Review Article

How to Crack the Sugar Code

(adhesion / glycosylation / lectin / proliferation / signalling)

H.-J. GABIUS

Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Munich, Germany

Abstract. The known ubiquitous presence of glycans fulﬁls an essential prerequisite for fundamental roles in cell sociology. Since carbohydrates are chemically predestined to form biochemical messages of a maximum of structural diversity in a minimum of space, coding of biological information by sugars is the reason for the broad occurrence of cellular glycoconjugates. Their glycans originate from sophisticated enzymatic assembly and dynamically adaptable remodelling. These signals are read and translated into effects by receptors (lectins). The functional pairing between lectins and their counterreceptor(s) is highly speciﬁc, often orchestrated by intimate co-regulation of the receptor, the cognate glycan and the bioactive scaffold (e.g., an integrin). Bottom-up approaches, teaming up synthetic and supramolecular chemistry to prepare fully programmable nanoparticles as binding partners with systematic network analysis of lectins and rational design of variants, enable us to delineate the rules of the sugar code.

Introduction

Cellular homeostasis and sociology critically depend on biochemical coding of information and translation of the corresponding messages into effects by molecular recognition and post-binding signalling. Naturally, cell surfaces are the interface for this communication, and its smooth efficiency directs attention to analysing the characteristics of the involved biomolecules. Biochemical detection of glycan-bearing lipids and proteins as common constituents of and on biomembranes (Eichwald, 1865; Thudichum, 1874; Klenk, 1942), histochemical monitoring visualizing the sugary coating of mammalian cells, termed glycocalyx (Gasic and Gasic, 1962; Bennett, 1963), and demonstration of the ubiquity of the surface presence of glycans as polysaccharides, glycolipids and glycoproteins (Sharon, 1975; Montreuil, 1995; Spiro, 2002; Kopitz, 2009, 2017; Corfield and Berry, 2015; Gabius, 2015; Tan et al., 2015; Corfield, 2017) converge to ignite the interest to reveal why glycans so abundantly reside at this strategically prominent place. Since space is limited there and a large panel of signals need to be accommodated, the building blocks of a code system must be able to form a maximal number of oligomers (signals) in a minimum of size. The coding capacity, the measure of structural complexity, must thus be high. That said, the question arises whether glycans are suited to serve as signals.

After carefully considering their chemical nature, the conclusion had been reached that “carbohydrates are ideal for generating compact units with explicit informational properties” (Winterburn and Phelps, 1972). Apparently, distinct features, which establish the basis of high-density information storage by glycans (Laine, 1997; Gabius and Roth, 2017), let them become “much more complex, variegated, and difﬁcult to study than proteins or nucleic acids” (Roseman, 2001). To gain a good knowledge of the rules of the carbohydrate-based language then inevitably took much longer than for the other two just mentioned types of biomolecules. It is therefore easy to understand that “only in recent years have we begun to appreciate how deeply glycan functions pervade all aspects of organismic biology, molecular biology, and biochemistry” (Hart, 2013). Due to the challenging character of work to crack the sugar code and the wide implications of the resulting insights it is timely to provide a primer to the concept of glycans in information storing (sugar code) together with describing the emerging principles of the role of their functional pairing with tissue lectins. We begin with an explanation why sugars deserve to be referred to as third alphabet of life, on par with nucleotides and amino acids.

Why sugars are ideal as biochemical signals

As noted above, nucleic acids and proteins differ from glycans fundamentally in their levels of structural complexity. Sequencing of those two classes of biomolecules is already complete, when a single parameter, i.e.,
the order of building blocks in a chain, is determined. In other words, their template-driven synthesis follows a strictly conserved assembly biochemistry, shown in Fig. 1a,b. Opportunities to increase the structural variability, for example by a 2',5'-linkage or isopeptide bonds, are only rarely used. This is vitally different for the letters of the sugar alphabet. Here, such opportunities are implemented in the structure of the building blocks and commonly used: as a consequence, the linkage between sugar units can vary with respect to i) the anomeric status (α or β), and ii) the position of the hydroxyl group involved in the glycosidic bond with the anomeric centre (Fig. 1c). Figure 1 thus explains the structural reason why sugar coding can reach the high density necessary for signals on the cell surface.

To realize the noted chemical potential for structural diversity, the machinery to produce glycans from monosaccharides must be equipped with a large set of enzymes, and this is the case when looking at the diversity of glycosyltransferases (Brockhausen and Schachter, 1997; Reuter and Gabius, 1999; Buddecke, 2009; Patsos and Corfield, 2009; Wilson et al., 2009; Zubert and Roth, 2009; Hennet and Cabalzar, 2015; Schengrund, 2015; Bhide and Colley, 2017). By using the letters of the third alphabet of life shown in Fig. 2, their activities make an unsurpassed level of structural complexity possible. Remarkably, changes in the expression of these enzymes and in substrate/acceptor availability convey manifold possibilities to direct glycan synthesis to different routes, even alter glycan (glycome) representation dynamically (McDonald et al., 2016; Neelamegham and Mahal, 2016). Site-specific introduction of substitutions such as a phosphate or sulphate group further adds to the structural multiformity and can give the respective letter a new biological meaning, for example by site-specific phosphorylation of mannose (see legend to Fig. 3a).

In sum, an oligosaccharide, the biochemical manifestation of a sugar-encoded message, is capable to harbour much more information than nucleic acids or proteins of equal size, as Figure 1 illustrates. To accomplish its structural analysis, each glycosidic linkage must thus be unambiguously defined in terms of two parameters, i.e., anomeric position and linkage points (see Fig. 1c), and the presence of substitutions must be mapped. Sophisticated methodology has been developed with ingenuity and perseverance to master this challenge so that even rare epitopes can now be reliably sequenced (Hounsell, 1997; Nakagawa, 2009; Alley et al., 2013; Novotny et al., 2013). Examples for structures of biologically relevant signals from a phosphorylated monosaccharide mentioned above to a pentasaccharide are provided in Fig. 3.

Showing these structures also serves a second purpose, which is to draw attention to another structural feature useful in ensuing information transfer: each glycan evidently presents a large number of chemical groups suited for hydrogen bonding. Their presence makes recognition by molecular complementarity readily feasible. Since this type of interaction is directional, natural epimer pairs such as glucose/mannose (2' position) or glucose/galactose (4' position) (Fig. 2) can easily be distinguished by suitably positioned amino acids. A respective contact to the axial 4'-OH group of galactose (not possible for equatorial positioning) is shown in the centre of Fig. 4, with a graphic illustration of its biorelevance. At the same time, the bottom side of this pyranose can engage in C-H/π-interactions with receptors, which adds to making contact formation specific (Fig. 4). Amino acids of proteins hereby become efficient means to mould docking sites for a carbohydrate.

Having reached the level of linear (e.g., the di- to pentasaccharides of Fig. 3b-d; first dimension of the sugar code) and branched glycans (e.g., histo-blood group ABH(0) epitopes; second dimension of the sugar code), a special property of glycans in the third dimension further favours their participation in biomolecular interactions. It concerns the extent of their intramolecular flexibility. Instead of a rather unrestricted movement of the sugar units around the glycosidic linkage, often encountered in peptides so that arresting them into a bioactive conformation requires their embedding into the structural context of a protein, oligosaccharides often adopt only few conformers. When schematically depicting the conformational space of a glycan in the equivalent of a topographical map, they represent energetically privileged ‘valleys’ (Carver, 1993; Hardy, 1997; von der Lieth et al., 1998; Roseman, 2001). Drawing on E. Fischer’s famous analogy of a lock-and-key recognition mechanism (shaped on the experimental basis of analysing glycoside and disaccharide hydrolysis by “Invertin und Emulsin”; Fischer, 1894), this constellation has also been likened to a glycan establishing a “bunch of keys” (conformers), “each of which can be selected by a receptor” (Hardy, 1997). As a consequence, the formation of a complex between a conformer and its receptor, driven by establishing a bonding pattern as shown in Fig. 4, will not incur a high entropic penalty, as it would be caused by a severe reduction of inherently high flexibility. In our context, this property in the third dimension combines well with the enormous structural variability and ability for contact formation to enable writing messages in the sugar code and let them be read. In order to attain fundamental physiological significance for this route of flow of biological information, receptors of sufficient number and specificity must have developed, as enzymes for signal generation have. This is indeed the case.

**Reading sugar-encoded signals: the lectins**

Historically, testing extracts of diverse sources for activity to agglutinate erythrocytes was a popular assay to detect presence of sugar receptors. The capacity of various plant extracts for selecting target cells according to their ABH(0) blood-group status, as human antibodies do, led W. C. Boyd to coin the term ‘lectin’ “from Latin lectus, the past principle of legere meaning to pick, choose or select” (Boyd, 1954) for such antibody-like
Fig. 1. Illustration of the linkage points for oligomer formation from nucleotides, amino acids and carbohydrates. The phosphodiester bond in nucleic acids (a) and the peptide bond in proteins (b) generate uniform structural modes to connect letters (building blocks) of these types of molecular alphabets to words and sentences (oligo- and polymers). In contrast, adding an activated sugar via its anomeric position (α or β) to an acceptor in chain elongation of glycans can involve any available hydroxyl group, as symbolized by the arrows (from Rüdiger and Gabius, 2009; with permission).

Fig. 2. Illustration of the letters of the sugar alphabet present in animal, plant and bacterial glycans. Arrows indicate known acceptor positions for chain elongation (from Rüdiger and Gabius, 2009; with permission).
activities, also called phyt(ohaem)agglutinins. Biochemically, members of this group of carbohydrate-binding proteins are separated from immunoglobulins, enzymes using carbohydrates as substrates, sensor/transport proteins for free mono- and oligosaccharides and the carbohydrate-binding modules associated to the catalytic centre of microbial glycoside hydrolases (Barondes, 1988; Gabius et al., 2011).

Corroborating the fundamental nature of the concept of the sugar code, the range of occurrence of lectins matches that of glycans. The development of ability to bind glycans in more than a dozen protein folds builds a firm and broad structural basis to generate the tools to read sugar-encoded signals (Gabius, 1997; Lis and Sharon, 1998; Loris, 2002; Fujimoto et al., 2014; see galleries of animal/human and plant lectins presented in Solís et al., 2015, and Manning et al., 2017a). By adding Ca\(^{2+}\) at strategic sites to lectins, contact building to carbohydrate ligands is made possible in ways beyond those involving amino acids, for example by coordination bonding of glycans in bacterial and in C-type lectins (Gabius, 2011). The resulting specificity of lectins for glycans is instrumental for many applications such as glycophenotyping of cells (Roth, 1978, 2011; Manning et al., 2017a). Its relevance for daily life is clinically apparent for example in infectious processes initiated by viral as well as bacterial lectins and toxins (Holgersson et al., 2009; Ströh and Stehle, 2014; Moonens and Remaut, 2017). Having initiated this line of research with plant and invertebrate lectins, especially for blood-group typing (Watkins, 1999), the increasing availability of mammalian lectins (the first purification using rabbit liver as starting material reported in 1974 by Hudgin et al.), now even reaching the level of complete lectin families, makes function-oriented studies possible, e.g., in host defence, inflammatory disorders or malignancy (Dawson et al., 2013; Brown and Crocker, 2016; Toegel et al., 2016; Weinmann et al., 2016; Mayer et al., 2017; Zivicová et al., 2017). Along this line of research, the size of the toolbox of tissue lectins and the characteristics of each protein are being thoroughly defined, as the glycome is being mapped.

Intra-family diversification starting from an ancestral gene is an evolutionary means to increase the number of lectins, which are then adapted to their individual functional missions. In the course of this process, two main
considered as equivalent of the highlighted chemical parameters of carbohydrates to attain diversity. The realization of this aspect of a molecular protein puzzle strengthens the conclusion that Nature has ample means to produce and to read sugar-encoded signals, enabling functional versatility.

Translating sugar-encoded signals: functional pairing of lectins and glycoconjugates

The current status of knowledge on the involvement of glycan-lectin interplay in cellular activities convincingly confirms this expectation. Already at the starting point of glycosylation in the endoplasmic reticulum, which is intimately connected with cotranslational N-glycosylation, the quality control of glycoprotein folding and the subsequent routing that separates misfolded and mature glycoproteins employ lectins (Roth, 2002; Roth and Zuber, 2017). Self/non-self discrimination, mediator release and pattern formation are among further processes with lectins as effectors, to name a few (Brown and Crocker, 2016; Gabius et al., 2016; Manning et al., 2017a). A central finding emerging from studying the course of reaction that begins with glycan binding to a lectin and leads to the triggered cellular effect is the intriguingly specific selection of glycoconjugate counterreceptor(s) by tissue lectins. Despite abundance of nominally fitting binding partners, e.g., the cell-binding RGD tripeptide in discoidin I (Gabius et al., 1985) or the canonical immunoreceptor tyrosine-based inhibitory motif in siglec-2 and the CD33-related siglecs, to modules such as epidermal growth factor-like protein sections in hyalectans (lecticans) or selectins (Iozzo and Murdoch, 1996) is a theme often encountered in lectin design.

How to Crack the Sugar Code

Fig. 4. Specific recognition of a carbohydrate by hydrogen bonding and C-H/π-interactions is a salient molecular means within human physiology (from Gabius, 2009; with permission).

Structural changes occur: the binding site is stepwisely altered to allow hosting new ligands, and the modular architecture can be tailored to cover new modes to present the carbohydrate recognition domain (CRD). Both structural features come into play when a lectin selects its counterreceptor(s) from the vast panel of cellular glycoconjugates. As illustrated for vertebrate β-sandwich-type ga(lactose-binding)lectins in Fig. 5, three types of display of their CRD have arisen in the course of phylogensis (Hirabayashi, 1997; Kaltner et al., 2017), and recent species-specific acquisitions by gene duplication and neofunctionalization such as the murine tandem-repeat-type galectin-6 document continuous dynamics on this level (Gitt et al., 1998; Houzelstein et al., 2008). Further attesting diversity on the lectin side, even up to 17 groups, establish the C-type lectin representation in mammals (Gready and Zelensky, 2009). In each group, several proteins can contribute to forming the network, with species-specific differences. Further gene duplication events followed by promoter diversification, seen in the case of a human galectin (Kaltner et al., 2013), increase the means of fine-tuning gene regulation. The permutations possible by combining a distinct type of CRD with different types of modular display can be considered as equivalent of the highlighted chemical parameters of carbohydrates to attain diversity. The realization of this aspect of a molecular protein puzzle strengthens the conclusion that Nature has ample means to produce and to read sugar-encoded signals, enabling functional versatility.

Translating sugar-encoded signals: functional pairing of lectins and glycoconjugates

The current status of knowledge on the involvement of glycan-lectin interplay in cellular activities convincingly confirms this expectation. Already at the starting point of glycosylation in the endoplasmic reticulum, which is intimately connected with cotranslational N-glycosylation, the quality control of glycoprotein folding and the subsequent routing that separates misfolded and mature glycoproteins employ lectins (Roth, 2002; Roth and Zuber, 2017). Self/non-self discrimination, mediator release and pattern formation are among further processes with lectins as effectors, to name a few (Brown and Crocker, 2016; Gabius et al., 2016; Manning et al., 2017a). A central finding emerging from studying the course of reaction that begins with glycan binding to a lectin and leads to the triggered cellular effect is the intriguingly specific selection of glycoconjugate counterreceptor(s) by tissue lectins. Despite abundance of nominally fitting binding partners, e.g., β-galactosides for galectins or sialylated glycans for siglecs, an endogenous lectin appears to home in on few distinct glycoconjugates to initiate signalling toward the cell-type-specific response.

Looking at a homodimeric (and thus cross-linking) effector such as galectin-1 shown in Fig. 5 (left), its binding of certain glycoproteins or gangliosides under-
lies its pro-anoikis/apoptosis activity: complex-type N-glycans on the fibronectin receptor (αβ-integrin) on pancreatic tumour cells made accessible by reduction of α2,6-sialylation controlled by a tumour suppressor, core 2 O-glycans on CD7 on activated T cells as well as ganglioside GM1 on neuroblastoma and activated effector T cells upregulated by desialylation of ganglioside GD1α are sugar-based messages for the galectin-mediated growth regulation (Kopitz et al., 2001; Cabrera et al., 2006; Wang et al., 2009; Amano et al., 2012; Ledeen et al., 2012; Smetana et al., 2013). In each case, both sides of the recognition system, that is i) cognate glycan on a distinct cellular glycoconjugate and ii) receptor, are often subject to orchestrated regulation, a common principle warranting explanation by a figure. As illustrated in Fig. 6, activation of effector/regulatory T cells leads to a coordinated response profile at the level of lectin, counterreceptor and response-driving factor, here an ion (Ca²⁺) channel. This co-regulation guarantees keeping the auto-aggressive activity of effector T cells at bay.

Which factors govern the functional pairing in all these cases is currently an open question (Gabius et al., 2015). The determinants underlying the target selection are assumed to be of biochemical and spatial nature, as summarized in Table 1, which lists the respective properties of glycans. Work with custom-made glycans and cluster glycosides up to the in vivo level has underlined that tissue lectins respond very sensitively to the ligand structure and spatial aspects of its presentation (Lee and Lee, 1994; Roy et al., 2016, 2017). Reaching a high local density of O-glycans in mucins or of the pentasaccharide of ganglioside GM1 or ganglioside-glycoprotein complexes as shown in Fig. 6 in microdomains can thus provide distinguishing marks of favoured, physiologically relevant docking sites. Since generation of genetic deficiencies in murine N- and O-glycosylation in vivo has led to compensations at the level of the glycome as a sign of remarkable cellular plasticity (Takamatsu et al., 2010; Ismail et al., 2011), animal models may be of limited relevance to pinpoint structure-function relationships. This situation raises interest in devising robust test platforms, whose level of complexity can be tailored in a bottom-up strategy. The design of cell-like models with fully programmable glycan presentation, along with that of variants of tissue lectins, is a means to gain access to tools for experiments with the aim to delineate rules of information transfer based on the sugar code.

An example for such a model system are glycodendrimersomes. They are the products of self-assembly of amphiphilic Janus glycodendrimers (Percec et al., 2013). Surface presentation of glycans can be varied in terms of structure and spatial parameters in a bottom-up manner, taking advantage of the enormous progress to produce glycans by synthetic carbohydrate chemistry (Oscarson, 2009; Krasnova and Wong, 2016). The turbidity increase by aggregation of these nanoparticles in solution provides an experimental read-out for crosslinking activities of lectins so that the influence of changes of surface characteristics becomes quantifiable (Zhang et al., 2015a; Xiao et al., 2016). On the lectin

![Fig. 6. Illustration of the galectin-1/ganglioside GM1-dependent communication between activated effector/regulatory T cells. Co-regulation on both sides, cross-linking of GM1/integrin complexes by the lectin and the hereby triggered signalling involving focal adhesion kinase leads to opening of Ca²⁺ channels (TRPC5), Ca²⁺ influx and effector T cell anergy/apoptosis (for details, see Wang et al., 2009).](image)

<table>
<thead>
<tr>
<th>Table 1. Characteristics of glycans relevant for the specificity of functional pairing of tissue lectins with distinct cellular glycoconjugatesa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Recognition of the core determinant of the glycan (mono- or disaccharide)</td>
</tr>
<tr>
<td>2. Recognition of structural extensions of the core determinant</td>
</tr>
<tr>
<td>3. Recognition of a distinct shape (differential conformer selection)</td>
</tr>
<tr>
<td>4. Influence of spatial parameters at the level of the glycoconjugate</td>
</tr>
<tr>
<td>a. modulation of shape by molecular switches not directly involved in recognition, e.g., core substitutions of complex-type N-glycans</td>
</tr>
<tr>
<td>b. cluster effect by different degrees of i) branching, e.g., in complex-type N-glycans, and ii) local glycan density, e.g., in mucins</td>
</tr>
<tr>
<td>5. Influence of spatial parameters at the level of neighbouring glycoconjugates, e.g., by complex formation (as shown in Fig. 6)</td>
</tr>
<tr>
<td>6. Influence of spatial parameters at the level of the cell surface (microdomains or glycosynapses)</td>
</tr>
</tbody>
</table>

* for exemplary graphical depictions of the six characteristics, see Fig. 3 in Gabius et al., 2015
side, testing natural proteins and engineered variants is a logical step on the way to fully exploit the potential of this model. Bringing the chemical versatility of tailoring the vesicle surface together with devising modular architectures beyond those shown in Fig. 5 enables us to answer fundamental questions on the structure-function relationships. Having galectin-responsive cells at hand, as referred to above, the scope of this work can be extended to measuring the impact of structural changes on cellular responses such as proliferation. The resulting concept is graphically depicted in Fig. 7 for the mentioned human galectin-1 as the test case.

An obvious question is what would happen if the two subunits were firmly linked, that is if the non-covalently associated homodimer was turned into a non-physiological tandem-repeat-type homodimer. When this has been done by engineering, it was possible to reveal that the protein’s capacity for trans-bridging was markedly enhanced (Zhang et al., 2015b). Non-covalent association of CRDs thus appears suited for cis-cross-linking operative in eliciting growth regulation or for transient trans-bridging. Of note, increases in the ability to sense very low extents of ligand display ensue conjugate formation to homodi- and tetramers, which are shown in Fig. 7 (Kopitz et al., 2017). The presence of galectin tetramers in oysters can thus be attributed to these invertebrates’ needs of protection against bacteria, addressed by producing highly potent agglutinins (Tasumi et al., 2007; Feng et al., 2013). In mammals, however, such high sensitivity for the presence of ligand would mean that a possibility for switching-on growth inhibition by up-regulating the signal presence (as shown in Fig. 6) would be lost, an insight obtained from following the experimental strategy shown in Fig. 7 (Kopitz et al., 2017). Taking advantage of the general applicability of this test system that allows evaluating a variety of sugar headgroups and further lectin variants, for example after modular transplantation of a CRD to a new scaffold (Fig. 8; Ludwig et al., 2016) or CRD shuffling to create a homodimer from a heterodimeric tandem-repeat-type galectin (Ludwig et al., 2017), opens manifold possibilities. Comprehensive study of the effect of distinct parameter changes on each side of the recognition system on the experimental read-out can hereby be performed.

Next, access to all members of a lectin network, as for example accomplished for the seven avian galectins (García Caballero et al., 2016; Kaltner et al., 2016; Manning et al., 2017b), enables us to proceed to testing lectins in mixtures. They are physiologically present so that such experiments will probe into the so far mostly uncharted territory of functional antagonism and synergism. Since a glycan determinant is often a target for diverse lectins, the core 1 O-glycan disaccharide (TF antigen or CD176) serving as instructive example (Krzeminski et al., 2011; Maestre-Reyna et al., 2012; Abed et al., 2016; Conover et al., 2016; Artigas et al., 2017), clinically relevant competition between bacterial/yeast and tissue lectins will become testable. Hereby, innovative ways to block infections are then explored, and variants such as the potently agglutinating homotetramer may eventually find applications as anti-infectious agent in acute situations.

Conclusion

Their chemical structure endows carbohydrates with unsurpassed capacity for structural diversity in glycans and the means for specific contact formation to receptors (lectins). Coding information by the sugar part of cellular glycoconjugates is being realized as a highly versatile biochemical platform that is engaged in diverse aspects of cell physiology. Teaming up analytical, synthetic and supramolecular chemistry with protein biochemistry and engineering produces tools for a bottom-

Fig. 7. Illustration of the principle of combining lectin engineering with assays using supramolecular cell models and lectin-responsive cells. The homodimeric (proto-type) galectin-1 (wild-type protein, WT) is turned into a (homo) tetramer, and the activities of these two proteins in trans-bridging to aggregate glycodendrimersomes or cis-cross-linking to trigger inhibition of cell proliferation can be comparatively determined (for details, see Kopitz et al., 2017).

Fig. 8. Illustration of an example of design of a new variant of a human galectin by protein engineering (modular transplantation). A CRD of a tandem-repeat-type galectin is connected to the N-terminal tail of galectin-3 to generate a bioactive hybrid (for details, see Ludwig et al., 2016).
up route of model building and testing. Adding assays with cell systems sets the stage to unveil the principles of translating the sugar-encoded information into physiological effects (cracking the sugar code), with potential for devising innovative therapeutic modalities. After all, the interplay of structural features of glycans and lectins with not yet clearly defined parameters of their in situ presentation apparently governs the selection process that lets certain glycoconjugates become functional counterreceptors. This functional pairing has been defined as the molecular basis of an already large number of physiological effects (cracking the sugar code), with potential for translating the sugar-encoded information into physiological effects (cracking the sugar code).

**Acknowledgements**

Inspiring discussions with Drs. B. Friday and A. Leddoz are gratefully acknowledged.

**References**


