Sensing Glycans as Biochemical Messages by Tissue Lectins: The Sugar Code at Work in Vascular Biology

Herbert Kaltner1  Hans-Joachim Gabius1

1Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität München, München, Germany

Address for correspondence Hans-Joachim Gabius, PhD, Institut für Physiologische Chemie, Tierärztliche Fakultät, Ludwig-Maximilians-Universität München, Veterinärstr. 13, D-80539 München, Germany (e-mail: gabius@tiph.vetmed.uni-muenchen.de).

Abstract

Although a plethora of players has already been revealed to be engaged in the haemostatic system, a fundamental consideration of the molecular nature of information coding can give further explorations of the mechanisms of blood clotting, platelet functionality and vascular trafficking direction. By any measures, looking at ranges of occurrence and of potential for structural versatility, at strategic positioning to influence protein and cell sociology as well as at dynamics of processing and restructuring for phenotypic variability, using sugars as an alphabet of life for generating the glycan part of glycoconjugates is a success story. The handiwork by the complex system for glycan biosynthesis renders biochemical messages of exceptionally high coding capacity available. They are read and translated into cellular effects by receptors termed lectins. The different levels of regulation on both sides, that is, glycan and lectin, establish an intriguingly fine-tuned capacity for functional pairing. The emerging insights into the highly branched routes of glycosylation, into lectin structures up to complete characterization in solution and the shape of lectin networks, first obtained for the three selectins, now extended to considering many other C-type lectins, galectins and siglec complexes, as well as into intra- and inter-family cross-talk and cooperations are sure to push boundaries in our understanding of the molecular basis of haemostasis.

Keywords
► C-type lectin
► galectin
► mucin
► selectin
► siglec
► von Willebrand factor

Introduction

The cell surface is the platform for a multitude of recognition processes. They can lead to selective cross-linking events within the membrane (lattice formation) and hereby, for example, trigger outside-in signalling. Moreover, bridging between cells or cells and the extracellular matrix can be facilitated. Mechanistically, the complementarity between surface epitopes and their receptors underlies the specificity of these cis/trans-interactions. Allegorically spoken, a molecular message is ‘read’ and ‘translated’ into a post-binding effect, that is, biochemically coded information is turned into a process and an outcome. Considering the stringent space limitations on the surface and the large size of the pool of signals relevant to cover all aspects of cell sociology, it immediately becomes apparent that the biochemical information coding is required to reach a high density. To do so, the building blocks (letters of an alphabet of life) of the molecular messages should be able to form many structural isomers (words) with as few constituents as possible. How is this chemically feasible?

In principle, ways of making connections between units to build a biopolymer chain are known from the 5′, 3′-phosphodiester of nucleic acids or peptide bonds of proteins. In each case, the same chemistry is applied in a uniform Lego-like manner so that exclusively the sequence matters for information coding. With focus on this aspect, that means that going beyond the spatial order of units linked to each other in exactly the same way will open opportunities to...
increase the informational contents of oligomers. Explicitly, if variations in the positions and topological properties of connecting points (and branching) become regular features of oligomers, then the extent of structural diversity will be greatly enhanced, and, indeed, one class of biomolecules offers to realize this potential: carbohydrates (for common structures, please see Fig. 1; for overviews on their diversity and properties, please see Gabius1). Our review introduces readers to the biochemical basis of the concept of the sugar code and respective recognition systems with relevance for vascular biology.

Known to be present abundantly and ubiquitously in nature, as polysaccharides and as conjugates with lipids and proteins, in these cases positioned strategically as ‘sugary coating of cells’2, glycans have the widespread occurrence profile and surface presentation expected of a basic coding system.3–14 Equally important to note, the positioning of the (glycosidic) bond that brings the anomic centre (marked for each carbohydrate shown in Fig. 1) of an activated sugar donor together with one of the hydroxyl groups of the acceptor during chain elongation has many more options than just one, what is the case in nucleic acids and proteins. Structurally, variability in linkage position and in anomic status of the glycosidic bond (α or β), the possibility for branching and also the introduction of various site-specific substitutions explain why, ‘among all biological molecules, carbohydrates, in a short sequence, can potentially display the largest number of ligand structures’.15

![Chemical structures of sugars](image)

Fig. 1 Main letters of the sugar alphabet present in vertebrate glycans. The anomeric centre, the site of conjugation to an acceptor for glycan chain elongation, is marked by a black dot, and the typical colour symbol is given for each sugar. A change in the position of a single OH group (from equatorial to axial) is sufficient to alter the character of the letter, i.e., by epimer formation (Glc to Gal or to Man), as is its exchange by an N-acetyl group (GlcNAc, GalNAc). Reduction to obtain a deoxy sugar (i.e., for the exocyclic methyl group in the C6 position in L-Fuc) or an introduction of an additional hydroxyl group in the side chain to obtain Neu5Gc from Neu5Ac are further means to increase the pool size of this alphabet of life. The type of anomic position for each compound is given in the monosaccharide’s name.
This assumption will gain impact when this pool size reaches impressive numbers. Calculated for amino acids and for carbohydrates, a set of 20 types of monomers will theoretically build 6.4 × 10⁷ hexapeptides but to as many as 1.44 × 10¹⁵ hexasaccharides. As concisely emphasized by Roseman, this exceptional level of structural diversity has its enormous pros but also imparts a practical con for analytical glycoscience: ‘glycoconjugates are much more complex, variegated and difficult to study than proteins or nucleic acids,’ nonetheless giving ample reason to be ‘intrigued as to whether these sugars might be arranged in specific sequences that function as information molecules in biological processes.’ Their capacity to store information and to transmit it by inter-molecular recognition is the basis of the sugar code.

Pioneering work to give substance to the concept of the sugar code applied enzymatically editing surface glycans of ³²P-labelled cells and following their in vivo routing. Using an enzyme fraction from Clostridium perfringens rich in L-fucosidase activity, ‘alteration of lymphocytes by glycosidases profoundly affects their fate in the body.’ Instead of their typical recirculation to lymph nodes, increased quantity of radioactivity was found in the liver. Interestingly, a similar phenomenon of re-programming of routing was discovered when measuring the kinetics of clearance of a ⁶⁴Cu-labelled glycoprotein (i.e. ceruloplasmin) from serum by the liver: enzymatic removal of terminal sialic acids from the glycan chains of ceruloplasmin, what caused presentation of intact galactose (Gal) residues, drastically shortened the period of asialoglycoprotein’s serum presence. Since the status of desialylation can be an indicator of a glycoprotein’s lifetime like a timer, a functional correlation between the level of Gal presence at terminal position and hepatic uptake of the processed glycoprotein makes sense physiologically. Obviously, the hepatocytes (somehow) ‘read’ the message encoded in Gal-terminated glycans, as removal of fucose (Fuc) moieties erases the signal for lymphocyte homing.

In both instances, a distinct glycan signal (on the cell surface or as part of a glycoprotein) appears to serve as molecular equivalent of a postal code (in lymphocyte homing surface or as part of a glycoprotein) appears to serve as a functional correlation between the level of Gal presence at terminal position and hepatic uptake of the processed glycoprotein makes sense physiologically. Obviously, the hepatocytes (somehow) ‘read’ the message encoded in Gal-terminated glycans, as removal of fucose (Fuc) moieties erases the signal for lymphocyte homing.

In contrast to homopolysaccharides such as chitin or cellulose with their uniform β1,4-linkage of Glc(NAc) residues, glycans of cellular glycoconjugates are the proof that the potential of carbohydrates to build oligomers of enormous diversity is actually exploited. The same substrate, that is, the sugar of its nucleotide derivative, can enter an acceptor at different positions depending on enzyme and acceptor presence. This principle that underlies the origin of glycan complexity is graphically illustrated by an example from the maturation of N-glycans of glycoproteins.

The pentasaccharide core of N-glycans shown in Fig. 2 is the common substrate for processing towards all forms of complex-type N-glycans. Six sites exist for branch initiation of N-acetylgalactosaminyltransferases (GnTs), as listed in Fig. 2. Linkages are in β1,2/4/6, and

![Fig. 2](image-url) The six modes of GlcNAc incorporation into the N-glycan core pentasaccharide by N-acetylgalactosaminyltransferases (GnTs). Their activity in the order of numbers given to each enzyme initiates production of up to penta-antennary complex-type N-glycans. For explanations of symbols for sugars, please see legend to Fig. 1.
each G0T has its specific acceptor, working in the assembly line at a position as indicated by the given numbers. Viewed systematically, large enzyme panels, too, are available for each other type of carbohydrate so that the toolbox of glycosyltransferases is well equipped for making glycan diversity and complexity possible.22–37 The connection of impairing certain enzymes of glycan biosynthesis to causing pathological phenotypes of mice deficient in the responsible genes underscores their relevance for cellular physiology.38 In these studies, important clues may even be missed, because compensatory mechanisms have been tracked down that maintain certain glycome features despite the engineered deficiency.39,40 The dynamic nature of availability of substrate, acceptor and enzyme accounts for the broad spectrum of the glycome and its spatiotemporal diversity.41–47

Overall, carbohydrates have the chemical features to make the envisioned high-density coding possible, and an enzymatic machinery has been developed to realize this potential. Variations in anomeric position and in linkage points enable the glycome to be composed of a huge number of structures—at the same time posing demanding challenges to structural glycan characterization. Progress in glycan analysis48–50 and synthesis51 has laid the foundation for aiming at delineating structure–function relationships. As important as the presence of the hydroxyl groups (in axial equatorial positions) is for structural variability, it then is for making directional hydrogen bonding possible, and there is more: in contrast to most peptides, glycans are not highly flexible. Within their conformational space, they often access few low-energy ‘valleys’ (conformers), a boon in the thermodynamic balance sheet when binding to a receptor.52 Hereby, the entropic penalty is rather low, favouring binding. In aggregate, the hypothesis for a flow of information from glycans as ligands (counter-receptors) via a recognition process (‘reading’) by receptors and a translation of the glycan-encoded information into biological effects by this functional pairing is attractive and testable.

Historically, an antibody-to-antigen-like association has been discussed approximately 70 years ago.53,54 Worded prophetically, ‘rather than trying to force all biological specificity into the immunological compartment, we might have to consider the latter as merely a special case of the more universal biological principle, namely, molecular key-lock configuration as a mechanism of selectivity’,54 here drawing on Fischer’s famous lock-and-key analogy derived from his work with the glycosidases ‘Invertin und Emulsin’.55 Within this concept, the two key questions that we must answer are as to whether (antibody-like) receptors for glycans exist and, if positive, whether their presence shows the necessary degree of diversification for accomplishing the assumed task of decoding of a large array of glycan-encoded signals.

The Sugar Code: Lectins as Readers/Translators

A simple and robust assay was crucial to detect lectin activity, that is, haemagglutination. The bridging of erythrocytes by at least bivalent proteins that leads to visible cell aggregation was the experimental read-out for the discovery of (haem) agglutinins, first in rattlesnake venom (1860), and then in plant extracts (1888). Mitchell put one drop of venom on a slide, ‘and a drop of blood from pigeon’s wounded wing (was) allowed to fall upon it. They were instantly mixed. Within three minutes the mass had coagulated firmly’,56 the landmark discovery of receptors bridging cells by ligand binding (an involvement of plasma compounds in the ‘coagulation’ was rigorously excluded ~40 years later,57 and the active protein biochemically characterized more than 100 years later).58

Giving these agglutinins further medical relevance, members of this family of effectors were described that could equal the ability of serum antibodies in distinguishing the blood group status of human donors (for historical survey, please see Refs.46,47 and 39–61). That constituents of plant extracts could thus ‘read’ surface determinants of erythrocytes as the blood group–specific antibodies do, which Landsteiner described as cause for fatal incompatibilities in blood transfusion,62,63 which inspired Boyd to shape a generic name for these activities: ‘it would appear to be a matter of semantics as to whether a substance not produced in response to an antigen should be called an antibody even though it is a protein and combines specifically with a certain antigen only. It might be better to have a different word for the substances and the present writer would like to propose the word lectin from Latin lectus, the past principle of legere meaning to pick, choose or select’.64

Instead of using a term like antibody-like substance with implicit structural comparison, the emphasis was wisely placed on a feature of ligand binding, ‘intending thus to call attention to their specificity without begging the question as to their nature’, while certainly being intrigued by the observations that ‘lectins imitate the behaviour of immune antibodies’.65 When asked (in 1973) how he would define a lectin, Boyd replied: ‘that although he had invented the word he had no right to dictate what it should mean. He said that once a word goes into general circulation it becomes independent of its originator and eventually becomes whatever people think it ought to mean! He added that he would like to use the word to mean a protein that had a more or less specific action and that there is no reason to think it is an antibody’.66 Over time, lectin became the most common name for agglutinating activities, although the synonym (phyto)haemagglutinin is still in use, for example, for two bean (Phaseolus vulgaris) lectins (PHA-E/L; their mixture holds a special place in lectinology, because the induction of mitogenesis in human leukocytes by PHA proved elicitor capacity of lectin binding for the first time67) and for a lectin of the influenza viral surface. Thus, lectins qualify as receptors, and investigating their specificity to cellular targets unveiled glycans as counter-receptors.

That their specificity is directed to glycans can historically be traced to reporting a ‘de-agglutination’ of erythrocyte aggregates that had formed in the presence of ricin or abrin preparations by using hog gastric mucin, a potent lectin binder.68,69 Blocking haemagglutination in the presence of blood group–specific lectins by a sugar, first Fuc inhibiting the activity of the lectin of eel serum to cross-link type O(H)
erythrocytes, proved carbohydrate binding beyond doubt, and this approach was systematically followed in transfusion medicine and in lectinology in general. Terminologically, carbohydrate binding is the first criterion for a lectin, a feature yet shared by other classes of proteins such as immunoglobulins (Igs), as the mentioned antibodies attest, and glycosyltransferases. Thus, besides carbohydrate-binding antibodies, lectins need to be strictly separated from any enzyme using a carbohydrate as substrate, from sensors/transporters for free mono- and oligosaccharides and carbohydrate-binding modules of bacterial and fungal glycoside hydrolases.

To have the implied broad impact on many aspects of cellular physiology, the range of glycan-lectin recognition must not be narrow. Implicitly, this means that lectins must not be a rare commodity in the proteome. The implied functional pairing is assumed to account for a co-evolution towards diversity. Indeed, matching diversity on the level of glycans, more than a dozen protein folds have developed capacity to interact with glycans (for three examples, please see – Fig. 3; complete illustration of the folds of the families of animal/human lectins is given in Fujimoto et al along with a detailed listing and description of methods to analyse complex formation of a lectin with its ligand in that paper’s table 1).

This is similarly seen on the level of glycogenes. They code for glycosyltransferases, glycosidases, glycan-binding enzymes such as sulfotransferases (please see below) and enzymes for carbohydrate biosynthesis such as for the two sialic acids shown in the bottom part of – Fig. 1 and for the ensuing generation of the activated carbohydrates as well as for transporters of activated sugars. Diversification of the gene for each ancestral lectin (or glycogene product) has led to families of genes. In the case of lectins, they code for homologous proteins with differences in fine specificity (due to the sequence alterations within the carbohydrate recognition domain [CRD]) and in the architecture in terms of modular display. Intriguingly, a CRD can be embedded into a complex structural context, with often not yet resolved physiologic relevance.

In fact, the CRD can be part of a multi-modular protein, and individual regions can functionally cooperate, for example, to form aggregates (such as galectin-3; please see below) or to guide a lectin to the extracellular space, where it then mediates ordered cell migration. When thoroughly examining occurrence of proteins with any of the three folds illustrated in – Fig. 3 as examples, it turned out to be a common theme within phylogensis that lectin diversity is truly attained commonly on the level of families. This apparent co-evolution with glycans resulting in large lectin numbers strongly argues in favour of the assumed role of lectins as translators for the large panels of glycan-encoded messages. This intra-family divergence is described in further detail for C-type lectins (please see reviews87–89), for sialic acid-binding Ig superfamally lectins (siglecs) (please see90,91 and for ga (lactose-binding)lectins. Notably, in all three cases, species differ in qualitative aspects of gene display and also with respect to gene number and organization, for example, for mammalian galectins92,93 so that extrapolations from animal models to the clinical situation should always be performed with adequate caution.

Having identified this wide range of carbohydrate-binding motifs and detected a large number of endogenous lectins, the two fundamental questions posed at the end of the previous chapter are convincingly answered with a ‘Yes’. The current challenge therefore is to proceed from completing the phase of lectin discovery to entering the era of building a functionally meaningful puzzle from the individual proteins. First, a lectin must be assigned to its binding partners and pairing is very selective. Next, respective studies are revealing that members of lectin families appear to be expressed in networks. This raises the possibility for functional crosstalk (agonism or cooperation) in cis/trans-interactions and in sensing non-self-signals. Obviously, the concept of lectins expressed in networks deserves to become a paradigm. The multiplicity of lectin representation is, for example, in certain cell types or tissues, attested by discovering a large group of myeloid C-type lectins94 and by revealing overlapping roles of C-type lectins in anti-microbial and anti-fungal immunity92,93.

In Fig. 3 illustration of three examples of basic folds of mammalian lectins with bound ligand, that is, C-type lectin E-selectin (A; Protein Data Bank [PDB] code: 1G1T), siglec-1 (sialoadhesin) (B; PDB code: 1QFO) and galectin-1 (C; PDB code: 1GZW).
First detected by immunohistochemistry with non-cross-reactive antibody preparations, galectins analysed up to the level of full-scale monitoring of the complete set of family members are expressed in distribution profiles with individual features and overlaps.\(^\text{94-98}\) These results, of course, have immediate relevance for the strategy how to analyse lectin activities, that is, in mixtures. Initial comparative study of galectin activities alone and in mixtures has been performed by testing surface-tailored glycodendrimersomes in bridging assays\(^\text{99,100}\) and by using cell-based functional assays.\(^\text{101-103}\) The obtained evidence for functional interplay encourages to study in more detail what will happen when lectins are tested alone and in combinations simulating in vivo conditions, and this by using members of the same family and of different families.

In summary, presence of sophisticated machineries for lipid and protein glycosylation, producing the ligand side, and for glycan recognition, the complementary side, clearly supports the concept of an information transfer by glycan–protein interactions (sugar code) and gives cracking the sugar code direction.\(^\text{104}\) The aim is to define the meaning of a glycan in its cellular context. Following this concept, distinct carbohydrate epitopes of glycan chains of cellular glycoconjugates can be expected to engage in functional pairing with certain tissue (endogenous) lectins. Any structural processing (for example, stepwise chain elongation) should then change the glycan’s ligand properties, that is, its biochemical meaning. By looking exemplarily at main products of N- and mucin-type O-glycan biosynthesis, the hypothesis that a series of functional pairings occurs along the biosynthetic pathways is next put to the test, first for branch ends of complex-type N-glycans. At the same time, the following description of routes of glycan processing itself documents that many branch points exist to let glycan diversity reach an amazing level.

**Functional Pairing: N-Glycans**

N-glycosylation is a co-translational process that is started at the signal-peptide-dependent entry of a nascent protein into the endoplasmic reticulum (ER). The Glc₃Man₉GlcNAc₂ oligosaccharide is transferred from its dolichol pyrophosphate donor to the N-atom of an asparagine (Asn) moiety within a protein’s N-glycosylation (sequon) motif [i.e. Asn-Xaa (no Pro)-Ser/Thr-Xaa (no Pro)].\(^\text{29}\) During the early stage of the pathway of N-glycan processing and maturation, initial removal of two to three Glc moieties makes a routing signal with this glycan accessible. It is a molecular signal for transportal function of the respective glycoprotein from the ER to the Golgi. Delivery is performed by a cargo transporter (Mr: 53,000) present in the ER-Golgi intermediate compartment (ERGIC), called ERGIC-53.\(^\text{105,106}\) Impairment of its binding to partially or fully deglucosylated N-glycans of target glycoproteins by mutations in the ERGIC-53 gene or the gene of its luminal interaction partner, that is, the multiple coagulation factor deficiency protein 2, is cause of an autosomal recessive bleeding disorder characterized by combined reduction of blood levels of coagulation factors V and VIII.\(^\text{107-109}\) This case study of combined factors V and VIII deficiency underlines the emerging relevance of glycan/lectin recognition for haemostasis.

Since the ER and the route from the ER to the Golgi are equipped with quality controls by lectins to sort out misfolded glycoproteins, glycans have more than one meaning already at this stage.\(^\text{110-112}\) Following the glycoprotein further along the Golgi route, each N-glycan can be trimmed to the pentasaccharide core as shown in \(-\text{Fig. 2}\) so that, for example, complex-type structures arise from the ensuing maturation. In the Golgi, as shown in \(-\text{Fig. 2}\), six GnTs are capable to add a GlcNAc residue at specific sites of the core pentasaccharide to initiate their synthesis. Incomplete maturation or restructuring by stepwise degradation of mature N-glycan branches are the means to let GlcNAc become presented for protein binding at branch ends.

Relevant for a route of platelet clearance, GlcNAc-terminating N-glycans of the GPIbα sub-unit of the von Willebrand factor (VWF) complex associate with the α₄β₃-integrin of phagocytic cells (Mac-1, CD11b/CD18, CR3; Mac-1 was defined as granulocyte- and monocyte-specific antigen by a monoclonal antibody [clones M1/70] raised against mouse spleen cells\(^\text{113}\); for details on the Mac-2 antigen [clones M3/31 and M3/38], please see below).\(^\text{114,115}\) Terminal βGlcNAc in N-glycans (α-anomeric linkage is known from mucins)\(^\text{116}\) is not only a ligand for an integrin but also for the C-type lectins langerin (CD207),\(^\text{117,118}\) the liver and lymph node sinusoidal endothelial cell lectin (LSECtin)\(^\text{119}\) and its close relative human dandritic cell-specific ICAM-3-grabbing nonintegrin-related protein (DCSIGN, CD299)\(^\text{120}\) or its murine homologue SIGN-R2.\(^\text{121}\) The first epitope along the illustrated route of N-glycan processing therefore fulfils the expectation for functional pairing.

Elongation of the N-glycan chain by adding a Gal moiety [in β1,4-linkage, as shown in the centre of \(-\text{Fig. 4}\)] drastically changes the ligand characteristics, underscoring the required specificity. As noted above, accessible Gal units are docking sites for hepatic clearance, and this also applies to coagulation factor VIII and to VWF to or tissue plasminogen activator.\(^\text{122}\) This type of interaction between desialylated glycoproteins on the surface of (senile) platelets and the hepatic asialoglycoprotein receptor is more than a clearance process. In fact, it is the long-elusive physiological ligand–receptor pair regulating hepatic thrombopoietin messenger ribonucleic acid production.\(^\text{124}\) Since the circulatory lifespan of platelets is shortened during systemic infection due to the activity of bacterial sialidases on cell surface glycans, the lectin-mediated clearance is a means of haemostatic adaptation limiting the severity of disseminated intravascular coagulation during sepsis.\(^\text{125}\) Physiologically, one means to mask this determinant (postal code) for hepatic uptake is by α₂,3-sialylation (\(-\text{Fig. 4}\), top, left). Fittingly, engineered deficiency in a respective glycosyltransferase, that is, α₂,3-sialyltransferase-IV, accelerated clearance of these glycoproteins and thereby (likely) accounts for prolonged bleeding and coagulation times in the knockout (KO) mice.\(^\text{126}\) In contrast, the alternative to α₂,3-sialylation, that is, α₂,6-sialylation of N-acetyllactosamine (LacNAc) shown in \(-\text{Fig. 4}\), can be tolerated by the hepatic C-type lectin, as shown with neo-glycoproteins and natural glycoproteins.\(^\text{127-129}\) Considering the frequent occurrence of LacNAc at the terminal position of...
N-glycans and its processing by these two different routes of sialylation shown in Fig. 4 (top part), it is no surprise that, in addition to the C-type lectin fold, other types of lectins can accommodate these epitopes, especially the galectin and siglec folds shown in Fig. 3.

LacNAc is the canonical ligand for galectins. Its α2,3-sialylation can enhance galectin affinity depending on the protein, most prominently for galectin-8. Of note, this galectin binds VWF and coagulation factor V, the latter then imported into megakaryocytes by endocytosis, and also to the platelet integrin αIIbβ3, a new entry to the list of platelet activators. Illustrating the marked impact of the site of sialic acid conjugation on ligand features, α2,6-sialylation switches off affinity to galectins, which bind LacNAc via hydrogen bonds to the axial 4'-OH and the exocyclic 6'-OH groups of the Gal moiety (please see Fig. 1, second row for illustration of these positions, relative to the 3'-OH group). In the case of α2,6-sialylation of LacNAc shown in Fig. 4 (top, right), the contact point at C6 is therefore occupied by a sialic acid residue. What is detrimental to galectin binding, though, is essential for association to siglec. Presence of such a residue in α2,3-linkage is suited for siglec-4 (myelin-associated glycoprotein); when in α2,6-linkage, it is the docking site for siglec-2 (CD22). In sum, α2,6-sialylation of LacNAc extension thus abolishes affinity to galectins, can be tolerated by a C-type lectin and is the primary contact to a siglec. By the way, sialic acids can hereby implement recognition sites on glycoproteins for a siglec so that a contribution to the regulation of their plasma levels is possible, as discussed for VWF and coagulation factor VIII by binding to siglec-5 on macrophages.

The two other illustrated products of terminal N-glycan tailoring shown in Fig. 4 (bottom), that is, the histo-blood group A determinant ([s]Lea, CD15(s)) or (sialyl) Lewis^x (CD209) a receptor of epitopes of the Lewis blood group system such as Le^a or Le^x and the selectins bind the shown sLex, an interaction of pivotal significance for the interaction of leukocytes and platelets with the vasculature under conditions of activation (please see below). Turning Le^x not into sLe^a but alternatively into Le^y (CD174) by terminal α1,2-fucosylation instead of the α2,3-sialylation shown in Fig. 4 renders the respective N-glycan capable to interact with thrombomodulin's C-type lectin-like domain, implicated in inhibiting angiogenesis.
The three selectins mentioned above are classical examples of lectins with modular architecture. Their design is shown in Fig. 5. The terminal C-type CRD, as is the case for the V-set IgG domain of siglec rods, is optimally presented on the membrane to bridge cells. First detected by function blocking or localization using monoclonal antibodies (Mel-14) for the lymphocyte homing receptor L-selectin (CD62L), H18/7 and H4/18 for the chemokine- and endotoxin-inducible endothelial-leukocyte adhesion molecule-1 (ELAM-1; E-selectin, CD62E) and S12 for an α-granule membrane protein of M, 140,000 (GMP-140) that is redistibuted to the plasma membrane upon platelet stimulation, what gave GMP-140 a new name: platelet activation-dependent granule to external membrane protein (PADGEM; P-selectin, CD62P). The spatial accessibility of the glycan counter-receptors at branch ends, rapid on-rates during contact formation by mostly ionic interactions and catch bonding are factors that underlie their role as selective cell adhesion molecules (explaining the origin of the term selectin), operating as anchors for cells in the blood flow. Like the mentioned enzymatic removal of Fuc from the lymphocyte surface raised evidence for the concept of glycan as postal codes in routing and delivery, the importance of a second sugar type present in sLeα/β epitopes, that is, sialic acids, has been revealed by treatment of sections of lymphoid organs with bacterial sialidase. These reports converge to support the involvement of glycans in lymphocyte homing, as the blocking experiments with selectin-binding monoclonal antibodies did for the protein side. Intriguingly, α-1-fucosidase is suggested to limit leukocyte migration at late stages of inflammation (tested in murine experimental autoimmune uveoretinitis) upon induction by chemokines CCL3/5 by reducing P-selectin binding. Besides presentation by N-glycans, the sLeα/β epitopes are characteristically also a part of certain types of mucin-type O-glycans, here reaching a high density that favours to build and maintain selectin–glycoprotein contacts. In addition to the negative charge of the sialic acid, sulphate groups at C6 of Gal and/or GlcNAc moieties increase the ligand’s capacity for rapid contact building via ionic charge complementarity under conditions of flow. Suflation is also conducive to ‘write’ a routing signal for glycoprotein clearance, if GlcNAc is not conjugated with Gal but with N-acetylgalactosamine (GalNAc), a further mode of N-glycan tailoring.

Before turning to mucin-type O-glycans in more detail, a second salient lesson should be drawn from Fig. 4, besides the documented sets of structure–ligand relationships; a glycan epitope (such as LacNAc), that is, a certain acceptor, can be converted into more than one product. This opens opportunities for highly versatile regulation of glycome representation by modulating enzyme and/or substrate availability. Exemplifying the intimate connection of a distinct biomedically relevant process with glycosylation, the malignancy-associated high-level sialylation can be interpreted, among other implications, as a protection against growth inhibition. Along this line, the tumour suppressor p16INK4a has been shown to counteract malignancy at this level. This protein is essential to induce anoikis in human pancreatic carcinoma (Capan-1) cells. It works via orchestrated down-regulation of α2,6-sialylation (by reducing sialic acid biosynthesis) teamed up with up-regulation of homobivalent galectin-1, which is the functional receptor ‘reading’ the increase in terminal LacNAc and cross-linking the similarly up-regulated glycoprotein counter-receptor αβ1-integrin via this recognition to trigger caspase-8 activation.

The signal (access to LacNAc), the ‘reader’ (galectin-1) and the downstream effector (αβ1-integrin) are thus co-regulated towards anoikis induction, giving the glycophenotypic change a functional meaning. Of note, a similar team building to attain T cell apoptosis has been reported for α2,6-sialyltransferase, CD45 and galectin-1. The respective delineation of regulatory pathways for generating selectin counter-receptors has proven to be a hallmark of supporting the concept of functionally dynamic glycomics in lymphocyte homing and platelet aggregation. On the level of O-glycans, counter-receptor occurrence can be as tightly controlled in a coordinated manner by competition between different routes and terminal tailoring, following the same principles as shown in Fig. 4 for N-glycans. Routes of mucin-type O-glycan processing after the initial incorporation of a GalNAc residue into the protein are illustrated in Fig. 6. The presence of the α-linked GalNAc residue

![Fig. 5 Modular organization of the three selectins characterized by the C-type lectin domain (centre; please see also Fig. 3A) at the most prominent position for intercellular contact, followed by an epidermal growth factor (EGF)-like domain and two to nine complement-binding consensus repeats.](image-url)
Fig. 6  Illustration of six different routes of biosynthetic elongation of the first product of mucin-type O-glycosylation, i.e., Tn antigen (centre). α2,6-Sialylation of the GalNAc residue results in the sialyl (s)Tn determinant that cannot undergo any further processing. The activity of the respective sialyltransferases underlies occurrence of sTn at the expense of all other O-glycans. Extension of the Tn antigen towards branched structures can alternatively proceed along two routes that establish either the core 1 or the core 3 structures. Enzymatic β1,3-galactosylation produces the core 1 disaccharide (T antigen) that can be sialylated stepwise in up to two positions, first in α2,3-linkage at the Gal moiety to yield the sialyl (s)T antigen, finally in α2,6-linkage at the GalNAc residue to produce disialylated T (top part, left). Alternatively, the T antigen can be subject to core 2 trisaccharide generation by adding GlcNAc at the central GalNAc residue in β1,6-linkage (top part, right). As shown in ►Fig. 2 and in ►Fig. 4 for N-glycans, this GlcNAc residue is the starting point of branch elongation that can encompass LacNAc repeats and the Le/Le determinant. This trisaccharide therefore is the platform for obtaining various types of oligosaccharides, depending on the actual availability of enzymes and substrates for the competing routes of processing. Mechanistically, the same holds true for the case of transition of the Tn antigen to the core 3 disaccharide and then core 4 structures (bottom part, right), in close analogy to core 1 and core 2 synthesis. Correspondingly, without branch introduction, the core 3 disaccharide can be extended in two steps to a sialylated tetrasaccharide (bottom part, centre), in analogy to core 1 oligosaccharide processing. Occurrence of α2,3-sialylated LacNAc in fully processed core 1/3 structures resembles that of a complex-type N-glycan branch ends (please see ►Fig. 4, top part, left). When considering that the processing of a Tn antigen can take diverse routes, it becomes obvious that the system of mucin-type O-glycosylation has manifold possibilities to generate a large number of O-glycans by regulating acceptor, enzyme and/or substrate levels. For explanation of symbols for sugars, please see legend to ►Fig. 1.
defines the Tn antigen (n for nouvelle; CD175), the target of serum auto-antibodies en route to spontaneous polyagglutinability of erythrocytes in the cold causing the Tn syndrome.161-163 This determinant is not only a docking site for antibodies but also for human lectins.

Functional Pairing: Mucin-Type O-Glycans

This type of protein glycosylation is characterized by reaching a high local density, ideal for a defensive barrier and also for high-affinity receptor binding.164-167 The committed step for mucin-type O-glycosylation is performed by members of the family of polypeptide N-acetylgalactosaminyltransferases (GalNAcTs), most of the 20 enzymes harbouring a β-trefoil lectin domain that binds GalNAc in addition to the enzymatically active protein part.32,33,168 They cooperate to let GalNAc incorporation reach diverse levels of local density in their substrates, and more than 80% of the proteins passing through the Golgi are predicted to be processed this way. Interestingly, each enzyme can well fulfill a particular functional role, as revealed by KO mouse analysis. Despite integrity of all other enzymes, a deficiency in one of the N-acetylgalactosaminyltransferases, that is, GalNAcT-1, caused a moderate to severe bleeding disorder due to reduced blood levels of coagulation factors and decreased level of recruitment of leukocytes to sites of inflammation, highlighting O-glycan relevance to haemostasis and vascular biology.169 Binding studies of Tn-presenting glycopeptides demonstrated that O-linked GalNAc is a ligand for endogenous lectins. The macrophage Gal-(binding C)-type lectin (MGL, CD301, CLEC10A) and galectin-4 are tissue receptors of the Tn antigen.170,171 Present in Kupffer cells that also express the mentioned α2β1-integrin for β-GlcNAc-dependent platelet clearance, a C-type lectin (CLEC4F) is responsible for efficient uptake of Tn-presenting mucins and of platelets of genetically engineered mice made defective in O-glycan maturation so that this process is entirely arrested at the stage of O-GalNAc presentation.172,173 Physiologically, this epitope is obliterated for this process by α2,6-sialylation. The resulting disaccharide sialyl Tn (STn; CD175s) shown in -Fig. 6 maintains affinity for human MGL,174 as is the case for α2,6-sialylated LacNAc at N-glycan termini and the closely related hepatic C-type lectin (please see above). At the same time, the sialylation step enables a gain-of-function, because it confers binding to siglec-15 on macrophages.175,176 A V-set module resembling that of siglec in a paired Ig-like receptor, that is, PILRα, also binds the STn antigen.177,178

Sialylation of the Tn α2,3-epitope is not the only way to use it as substrate. As shown in -Fig. 6, an alternative pathway, competing with the synthesis of the STn α2,6-antigen, is the generation of the T(F) disaccharide (CD176); first described as antigen by O. Thomsen and his assistant V. Friedenreich, thus termed Thomsen–Friedenreich antigen179), the core 1 disaccharide of mucin-type O-glycans. Dense clustering favours binding of galectins-3 and -4.180,181 Its sialylation in α2,3-linkage renders binding to galectin-8182 and siglec, especially sialoadhesin (siglec-1)182 and siglec-9,183 possible, disialylation (at the Gal and GalNAc moieties) to siglec-4.184 In the framework of the system of O-glycosylation, the α2,3-sialylation of the core 1 disaccharide precludes its branching to yield core 2 glycans, as core 1 disaccharide synthesis makes core 4 production impossible (-Fig. 6). Of fundamental importance, the decisions made by committing steps at the various stages highlight the broadness of the glycan panel and the enormous adaptability of the glycan profile, strongly suggesting their influence on regulating cellular activities. Core 2 branching and sLeα production, as shown in -Fig. 6, are steps to build selectin counter-receptors such as sLea noted above. Interestingly, the Golgi-resident sialyltransferase-IV-dependent final step of sLeα synthesis has also been revealed to be important for chemokine (CCL3 and CCL4) binding to the CCR5 receptor185 and for CXCR2-triggered firm leukocyte arrest after CXCL1/CXCL8 injection in cremaster muscle venules.186,187

Looking at regulation of galectin binding to mucin-type O-glycans more closely, an analogy can be drawn to the concerted actions described above for N-glycan α2,6-sialylation and galectin-1. Decisions between routes of synthesis explain how growth-regulatory processes are switched on or off.181,188,189 The role as a metastasis suppressor of the glycosyltransferase N-acetylgalactosaminide α2,6-sialyltransferase-2 (ST6GalNAcT-2) in breast cancer had been attributed to reduced T(F) disaccharide presence that lowers breast tumour cell aggregation and retention at metastatic sites via galectin-3.190 Effectively, this type of α2,6-sialylation diverts product generation from reaching, for example, core 1 to STn, as shown in -Fig. 6 and already mentioned above. Convergent effects on extent of presence of this disaccharide by altering its status of α2,6-sialylation and core 2 branching, too, favour lung cancer formation.191

Next, like selectins, galectins can home in on branched O-glycans, selectins to the sLeα/x termini, especially galectins-1 and -3 to the internal LacNAc repeats of the β1,6-branch. Susceptibility of activated T cells,192 lymphoma cells193 and prostate (LNCaP) carcinoma cells194 to apoptosis induction by galectin-1 and impairment of natural killer cell activation by the binding of galectin-3 to tumour cells195 are examples for the significance of galectin pairing with this part of the β1,6-linked branch of extended core 2 O-glycans shown in -Fig. 6. Branch extension (and terminal tailoring by fucosylation, sialylation and sulphation to obtain sulphated sLeα/x epitopes) is thus crucial to convey the signal for lectin binding.

In aggregate, any modulation in enzyme, substrate and acceptor availability can shift the relative proportions of the products, as can be deduced from -Figs. 4 and 6 (for gangliosides as counter-receptors of tissue lectins, please see Ref. 196; for glycan–lectin interactions, please see Ref. 197). Considering occurrence of competition for an acceptor at different sites of the illustrated pathways of N- and mucin-type O-glycosylation, the extent of impact is not fixed but regulatable. As a consequence, cellular responsiveness to tissue lectins is intimately tunable, giving a functional dimension to the noted complexity of the glycome. Clearly, the potential for regulatory events will be broadened, if the lectin side is also a platform for fine-tuning, at the levels of presence and of structural design. The mentioned diversification of an ancestral gene is a main route towards this aim, and -Fig. 5 documents changes in the length of the stalk...
among the three selectins. Molecular diversity can also be achieved (1) by sequence changes in the CRD affecting the profile of ligands among members of a family and (2) by alterations of the modular design. Hereby, emergence of functional antagonism due to differences in modular design is possible. For this case, galectins-1 and -3 serve as role model, as already mentioned above.\textsuperscript{101,198} This emerging concept of functional crosstalk is a driving force to characterize in detail the lectins’ structures, and this work raises awareness to address a fundamental question: what is the significance of modular architecture? Whereas lectins with trans-membrane sections such as selectins (please see \textsuperscript{\ref{fig:selectins}}) are ideal for cell-cell bridging, the situation is much less clear for lectins without such a module.

**Lectin Diversity: The Example of Galectins**

Vertebrate adhesion/growth-regulatory galectins are organized in three types of modular architecture (\textsuperscript{\ref{fig:modules}}). The manifestation of a strict selection process towards the three illustrated forms serves as instructive case for further work to answer the given question. As common approach of structural analysis, crystallography has been brought to its limits for the galectins.

Whereas the prototype (homodimeric) galectins and most CRDs obtained by engineering could readily be analysed by crystallography, often in complex with ligands and even within a broad range of pH values,\textsuperscript{133,199} intra-molecular dynamics of the tandem-repeat- and chimera-type proteins is a likely reason that precluded obtaining crystals for the full-length proteins of these two groups. Recently, partial truncation of the N-terminal segment of human galectin-3 by engineering facilitated to take structural analysis of crystals beyond the CRD.\textsuperscript{200,201} Like the three selectins, this galectin had first been detected by monoclonal antibodies (clones M3/31 and M3/38),\textsuperscript{202} immunohistochemically present in macrophages, dendritic cells and in epithelium, thus termed the Mac-2 antigen.\textsuperscript{203–205} Its ligand spectrum covers glycans and proteins,\textsuperscript{206} its trimeric design shown in \textsuperscript{\ref{fig:modules}} enables aggregation to oligomers via contacts between the N-terminal tails and/or the CRDs in the presence of multivalent ligand\textsuperscript{207–210}; please see Flores-Ibarra et al\textsuperscript{201} for review of the literature). This manifestation of a strict selection process towards the three forms shown in \textsuperscript{\ref{fig:modules}} is not only a suited test model, but also inspires routes to redesign nature to clarify structure–activity relationships.

Looking at an involvement of galectin-3 in thrombosis, together with its counter-receptor galectin-3-binding protein (Mac-2 BP/90K; an adhesion mediator interacting with β₁-integrin sub-units, fibronectin, nidogen and collagens IV–VI),\textsuperscript{211–213} it has been suggested to play a critical role in venous thrombosis.\textsuperscript{214} The appearance of Mac-2 BP/90K as contaminant in preparations of recombinant coagulation factor IX was therefore judged to be cause of ‘unforeseen consequences’.\textsuperscript{215} Its pro-inflammatory signalling function is not only active in osteoarthritis pathogenesis,\textsuperscript{102} but, for example, also locally in unstable plaque regions of carotid endarterectomy\textsuperscript{216} or rheumatoid synovium.\textsuperscript{217} The molecular and topological nature of the active galectin-counter-receptor lattice for outside-in signalling remains to be defined. As the role of the recently discovered appearance of galectin-3’ CRD in heterodimers with prototype CRDs\textsuperscript{218} warrants elucidation.

When comparing the design of the homo-/hetero-bivalent family members, the structural difference between the non-covalent homodimers and the heterodimers with linker peptides between the two different CRDs is clear. However, the relationship from structure (e.g. type of CRD or length of linker) to function is again largely unexplored. Here, the combination of rational protein engineering, for example, turning galectin-1 into a covalently associated variant connected by the linker of a tandem-repeat-type galectin,\textsuperscript{219} with activity assays, for example, platelet activation with galectin-8 as positive control, offers an attractive approach

**Fig. 7** Illustration of the three types of modular architecture of vertebrate galectins, i.e., non-covalent association of two identical carbohydrate recognition domains (CRDs) to a homodimer (prototype), covalent association (by a linker of distinct length; it differs between galectins, and alternative splicing brings about length variation in certain cases, e.g. galectin-8 with two lengths at 33 or 75 amino acids) of two different CRDs (tandem-repeat type) and the trimodular product of N-terminal association of a peptide with two sites for serine phosphorylation, a repeat region of non-triple-helical collagen-like peptides (nine in human galectin-3) and the canonical CRD, termed chimera type, with listing of the respective numbers/acronyms in the galectin nomenclature in each group (right side).
for thorough analysis, with the perspective to develop innovative antagonists or agonists on the platform of human galectins and beyond.

**Conclusion**

In retrospect, there has been ‘the widely held belief that carbohydrates are dull compounds, and that they serve only as structural or protective materials (e.g. cellulose in plants and chitin in insects) and as an energy source (glycogen in animals), but lack any biological specificity. The possibility that living organisms form a myriad of compounds in which carbohydrate is covalently linked to protein, with the carbohydrate having manifold functions, was ignored by most chemists and biochemists alike.220’ Ironically, when it was realized, the exceptionally demanding character of the task to handle analysis of glycans, called complex carbohydrates for a long time, required enormous efforts to begin reaching firm conclusions on structure–activity relations.

Our introductory focus on the amazing structural talents of carbohydrates to accomplish high-density coding has laid the foundation to understand the paradigmatic change in the way glycans are viewed. The convergence of the great strides made in glycan analysis with discovering a matching diversity of glycans are viewed. The concept of their functional pairing with endogenous receptors is becoming a gateway towards understanding the workings of this aspect of the sugar code. The case study of the lectins has already taught the amazing lesson that platelet activation and inflammatory stimuli set genetic reprogramming and transport processes in motion to let functional pairing happen at the right place and at the right time, and selectins as representatives of C-type lectins are not alone.

The range of tissue lectins revealed to be involved in the mentioned processes is steadily growing, underscoring the nature of glycans as biochemical messages, as noted for inflammation.222–227 The co-evolution of messages (glycans) and readers (lectins) as well as their apparently intimately orchestrated expression justify the term ‘third alphabet of life’ for carbohydrates shown in – Fig. 1.

With these central points made, it is reasonable to advise that the possibility of functional activity of the glycans should be considered, when dealing with a glycoprotein. Looking, for example, at the glycoprotein VWF, C-type lectins, a siglec and galectins are capable to home in on its glycan(s) for overview of VWF binders, please see Ref. 229), and new aspects in the field of lectins are being connected to haemostasis, for example, the lectin pathway of complement activation in crosstalk to coagulation.230,231 Among the open challenges thus are a delineation of counter-receptors for lectins as well as of lectin networking within and between families, hereby deservedly putting members of families beyond selectins firmly on the map of receptors acting in concert in haemostasis and thrombosis.

**Conflict of Interest**

None declared.

**Acknowledgements**

We gratefully acknowledge inspiring discussions with Drs. B. Friday and A. Leddoz.

**References**


Marathe DD, Chandrasekaran EV, Lau JT, Matta KL, Neelamegham S. Systems-level studies of glycosyltransferase gene expression and enzyme activity that are associated with the selectin binding function of human leukocytes. FASEB J 2008;22(12):4154–4167


Dam TK, Gerken TA, Brewer CF. Thermodynamics of multivalent carbohydrate–lectin cross-linking interactions: importance of entropy in the bind and jump mechanism. Biochemistry 2008;48(18):3822–3827


Lei Q, Lu G, Qi J, et al. PIRRx and PIRRY have a siglec fold and provide the basis of binding to sialic acid. Proc Natl Acad Sci U S A 2014;111(22):8221–8226


Blostein M, Cuerquis J, Galipeau J. Galectin-3-binding protein is a potential contaminant of recombinantly produced factor IX. Haemophilia 2007;13(06):701–706


Damjanov I. Lectin cytochemistry and histochemistry. Lab Invest 1987;57(01):5–20

Spicer SS, Schulte BA. Detection and differentiation of glycoconjugs in various cell types by lectin histochemistry. Basic Appl Histochem 1988;32(03):307–320


