättegård R., Zavialov A., Choudhury D., Langermann S., Hultgren S. J., Wyns L., Klemm P., Oscarson S., Knight S. D., De Greve H., *Mol. Microbiol*., **55** (2005), 441–455.

Boyd E. F., Hartl D. L*., J. Bacteriol.,* **181** (1999), 1301–1308.

Brewer C. F., *Glycoconj. J*., **19** (2004), 459–465.

Brewer C. F., Micell M. C., Braun L. G., *Curr. Opin. Struct. Biol.,* **12** (2002), 616–623.

Chabre Y. M., Roy R., *Adv. Carbohydr. Chem. Biochem*., **63** (2010), 165–393.

Chabre Y. M., Roy R., *Chem. Soc. Rev*., **42** (2013), 4657–4708.

Chandrasekaran V., Jacob H., Petersen F., Kathirvel K., Tuczek F., Lindhorst T. K., *Chem. Eur. J*., **20** (2014), 8744–8752.

Ciuk A. C., Lindhorst T. K., *Beilstein J. Org. Chem.*, **11** (2015), 668–674.

Crocker P. R., Varki A., *Trends Immunol*., **22** (2001), 337–342.

Dahms N. M., Olson L. J., Kim J.-J. P., *Glycobiology*, **18** (2008), 664–678.

Dam T. K., Brewer C. F., *Glycobiology,* **20** (2010), 270–279.

Dam T. K., Roy R., Pagé D., Brewer C. F., *Biochemistry*, **41** (2002), 1351–1358.

De Greve H., Wyns L., Bouckaert J., *Curr. Opin. Struct. Biol.,* **17** (2007), 506–512.

de la Fuente J., Penadés, S., *Biochim. Biophys. Acta,* **1760** (2006), 636–651.

Dennis J. W., Pawling J., Cheung P., Partridge E., Demetriou M., *Biochim. Biophys. Acta,* **1573**, (2002) 414–422.

Drickamer K., *Curr. Opin. Struct. Biol*., **9** (1999), 585–590.

Drickamer K., A genomics resource for animal lectins (2014), http://www.imperial.ac.uk/research/animallectins (accessed August 12, 2015).

Drickamer K., Taylor M. E., *Annu. Rev. Cell Biol*., **9** (1993), 237–264.

AU: The text contains the citation Drickamer (2014), but the list contains the details of Drickamer (2015). Please check and advise.

Dubber M., Lindhorst T. K., *Synthesis,* 2001, 327–330.

Dubber M., Sperling O., Lindhorst T. K., *Org. Biomol. Chem*., **4** (2006), 3901–3912.

Dwek, R. A. *Chem. Rev*., **96** (1996), 683–720.

Eriksson M., Lindhorst T. K., Hartke B., *J. Chem. Phys.*, **128** (2008), 105105.

Ernst B., Magnani, J. L., *Nat. Rev. Drug Discov.*, **8** (2009), 661–677.

Feizi T., Chai W., *Nat. Rev. Mol. Cell Biol*., **5** (2004), 582–588.

Fernández-Alonso M. C., Cañada F. J., Jiménez-Barbero J., Cuevas G., *J. Am. Chem. Soc*., **127** (2005), 7379–7386.

Forero M., Yakovenko O., Sokurenko E. V., Thomas W. E., Vogel V., *PLoS Biol.*, **9** (2006), 1509–1516.

Gabius H.-J., *Crit. Rev. Immunol*., **26** (2006), 43–80.

Gabius H.-J., Siebert H. C., André S., Jiménez-Barbero J., Rüdiger H., *ChemBioChem*, **5** (2004), 740–764.

Gamblin D. P., Scanlan E. M., Davis B. G., *Chem. Rev.*, **109** (2009), 131–163.

Gamblin S. J., Skehel J. J., *J. Biol. Chem*., **285** (2010), 28493–28409.

Gómez-García M., Benito J. M., Rodríguez-Lucena D., Yu J.-X., Chmurski K., Ortiz Mellet C., Gutiérrez Gallego R., Maestre A., Defaye J., García Fernández José M., *J. Am. Chem. Soc.*, **127** (2005), 7970–7971.

Grabosch C., Kleinert M., Lindhorst T. K., *Synthesis,* 2010, 828–836.

Grabosch C., Kind M., Gies Y., Schweighöfer F., Terfort A., Lindhorst T. K., *Org. Biomol. Chem*., **11** (2013), 4006–4015.

Grewal P. K., Uchiyama S., Ditto D., Varki N., Le D. T., Nizet V., Marth J. D., *Nat. Med.,* **14** (2008), 648–655.

Hansen S., Holmskov U., *Immunobiology,* **199** (1998), 165–189.

Hartmann L., Becer R. (eds.), *Glycopolymer Code: Synthesis of Glycopolymers and Their Application,* RSC Publishing 2015, RSC Polymer Chemistry Series (ISBN: 9781849739788.

Hartmann M., Horst A. K., Klemm P., Lindhorst T. K., *Chem. Commun.,* **46** (2010), 330–332.

Heidecke C., Lindhorst T. K., *Chem. Eur. J.,* **13** (2007), 9056–9067.

Herzner H., Reipen T., Schultz M., Kunz H., *Chem. Rev.*, **100** (2000), 4495–4538.

AU: Please provide the volume of refs. Grabosch et al. (2010); Roy et al. (1993); Schierholt et al. (2010).

Hilaire P. M. S., Meldal M., *Angew. Chem*., **112** (2000), 1210–1228; *Angew. Chem. Int. Ed*., **39** (2000), 1162–1179.

Horst A. K., Wagener C., *Top. Curr. Chem*., **288** (2009), 139–156.

Houseman, B. T., Mrksich, M., *Chem. Biol.,* **9** (2002), 443–454.

Imberty A., Chabre Y. M., Roy R., *Chem. Eur. J.*, **14** (2008), 7490–7499.

Iobst S. T., Wormald M. R., Weis W. I., Dwek R. A., Drickamer K., *J. Biol. Chem.,* **269** (1994), 15505–15511.

Jiménez–Barbero J., Asensio J. L., Cañada F. J., Poveda A., *Curr. Opin. Struct. Biol*., **9** (1999), 549–555.

Jiménez-Barbero J., Asensio J. L., Cuevas G., Canales A., Fernandez-Alonso M. C., Cañada F. J., *Biocatal. Biotransform*., **24** (2006), 13–22.

Jonathan F., Popplewell M. J., Swann Y. A., Turnbull J. E., Fernig D. G., *ChemBioChem*, **10**, (2009) 1218–1226.

Kau A. L., Hunstad D. A., Hultgren S. J., *Curr. Opin. Microbiol.,* **8** (2005), 54–59.

Kelm S., Pelz, A., Schauer R., Filbin M. T., Tang S., de Bellard M. E., Schnaar R. L., Mahony J. A., Hartnell A., Bradfield P., Crocker P. R., *Curr. Biol*., **4** (1994), 965–972.

Kiessling L. L., Gestwicki J. E., Strong L. E., *Angew. Chem.,* **118** (2006), 5418–5422; *Angew. Chem. Int. Ed. Engl.*, **45** (2006) 2348–2368.

Kilpatrick D. C., *Biochim. Biophys. Acta*, **1572** (2002), 187–197.

Kind M., Wöll C., *Prog. Surf. Sci.* **84** (2009), 230–278.

Kirschner K. N., Woods R. J., *PNAS*, **98** (2001), 10541–10545.

Kitov P. I., Bundle D. R., *J. Am. Chem. Soc*., **125** (2003), 16271–16284.

Kleinert M., Röckendorf N., Lindhorst T. K., *Eur. J. Org. Chem*., 2004, 3931–3940.

Kleinert M., Winkler T., Terfort A., Lindhorst T. K., *Org. Biomol. Chem*., **6** (2008), 2118–2132.

Klemm P., Schembri M., *Int. J. Med. Microbiol*., **290** (2000), 27–35.

Kline K. A., Fälker S., Dahlberg S., Normark S., Henriques-Normark B., *Cell Host Microbe*, **5** (2009), 580–592.

Knight S. D., Berglund J., Choudhury D., *Curr. Opin. Chem. Biol.*, **6** (2000), 653–660.

Knight S. D., Bouckaert J., *Top. Curr. Chem*., **288** (2009), 67–107.

Köhn M., Benito J. M., Ortiz Mellet C., Lindhorst T. K., García Fernández J. M., *ChemBioChem*, **5** (2004), 771–777.

Kogan T. P., Dupré B., Bui H., McAbee K. L., Kassir J. M., Scott I. L., Hu X., Vanderslice P., Beck P. J., Dixon R. A. F., *J. Med. Chem*., **41** (1998), 1099–1111.

Krishnamoorthy L., Mahal L. K., *ACS Chem. Biol.*, **9** (2009), 715–732.

Lahmann M., *Top. Curr. Chem*., **288** (2009), 17–65.

Lau K. S., Partridge E. A., Grigorian A., Silvescu C. I., Reinhold V. N., Demetriou M., Dennis J. W. *Cell*, **129** (2007), 123–134.

Lee C. C., MacKay J. A., Fréchet J. M. J., Szoka F. C., *Nat. Biotechnol.,* **23** (2005), 1517–1526.

Lee Y. C., Lee R. T., *Acc. Chem. Res*., **28** (1995), 321–327.

Leffler H., Carlsson S., Hedlund M., Qian Y., Poirier, F., *Glycocon. J*., **19** (2002), 433–440.

Lemieux R. U., *Acc. Chem. Res*., **29** (1996), 373–380.

Ley K., Kansas G. S., *Nat. Rev. Immun*., **4** (2004), 1–11.

Lindhorst T. K., *Nachr. Chem. Techn. Lab.,* **44** (1996), 1073–1079.

Lindhorst T. K., *Top. Curr. Chem.,* **218** (2002), 201–235.

Lis H., Sharon N., *Chem. Rev*., **98** (1998), 637–674.

Liu F. T., Rabinovich G. A., *Ann. N. Y. Acad. Sci*., **1183** (2010), 158–82.

Liu Y., Palma A. S., Feizi T., *Biol. Chem.,* **390** (2009), 647–656.

Loris R., Hamelryck T., Bouckaert J., Wyns L., *Biochem. Biophys. Acta*, **1383** (1998), 9–36.

Love J. C., Estroff L. A., Kriebel J. K., Nuzzo R. G., Whitesides G. M., *Chem. Rev*., **105** (2005), 1103–1170.

Lowe J. B., *Cell*, **104** (2001), 809–812.

Lowe J. B., *Curr. Opin. Cell Biol*., **15** (2003), 531–538.

Lundquist J. J., Toone E. J., *Chem. Rev*., **102** (2002), 555–578.

Macauley M. S., Crocker P. R., Paulson J. C., *Nat. Rev. Immunol*., **14** (2014), 653–666.

Mammen M., Choi S.-K., Whitesides G. M., *Angew. Chem*., **110** (1998), 2908–2953; *Angew. Chem. Int. Ed.*, **37** (1998), 2754–2794.

McEver R. P., *Curr. Opin. Cell Biol.,* **14** (2002), 581–587.

Mitchell E. P., Sabin C., Šnajdrová L., Pokorná M., Perret S., Gautier C., Hofr C., Gilboa-Garber N., Koča J., Wimmerová M., Imberty A., *Proteins: Struct. Funct. Bioinformatics,* **58** (2005), 735–746.

Möckl L., Horst A. K, Kolbe K., Lindhorst T. K., Bräuchle C., *ChemBioChem*, **16** (2015), first published online: 20 Aug 2015, DOI: 10.1002/cbic.201500361.

Morris S., Ahmad N., André S., Kaltner H., Gabius H.-J., Brenowitz M., Brewer F., *Glycobiology*, **14** (2004), 293–300.

Mulder A., Huskens J., Reinhoudt D. N., *Org. Biomol. Chem.,* **2** (2004), 3409–3424.

Mulholland E. K., Adegbola R. A., *N. Engl. J. Med*., **352** (2005), 75–77.

Naismith J. H., Field R. A., *J. Biol. Chem*., **271** (1996) 972–976.

Natchiar K. S., Srinivas O., Mitra N., Surolia A., Jayaraman N., Vijayan M., *Acta Cryst*., **D62** (2006), 1413–1421.

Nieminen J., Kuno A., Hirabayashi J., Sato S., *J. Biol. Chem*., **282** (2007), 1374–1383.

Ohlsen K., Oelschlaeger T. A., Hacker J., Khan A. S., *Top. Curr. Chem*., **288** (2009), 109–120.

Ortiz Mellet C., Defaye J., García Fernandez J. M., *Chem. Eur. J*., **8** (2002), 1983–1990.

Osrin D., Vergnano S., Costello A., *Curr. Opin. Infect. Dis.,* **17** (2004), 217–224.

Parker R. B., Kohler J. J., *ACS Chem. Biol.*, **5** (2010), 35–46.

Peréz S., Sarkar A., Breton C., Drouillard S., Rivet A., Imberty A., Glyco3D: A Portal for Structural Glycoscience (2013), http://glyco3d.cermav.cnrs.fr.

Piontek C., Ring P., Harjes O., Heinlein C., Mezzato S., Lombana N., Pöhner C., Püttner M., Varón Silva D., Martin A., Schmid F. X., Unverzagt C., *Angew. Chem*., 121 (2009a), 1968–1973; *Angew. Chem. Int. Ed*., **48** (2009a), 1936–1940.

Piontek C., Varón Silva D., Heinlein C., Pühner C., Mezzato S., Ring P., Martin A., Schmid F. X., Unverzagt C., *Angew. Chem.*, **121** (2009b), 1974–1978; *Angew. Chem. Int. Ed.*, **48** (2009b), 1941–1945.

Rabinovich G. A., Baum L. G., Tinari N., Paganelli R., Natoli C., Liu F. T., Iacobelli S., *Trends Immunol*., **23** (2002), 313–320.

Rabinovich G. A., Toscano M. A., Ilarregui J. M., Rubinstein N., *Glycoconj. J.,* **19** (2004), 565–573.

Rabinovich G. A., Toscano M. A., Jackson S. S., Vasta G. R., *Curr. Opion. Struct. Biol*., **17** (2007), 513–520.

Ramírez-Gualito K., Alonso-Ríos R., Quiroz-García B., Rojas-Aguilar A., Díaz D., Jiménez-Barbero J., Cuevas G., *J. Am. Chem. Soc*., **131** (2009), 18129–18138.

Ravishankar R., Ravindran M., Suguna K., Surolia A., Vijayan M., *Curr*. *Sci.*, **72** (1997), 855–861.

Reynolds M., Fuchs A., Lindhorst T. K., Peréz S., *Mol. Simul.*, **34** (2008), 447–460.

Rillahan C. D., Paulson J. C., *Annu. Rev. Biochem*., **80** (2011), 797–823.

Rini J. M., *Ann. Rev. Biophys. Biomol. Struct.*, **24** (1995), 551–577.

Rinnbauer M., Ernst B., Wagner B., Magnani J., Benie A. J., Peters T., *Glycobiology*, **13** (2003), 435–443.

Röckendorf N., Lindhorst T. K., *Top. Curr. Chem.*, **217** (2001), 201–238.

Röckendorf N., Sperling O., Lindhorst T. K., *Austr. J. Chem*., **55** (2002), 87–93.

Rosen S. D., *Ann. Rev. Immunol*., **22** (2004), 129–156.

Rostovtsev V. V., Green L. G., Fokin V. V., Sharpless K. B., *Angew. Chem.*, **114** (2002), 2708–2711; *Angew. Chem. Int. Ed.*, **41** (2002), 2596–2599.

Roy R., Shiao T. C., *Chem. Soc. Rev.*, **44** (2015), 3924–3941.

Roy R., Zanini D., Meunier S. J., Romanowska A., *J. Chem. Soc. Chem. Commun*., 1993, 1869–1872.

Sadalapure K., Lindhorst T. K., *Angew. Chem.,* **112** (2000), 2066–2069; *Angew. Chem. Int. Ed.,* **39** (2000), 2010–2013.

Sato S., *Trends Glycosci. Glycotechnol*., **14** (2002) 285–301.

Sato S., Nieminen J., *Glycoconj. J.*, **19** (2004), 583–591.

Schierholt A., Hartmann M., Schwekendiek K., Lindhorst T. K., *Eur. J. Org. Chem*., 2010, 3120–3128.

Schmidt-Lassen J., Lindhorst T. K., *MedChemCommun*, **5** (2014) 1218–1226.

Sgouros S. N., Bergele C., *Postgrad. Med. J.,* **82** (2006), 338–342.

Sharon N., Lis H., *J. Agric. Food Chem*., **50** (2002), 6586–6591.

Simons J. P., Davis B. G., Cocinero E. J., Gamblin D. P., Stanca-Kaposta E. C., *Tetrahedron: Asymmetry*, **20** (2009), 718–722.

Somers W. S., Tang J., Shaawa G. D., Camphausen R. T., *Cell*, **103** (2000), 467–479.

Song E.-J., Pohl N. L. B., *Curr. Opin. Chem. Biol*., **13** (2009), 626–632.

Sumner J., *J. Biol. Chem*., **37** (1919), 137–141.

Taylor M. E., Drickamer K., *Curr. Opin. Cell Biol*., **19** (2007), 572–577.

Thomas W. E., Trintchina E., Forero M., Vogel V., Sokurenko E. V., *Cell*, **109** (2002), 913–923.

Timmer M. S. M., Stocker B. L., Seeberger P. H., *Curr. Opin. Chem. Biol*., **22** (2007), 59–65.

Tornøe C. W., Christensen C., Meldal M., *J. Org. Chem*., **67** (2002), 3057–3064.

Touaibia M., Roy R., *Mini Rev. Med. Chem*., **7** (2007), 1270–1283.

Turnbull W. B., Stoddart J. F., *Rev. Mol. Biotechnol*., **90** (2002), 231–255.

van den Hamer C. J. A., Morell A. G., Scheinberg I. H., Hickman J., Ashwell G., *J. Biol. Chem*., **245** (1970), 4397–4402.

van Eerde A., Grahn E. M., Winter H. C., Goldstein I. J., Krengel U., *Glycobiology,* **25** (2015)*,* 492–501.

Varki A., *Nature*, **446** (2007), 1023–1029.

Vijayan M., Chandra N., *Curr. Opin. Struct. Biol*., **9** (1999), 707–714.

Virji M., *Top. Curr. Chem.,* **288** (2009), 1–15.

von der Lieth C.-W., Frank M., Lindhorst T. K., *Rev. Mol. Biotech.*, **90** (2002), 311–337.

Wang S.-K., Liang P.-H., Astronomo R. D., Hsu T.-L., Hsieh S.-L., Burton D. R., Wong C.-H., *PNAS*, **105** (2008), 3690–3695.

Weber T., Chandrasekaran V., Stamer I., Thygesen M. B., Terfort A., Lindhorst T. K., *Angew. Chem*., **126** (2014), 14812–14815; *Angew. Chem. Int. Ed*., **153** (2014), 14583–14586.

Weis W. I., Drickamer K., *Structure*, **15** (1994), 1227–1240.

Weis W. I., Drickamer K., *Annu. Rev. Biochem*., **65** (1996), 441–473.

Wellens A., Garofalo C., Nguyen H., Van Gerven N., Slättegård R., Hernalsteens J. P., Wyns L., Oscarson S, De Greve H, Hultgren S, Bouckaert J., *PLoS One*, **3** (2008), e2040.

Werz D. B., Ranzinger R., Herget S., Adibekian A., von der Lieth C.-W., Seeberger P. H., *ACS Chem. Biol.*, **2** (2007), 685–691

Zhu X., Schmidt R. R., *Angew. Chem*., **121** (2009), 1932–1967; *Angew. Chem. Int. Ed*., **48** (2009), 1900–1934.