Feature Review

From lectin structure to functional glycomics: principles of the sugar code

Hans-Joachim Gabius1, Sabine André1, Jesús Jiménez-Barbero2, Antonio Romero2 and Dolores Solís3,4

1 Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80339 München, Germany
2 Chemical and Physical Biology, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maelzu 9, 28040 Madrid, Spain
3 Instituto de Química Física Rocasolano, CSIC, Serrano 119, 28006 Madrid, Spain
4 Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Bunyola, Mallorca, Illes Balears, Spain

Lectins are carbohydrate-binding proteins which lack enzymatic activity on their ligand and are distinct from antibodies and free mono- and oligosaccharide sensor/transport proteins. Emerging insights into the functional dimension of lectin binding to cellular glycans have strongly contributed to the shaping of the ‘sugar code’. Fittingly, over a dozen folds and a broad spectrum of binding site architecture, ranging from shallow grooves to deep pockets, have developed sugar-binding capacity. A central question is how the exquisite target specificity of endogenous lectins for certain cellular glycans can be explained. In this regard, affinity regulation is first systematically dissected into six levels. Experimentally, the strategic combination of methods to monitor distinct aspects of the lectin–glycan interplay offers a promising perspective to answer this question.

Glycosylation: abundant and frequent
Post-translational modifications broaden the range of functionality of many proteins. Unraveling structural aspects and implications of such a conjugate formation naturally depends on the complexity of the attached compounds. Phosphorylation represents a relatively simple case; however, in terms of degree of structural complexity, glycosylation tops this list. Starting from a single sugar unit, glycan chains in glycoproteins (and also glycolipids) can be enormously variegated in terms of branching and length so that the glycome (a generic term such as genome and proteome) covers a wide range of structures. Thus, it is no surprise that glycome mapping (glycomics) and ensuing functional analyses represent rapidly growing areas of research. Our review outlines the current paradigm of functional glycomics and introduces the crucial impact of topological parameters in recognition processes with glycan receptors (lectins; see Glossary). We begin by documenting the widespread occurrence of glycosylation, referring to classical morphological observations.

Microscopic study of eukaryotic cells (as well as bacteria) has revealed a common structure described as ‘sugary coating of cells’, in the form of basement membranes or cell walls, termed ‘glycocalyx’ [1]. Initially, this so-called polysaccharide coating was suggested to primarily protect cells and act as a filtration/selective-binding device, thereby shaping the microenvironment above the plasma membrane. The animal glycocalyx is composed of a complex network of glycoconjugates with extensive glycosylation, both in frequency and chain size, including the presence of glycosaminoglycans with their characteristic profile of anionic charge distribution [2]. With the growing awareness that ‘the significance of glycosyl residues’ is to impart a discrete recognition role to the protein’ [3], the first, mostly structural view has gradually given way to consider glycans as bioactive signals with information-coding ability hitherto assigned exclusively to proteins. As a consequence, and reflecting the microscopic observations on the biochemical level, conjugation of sugars to proteins occurs throughout the entire phylogenetic spectrum, with a

---

Glossary

Agglutinins: proteins with the capacity to agglutinate cells, an activity akin to blood group-specific antibodies but biochemically separate from the immunoglobulins. If agglutination is inhibited by sugars, the term is synonymous with lectin.

Carbohydrate-binding modules (CBMs): non-catalytic domains linked to the catalytic part of glycoside hydrolases, mostly in fungi and bacteria, and involved in degradation of cell wall and storage polysaccharides. CBMs are classified in currently up to 54 families. Based on their ligand specificity they are divided into type A (binding insoluble, highly crystalline cellulose or chitin to a flat surface rich in aromatic amino acids), type B (binding oligosaccharides as part of a polymer by stacking via aromatic amino acids, up to contact by both sides as in a sandwich, and directional hydrogen bonds), and type C CBMs (binding mono- to trisaccharides, with increase of hydrogen bonding relative to type B).

Lectins: members of a superfamily of (glyco)proteins with the capacity to bind carbohydrates, and distinguished from antibodies, enzymes that alter the structure of the bound glycan, and sensor/transport proteins for free mono- to oligosaccharides. If active in cell agglutination assays, synonymous terms such as agglutinin, hemagglutinin (if red blood cells are agglutinated) or phytohemagglutinin (if the lectin originates from plant material) are used. W.C. Boyd originally coined the term ‘lectin’ in 1954 to underscore that blood group-specific proteins from plants are entities separate from common antibodies.

Phytohemagglutinins: (glyco)proteins from plants agglutinating red blood cells, often used as designation for the two isolecitins from Phaseolus vulgaris. Also a generic name for plant lectins.

Sensor/transport proteins: proteins that bind free mono- to oligosaccharides and act as high-affinity sensors, for example for oligosaccharides in defense reactions (plants) or chemotaxis, and as vehicles for active transport. Periplasmic receptors of Gram-negative bacteria were originally classified as type 1 carbohydrate-binding proteins, set apart from the type 2 comprising antibodies, CBMs and lectins [13].
known total of 13 monosaccharides and 8 amino acids forming at least 41 types of glycosidic linkages [4,5]. Further pointing to physiological importance, a group of 250–500 genes is devoted to the synthesis and remodeling of glycan chains (glycogenes). The frequent occurrence of N-glycosylation, whose well-known entry site is the Asn-X-Ser/Thr sequon (X: all amino acids except proline) [6], in extracellular proteins was first predicted by database searches and later confirmed by high-accuracy mass spectrometric mapping [7,8]. Proteins from the extracellular side of the cell membrane and from along the route of trafficking from the endoplasmic reticulum (ER) to the cell membrane or to lysosomes bear N-glycans mostly in loops, turns and β-sheets; these glycoproteins constitute more than 10% of the murine proteome [8]. The intriguing complexity among the N-glycans (over 100 different structures have been identified in the nematode Caenorhabditis elegans [5]) and the high level of sophistication of the enzymatic machinery for this and other types of glycosylation (e.g. 20 isoenzymes are known to initiate human mucin-type O-glycosylation [9]) intimate more than passive effects of protein glycosylation (e.g. on solubility). In fact, these investments suggest the existence of a coding system on the basis of carbohydrates, embodied by the term ‘sugar code’. Indeed, carbohydrates ‘are ideal for generating compact units with explicit informational properties’ [3]. This judicious statement rests on the unsurpassed coding capacity reached by carbohydrates, by virtue of their special chemical properties. Which features turn carbohydrates into ‘ideal’ hardware for information coding?

**Carbohydrates: the sweet side of biorecognition**

When directly compared with oligonucleotides and peptides in terms of coding capacity, the number of ‘words’ (oligosaccharide isomers) built from a set of ‘letters’ (monosaccharides) is several orders of magnitude larger [theoretical number of all possible hexamers: 4096 (oligonucleotides), $6.4 \times 10^6$ (peptides) and $1.44 \times 10^{15}$ (saccharides)] [10]. Possibilities for two anomic configurations ($\alpha/\beta$), bond formation via different linkage positions (1–1, 2, 3, 4, 6 for hexopyranose), change in ring size (pyranose/furanose) as well as introduction of branching and additional site-specific substitutions such as acetylation, phosphorylation or sulfation underlie this exceptional structural diversity. Isomers with different ‘meaning’ can thus be established in ‘compact units’. Their hydroxyl groups make the sugars suited for directional acceptor/donor hydrogen (H) bonds (or coordination of Ca$^{2+}$ ions). At the same time, the set of C–H bonds is suited for van der Waals interactions and stacking, on the grounds of their inherent polarization even enabling C–H/π-interactions, for example with Trp or Tyr (including the perturbation of solvent structure in their vicinity with enthalpic/entropic consequences [11]). Because the main carbohydrate constituents of glycans differ only in the relative positioning of one or two hydroxyl groups, epimerization must have substantial consequences for biorecognition (Figure 1).

The illustrated structural differences between β-galactose and α-mannose alter the topological displays of hydroxyl groups and hydrophobic patches in a coordinated manner, shaping a characteristic profile of potential contacts [12]. These contact signatures can be recognized by particular amino acid side chains. Establishment of cooperative H-bonding, simultaneously engaging donor/acceptor properties of a hydroxyl group, and bidentate interactions with planar polar amino acid side chains intuitively position Arg, Asp and Glu, along with the aromatic amino acids for stacking, as expected interaction partners. Indeed, analysis of the amino acid distribution within receptor domains verified this assumption [12–15]. Overall, the occurrence of topological complementarity at the protein level will determine whether a match is feasible for either the axial 4-hydroxyl group (galactose; Figure 1a) or its equatorial presentation in mannose/glucose (Figure 1b,c) [16]. A natural alternative to amino acid side chains to sense the spatial distribution of hydroxyl groups is a strategically presented Ca$^{2+}$ ion (Figure 1c; for
overview on Ca\textsuperscript{2+} functionality in lectin activity, see Table 1 and [17].

The assumption of a far-reaching biological significance of carbohydrate-protein interactions would receive strong support if this common theme of contact complementarity is found not just in one or a few, but in diverse types of topologies for receptor-ligand complexes. Indeed, all modes of spatial ligand accommodation that can be envisioned have developed in evolution, starting from flat surfaces and shallow grooves and progressing to a firm grip in a deep pocket. Initially, classification in type I/type 2 proteins followed this structural parameter [13].

### Binding the carbohydrate: from shallow grooves to deep pockets

In general, the common principles of recognition outlined above are operative with different degrees of engagement in all types of carbohydrate-binding proteins. To support teaching this salient lesson, examples of contact profiles are presented in the order of increasing degrees of H-bonding along the route to reach high affinity for monosaccharides (Figure 2). The first two cases document the epimer selection for a human lectin (galectin-1) and the leguminous lectin concanavalin A (Figures 1 and 2a,b). In the first illustration, the probing for the axial 4-hydroxyl and 4-equatorial (mannose) or equatorial/axial (galactose) hydroxyls (Figure 2c,d). This prominent role of the ion together with a particular folding of the carbohydrate-recognition domain (Figure 3a) led to this lectin class being referred to as C-type lectins, with a remarkable diversification in phylogenesis and even development of cases that lack a direct Ca\textsuperscript{2+}-carbohydrate contact [18,19]. Stacking of an aromatic ring (tyrosine) against a fitting sugar (mannose, N-acetylgalactosamine, fucose) can additionally occur in this recognition mode, thus contributing approximately 25% of total free binding energy, and van der Waals forces help to distinguish between a carbohydrate letter and its N-acetylated derivative (the chemical equivalent of an ‘Umlaut’). Such a contact between the methyl part of the acetamido group and Cc-1/Cc-2 atoms of His202, complementing the H-bonding (Figure 2d), is tied to the 60-fold higher affinity of the rat C-type lectin for N-acetylgalactosamine than for galactose [20]. Remarkably, these two sugars with identical hydroxyl group positioning can be differentiated to an even far greater extent: joining van der Waals attraction forces to two amino acid side chains with two carbonyl oxygen-mediated H bonds to the N-acetyl group brings the relative affinity difference to up to ~1000-fold in an invertebrate C-type lectin [21].

High-affinity monosaccharide binding can also be accomplished by an intricate H-bonding network (Figure 2e). Every hydroxyl group of N-acetylgalactosamine participates in binding to generate an affinity level at the upper end for carbohydrate-specific antibodies [22]. By contrast, in carbohydrate-binding modules (CBMs) of glycosyl hydrolases, intensifying contacts take place via a shift from H bonding to stacking; this binding mode dominates in CBMs targeting insoluble, highly crystalline cellulose and/or chitin, typically presenting a flat platform rich in aromatic amino acids for the entropically favored complex formation (type A) [23–25]. Although also found in certain lectins such as the cation-dependent mannose-6-phosphate receptor (the presence of the 6-phosphate group enhances affinity approx. 10 000-fold) [26], the formation of tight complexes is the hallmark of sugar transporters such as the L-arabinose-binding protein [13,27]. The network of H bonding in its complex with galactose (Figure 2f) is supplemented by stacking of the B-face of the sugar to Trp16, resulting in a $K_d$ value (galactose) in the below μM range and a nearly complete (approx. 98%) engulfment. Similar

### Table 1. The strategic roles of Ca\textsuperscript{2+} in lectin activity\textsuperscript{a}

<table>
<thead>
<tr>
<th>Function</th>
<th>Lectin type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural role in stabilizing the lectin domain or organizing the site for ligand binding (no direct contact to ligand), oligomerization of subunits</td>
<td>Leguminous lectins homologous to concanavalin A, lectin chaperones involved in quality control (calnexin, calreticulin), animal lectin-type cargo receptors (i.e. ERGIC53\textsuperscript{b} and VIP36\textsuperscript{d}), Anguilla anguilla agglutinin, discoidin I</td>
</tr>
<tr>
<td>Structural role and direct contact to anionic group(s) of ligand or neutralization of repulsive forces between anionic charges in ligand and lectin</td>
<td>Pentraxins, laminin G-like module, annexin A2, cation-dependent mannose-6-phosphate receptor</td>
</tr>
<tr>
<td>Direct contact to neutral group(s) of ligand with/without structural role</td>
<td>Most C-type lectins, Cucumaria echinata lectin III (β-trefoil fold), Pseudomonas aeruginosa lectin I (two coordination bonds) and lectin II (four coordination bonds)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Adapted from [17], with permission.

\textsuperscript{b}ER-Golgi intermediate compartment protein (lectin) ($M_w = 53$ kDa).

\textsuperscript{c}Vesicular-integral (membrane) protein (lectin) ($M_w = 36$ kDa).
to most lectins and the CBMs interacting with soluble ligands (types B/C), ligand loading is an enthalpically driven process with some inherent entropic penalty [12,23,27]. As a consequence of the intimate contact in the cleft between the two globular domains of the periplasmic monosaccharide transporter, the presence of the sugar triggers a substantial compaction of the protein (from 21.22 Å to 20.28 Å) by cleft closure [28]. Ligand-induced changes with a similar extent of compaction measured by small angle neutron scattering or fluorescence correlation spectroscopy (Figure 4b) [29,30] or with alteration of quaternary structure [31] and reorganization of the pocket [26,32] have also been recorded for animal and human lectins. In contrast to the transporters of free saccharides, lectins deal with ligands with anomeric extensions, that is particular sections of glycan chains in cellular glycoconjugates, precluding an engulfment. That at least 14 different types of folds are present in lectins [33], with phylogenetic diversification in each group (see [18,34,35] for respective information on C-type lectins and galectins), attests that emergence of lectin sites is not a rare event restricted to a particular fold.

The lectin families, each identified by its structural traits (Figure 3), are capable of realizing the coding
potential of cellular glycans. The information they convey is translated by endogenous lectins into firm/transient adhesion or triggers signaling in cells via diverse pathways, for example leading to cell cycle arrest or to the induction of anoikis/apoptosis [33,36,37]. However, given the enormously large number of cellular glycoconjugates, it is unclear how the effectors (lectins) achieve specificity for their cognate counter-receptors (glycans). To delineate the details of the high-precision recognition process, it will be necessary to define the structures of both interaction partners separately, then to work with carbohydrate epitopes, glycoconjugates and cells, also measuring kinetic and mechanical properties of binding and complex dissociation. In other words, a strategic combination of different approaches and integration of the results should provide the clues needed to uncover the origin of specificity. To give a graphic example for binding preferences, the homodimeric human lectin galectin-1 (Figure 2a, Figure 3d) homes in on ganglioside GM1 of neuroblastoma cells (SK-N-MC) [38], whereas the fibronectin receptor α5β1-integrin and tissue plasminogen activator are among the cell type-specific (glycoprotein) counter-receptors in other cases (reviewed in [33]). As programmatically outlined above, respective analyses have been carried out for this adhesion/growth-regulatory lectin (Figure 4a–e) [29,30,39–47] (Box 1). The way the results are connected guides the reader from work with the protein (Figure 4a,b) to monitoring different aspects of the interaction with glycans (Figure 4c,d), up to the level of cells and organs (Figure 4e). Hence, the current challenge is to give the
Box 1. Experimental approaches to determine lectin structure and complex formation

| Atomic force microscopy | measurement of binding strength under force for surface-presented lectin/ligand pairs indicates lectin potential to engage in transient or firm contacts. |
| Biodistribution | determination of clearance from serum and organ uptake of radiolabeled lectin after i.v. injection. |
| Chemical mapping | analysis of inhibitory potency of ligand derivatives (mostly deoxy, fluoro or O-methyl) in a binding assay yields information on hydrogen bonding contribution and steric aspects at each position. |
| Circular dichroism | monitoring yields insights into secondary structure elements; ellipticity changes at distinct wavelengths reflect involvement of aromatic amino acid residues in ligand binding. |
| Fluorescence-activated cell scanning (cytofluorimetry) | binding of labeled lectin to cell surfaces (measuring mean fluorescence intensity and percentage of positive cells) characterizes the glycophenotype; when performed with a panel of lectins (Table 3), this is termed glycophenotyping. |
| Fluorescence correlation spectroscopy | monitoring of translational diffusion of a fluorescent protein in solution indicates shape properties and detects ligand-induced changes. |
| Gel filtration | elution profile yields information on hydrodynamic behavior as an indication of quaternary structure and respective influence of the ligand. |
| Hemagglutination | erythrocyte agglutination to determine inhibitory potential by saccharides (classical assay for lectin activity). Note: the glycomic profile on erythrocytes varies. |
| Isothermal titration calorimetry | measurement of heat released or absorbed by ligand binding enables determination of association constant (Kd), stoichiometry (n) and enthalpy of binding (∆H°), thereby facilitating calculation of entropy (∆S°). |
| Lectin cyto- and histochemistry | localization of lectin-reactive sites in cell preparations and frozen/fixed tissue sections. |
| Saturation transfer difference NMR spectroscopy | magnetization transfer (spin diffusion) from, in this example, saturated protein to ligand protons maps spatial vicinity (below 5 Å). |
| Small angle neutron/X-ray scattering | spectrum yields information on quaternary structure and shape and detects ligand-induced changes. |
| Surface plasmon resonance | measurement of changes in refractive index near a planar surface presenting lectin/ligand in resonance units (RU); 1000 RU equals 1 ng of mass per mm² leads to equilibrium and kinetic constants. |
| Transferred nuclear Overhauser effect spectroscopy | signal generation by through-space dipolar interaction between two ligand protons, preferably separated by a glycosidic bond, helps to define bound-state topology. |
| Ultracentrifugation | relative protein sedimentation reflects average molecular mass (equilibrium analysis) and hydrodynamic parameters (velocity analysis). |
| X-ray crystallography | diffraction pattern yields electron density map; crystallization might require unphysiological conditions, packing can cause artifacts and will preclude the separation of static from dynamic disorder. |

To give this research a clear direction, we dissect affinity regulation into six levels. After starting on level one, in which mono- or disaccharides are distinguished in the previously described manner, moving on takes us to a series of five additional levels (Table 2).

The six levels of affinity regulation: from mono- to oligosaccharides

Given the relatively low affinity of lectins with shallow grooves for mono- or disaccharides (Table 2, level 1), a spatial extension of the binding site, referred to as subsite multivalency to signal its combinatorial nature [14], is required. The contact site for the carbohydrate unit, which is the haptenic inhibitor in an activity assay (classically hemagglutination; Figure 4c), will thus be embedded in a more complex sugar contact region (Table 2, level 2). This principle of modularity efficiently operates at the level of small lectins. The 43-amino acid protein hevein, a receptor for the polysaccharide chitin and its building blocks (reviewed in [50]) and the first lectin structurally defined in complex with its ligand in solution by NMR spectroscopy [51], provides compelling evidence at a minimum of size. Binding is enthalpically favored (determined by titration calorimetry; Figure 4c) and, when testing oligomers of N-acetylglucosamine with increasing chain length, the ∆G-values increase from −3.8 kcal/mol for the dimer to up to −7.8 kcal/mol for the pentamer [52]. The translation of NMR-derived information, such as the chemical shift perturbation by ligand presence, into structural models revealed a preorganized binding area with five sections for accommodating the different parts of the linear oligomers (Figure 5a). Remarkably, the molecular fit does not distort their low-energy conformations. A dynamic interchange

Table 2. Six levels of regulation of affinity of glycan binding to a lectin

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mono- and disaccharides (including anomeric position and substitutions)</td>
</tr>
<tr>
<td>2</td>
<td>Oligosaccharides (including branching and substitutions)</td>
</tr>
<tr>
<td>3</td>
<td>Spatial parameters of oligosaccharides</td>
</tr>
<tr>
<td>a</td>
<td>Shape of oligosaccharide (differential conformer selection)</td>
</tr>
<tr>
<td>b</td>
<td>Conformational flexibility differences between isomers</td>
</tr>
<tr>
<td>4</td>
<td>Spatial parameters of glycans in natural glycoconjugates</td>
</tr>
<tr>
<td>a</td>
<td>Shape of glycan chain (examples: modulation of conformation by substitutions not acting as lectin ligand, such as core fucosylation or introduction of bisecting GlcNAc in N-glycans, influence of protein part)</td>
</tr>
<tr>
<td>b</td>
<td>Cluster effect with bi- to pentaantennary N-glycans or branched O-glycans (including modulation by substitutions, please see a)</td>
</tr>
<tr>
<td>5</td>
<td>Cluster effect with different, but neighboring, glycan chains on the same glycoprotein (e.g. in mucins) or a glycoprotein-glycolipid complex (e.g. integrin-ganglioside complexes)</td>
</tr>
<tr>
<td>6</td>
<td>Cluster effect with different glycoconjugates on the cell surface in spatial vicinity forming microdomains</td>
</tr>
</tbody>
</table>

*Adapted from [33], with permission.
(gliding) between subsites and sugar residues in the polymer chitin is thought to contribute to the binding properties for natural ligands and the overall defense function [52] (this factor of local density in a polyvalent ligand will be discussed below). Polymer binding with a strong contribution of aromatic residues is likewise operative in CBMs homing in on crystalline or soluble polysaccharides [23]. In chitinase-like proteins such as the human cartilage glycoprotein-39 [53], a series of up to six subsites ensures tight binding of chitin fragments. One of its pyranose rings then adopts the boat conformation, a distortion not seen for hevein but known from bacterial chitinases [52,53]. The next example for an extended binding site fulfills the expectation that the rapid binding kinetics to facilitate initial contact between endothelium and leukocytes in the blood flow during inflammation [37] stems from electrostatic recognition.

The binding of the tetrascarhide sialyl Lewis\(^s\) to the C-type lectin P-selectin is mostly governed by a network of such interactions including the Ca\(^{2+}\) ion (Figure 5b). Spatial complementarity between a low-energy conformation of the sugar and the preorganized binding site of the selectin can thus be reached without a delay owing to lengthy structural adaptations, engendering rapid binding kinetics. Fucose as well as galactose/sialic acid account for contacts (Figure 5b). The two sulfated tyrosine moieties (Tys7 and Tys10) of the P-selectin glycoprotein ligand-1, too, are intimately engaged in binding in the crystal, via protein–peptide interactions, and help to achieve the very fast on-rate (>10\(^{9}\) M\(^{-1}\) s\(^{-1}\)), broadening the concept of subsite multivalency to non-carbohydrate determinants (Figure 5c) [54]. There is more: the selectins, C-type lectins with trimerodular design [33,37], are special in that an external force below a critical threshold can lengthen the lifetime of ligand association (catch-bond principle, facilitating rolling under shear stress en route to firm attachment) [55,56]. Such sensitivity appears to arise from the relative orientation of two domains (C-type lectin and epidermal growth factor-like modules) and/or the extended binding site [56]. Although mono- and disaccharides naturally can inhibit lectin activity in many cases, this feature of a much more extended contact area for the sugar is common in lectin families, including galectins.

**The six levels of affinity regulation: shape and dynamics**

In the case of galectin-1 binding to the branched pentasaccharide of ganglioside GM1 (Figure 6), mentioned above as an example for cell type-dependent specificity [38,57], the association process probably encompasses more than a galactose unit or lactose. Teaming up NMR techniques with knowledge-based molecular modeling (Figure 4d), the molecular rendezvous between the lectin and the branched pentasaccharide of the ganglioside was structurally defined in solution [58]. Both branches are in contact with galectin-1 (Figure 6a,b). Because the α2,3-sialylgalactose linkage can adopt three low-energy conformers in solution (Figure 4d), it came as no surprise that one of these energetically favored conformers of the glycans (‘key’) appears to fit well into the binding site ‘lock’ [58]. In other words, the lectin selects one of the three constellations of this linkage at the branch point. Technically, the combination of NMR with other spectroscopic techniques (Figure 4b) not only pinpoints conformer selection, but can also detect even small ligand-induced structural changes in solution [40]. In considering this process of selecting one from several shapes (the third dimension of the sugar code), the question arises whether bioactivity will always reside in a certain conformer.

This question is answered by the comparison of this bound-state conformation to the way the carbohydrate headgroup of the ganglioside is positioned in the binding site of cholera toxin, a member of the bacterial AB\(_5\) toxin family [59]. In this group, the five B-subunits are built as a five-stranded β-barrel flanked by an α-helix and associate to form a homopentameric ring; all five contact sites for glycans face cell surfaces to allow this ring surpass homodimeric galectin-1 in affinity to neuroblastoma cells by more than one order of magnitude [38]. Presenting the two conformations of the pentasaccharide when bound side-by-side unveils a conspicuous case of differential conformer selection (Figure 6b). Ganglioside GM1-specific receptors can thus be further classified based on conformer reactivity. The reasoning, when discussing the biological role of glycoproteins, that ‘presumably the specificity would reside in the tertiary conformation of the carbohydrate unit’ [60] has turned out to be truly prophetic. Equally important, these results identify a special property of oligosaccharides, that is their potential for limited flexibility. It is enormously favorable for acting as ligand, in stark contrast to highly flexible peptides, hereby avoiding entropic penalties. Taken together, an oligosaccharide (Table 2, level 2) can thus adopt more than one low-energy conformation (Table 2, level 3a). In other words, ‘the carbohydrate moves in solution through a bunch of shapes each of which can be selected by a receptor’ [61], with obvious perspectives for the design of bioinspired conformationally arrested lectin blockers. The combined application of chemical mapping with deoxy and fluoro sugar derivatives, also a valuable tool to delineate differences between homologous proteins such as galectins (Figure 4c) [62], and docking procedures, which are especially useful when retaining the full flexibility of lectin and ligand [63], will guide the selection of functional groups for the required intramolecular bridging or the introduction of synthetic substitutions to enhance affinity [64,65].

This pertinent principle of conformer selection begins to be operative at the disaccharide level. Proton pairs tied together by through-space magnetization transfer act as sensitive reporter units for the presence of conformers [12]. Because each of the three low-energy conformers of a non-hydrolyzable lactose derivative, that is C-lactose, is characterized by such an exclusive proton–proton contact [45], a diagnostic means is at hand for conformer identification in the bound state. The NMR analysis of receptor–ligand complexes (Figure 4d) could thus exploit this parameter reliably separating conformers to disclose selection of each conformer by a fitting receptor (galectin-1, a plant toxin and bacterial β-galactosidase) [45]. On the thermodynamic balance sheet, as stated above, this process incurs much less entropic debt than arresting a fully flexible ligand. As an added value, the different shapes are bioactive. Consequently, substantial shifts in conformational equilibria...
Figure 4. (Continued).
The page contains a diagram and text discussing lectin-ligand interaction, specifically focusing on galectin-1. The text explains various analytical aspects, topological aspects, and cellular aspects of the interaction. The diagram illustrates different techniques used for analysis, such as hemagglutination, fluorescence titration, and surface-based binding assays. The text mentions the detection of ligand-dependent changes in characteristic signals by circular dichroism and fluorescence titration. Additionally, it discusses the structural aspects of galectin-1, including its jelly-roll topology and the binding site, as well as its hydrodynamic shape and extended by small angle neutron/X-ray scattering (SANS/SAXS). The page also includes figures showing the results of these analyses, including a bar graph displaying the percentage of injected dose bound to tissue and a plot showing the force in pN as a function of loading rate. The text concludes with a discussion on the strategic integration of experimental techniques for the analysis of lectin structure and lectin-ligand interaction.
provide the opportunity for significant affinity regulation (e.g. in clustered glycan displays). Offering an alternative towards the same end, conformational characteristics are modulated by positioning certain isomers at specific sites (Table 2, level 3b).

Isomer synthesis is a chemical means to increase the range of code word generation by a given set of carbohydrate letters. The structural changes that occur by altering linkage positions can have a considerable impact on flexibility. A prominent case is sialylation at the spatially accessible termini of N-glycans, which non-randomly engages either the 3- or 6-hydroxyl groups of galactose (Figure 7a,b) [6]. The participation of the exocyclic C5–C6 bond of galactose in the three-linkage system confers increased flexibility to the α,2,6-connection, and the meaning of the two isomers, that is α,2,3/α,2,6, is thus drastically changed, as apparent from their differential reactivities with bacterial toxins and viral adhesins [66,67]. Next, α,2,6-sialylation blocks the 6-hydroxy group, which is crucial as a contact site (e.g. for galectins) (Figure 1a, Figure 2a) [68]. The presence of the substitution thus abrogates galectin reactivity, the second aspect of the importance of the sialylation status for physiological processes. These functional implications explain the interest for its determination. Glycophenotyping of cell surfaces by cytofluorometry affords corresponding data on sialylation (Figure 4e). Experimentally, plant and, recently, mammalian lectins, combined with sialidase treatment, are sensitive tools of choice (Table 3; [69]). This method will also identify shifts in the sialylation status upon a specific genetic/molecular change. By example, reconstituting the expression of the tumor suppressor p16INK4a (in pancreatic carcinoma cells) or of genes lost owing to microsatellite instability (in colon carcinoma cells) strongly affects the extent of surface sialylation [48,70]. In addition to plant agglutinins, human lectins such as galectins are capable of detecting such changes, with notable implications for functionality. Because p16INK4a also regulates the cell surface presence of galectin-1, the stage is set for suppressor-induced induction of anoikis, via the orchestrated increase of ligand–receptor complexes active as elicitors of downstream signaling [48]. Does α,2,6-sialylation then exclusively mask glycans?

Sialic acid bioactivity in this linkage type rests on a sufficient number of contacts to its functional groups in the highly flexible α,2,6-sialylgalactose disaccharide without disrupting the conformational dynamics around the three connecting bonds [66,71]. α,2,3,6-Sialylation is thus not simply a minor structural disparity; instead, it conveys non-synonymous bioinformation at branch ends [66]. The terminal position can thus be labeled by two distinct molecular signals with the same sugar letters. This principle is also exploited by turning D-glucuronic acid into the 5-epimer L-iduronic acid in glycosaminoglycans. Owing to this process, a molecular hinge for shape adaptations is installed, made possible by the access of the epimer to the skew-boat conformation [2]. The array of such molecular ‘badges’ further includes the products of introduction of sulfate groups, into N-acetylgalactosamine/glucosamine in certain N,O-glycan termini (di-N-acetylated lactose in pituitary glycoprotein hormones is a marker for serum clearance by the β-trefoil domain of the hepatic endothelial cell based binding assays, three technical approaches are illustrated: (iv) differential inhibition of galectin binding to N-glycans of a glycoprotein (asialofibrin) coated onto the plastic surface of microtiter plate wells by methyl β-lactoside (I) as well as its 3-deoxy derivatives in the galactose (C) or in the glucose (A) units, respectively [42]; (v) a sensorgram of a surface plasmon resonance (SPR) analysis with avian galectin-1 in solution [43]; (vi) a plot of an atomic force microscopy (AFM) analysis showing linear dependence of the rupture force with the natural logarithm of the loading rate [44]; (d) Hydroxyl groups of the ligand will displace bound water molecules in the process of ligand accommodation (see also Figure 2a). Vicinity of lactose to the Trp residue in the crystal (i) is in full accordance with data of CD spectroscopy and fluorescence titrations. Information on the bound-state conformation of the ligand in solution is obtained by (ii) transferred nuclear Overhauser effect spectroscopy (trNOESY) detecting the H1Gal-H4Glc contact diagnostic for the low-energy syn conformation (see energy contour map; [45]); (iii) saturation transfer difference (STD) NMR spectroscopy mapping ligand contact after selection of one low-energy conformer of α,2,3-sialylactose (structure and energy contour map given) by picking up signals from distinct carbohydrate protons after saturation of protein resonances in the aromatic region (top: one-dimensional spectrum of ligand prior to saturation for assignment of signals to protons; bottom: STD spectrum after saturation revealing carbohydrate proton signals). Computational ligand docking then rationalizes the measured contact profile. On the level of the protein, SANS and FCS disclosed compaction of human galectin-1 by ligand binding [29,30]. (e) (i) Fluorescence-activated cell scanning (FACSscan) reveals binding of human galectin-1 to T leukemic Jurkat cells (black line: no lactose = 100% value; green line: presence of 0.1 M lactose; gray area: no lectin added = 0% value; percentage of positive cells and mean fluorescence intensity given). (ii) To determine cellular localization, fixed cells, here FaDu cells, a hypopharyngeal squamous cell carcinoma line, display nuclear galectin-1 reactivity (46). (iii) In vivo analyses can be utilized to measure biodistribution after i.v. injection [47].

<table>
<thead>
<tr>
<th>Table 3. Panel of lectins applied for glycophenotypinga</th>
<th>Latin name (common name)</th>
<th>Abbreviation</th>
<th>Monosaccharide specificity</th>
<th>Potent oligosaccharide</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maackia amurensis</strong> (leukoagglutinin)</td>
<td>MAA I</td>
<td>b</td>
<td>Neu5Acα3Galβ4GlcNAc/Glc, 3'-sulfation tolerated</td>
<td></td>
</tr>
<tr>
<td><strong>Sambucus nigra</strong> (elderberry)</td>
<td>SNA</td>
<td>Gal/Gal/NAc</td>
<td>Neu5Acα6Gal/Gal/NAc, clustered Tα-carbohydrates, 9'-O-acetylation tolerated</td>
<td></td>
</tr>
<tr>
<td><strong>Siglec-2</strong></td>
<td>CD22</td>
<td>Gal</td>
<td>Neu5Acα6Galβ3(4)GlcNAc, 9'-O-acetylation blocks binding</td>
<td></td>
</tr>
<tr>
<td><strong>Arachis hypogaea</strong> (peanut)</td>
<td>PNA</td>
<td>Gal</td>
<td>Galβ3GalNAc/β</td>
<td></td>
</tr>
<tr>
<td><strong>Artocarpus integroflora</strong> (jack fruit)</td>
<td>Jacalin (JAC)</td>
<td>Gal/Gal/NAc</td>
<td>Galβ3GalNAcβ, tolerates sialylation of Tα-carbohydrates</td>
<td></td>
</tr>
<tr>
<td><strong>Phaseolus vulgaris</strong> (erythroagglutinin, kidney bean)</td>
<td>PHA-E</td>
<td>b</td>
<td>Biased complex-type N-glycans: Galβ4GlcNAcβ2Manβ6GlcNAcβ2Manβ3GlcNAcβ2Manβ4Glc</td>
<td></td>
</tr>
<tr>
<td><strong>Pisum sativum</strong> (garden pea)</td>
<td>PSA</td>
<td>Man/Glc</td>
<td>N-Glycan binding enhanced by core fucosylation</td>
<td></td>
</tr>
<tr>
<td><strong>Phaseolus vulgaris</strong> (leukoagglutinin, kidney bean)</td>
<td>PHA-L</td>
<td>b</td>
<td>Tetra- and triantennary N-glycans with β6-branched</td>
<td></td>
</tr>
</tbody>
</table>

aWith an emphasis on status of sialylation and presence of core substitutions of N-glycans (see Figure 7 for structural details of an N-glycan with α,2,3/6-sialylation and core substitutions).

bNo monosaccharide known as ligand.
Figure 5. Examples for extended binding sites (subsite multivalency) of lectins structurally characterized by NMR spectroscopy or crystallography. (a) Illustration of the structure of the Asn-linked core trisaccharide of N-glycans (please see Figure 7 for a complete complex-type N-glycan), a ligand for hevein, the 43-amino acid lectin from laticifers of the rubber tree (*Hevea brasiliensis*) (i) and the chemical shift perturbation analysis as a diagnostic test for the influence of ligand association on proton signals which pinpoints reactivity mostly in the region between Ser19 and Tyr30 (ii). The extent of chemical shift difference ($\Delta\delta$, ppm) is plotted for the $\text{H}^a$ and $\text{H}^N$ protons of each...
receptor (Figure 3e) and Lewis epitopes become potent addressins in selectin-mediated lymphocyte homing [33,37], or into galactosylerceramide for conversion into sulfatide, a binding partner of galectin-4 in the clustering of lipid rafts for apical delivery [72,73], or of addition of acetyl groups to sialic acids (Table 3), or the transient presence of sugar moieties such as the three glucose moieties in the \( \alpha1,3 \)-arm during N-glycan maturation [6]. The latter is strictly confined to the ER, where a sugar-based quality control system operates with the glucose-binding lectins malectin, calnexin and calreticulin [6,74,75]. Notably, it is not only branch-end positioning and direct binding sites that matters for glycan reactivity to lectins (Table 2, level 4).

The six levels of affinity regulation: from core substitutions to microdomains

Two core substitutions are typical for animal and human N-glycans (Figure 7a,b). They are not added to the maturing chain randomly, but at particular positions in the course of assembly [5,6,76]. Bean/pea lectins (phytohemagglutinins) are common tools to detect such substituted N-glycans (Table 3), and the presence of these substitutions markedly affects glycan properties; in fact, they might not necessarily be in contact with the particular lectin being studied to make their presence felt. Molecular modeling and systematic bioassays with neoglycoproteins carrying synthetic biantennary N-glycans [measuring cell binding and blood level after intravenous (i.v.) injection; Figure 7c,d] disclosed that these moieties act as molecular switches for the shape of the glycans (Table 2, level 4a) [77,78]. To achieve prolonged presence of pharmaceutical glycoproteins in circulation after i.v. injection, tailoring of N-glycans to contain both substitutions and \( \alpha2,3 \)-sialylation appears to be most advantageous (Figure 7d).

Branching, too, introduces molecular switches, most impressively seen in the glycoside cluster effect. A numerical valency enhancement translates into a geometrical increase in affinity, with the type of branching exerting a pronounced effect (Table 2, level 4b) [79–81]. When examining triantennary N-glycans, the presence of the \( \beta1,6 \)-branch, detectable by the bean lectin PHA-L (Table 3, Figure 7a,b), leads to relatively reduced hepatocyte uptake via the hepatic asialoglycoprotein receptor [80]. The topological match between the branch-end galactosides and the subunits of this endocytic C-type lectin (homo- and heterooligomers of two closely related types of subunit) explains the glycoside cluster effect. This principle of spatially arranging lectin sites for affinity regulation is of general importance. The evolution-

---

Figure 6. Differential conformer selection of the pentasaccharide of ganglioside GM1 by human galectin-1 and the cholera toxin B-subunit. The \( \alpha2,3 \)-sialylgalactose linkage at the branch site (a) can adopt three low-energy structures (see Figure 4d, section on topological aspects (STD), for energy contour map). Human galectin-1 (b, i) and the B-subunit of the cholera toxin (PDB 2CHB, 3CHB) (b, ii) bind different conformers (‘keys’) so that the same pentasaccharide is bioactive in two distinct shapes [58].
ary diversification of lectin families not only includes divergence in extended binding sites but also in topology of binding site presentation, termed subunit multivalency [14]. Clusters of binding sites might not only be generated with non-covalent aggregates via an α-helical coiled coil stalk (Figure 3a) but also by tandem repeat display in one protein, with identical or disparate carbohydrate recognition domains, or by non-covalent association in homo- or heterodimers or larger aggregates in solution [14,33]. Foreign glycosignatures or malignant aberrations characterized by an unusually high density of mannose or fucose units (on yeast cells or as Lewisαβ epitopes of tumor cells) can be readily traced by such a multimeric lectin, the mannose (mannan)-binding C-type serum lectin (Figure 2c), to initiate a defense reaction via complement activation [82]. Viewed from the perspective of identifying glycobiomarkers, this sensitivity of the endogenous lectin is of interest and might even be enhanced by mutational engineering [83,84].

An intriguing consequence of this diversity in lectin site presentation is functional competition. Although they target the same ligand, lectins with different crosslinking capacity will trigger non-identical post-binding reactions or block signal elicitation. As an example of such antagonism, the growth regulatory activity of homodimeric galectin-1 is neutralized by saturating binding sites with a molar excess of galectin-3, a monomer in solution that aggregates when associating with multivalent ligands [85,86]; pharmaceutically, the design of custom-made synthetic glyoclusters combining headgroup tailoring with optimal (discriminatory) spatial parameters to exploit intrafamily divergence hereby becomes a reasonable goal [65,87,88].

Of emerging significance, the regulatory potency residing in natural cluster formation of lectin-reactive epitopes is beginning to be unraveled. High local density can be attributed to branches of the same glycan chain (Table 2, level 4b) or vicinity of neighboring glycans (Table 2, level 5). After all, the affinity for lectin binding is not a constant feature when the ligand is multivalent. When testing the glycoprotein asialofetuin (three reactive N-glycans, mostly triantennary, thus nine potential valencies) in solution, the titrations with galectins, as a test case, disclosed negative cooperativity with a gradient of decreasing microaffinity constants [89]. The initial, spatially privileged steps to saturation apparently harbor very high affinity, a feature considered to be relevant for understanding binding preferences on the cell surface [89,90]. Overall, the ligand density in two (polymers) or three (glycoalyx, extracellular matrix) dimensions will thus increase the propensity for lectin association [52,90]. That such spatial factors can be assumed to play pivotal regulatory roles is supported by the occurrence of compensatory mechanisms within the glycosylation machinery in response to engineered deficiencies.
in N- and O-glycan branching [91,92]. The nature of the structural context (type of glycosylation and branching as well as local vicinity) therefore can endow cells with preferential contact sites, subject to spatiotemporal dynamics [93,94]. Admittedly, these assumedly active constellations with distinct chains from separate glycoconjugates on cell surfaces are not easy to purify and characterize experimentally, thus providing a challenge for further research. Microdomain integrity, however, can be perturbed rather easily to delineate the importance of this mode of clustering for affinity regulation (Table 2, level 6). Applying reagents for cholesterol depletion (a cyclodextrin derivative and filipin III), aimed at harming microdomain integrity, effectively reduced the affinity of galectin-1 binding to neuroblastoma cells by a factor of 10–12, thus revealing the impact of this type of ordered ligand presentation [57]. Together with the equally drastic effects of inhibitors of glucosylceramide synthesis or cell surface ganglioside sialidase in impairing ganglioside GM1 availability [38,57], these results identify the topology of presentation of this ganglioside as determinant for galectin-1-mediated growth regulation.

Concluding remarks
The glyocalyx and the extracellular matrix are the biochemical platforms for a plethora of regulatory and recognition processes and hereby guarantee the integrity and operativeness of cell sociology. Given the obvious space limitations, high-density coding is required. Owing to the unsurpassed capacity of carbohydrates in this respect, the abundant nature of glycosylation as well as the complexity and dynamics of the glycophenotype, glycans qualify to properly endow cell surfaces with the toolbox to generate biosignals in necessary quality and quantity. The elaborate enzymatic machinery for synthesis and dynamic remodeling makes sure that the coding potential inherent to oligosaccharides is realized [5,6,9,76,95]. Drawing on Goethe’s wisdom that it is in her moments of abnormalities that Nature reveals her secrets, the insights from the study of congenital diseases of glycosylation and genetically engineered animal models signify compelling support for the implied glycan functionality [91,92,95,96]. This evidence for functional glycomics is flanked by the matching complexity of glycan receptors (lectins). Indeed, lectins are earning a reputation as translators of the sugar code (Table S1 in the supplementary material online). Fittingly, the detailed analysis of their binding sites is uncovering structural motifs from nearly flat surfaces to deep pockets embedded in diverse types of folds. With advanced knowledge on how glycans enter these sites, we gain a greater appreciation of the significance of shape and clustering. In other words, the systematic dissection of affinity regulation into six levels (Table 2) guides the way from structural to topological factors, for example dynamic changes of conformation and local density, up to the level of microdomains with special glycolipids as contact sites and an influence of the lipid anchor and neighboring cholesterol [57,72,97,98]. An integrated approach with strategically combined techniques (Figure 4), which has proven its eminent value for the complete structural elucidation of N-glycans [99], will be essential for addressing the arising challenge of ascribing meaning to topological factors, along with the inspiring perspective for devising novel routes for biomarker discovery and therapeutic interventions. The illustrated or cited cases for differential conformation selection and modulation of lectin affinity via core substitutions or impairment of integrity of glycan-presenting microdomains thus add convincing support to the paradigm for glycans serving as biochemical basis of a versatile means of information storage and transfer, that is the sugar code [100].

Acknowledgements
We are grateful for generous funding to CIBERES (ISCIII), to EC GlycoHIT program, the Spanish Ministry of Science and Innovation (grants BFU2008-02595, BFU2009-10052 and CSD2009-00068) and the Verein zur Förderung des biologisch-technologischen Fortschritts in der Medizin e.V. as well as to Drs. B. Friday and J. Kopitz for inspiring discussions.

Appendix A. Supplementary data
Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tibs.2011.01.005.

References
1 Bennett, H.S. (1963) Morphological aspects of extracellular polysaccharides. J. Histochem. Cytochem. 11, 14–23
38 Kopitz, J. et al. (1998) Galectin-1 is a major receptor for ganglioside GM1, a product of the growth-controlling activity of a cell surface ganglioside sialidase, on human neuroblastoma cells in culture. J. Biol. Chem. 273, 11205–11211
48 André, S. et al. (2007) Tumor suppressor p16INK4a, modulator of glycoprotein profile and galectin-1 expression to increase susceptibility to carbohydrate-dependent induction of anoikis in pancreatic carcinoma cells. FEBS J. 274, 3233–3256
52 Asensio, J.L. et al. (2000) Structural basis for chitin recognition by defense proteins: GlcNac residues are bound in a multivalent fashion by extended binding sites in hevein domains. Chem. Biol. 7, 529–543
54 Somers, W.S. et al. (2000) Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLAα and PSGL-1. Cell 103, 467–479
57 Kopitz, J. et al. (2010) How adhesion/growth-regulatory galectins-1 and -3 attain cell specificity: case study defining their target on neuroblastoma cells (SK-N-MC) and marked affinity regulation by affecting microdomain organization of the membrane. IUBMB Life 62, 624–628
64 Ribeiro, J.P. et al. (2010) Lectin-based drug design: combined strategy to identify lead compounds using STD NMR spectroscopy, solid-phase
assays and cell binding for a plant toxin model. ChemMedChem 5, 415–419


Siebert, H.-C. et al. (2006) α2,3α2,6-Sialylation of N-glycans: non-synonymous signals with marked developmental regulation in bovine reproductive tracts. Biochimie 88, 399–410


Dam, T.K. et al. (2005) Galectins bind to the multivalent glycoprotein asialofetuin with enhanced affinities and a gradient of decreasing binding constants. Biochemistry 44, 12564–12571


