

Chemical Biology of the Sugar Code

Hans-Joachim Gabius,^{*[a]} Hans-Christian Siebert,^[a] Sabine André,^[a]
Jesús Jiménez-Barbero,^[b] and Harold Rüdiger^[c]

In respectful and thankful memory of Professor Friedrich Cramer, who died three months before his 80th birthday

A high-density coding system is essential to allow cells to communicate efficiently and swiftly through complex surface interactions. All the structural requirements for forming a wide array of signals with a system of minimal size are met by oligomers of carbohydrates. These molecules surpass amino acids and nucleotides by far in information-storing capacity and serve as ligands in biorecognition processes for the transfer of information. The results of work aiming to reveal the intricate ways in which oligosaccharide determinants of cellular glycoconjugates interact with tissue lectins and thereby trigger multifarious cellular responses (e.g. in adhesion or growth regulation) are teaching amazing lessons about the range of finely tuned activities involved. The ability of enzymes to generate an enormous diversity of biochemical signals is matched by receptor proteins (lectins), which are equally elaborate. The multiformity of lectins ensures accurate signal decoding and transmission. The exquisite refinement of both sides of the protein-carbohydrate recognition system turns the

structural complexity of glycans—a demanding but essentially mastered problem for analytical chemistry—into a biochemical virtue. The emerging medical importance of protein-carbohydrate recognition, for example in combating infection and the spread of tumors or in targeting drugs, also explains why this interaction system is no longer below industrial radarscopes. Our review sketches the concept of the sugar code, with a solid description of the historical background. We also place emphasis on a distinctive feature of the code, that is, the potential of a carbohydrate ligand to adopt various defined shapes, each with its own particular ligand properties (differential conformer selection). Proper consideration of the structure and shape of the ligand enables us to envision the chemical design of potent binding partners for a target (in lectin-mediated drug delivery) or ways to block lectins of medical importance (in infection, tumor spread, or inflammation).

1. Introduction

Biological information storage and transfer are commonly described to be based solely on nucleic acids and proteins. In contrast to nucleotides and amino acids, the most abundant type of biomolecule in nature, the carbohydrate molecule, has been almost completely sidelined in this respect. Sugar molecules have been nearly exclusively assigned as building blocks of protective cell wall constituents (for example cellulose and chitin) or as biochemical fuel in energy metabolism. This paradigm, which is reflected in textbooks, has been questioned occasionally over the years. An exemplary quotation from 1972 points out that glycans do matter more than originally assumed: "The polysaccharides of mammalian connective tissue, and glycoproteins, begin to make biochemical sense for the first time ever. So many exciting developments have occurred that this period seems to have moved us out of a dark age to see polysaccharides in quite a new light. They have become interesting molecules to contemplate in relation to the life of a cell. The ugly ducklings have begun to look a little more like swans. In this sense, polysaccharides begin to appear attractive molecules, shapely molecules."^[1] With hindsight, the answer to the question of why the exceptional talents of carbohydrates have remained nearly unnoticed for so long appears to be rather simple; in essence, this neglect occurred because "glycoconjugates are much more complex, variegated, and difficult to study than proteins or nucleic acids."^[2] Viewed from the perspective of bioinformatics, however, this structural property in fact makes oligomers of saccharides "ideal for generating com-

pact units with explicit informational properties."^[3] This argument and other reasons listed below explain why it is justified to portray individual monosaccharides as letters of an alphabet. These letters form biochemical code words. The coining of terms such as *sugar code* or *glycomics* helps condense the concept into keywords. However, it goes without saying that the reader can expect us to carve out a distinctive image of this fundamental functionality of glycans.^[4]

2. The Hardware of the Sugar Code

Carbohydrates have several exceptional features at their disposal. These features make a strong case for a prominent role of D-glucose and its relatives in information handling. Fore-

[a] Prof. Dr. H.-J. Gabius, Priv.-Doz. Dr. H.-C. Siebert, Dr. S. André
Institut für Physiologische Chemie, Tierärztliche Fakultät
Ludwig-Maximilians-Universität
Veterinärstraße 13, 80539 München (Germany)
Fax: (+49) 89-2180-2508
E-mail: gabius@tiph.vetmed.uni-muenchen.de
gabius@lectins.de

[b] Prof. Dr. J. Jiménez-Barbero
Centro de Investigaciones Biológicas
CSIC, Ramiro de Maeztu 9, 28040 Madrid (Spain)

[c] Prof. Dr. H. Rüdiger
Institut für Pharmazie und Lebensmittelchemie
Julius-Maximilians-Universität
Am Hubland, 97074 Würzburg (Germany)

Hans-Joachim Gabius was born in Bad Bevensen (Germany). He studied biochemistry in Hannover (as a fellow of the Studienstiftung des deutschen Volkes) and obtained an MSc. in 1980, then a PhD in 1982 for chemical and biochemical studies on the proofreading mechanisms of aminoacyl-tRNA synthetases, under the direction of F. Cramer, Max Planck Institute for Experimental Medicine, Göttingen.



He spent most of 1981 investigating tRNA splicing in the laboratory of J. Abelson (Department of Chemistry) at UC San Diego. After starting work in tumor lectinology in 1983 at the Max Planck Institute in Göttingen, he went on to a post-doctoral research post in the group of S. H. Barondes at UC San Diego (1984–1985) and appointments as assistant professor for biochemistry at the Max Planck Institute for Experimental Medicine in Göttingen (1987), as associate professor for pharmaceutical chemistry at the University of Marburg (1991), and as head of the Institute for Physiological Chemistry, Faculty of Veterinary Medicine, University of Munich (1993). His research awards include the Otto-Hahn-Medal (1983), the Award of the Dr. Carl Duisberg Foundation (1988), and the Award of the Paul Martini Foundation (1990). His research interests comprise chemical, biophysical, and biochemical analysis of protein–carbohydrate interactions relevant to the biological and medical fields, such as the development of glycoscientific strategies for tumor diagnosis and therapy and the elucidation of the functions of mammalian lectins. He was prominently placed in the ranking of researchers by number of hot papers produced by the Institute of Scientific Information in 1998 (http://www.the-scientist.library.upenn.edu/yr1999/june/hotresearch_p1_990621.html; see also, www.lectins.de).

Jesús Jiménez-Barbero was born in 1960 in Madrid (Spain) and is Professor at the Center for Biological Research of the CSIC. He obtained his PhD in Organic Chemistry in 1987 for synthesis work and conformational studies on saccharides at the Institute of Organic Chemistry of the CSIC in Madrid under the supervision of M. Bernabé and M. Martín-Lomas. Following post-doctoral training in molecular mechanics and NMR methodology from



1986 to 1988 at CERMAV-CNRS in Grenoble, the University of Zürich, and the National Institute for Medical Research at Mill Hill, he received tenure in the CSIC in 1988. He was Visiting Scientist at the Department of Chemistry of Carnegie Mellon University at Pittsburgh between 1990 and 1992 and then started work on the application of NMR methodology to the study of interactions between carbohydrates and proteins and conformational and structural studies of oligo- and polysaccharides in Madrid. He was promoted to Senior Research Scientist in 1996 and to Full Professor in 2002, when he moved from the Institute of Organic Chemistry to the Center for Biological Research of the CSIC. He is mostly interested in obtaining a 3D view of the molecular recognition processes in which carbohydrates are involved, in particular by the application of NMR spectroscopy and modeling methods. He is a member of the editorial boards of several international journals and has published almost 200 scientific papers, reviews, and book chapters on the above-mentioned topics. He has also given more than sixty lectures at international conferences and research institutions.

Hans-Christian Siebert was born in 1960 in Kiel (Germany) and studied physics in Kiel (1980–1983) and in Heidelberg (1983–1987). He obtained a diploma (1987) for work on dynamic NMR spectroscopy and a PhD (1990) for conformational studies of gangliosides by NMR spectroscopy and computational methods in J. Dabrowski's group at the Department of Organic Chemistry, Max Planck Institute for Medical Research, Heidelberg. Following post-doctoral research in J. F. G. Vliegthart's and R. Kaptein's groups at the Bijvoet Center for Biomolecular Research, Utrecht University, he joined H.-J. Gabius' research team at the Institute of Pharmaceutical Chemistry in Marburg in 1992 and moved with him to Munich, where he became Dr. med. vet. habil. and received the Venia legendi in Biochemistry in 1999. His research interests include NMR spectroscopy and molecular modeling structural studies of carbohydrate–protein interactions with biomedical relevance.



Harold Rüdiger was born in Stolp (Germany). He studied chemistry and biochemistry at the University of Würzburg, where he earned his PhD in 1963 for studies on the kinetics of peptide synthesis in yeast. He moved to the Biochemistry Institute of the University of Uppsala (Sweden), then headed by Nobel Prize winner A. Tiselius, to do postdoctoral research in J. Porath's group, where he studied modern biochemical analytical and separation techniques and worked out a purification protocol for a plant lectin. Upon returning to Germany in 1966, he joined L. Jaenicke's group at the Biochemistry Institute of the University of Cologne, where he studied the bacterial cobalamin-dependent methionine biosynthesis. In 1971, he became lecturer in biochemistry and in 1974 he moved to Würzburg University to take up a position as a professor of biochemistry. His research interests center on plant lectins and their interaction with various ligands.

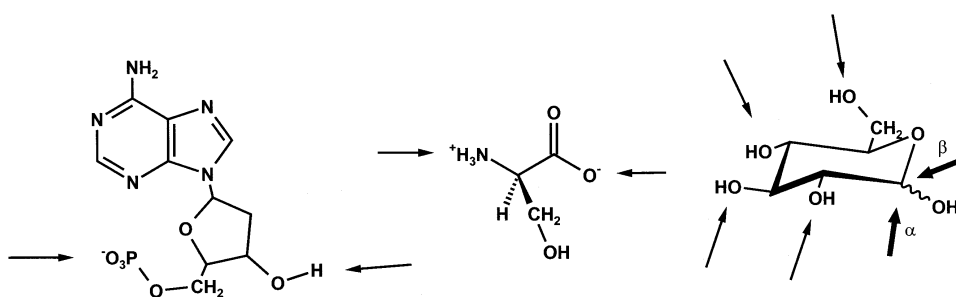


Sabine André was born in 1966 in Bad Hersfeld (Germany) and studied biology in Göttingen (1986–1991). In 1991 she joined H.-J. Gabius' group at Philipps University in Marburg. She obtained a diploma (1992) and a PhD (1996) in biochemical and cell/molecular biological analysis of protein–carbohydrate interactions. She gained experience in epidermal growth factor receptor signaling research in A. Villalobo's group at the Higher Council for Scientific Research of Spain (CSIC) in Madrid in 1994. Her research gives special emphasis to the study of glycoclusters and glycomimetics for lectin-targeted drug design, of lectin functions with medical relevance by using cell biological models produced with up-to-date technology for rationally manipulating galectin/glycoconjugate expression, and developing new diagnostic tools for histopathology.

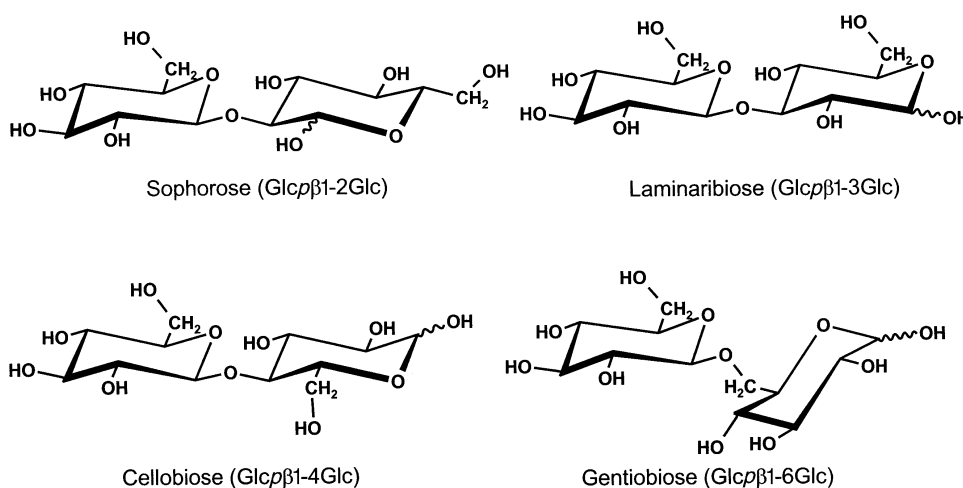


biose (Glc β 1-4Glc) is the basic structural unit of the most common carbon compound in nature, cellulose. Gentiobiose (Glc β 1-6Glc) occurs as the bitter-tasting ingredient of extracts taken from the roots of *Gentiana lutea*. This diglucoside also forms the carbohydrate part of various plant glycosides, among them amygdalin, the glycoside found in bitter almonds. This compound was instrumental in experiments delineating the famous "lock-and-key" principle. In 1894, E. Fischer investigated the stereospecificity of the enzyme emulsin and reported that it shows "eine kräftige Wirkung auf Amygdalin" (a strong effect on amygdalin; p. 2990, ref. [5]). He continues (p. 2992): "Invertin und Emulsin haben bekanntlich manche Ähnlichkeit mit den Proteinstoffen und besitzen wie jene unzweifelhaft ein asymmetrisch gebautes Molekül. Ihre beschränkte Wirkung auf die Glucoside liesse sich also auch durch die Annahme erklären, dass nur bei ähnlichem geometrischem Bau diejenige Annäherung der Moleküle stattfinden kann, welche zur Auslösung des chemischen Vorganges erfor-

most is the unsurpassed capacity of these molecules to form isomers. In contrast to nucleotides or amino acids, saccharides contain several approximately chemically equivalent sites for chain elongation and, notably, even for branching (Scheme 1). As illustrated for β -linked diglucosides in Scheme 2, chemically distinct compounds are generated when the attachment point for the unit at the reducing end is moved stepwise from the 2' to the 3', 4', or 6'-hydroxy group: Sophorose (Glc β 1-2Glc) is a constituent of plant glycosides such as the sweetener stevioside from the Composita *Stevia rebaudiana*, which is popular in Japan, or the glycosides found in root extracts of *Uzara* (South African *Xysmalobium* and *Pachycarpus* species of the Asclepiadaceae family). Laminaribiose (Glc β 1-3Glc) is a product of the partial hydrolysis of laminarin, an algal polysaccharide from *Laminaria* (seaweed; Chrysophyceae/Phaeophyceae). Laminarin also occurs in immunomodulatory fungal β 1-3/1-6-linked polysaccharides such as schizophyllan (from *Schizophyllum commune*) or lentinan (from *Lentinus edodes*). Cello-



Scheme 1. Illustration of the exquisite chemical versatility of a monosaccharide as a module for chain initiation and elongation. Whereas nucleotides (left; deoxyadenosine monophosphate) or amino acids (center; serine) form linear oligo- and polymers by 5',3'-phosphodiester-dependent or peptide-bond-dependent elongation (positions marked by arrows), monosaccharide (right; α/β -D-glucose) addition to a growing oligomer can proceed through the four hydroxy groups at C2, C3, C4, and C6 and the two anomeric hydroxy positions (see Schemes 2 and 3 for the structures of the resulting diglucosides).

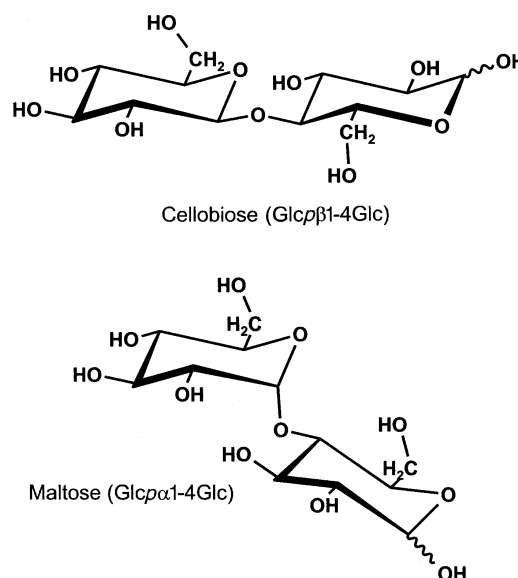


Scheme 2. Illustration of the structural series of β -diglucosides derived by shifting the position of the β 1-linked hydroxy group of the reducing-end glucose moiety from the 2' to the 3', 4', or 6'-site (shown by arrows in Scheme 1). The biological relevance of this variability is underscored by the examples of the natural occurrence of each diglucoside given in the text.

derlich ist. Um ein Bild zu gebrauchen, will ich sagen, dass Enzym und Glucosid wie **Schloss** und **Schlüssel** zu einander passen müssen, um eine chemische Wirkung auf einander ausüben zu können." (It is known that invertin and emulsin have several features in common with proteinaceous compounds and undoubtedly also harbor an asymmetrically built molecule. Their limited effect on the glucosides might thus also be explained by the assumption that only molecules with a similar geometrical design can approach one another as required for a chemical process to occur. To use a metaphor, I wish to say that enzyme and glucoside must fit like **lock** and **key** to be able to exert a chemical effect on each other).^[5] Moving on from this case study of a disaccharide, the structural features described above raise the expectation that carbohydrates will be second to no other biomolecule class in the diversity of the isomers they form, and this expectation is completely vindicated by diversity calculations. The theoretical limit of isomer diversity, that is, the total number of hexamers that can be formed with 20 different building blocks, differs tremendously between types of monomer: 6.4×10^7 hexapeptides are possible versus as many as 1.44×10^{15} hexasaccharides.^[6]

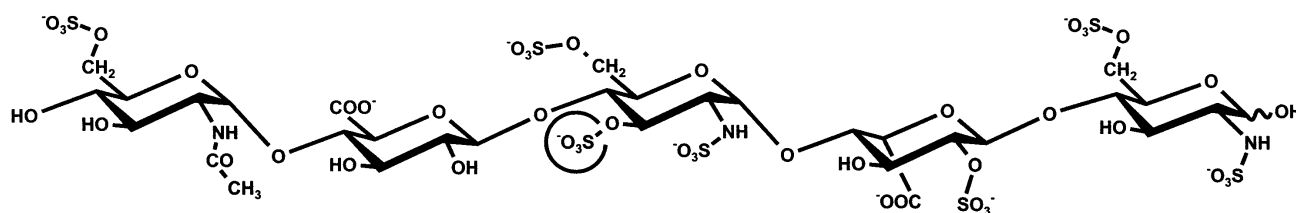
These calculations include the noted variability of the attachment point for glycosides. Moreover, the level of diversity introduced by the occurrence of the two anomeric variants at each glycosidic linkage is taken into account. The two structures in Scheme 3 illustrate that the seemingly rather minor difference in only one structural parameter between the diglucosides cellobiose and maltose effectively translates into the widely disparate properties of the polymers cellulose and starch/glycogen. Thus, in order to characterize a glycosidic linkage precisely, not one (i.e. the sequence) but three independent parameters are necessary. These parameters (for the first and second dimensions of structural diversity; for the third dimension, see Section 6) are: a) the sequence of the individual monomers, b) the individual linkage points, and c) the anomeric configuration. Amazingly, the potential for structural diversity at the level of the sequence does not end at this point.

In structural terms, a further level of diversity is accessible through the introduction of substituents. Glycosaminoglycan chains of proteoglycans found in the extracellular matrix, such as heparan sulfates, provide a telling example of how even a branchless backbone with repeating disaccharide units whose main function has been thought of as passive structural scaffold



Scheme 3. Illustration of the structural impact of anomer variation on the two otherwise structurally identical diglucosides cellobiose (building block of cellulose; see also Schemes 1 and 2) and maltose (building block of starch and glycogen).

is turned into a chain of biologically distinct microdomains through substitution.^[7] The presentation of substituents facilitates versatile multicontact recognition relevant for the coordination of cell–matrix interactions. Site-specific introduction of sulfate substituents to hydroxy/amino groups and the epimerization of D-glucuronic acid to L-iduronic acid in the basic repeating unit (GlcN-HexA)_n are the key to this *heparanomic* complexity.^[8] From the repeating core unit of the initial enzymatic polymer formation, a total of 48 different disaccharides can theoretically be formed by the ensuing modifications. A particular and rare modification pattern results in the anticoagulant pentasaccharide determinant of heparin (Scheme 4), an example of a carbohydrate compound currently used in clinical applications and an object of chemical refinement toward an optimal design.^[9] A synthetic pentasaccharide comprising the same features is now commercially available (Table 1). As alluded to above, the three-dimensional shape of the molecule comes into play too. L-Iduronic acid can undergo conformer interconversion (¹C₄ chair, ²S₀ skew boat).^[10] By adopting different



Scheme 4. Illustration of the structure of the heparin-derived anticoagulant pentasaccharide that binds antithrombin III with high specificity. The introduction of the 3'-O-sulfate group (circled) into the central substituted (N- and O-sulfated) D-glucosamine residue by a 3-O-sulfotransferase is essential for pharmacologic activity. This rare natural carbohydrate is used as a model for the development of drugs for preventing and treating venous and arterial thromboembolism.

Table 1. Examples of sugar compounds used as pharmaceuticals.

Compound	Target	Disease
acarbose	α -glucosidases (amylases)	diabetes mellitus
heparin/heparinoids	antithrombin III	thrombosis
heparin pentasaccharide (Fondaparinux)	antithrombin III (factor Xa)	thrombosis
derivatives or mimetics of 2-deoxy-2,3-dehydro-N-acetylneuraminic acid	neuraminidase	viral infection
N-butyldeoxyjirimycin	α -glucosidases (N-glycan processing)	viral infection
derivatives or mimetics of milk oligosaccharides	adhesins and toxins (lectins)	bacterial infection
GlcN-(2-O-hexadecyl) phosphatidylinositol	GPI-mannosyltransferase I	protozoan infection (e.g. African sleeping sickness)
derivatives or mimetics of sialylated/sulfated Le ^{a/x} epitopes	selectins	inflammatory reaction
D-Man	phosphomannose isomerase deficiency	congenital disorder of glycosylation Ib
L-Fuc	GDP-fucose transport	congenital disorder of glycosylation IIc (LAD II)
N-butyldeoxygalactonjirimycin and properly glycosylated β -gluco(galacto) cerebrosidase	glycosphingolipid synthesis and enzymatic degradation	glycosphingolipid storage disorders

conformations at these flexible hingelike sites, the topological display of the neighboring substituents can be easily modulated, and this substituent pattern has an impact on the contact sites for interaction with receptors, for example, antithrombin III and fibroblast growth factors (see Section 6 for further information). Needless to say, aberrations in proteoglycan synthesis and modification have been linked to developmental dysregulation in model organisms, for example, *sqv* (squashed vulva) genes in the nematode *Caenorhabditis elegans* or *sfl* (sulfateless)/*pipe* genes in the fruitfly *Drosophila melanogaster*.^[11] It is of note that sulfation has also found frequent use as a tool to form "Umlaut-like" letters in the sugar alphabet in the N- and O-glycans of glycoproteins and in glycolipids. For example, GalNAc-4-sulfate (but not GalNAc) is important for routing pituitary glycoprotein hormones, as are GalNAc/Gal/GlcNAc-6-sulfates for lymphocyte homing.^[12] So far, 31 carbohydrate sulfotransferases have been described along with their individual ligand spectrum, which underlines the sophisticated ramifications of this type of modification.^[12]

Coming back to the basic chemical features that are favorable for a role in information transfer, the amphiphilic character of carbohydrates is a boon for intermolecular interactions. This property affords multiple donor/acceptor sites for directional hydrogen bonds.^[13] Moreover, a set of suitably positioned polarized C–H bonds can be engaged in C–H/ π -electron and stacking interactions in certain cases (e.g. to bring about intimate D-Gal–Trp contact).^[13] This principle has been seen to work in organisms from various branches of the evolutionary

tree. The enterotoxin of *Escherichia coli*, plant lectins such as ricin, the agglutinin from *Erythrina corallo-dendron*, and animal galectins exploit such a contact between the ligand and an aromatic side chain, preferably that of tryptophan. This duty to make thermodynamically favorable contact to a ligand in the binding site (see Section 6 and illustrations therein for a view of a receptor–ligand complex with such a contact) is one likely reason why tryptophan is indispensable in the panel of proteinogenic amino acids. In the course of establishing contact between the sugar ligand and the aromatic ring, water molecules are dispelled from the rather hydrophobic patches of the carbohydrate, a process that makes a sizable contribution to the thermodynamic driving force of the binding process.^[13b] The selectivity of the molecular rendezvous, achieved through a combination of the above-mentioned factors, explains the exquisite way in which individual code letters, for example D-glucose and its 4-epimer D-galactose, are distinguished. Thus, a change of only one hydroxy group from the equatorial to the axial orientation keeps perturbation of the favored "tridymite" water structure minimal and is sufficient to establish distinct letters.^[14]

In summary, the structural variability introduced by changes in linkage points, anomeric position, and placement of substituents endows carbohydrates with the features necessary for a high-density coding system. In fact, not only glycosaminoglycans but all N- and O-glycans and glycolipids are representatives of the chemical diversity realized by the enzymatic machinery of glycan production.^[15] Good reasons for development of a new paradigm that views oligosaccharides as "multipurpose tools" are as follows: a) the strategic placement of glycan chains in the glycocalyx so that they reach out into the extracellular space like sensors or tentacles, b) the existence of more than 1000 known N-glycan structures (this list is continuously growing), c) the ways of marking proteins with a distinct sugar signal likened to a postal code (e.g. Man-6-phosphate and GalNAc-4-sulfate for routing of lysosomal enzymes and pituitary glycoprotein hormones, respectively) that have already been detected, and d) the overall complexity of the families of glycosyltransferases and glycan-modifying enzymes such as the sulfotransferases mentioned above.^[4a, 15a] According to recent accounts, the number of glycosyltransferase-related sequences identified has grown to more than 7200, distributed over 65 distinct sequence-derived families; about 1% of the open reading frames of each metazoan genome is calculated to be devoted to building up glycans.^[15d] It now looks like a foregone conclusion that these determinants equip cells with attractive sensor points/areas for intermolecular contact. If matched on the level of receptor proteins, the well-elaborated processes of code word generation would make sense as a way to establish a versatile communication mode involved in biosignaling, cell-specific targeting, and host-defence pathways.^[16] Families of proteins capable of "reading" the sugar-encoded messages—

besides the enzymes that tailor glycan determinants—are the missing link needed to turn structure into biological response. This assumption about information transfer pathways has been shown to be correct by the discovery of lectins.

3. Lectins: Tools to Read Sugar-Encoded Messages

The first documented observations of lectin activity were made on clumping red blood cells (Table 2). In 1860, the venom of the rattlesnake proved active in this respect: “one drop of venom was put on a slide and a drop of blood from a pigeon’s wounded wing allowed to fall upon it. They were instantly mixed. Within three minutes the mass had coagulated firmly, and within ten it was of arterial redness.”^[17] The concern that the term “coagulation” used by Mitchell might reflect the action of procoagulants but not cell agglutination was satisfactorily addressed by deliberately repeating the experiments with washed erythrocytes.^[18] The paper by Flexner and Noguchi in which this work is reported was in fact introduced by Mitchell who commented on it as follows: “I have long desired that the actions of venoms upon blood should be further examined. I finally indicated in a series of propositions the direction I wished the inquiry to take. Starting from these the following very satisfactory study has been made by Professor Flexner and Dr. Noguchi. My own share in it, although so limited, I mention with satisfaction.”^[18] Only a few years later, in 1906, an agglutinating activity of activated-complement-coated erythrocytes was detected in bovine serum. To allow the reader to follow the major historical events in this field, we have listed this finding in Table 2. As with snake venom, the active protein of bovine serum was later biochemically characterized as a C-type lectin (see also Section 4), in this case from the subgroup of collectins named conglutinin, which binds to the Man₈/Man₉ N-glycan of human iC3b at Asn917 in the α chain of complement glycoprotein C3.^[19]

The assay used to look at haemagglutination was also instrumental to the discovery of the cell-bridging capacity of proteins in plant extracts, initially that of toxic castor bean extract.^[20] Stillmark remarked in his M.D. thesis, published in 1888: “Das Ricin bewirkt in defibriertem serumhaltigem Blute eine Zusammenballung der rothen Blutkörperchen unter Bildung einer fibrinähnlichen Substanz.” (Ricin causes a conglomeration (or agglutination) of the red blood corpuscles in defibrinated serum-containing blood that yields a fibrin-like substance).^[20] The discovery that plant extracts are rich sources of agglutinins made possible the first purification of such a protein (named concanavalin A) by crystallization (“If jack bean extracts are covered with toluene and simply allowed to stand exposed to the air for several weeks, this protein is precipitated as beautifully formed crystals having a diameter of about 0.1 mm”)^[21] and the

Table 2. Brief historical account of lectinology.^[a]

1860	Observation of blood “coagulation” by rattlesnake venom (S. W. Mitchell)
1888	Detection of erythrocyte agglutination by protein fractions from castor beans and other plant seeds (H. Stillmark)
1891	Toxic plant agglutinins applied as model antigens (P. Ehrlich)
1898	Introduction of the term “haemagglutinin” or phytohaemagglutinin for plant proteins that agglutinate red blood cells (M. Elfstrand)
1902	Detection of bacterial agglutinins (R. Kraus)
1902	Demonstration that blood “coagulation” by snake venom (later shown to depend on a C-type lectin) observed in 1860 was not caused by blood clotting but by cell agglutination (S. Flexner, H. Noguchi)
1906	Detection of an agglutinin in bovine serum (later characterized as the C-type lectin conglutinin) that acts on activated complement-coated erythrocytes (J. Bordet, F. P. Gay)
1907	Detection of nontoxic agglutinins in plants (K. Landsteiner, H. Raubitschek)
1913	Use of intact cells for the purification of lectins (R. Kobert)
1919	Crystallization of a lectin, concanavalin A (J. B. Sumner)
1936	Precipitation of starch, glycogen, and mucins by concanavalin A and its interaction with the stromata of erythrocytes define the carbohydrate as a ligand (J. B. Sumner, S. F. Howell)
1941	Detection of viral agglutinins (G. K. Hirst)
1947–1948	Detection of lectins specific for human blood groups (W. C. Boyd, K. O. Renkonen)
1952	Carbohydrate nature of blood group determinants proven by lectin-mediated agglutination and its sugar-dependent inhibition (W. M. Watkins, W. T. J. Morgan)
1954	Introduction of the term “lectin” for plant agglutinins, primarily for those that are blood-group specific (W. C. Boyd)
1960	Detection of the mitogenic potency of lectins toward lymphocytes (P. C. Nowell)
1965	Application of affinity chromatography for the isolation of lectins (I. J. Goldstein, B. B. L. Agrawal)
1972	Determination of the amino acid sequence and three-dimensional structure of a lectin, concanavalin A (G. M. Edelman, K. O. Hardman, C. F. Ainsworth et al.)
1972–1977	Discovery of impaired synthesis of a marker for glycoprotein (lysosomal enzymes) routing as the cause of a human disease (mucopolysaccharidosis II) and identification of the marker as Man-6-phosphate, the ligand for P-type lectins (E. F. Neufeld et al.; W. S. Sly et al.)
1974	Isolation of a mammalian Gal/GalNAc-specific lectin from the liver (G. Ashwell)
1978	First conference focusing on lectins and glycoconjugates, termed Interlec (T. C. Bøg-Hansen)
1979	Detection of endogenous ligands for plant lectins (H. Rüdiger)
1983	Detection of the insecticidal action of a plant lectin (L. L. Murdock)
1984	Isolation of lectins from tumors (H.-J. Gabius; R. Lotan, A. Raz)
1985	Discovery of immobilized glycoproteins as pan-affinity adsorbents for lectins (H. Rüdiger)
1987	Introduction of neoglycoconjugates for localization of tissue lectins for tumor diagnosis (H.-J. Gabius et al.)
1989	Detection of the fungicidal action of a plant lectin (W. J. Peumans)
1992–1993	Identification of impaired synthesis of lectin (selectin) ligands by defective fucosylation as the cause for leukocyte adhesion deficiency type II (A. Etzioni et al.)
1995	Structural analysis of a lectin–ligand complex in solution by NMR spectroscopy (J. Jiménez-Barbero et al.)
1996–1998	Detection of differential conformer selection by plant and animal lectins (H.-J. Gabius et al.; L. Poppe et al.)
2001–2002	Advances in lectinology and glycosciences honored by dedication of special issues of <i>Biochim. Biophys. Acta</i> , <i>Biochimie, Biol. Chem.</i> , <i>Cells Tissues Organs</i> , <i>Chem. Rev.</i> , <i>Curr. Opin. Struct. Biol.</i> , <i>J. Agric. Food Chem.</i> (Liener symposium), and <i>Science</i> to these topics

[a] Extended and modified from ref. [16d].

demonstration of its interaction with carbohydrate groups (Table 2). As summarized by Sumner and Howell in 1936, "concanavalin A unites with some constituent of the stromata and, since concanavalin A unites with starch, glycogen, mucins, etc., it is possible that this may be a carbohydrate group in a protein."^[22] Since the activities of several plant and animal haemagglutinins towards reaction with erythrocytes of different ABO blood group status resemble those of serum antibodies initially observed by Creite (1869) and Landois (1875) and referred to as isoagglutinins by Landsteiner in 1900,^[23] the resulting classification of the haemagglutinins as antibody-like substances sounds logical. The following quotation explains Boyd's reason for introducing the term *lectin* in 1954, a term that continues to be commonly used today: "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."^[24] By building on the pioneering observations made by Sumner and Howell, on the detection of haptenic inhibition of antibody-antigen reactions by Landsteiner and van der Scheer, and on the description of blood-group-specific lectins by Renkonen and Boyd (cited above),^[22,23d-f] a milestone of lectin application was established shortly before 1954 (see Table 2). This breakthrough was the inhibition of haemagglutination, mediated by eel (*Anguilla anguilla*) serum and seed extracts of the Leguminosa *Lotus tetragonolobus*, by L-fucose. These key experiments led to the determination of "the biochemical basis of blood group ABO and Lewis antigenic specificity"^[25e] (for further listings of the course of lectin research history, see Table 2).

To reach the present version of the term *lectin*, its definition had to be subjected to several refinements. The experimental focus on agglutination, which requires at least bivalency for the bridging of two cell surfaces, was dropped completely in the course of this process. The three criteria that must currently be met by a (glyco)protein for it to qualify as a member of the lectin family are given below.^[26]

a) Carbohydrate-binding activity

Assays monitoring binding to carrier-immobilized carbohydrate ligands of (neo)glycoconjugates are now commonly used to detect and quantify lectin activity, irrespective of the presence of bridging functionality.^[4a,27] The presence of a carbohydrate recognition domain (CRD) linked with other bioactive modules in a mosaic-like protein (see also Sections 4 and 5) makes it possible to assign bi- and multifunctional proteins to different protein families.

b) Distinction from immunoglobulins

In the original definition of a lectin given by Boyd in 1954,^[24] the groups of immunoglobulins (Ig), such as IgG or IgM, are

deliberately excluded. It should be noted that the animal lectins of the I-type class with a distal V-set module and C2-set domains belong to the Ig superfamily and that various lectins, such as galectins, as well as C- and I-type lectins, are produced from lymphocytes along with antibodies.^[28]

c) Distinction from enzymes tailoring free saccharides/glycan chains of glycoconjugates, and from sensor or carrier proteins for free mono- or oligosaccharides

Any glycosyltransferase, glycosidase, or enzyme that modifies its cognate carbohydrate (e.g. the sulfotransferases or epimerases), as well as transport/chemotaxis receptors for free mono-, di-, or oligosaccharides are excluded from the lectin family.

With this explanation of the generic name for (glyco)proteins that read sugar-encoded messages in mind, it is instructive to examine the diversity of these proteins in plants and animals. If lectins were rare inventions of nature, then communication with sugar code words would surely be restricted to only a few messages that can be decoded.

4. Plant Lectins: Occurrence, Functions, and Applications

The richest sources of plant lectins are the seeds or, more generally, the storage organs of plants. For most plants studied so far, lectins have been prepared from the seeds, but roots, tubers, bulbs, bark, or leaves have also served as starting materials for the isolation of lectins.^[29] As emphasized above in the context of the inter- and intrafamily diversity of glycosyltransferases (see the last paragraph of Section 2), the wide distribution of lectins is a strong argument for their physiological relevance. Table 3 lists families of higher plants, as defined by the rules of botanical taxonomy, with the numbers of lectin-bearing species in each family. Algal, fern, and fungal lectins are not included. Since an activity assay solely with haemagglutination without proper controls can yield false-positive results, we limited the compilation to those cases for which further unambiguous evidence for lectin presence is available. The overwhelming majority of lectins characterized up to now has been found in the Angiospermae section. Among these, about three-quarters of the lectin-bearing species belong to the Dicotyledoneae and almost 90% of these to the Archichlamydeae subclass of the dicot class. Leguminosae played an important role in the early history of lectinology, as outlined above (see also Table 2), and still hold a prominent position in the field. However, to avoid misinterpretation of our systematic compilation, we must add that the search for lectins has not really been carried out strategically by following the rules of botanic systematics and searching species by species. It is thus likely that the literature-based numbers given in Table 3 will promptly increase when researchers begin doing so. Studies have so far often focused on economically relevant plants. Besides the advantage of easy access to the starting material, reports on compounds from plants of nutritional value are sure to find a wide readership. Consequently, the occurrence of lectins in plants outside the remit of modern agriculture is proba-

Table 3. Systematic coverage of the occurrence of plant lectins.

Section	Class	Subclass	Order	Family	Number of known lectin-bearing species	
Angiospermae	Dicotyledoneae	Archichlamydeae	Salicales	Salicaceae	2	
			Fagales	Fagaceae	2	
			Urticales	Cecropiaceae	1	
				Urticaceae	1	
				Moraceae	18	
			Santalales	Loranthaceae	1	
				Viscaceae	3	
			Centrospermae	Amaranthaceae	7	
				Caryophyllaceae	1	
				Chenopodiaceae	1	
				Phytolaccaceae	1	
			Cactales	Cactaceae	2	
			Magnoliales	Lauraceae	2	
			Ranunculales	Ranunculaceae	1	
			Guttiferales	Theaceae	1	
			Papaverales	Cruciferae	4	
				Papaveraceae	1	
			Rosales	Crassulaceae	1	
				Leguminosae	140	
				Saxifragaceae	1	
			Geraniales	Euphorbiaceae	13	
		Sapindales	Sapindaceae	3		
		Celastrales	Celastraceae	2		
		Rhamnales	Rhamnaceae	1		
			Vitaceae	1		
		Thymelaeales	Eleagnaceae	1		
		Violales	Passifloraceae	2		
		Cucurbitales	Cucurbitaceae	22		
		Myrtiflorae	Myrtaceae	1		
		Umbelliflorae	Araliaceae	1		
			Umbelliferae	3		
			Ebenaceae	1		
		Metachlamydeae	Tubiflorae	Solanaceae	5	
				Lamiaceae	9	
				Convolvulaceae	5	
				Pedaliaceae	1	
				Verbenaceae	1	
			Dipsacales	Caprifoliaceae	5	
			Campanulales	Compositae	2	
			Monocotyledoneae	Helobiae	Alismataceae	1
				Liliiflorae	Alliaceae	6
					Amaryllidaceae	11
					Dioscoreaceae	1
	Iridaceae	6				
	Liliaceae	20				
Graminales	Gramineae	9				
Spathiflorae	Araceae	17				
Cyperales	Cyperaceae	1				
Scitamineae	Musaceae	2				
Microspermae	Orchidaceae	6				
	Araucariaceae	1				
Gymnospermae	Coniferopsida	Coniferae	Araucariaceae	1		
			Pinaceae	3		

bly underestimated. Another factor that may have played a role in the count is that such plants often have tiny seeds.

The most popular method for tracing lectin presence is still to test plant extracts for their ability to agglutinate cells, usually human or other mammalian erythrocytes. This classical method, already used more than a century ago by Mitchell^[17] and Stillmark,^[20] excels because of its simplicity. However, the technique suffers from several noteworthy disadvantages. Plant extracts can contain active material such as tannins that

leads to the above-mentioned false-positive results. The presence of lipids can also lead to misinterpreted results, and erythrocytes tend to agglutinate spontaneously in the presence of only moderate concentrations of bivalent metal ions. Erythrocytes are sensitive to surface-active substances such as saponins, so lectins may easily be overlooked in their presence. Moreover, an agglutination assay only detects lectins that are at least bivalent and can therefore link cells, a factor noted above in criterion (a) of the lectin definition (see Section 3).

Therefore, screening methods have been developed that capitalize on the carbohydrate-binding capacity of lectins. Chemically tailored neoglycoconjugates present carrier-immobilized carbohydrate ligands for interaction, and complex formation is then picked up analytically by using any suitable label.^[27,30] To avoid radioactive labeling, the intrinsic activity of enzymes that are naturally glycosylated, such as horseradish peroxidase, or that can serve as acceptors for glycans through chemical conjugation, such as *E. coli* β -galactosidase, is used to detect any lectin-like activity presented on a matrix.^[31] Recently developed methods employ the microarray technology. This approach can be combined with combinatorial synthesis, which illustrates the emerging importance of the interface between lectin research and carbohydrate chemistry.^[32] Needless to say, arrays will prove instrumental in the definition of ligands with optimal affinity and selectivity, a factor of relevance for research aiming to extend the contents of Table 1 in the future. However, at present these methods are too sophisticated for general use. Expensive equipment and considerable expertise are required to master the chemical syntheses and analytical evaluation techniques. In consequence, even recent studies dealing with newly discovered lectins rely on cell agglutination as an analytical tool. The shortcomings of this approach are generally addressed by using controls to prove inhibition of the agglutination by haptenic sugar, as in the elucidation of the determinants of the AB0 histo-blood group epitopes fifty years ago.^[25]

Once lectin activity had been detected, the next step in the characterization pathway, regardless of the source of the material, is isolation of the lectin(s). This step can certainly be performed by standard protocols for protein purification, which include ion exchange, size exclusion, and hydrophobic chromatography. Investments of time and effort are reduced by taking advantage of the highly efficient method of affinity chromatography on immobilized carbohydrates or glycoconjugates. A very simple means of applying this technique is to use naturally occurring polysaccharides such as dextrans. These compounds are high-affinity adsorbents for glucose-binding lectins such as concanavalin A and pea or lentil agglutinins when the polymer chains are cross-linked. Surprisingly, the enormous potential of this method was not initially realized. As was recently pointed out in a commentary on the path of the lectins "from obscurity into the limelight" by Sharon,^[33] the manuscript pioneering this approach did not find favorable review at first: "Irwin J. Goldstein from the University of Michigan at Ann Arbor, a leading lectin researcher to this very day, tells that when he sent a note, in 1963, to *Biochemical and Biophysical Research Communications* describing the purification of concanavalin A by affinity chromatography, it was rejected forthright because 'this represents a modest advance in an obscure area.' The note was eventually published in *Biochemical Journal*^[34a] and affinity chromatography soon became the method of choice for lectin isolation" (see Table 2). Among the procedures used to conjugate a saccharide to the matrix, we found divinyl sulfone activation particularly easy in handling and efficient in terms of final lectin yields.^[34b-d] To broaden the scope of one-step lectin purification, it is convenient to covalently couple

not only saccharides but also naturally occurring glycoproteins to the resin. For this purpose, hog gastric mucin or hen ovomucoid, both easily available in large amounts, was successfully employed.^[35] The prevailing method used to elute the lectin exploits the haptenic sugar as a competitive inhibitor. Problems arise when binding is directed to extended glycans, as is the case for *Phaseolus* bean lectins (or phytohaemagglutinins (PHAs), a formerly used generic name for plant lectins; see below). The presence of these lectins is the biochemical cause of the nausea that results from eating insufficiently cooked beans (see also Table 4). In such instances of binding to the extended glycans, lectin elution from the resin can be performed by lowering the pH value of the buffer. If the lectin is too sensitive to withstand an acidic medium, desorption with a borate-containing buffer offers a simple and affordable alternative. The elution profiles that result from the use of these two protocols are illustrated in Figure 1. Successive elution with haptenic sugar and borate was helpful for purification of distinct lectins from the same source that differ in carbohydrate specificity. Figure 1 A shows that isolectin family I (specific for Gal/GalNAc) found in *Griffonia simplicifolia* seed extracts can be easily separated from the type II lectin (specific for GlcNAc_n). Figure 1 B illustrates that this procedure even allows closely related isolectins such as the *Phaseolus* bean lectins to be resolved. When the concentration of the eluant borate was increased stepwise, it was possible to obtain the five isolectins L₄, L₃E, L₂E₂, LE₃, and E₄ in separate fractions.^[36] The isolectin L₄ (listed by commercial suppliers as PHA-L₄ or phytohaemagglutinin L₄) is a popular laboratory tool used as a mitogen for lymphocytes and the chromatographic method described gives remarkably easy access to pure material without contamination by the isoagglutinin E₄ or the other three forms, as explained in detail in the figure legend.

The members of the diverse group of plant lectins that are studied and used most frequently are listed in Table 4. The leading position is held by concanavalin A, the "classical" Man/Glc-binding lectin from Jack beans (see above and Table 2 for the central role of this lectin in the history of lectinology). The obtainable yield of concanavalin A from seed material is about 2 g per 100 g and it is chemically stable, key factors for its initial isolation by crystallization (see above). Once purified, the lectin can undergo numerous chemical modifications. All these properties are very favorable for chemical, biochemical, and biomedical applications (see Table 5 for a summary of research areas in which plant lectins are used as tools). These facts explain why this lectin has attained its status as a reliable and popular workhorse, especially for carbohydrate chemists looking for a lectin to use in an attempt to prove the ligand properties of a sugar compound attached to a new synthetic scaffold. The other lectins listed in Table 4 are capable of following the role model concanavalin A, although they are less prominently used in research. These compounds form a panel of probes for isolation and structural characterization of glycoconjugates (glycoproteins, glycolipids, or polysaccharides), as well as use in various assays in cell biology, histochemistry, and the medical sciences (Table 5).^[26g,34d,37] The size of the panel of lectins with related specificities (for a selection of frequently

Table 4. Examples of plant lectins to illustrate the inter- and intrafamily diversity of these proteins.^[a]

Plant species and name/ abbreviation of lectin	Family	Mono- or disaccharide specificity	Comments
<i>Canavalia ensiformis</i> (concanavalin A, ConA)	Leguminosae	Man/Glc	cheapest and most popular lectin; first lectin isolated by crystallization and demonstrated to interact with carbohydrate (see text and Table 2 for details)
<i>Ricinus communis</i> (ricin)	Euphorbiaceae	Gal	ribosome-inactivating protein, type II (RIP II), used for generating immunotoxins; biohazard
<i>Triticum vulgare</i> (WGA)	Gramineae	(GlcNAc) ₁₋₃ , Neu5Ac	potential function in plant defence mechanisms
<i>Phaseolus vulgaris</i> (PHA)	Leguminosae	no simple carbohydrate known	isolectin L ₄ is a strong mitogen for T-lymphocytes, isolectin E ₄ is a strong erythrocyte agglutinin (see Figure 1 B for chromatographic isolectin separation); distinguish between bisected and nonbisected N-glycans; cause of severe gastrointestinal irritation when ingested in insufficiently cooked beans
<i>Glycine max</i> (SBA)	Leguminosae	GalNAc/Gal	cell sorting, bone marrow purging
<i>Pisum sativum</i> (PSA)	Leguminosae	Man/Glc	binding of N-glycans enhanced by core fucosylation
<i>Viscum album</i> (VAA, viscumin)	Viscaceae	Gal	RIP II used for generating immunotoxins, constituent of proprietary mistletoe extracts (immunomodulatory and growth stimulatory for tumor cells in vitro and in vivo at low doses; see text for details)
<i>Arachis hypogaea</i> (PNA)	Leguminosae	Gal, Galβ3GalNAcα (TF-antigen)	very popular in histochemistry; separates immature from mature thymocytes
<i>Lens culinaris</i> (LCA)	Leguminosae	Man/Glc	binding of N-glycans enhanced by core fucosylation; lymphocyte mitogen
<i>Dolichos biflorus</i> (DBA)	Leguminosae	GalNAcα3GalNAc, GalNAc	cell sorting, agglutinates blood group A erythrocytes
<i>Griffonia simplicifolia</i> (GSA-I)	Leguminosae	Gal/GalNAc	isolectin GSA-I-A ₄ agglutinates blood group A erythrocytes, isolectin GSA-I-B ₄ blood group B erythrocytes
<i>Griffonia simplicifolia</i> (GSA-II)	Leguminosae	(GlcNAc) _n	insecticidal activity, potential defence role
<i>Artocarpus integrifolia</i> (jacalin)	Moraceae	Gal (Man, TF-antigen)	used for isolation of IgA ₁ and mucins, mitogenic for CD4 ⁺ T-cells
<i>Solanum tuberosum</i> (STA)	Solanaceae	(GlcNAc) _n	potential function in plant defence mechanisms
<i>Galanthus nivalis</i> (GNA)	Amaryllidaceae	Man	does not bind Glc as the Leguminosae lectins do, application for insect and nematode defence in transgenic crop plants tested, antiretroviral activity in vitro, selective agglutination of rabbit but not human erythrocytes
<i>Ulex europaeus</i> (isolectin UEA-I)	Leguminosae	L-Fuc	agglutinates blood group O(H) erythrocytes; selective marker for endothelial cells of primates
<i>Erythrina corallodendron</i> (ECA)	Leguminosae	Galβ4GlcNAc, Gal, GalNAc	mitogen for human lymphocytes
<i>Vicia faba</i> (VFA)	Leguminosae	Man/Glc	binding of N-glycans enhanced by core fucosylation
<i>Sambucus nigra</i> (SNA)	Caprifoliaceae	Neu5Acα6Gal/ GalNAc, (Gal/GalNAc)	probe for sialylated glycoconjugates, e.g. in thymocyte differentiation
<i>Abrus precatorius</i>	Leguminosae	Gal	RIP II used for generating immunotoxins
<i>Lotus tetragonolobus</i> (LTA)	Leguminosae	L-Fuc	agglutinates red cells of blood group O(H), instrumental to the definition of α-L-fucose as a crucial O(H) epitope (see Table 2)
<i>Lycopersicon esculentum</i>	Solanaceae	(GlcNAc) _n	potential function in plant defence mechanisms; marker of endothelium of small vessels in rats
<i>Phaseolus lunatus</i> <i>limensis</i>	Leguminosae	GalNAcα3[Fucα2]Gal, GalNAc	agglutinates blood group A erythrocytes
<i>Datura stramonium</i> (DSA)	Solanaceae	(GlcNAc) _n	potential function in plant defence mechanisms
<i>Maackia amurensis</i> (MAA)	Leguminosae	Neu5Acα3Gal/ GalNAc	probe for sialylated glycoconjugates
<i>Phytolacca americana</i> (PWM)	Phytolaccaceae	GlcNAc	known as pokeweed mitogen; detected in 1969 in the course of investigating a fatality associated with ingestion of pokeweed berries
<i>Bauhinia purpurea</i> (BPA)	Leguminosae	GalNAcβ3GalNAc, GalNAc	enrichment of B lymphocytes, isolation of T cells producing IL-2
<i>Urtica dioica</i> (UDA)	Urticaceae	(GlcNAc) _n	antifungal activity
<i>Hevea brasiliensis</i> (hevein)	Euphorbiaceae	(GlcNAc) _n	antifungal activity; allergen in rubber products of poor quality
<i>Maclura pomifera</i> (MPA)	Moraceae	T-antigen > Tn- antigen	mitogen for lymphocytes

[a] The order of the list reflects the share of attention given to each lectin in the literature.

used plant lectins, see Table 4) ensures that the optimal tool for a defined purpose can always be found. For example, LCA can be used when the cells to be desorbed from a lectin-con-

taining solid matrix must be handled under gentle conditions, in contrast to concanavalin A, with which harsher conditions are required since binding is comparatively tight.^[38] A frequent-

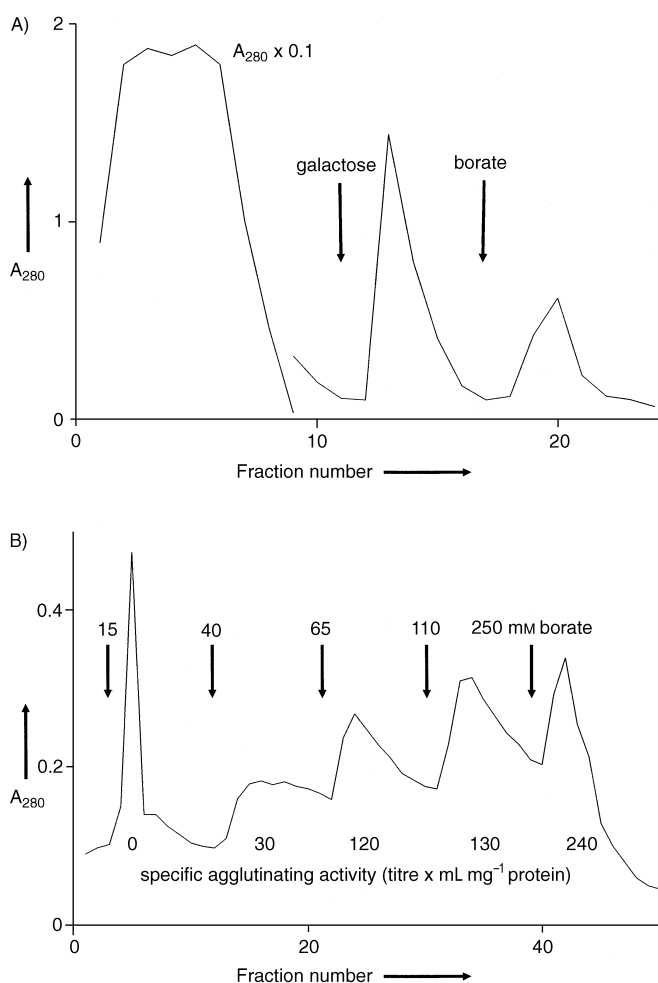


Figure 1. Illustration of the chromatographic purification and separation of plant lectins from the same species and source (see Table 4 for further information on these lectins) by using the glycan chains of immobilized glycoproteins as affinity ligands. A) Successive elution with 25 mM D-galactose and 50 mM borate from a column bearing desialylated hog gastric mucin as the affinity ligand and loaded with plant extract as previously described^[36] resulted in purification of Gal/GalNAc-specific *Griffonia simplicifolia* agglutinin I (GSA-I; subunit $M_r = 30/32$ kDa) and GSA-II ((GlcNAc)_n-specific; subunit $M_r = 28$ kDa). B) Stepwise increases in the borate concentration in the elution buffer resulted in desorption of the five *Phaseolus vulgaris* isoagglutinins from immobilized ovomucoid. Elution started with PHA-L₄ (subunit $M_r = 31$ kDa) at 15 mM borate and finally reached PHA-E₄ at 250 mM borate. Elution was monitored by measuring the absorption at 280 nm (A_{280}) and the agglutination activity, as described previously.^[35] The latter assays revealed that potency increases from E₁L₃ to E₂L₂ to E₃L₁, and finally E₄ (subunit $M_r = 34$ kDa), the strongest erythrocyte agglutinin. Lymphocyte stimulation increased from E₄ (20-fold at 37 $\mu\text{g mL}^{-1}$) to L₄ (24-fold at 8 $\mu\text{g mL}^{-1}$).

ly encountered application concerns the mitogenic activity of lectins (Table 5). The fact that plant lectins can affect lymphocyte activity and proliferation has led to suggestions that the laboratory tools could be introduced as immunomodulatory therapeutic agents in clinical applications. The example of the galactoside-specific mistletoe lectin (VAA, formerly ML-1), a constituent of proprietary extracts used in Austria, Germany, and Switzerland, shows that immune functions such as secretion of proinflammatory cytokines or priming of granulocytes/

Table 5. Versatility of plant lectins as research tools.^[a]

Biochemistry

detection of defined carbohydrate epitopes of glycoconjugates in blots or on thin-layer chromatography plates
 purification of lectin-reactive glycoconjugates by affinity chromatography
 glycan characterization by serial lectin affinity chromatography (lectin affinity capture)
 glycome analysis (glycomics)
 quantification of lectin-reactive glycoconjugates in enzyme-linked lectin-binding assays (ELLA)
 quantification of activities of glycosyltransferases/glycosidases by lectin-based detection of products of enzymatic reaction
 model reagents for the assessment of the ligand functionality of carbohydrate-presenting scaffolds (e.g. glycodendrimers)

Cell biology

characterization of intracellular assembly, routing, and cell surface presentation of glycoconjugates in normal and genetically engineered cells (glycomics profiling, spatially defined)
 selection of cell variants (mutants, transfectants) with altered lectin-binding properties as models for dissecting glycosylation machinery and glycan functionality (glycomics profiling, functionally defined)
 fractionation of cell populations
 modulation of the proliferation and activation status of cells and dissection of the involved signal pathways
 model substratum for study of cell aggregation, adhesion, and migration

Medicine

detection of disease-related alterations of glycan synthesis by lectin cyto- and histochemistry
 histo-blood group typing and definition of secretor status
 quantification of aberrations of cell surface glycan presentation, e.g. in malignancy
 cell marker for diagnostic purposes including marking infectious agents (viruses, bacteria, fungi, parasites)
 cell marker for functional assays to pinpoint defects in cell activities such as mediator release

[a] Extended and modified from ref. [26g].

activity of NK cells can indeed be stimulated at nontoxic doses of lectin (VAA concentration needed to elicit *in vivo* effects: 1–2 ng kg⁻¹ body weight, given subcutaneously).^[39] However, this immunomodulatory capacity is unlikely to have a clinical perspective because lectin-dependent increase in the proliferation (and also metastatic capacity) of tumor cells has likewise been described for cell lines, histocultures of human tumors, and animal models *in vivo* (*primum non nocere*).^[26g,40] Enhanced availability of proinflammatory cytokines might account for this effect. In more general terms, it is becoming evident that these immune factors can also trigger growth responses in malignant cells.^[41] Our understanding of how immune/inflammatory cells influence tumor growth and neovascularization is thus undergoing a paradigmatic shift. This development is reflected in the statement that these cells “conspire with cancer cells in promoting” (rather than inhibiting) these processes,^[41e] which has implications for the way we look at immunostimulation in cancer patients. As a consequence, immunomodulation by a lectin can exert a nonbeneficial influence on tumor parameters. Case studies, including an account of a study on melanoma patients in which treatment with a proprietary mistle-

toe extract appeared to decrease the lengths of overall survival and disease-free intervals of patients with lymph node metastases, underline concerns that herbal treatment modalities in alternative/complementary medicine may not be free of serious risk potential.^[42] A recent review on the controversial issue of the clinical use of *Viscum album* extracts in cancer treatment concluded that "mistletoe therapy has the potential to harm cancer patients."^[42d] These data also caution against intuitive expectations that in vitro modulation of one or more immune parameters (plant lectins are very active elicitors of such a response) will automatically be clinically beneficial.

While knowledge on the distribution of lectins in plants has taken enormous strides as a result of documentation of their widespread occurrence, it is difficult to produce a succinct compendium of their functions in situ. In principle, each lectin might have distinct functions at the site of expression and through interplay with binding partners in the cell and the extracellular environment, an idea also valid for animal lectins. One particular protein can thus take care of several tasks. Powerful techniques used to regulate lectin presence on the level of gene expression in vitro and in vivo that were a boon for the elucidation of lectin functions in animals are starting to be exploited in plants,^[43] so progress in refining and extending current knowledge of the functions of plant lectins will not be long in coming. Table 6 summarizes current concepts on this topic, together with examples of lectins with the activities concerned. Free oligosaccharides also convey biochemical messages and, although their binding partners do not fit the lectin definition given herein in every respect (see criterion (c) in Section 3), our survey would not be complete without paying tribute to this aspect of oligosaccharide behavior. Indeed, an emerging topic in the area of protein-carbohydrate interaction is the way in which oligosaccharide elicitors interact with their often ill-defined receptors.^[44] These elicitors are products of the degradation of plant/fungal cell walls or lipochitoooligosaccharides (*Nod* factors involved in the chemical cross-talk between nitrogen-fixing soil bacteria and their leguminous host plant). Of note is the observation that the rhizobial nodulation protein *NodC*, a glycosyltransferase responsible for GlcNAc incorporation within the synthetic pathway of the *Nod* factors, does not appear to be a unique invention of the evolutionary process because similar sequences have been found in *Xenopus*, zebrafish, and mouse proteins^[44e] (the alternative route to chitoooligosaccharides employs endochitinases). Members of this glycosylhydrolase family (no. 18), like many other enzymes involved in bacterial/fungal carbohydrate polymer degradation

Table 6. Functions of plant lectins.^[a]

	Activity	Example of lectin
external activities	protection from fungal attack	<i>Hevea brasiliensis</i> (rubber tree), <i>Urtica dioica</i> (stinging nettle), <i>Solanum tuberosum</i> (potato)
	protection from herbivorous animals	<i>Phaseolus vulgaris</i> (French bean), <i>Ricinus communis</i> (castor bean), <i>Galanthus nivalis</i> (snow-drop), <i>Triticum vulgare</i> (wheat)
	involvement in establishing symbiosis between plants and bacteria	<i>Pisum sativum</i> (common pea), <i>Lotononis bainesii</i> (miles lotononis), <i>Arachis hypogaea</i> (peanut), <i>Triticum vulgare</i> (wheat), <i>Oryza sativa</i> (rice)
Internal activities	storage proteins	valid for all lectins
	ordered deposition of storage proteins and enzymes in protein bodies and mediation of contact between storage proteins and protein body membranes	<i>Pisum sativum</i> (common pea), <i>Lens culinaris</i> (lentil), <i>Glycine max</i> (soybean), <i>Oryza sativa</i> (rice)
	modulation of enzymatic activities such as phosphatase activity	<i>Secale cereale</i> (rye), <i>Solanum tuberosum</i> (potato), <i>Pleurotus ostreatus</i> (oyster mushroom), <i>Glycine max</i> (soybean), <i>Dolichos biflorus</i> (horse gram)
	participation in growth regulation	<i>Medicago sativa</i> (alfalfa), <i>Cicer arietinum</i> (chick pea)
	adjustment to altered environmental conditions	<i>Triticum aestivum</i> (winter wheat)

[a] For further information on carbohydrate specificities, see Table 4. For a recent review, see ref. [26g].

and bacterial sialidases, often contain a second domain besides their catalytic section. This domain has exclusive carbohydrate-binding activity that allows it to guide and firmly position the hydrolytic center.^[45] This close cooperation of the two sites toward polysaccharide degradation (see criterion (c), Section 3) explains the reluctance of researchers to count these enzymes with a carbohydrate-binding module as lectins. Equivalent proteins that bring a catalytic and a carbohydrate-binding domain together are found in both plants (e.g. β -galactosidases and endo- β -1-4-glucanase in strawberry) and animals (see below).^[26g] A recent example of clinical interest implicates mutations affecting a putative glycogen-binding domain (CBD-4) of laforin in disease onset, which is supposedly a result of mispositioning of the phosphatase activity. This domain is the product of the EPM2A (epilepsy of progressive myoclonus type 2) gene, which is defective in Lafora disease.^[46] The detection of a chitinase-related receptor-like kinase (CHRK1) in tobacco and of receptor-like protein kinases with extracellular lectinlike domains in thale cress (*Arabidopsis thaliana*) and lombardy poplar (*Populus nigra* var. *italica*) suggests the existence of an outside/inside signaling route for the transfer of sugar-encoded messages into the plant cell.^[16c,47] Although it is tempting to draw analogies between plant and animal lectin functions, this approach should not be taken too far. The enzymatic apparatus of glycan synthesis is not identical in plants and animals, so the patterns of potential natural ligands for evolutionary adaptation diverge. For example, the structures of the core regions of complex-type *N*-glycans in plants differ from the structures in animals in that the plant glycans harbor two unique additions to the substitution pattern of the core region (the α 1-3-linked fucose attached to the proximal GlcNAc residue and the β 1-2-linked xylose in the core mannose residue). Mammalian cells, in contrast, have relatively abundant supplies of β 1-4-galactosyltransferases, α 1-2/6-fucosyltransferases, and sialyltransferases.^[48]

With regard to the lectins, the take-home messages of this section are clear: a) plant lectins are widely found, and b) these lectins are endowed with various functional activities through their carbohydrate-binding activity. The number of ways in which plant lectins are successfully applied as tools (summarized in Table 5) intimates that far-reaching opportunities would be missed if the system of complementary molecular interaction exploited with these laboratory tools were not naturally operative in animals. In the search for biochemical hardware for programmed "lock-and-key" interactions, lectins and glycans have thus been judged to be "reasonable candidates."^[49] To gauge the extent to which our knowledge of animal lectins has advanced over the last few decades (see also Table 2), it is informative to recall the scepticism with which this concept was confronted three decades ago. At that time, the view was held (as for antibodies and antigens) that lectins and oligosaccharides "are unlikely to provide a general mechanism of recognition and communication of the type postulated by Weiss^[50] because one member of each pair is probably not a common cell component. The known lectins generally originate from plants or invertebrates..."^[49] Isolation of the C-type hepatocyte asialoglycoprotein receptor in 1974 and the galectin (electrolectin) from the electric eel (*Electrophorus electricus*) in 1975, along with later work in 1980 leading to the biochemical verification of the presence of lectins in snake venom (originally discovered by Mitchell in 1860),^[17,51] as well as the ensuing work has markedly changed this view.^[28c,e] The next section is a brief survey of the current status of knowledge of lectin occurrence and functions in animals.

5. Animal Lectins: Occurrence, Functions, and Applications

The complexity of glycosylation reactions in animals, especially mammals, and of the resulting glycans which form the cellular glycome is matched by that of proteins with a carbohydrate recognition domain that meet the criteria for classification as a lectin given at the end of Section 3.^[4,28c,e] The great strides taken in sequence and three-dimensional analysis of lectins have enabled researchers to pinpoint modules that accommodate glycan epitopes with great precision.^[28c,e,52] A minimum of five lectin families has been solidly defined, the C-, I-, and P-type lectins, pentraxins, and galectins.^[28c,e] New additions to this list will very likely include: a) the two molecular chaperones calnexin and calreticulin, which have a folding pattern resembling that of leguminous lectins, b) a mannose-binding lectin from the pufferfish *Fugu rubripes* with sequence similarity to the agglutinins of monocotyledonous plants with the same binding specificity, c) tachylectin 5A/ficolin, with their fibrinogen-like binding sites, d) the "chitinase-like" Ym1 lectin with its TIM barrel, e) fucose-binding eel lectins, which have a β -barrel with jelly-roll topology, and f) glycosaminoglycan-binding receptors/adhesion molecules.^[19f,28c] This subclassification is evocative of that of glycosyltransferases and each lectin family encompasses more than one member. Table 7 gives an idea of the degree of intrafamily diversity. The table shows the current status of the family of mammalian galectins (Ca²⁺-independent

animal lectins with specificity for β -galactosides and derivatives thereof, a jelly-roll-like folding pattern, and a set of invariant amino acids in the site of contact with the ligand that includes a central Trp residue; see Section 2 for the role of the indolyl side chain). Scouring genome databases for respective hits is thus a worthwhile activity, and homology-based database mining is becoming a valuable tool for the detection of new family members.^[53] A further striking example of intrafamily diversity is the proteins containing the C-type domain (115–130 amino acids with four invariant Cys residues and a characteristic consensus sequence). This domain is often found in mosaic-like proteins with functions involved in cell adhesion (e.g. the selectins in lymphocyte recirculation) or organization of the extracellular matrix (e.g. the hyalectans/lecticans), and in proteins involved in glycan endocytosis. The gene encoding the C-type domain is placed seventh in frequency amongst the 19099 predicted genes of the nematode *Caenorhabditis elegans* and thus surpasses even the epidermal-growth-factor-like and Ig-superfamily domains in ranking.^[54] With 165 or 183 open reading frames (according to separate calculations), this motif, typical for a member of the family of animal lectins, is well-represented in the genome of the model organism.^[52c,55] To date, over a hundred human proteins with C-type lectinlike domains have been described, which establishes this group of domains as a lectin family. These lectins are divided into six subgroups based on their individual modular and quaternary structures.^[52c] These numbers reflect a complex evolutionary genealogy and intimate fine-tuning of ligand specificity for distinct functions.

This type of lectin and also members of several other families take advantage of the elaborate enzymatic process line that specifically tailors the branch ends of glycan chains by preferentially targeting the spatially accessible tips of the sugar antennae. That lectins, through binding to their distinct glycan determinants, are indeed able "to provide a general mechanism of recognition and communication"^[49] (the widespread presence of lectins in animals has already convincingly dispelled the concerns quoted above) is proven by the accrued knowledge presented in Table 8. It is immediately clear from the entries in this table that these insights into lectin function offer enormous potential for applied research in chemical biology. Endocytic receptors of the C-type lectin family, with their fixed geometry of binding sites, are ideal as targets for synthetically tailored drug carriers. These receptors render uptake into cells such as hepatocytes feasible.^[56] Antiviral drugs can thus be delivered to hepatocytes, for example by using triantennary *N*-glycans with GalNAc in the terminal position as a post code. Conversely, lectin-dependent clearance of glycosylated pharmaproteins is therapeutically disadvantageous as it reduces the bioavailability of the drug. It is reasonable in this case to modify the glycan structure to reduce or even avoid lectin binding. Integration of chemoenzymatic *N*-glycan synthesis and bioassays toward this aim has spawned progress in this field.^[57] To be specific, biantennary complex-type *N*-glycans with α 2-3(6)-sialylation and/or bisecting GlcNAc or core fucosylation in the bioactive part of the neoglycoproteins have been studied. α 2-6-Sialylation of a biantennary complex-type *N*-

Table 7. Members of the galectin family of mammalian lectins.^[a]

Name	Occurrence	Structural features
galectin-1 (galaptin, L-14)	many cell types	homodimer; one CRD per subunit (14–15 kDa); proto type
galectin-2	gastrointestinal tract; clone from human hepatoma	homodimer; one CRD per subunit (43% sequence identity to galectin-1; 14 kDa); proto type
galectin-3 (CBP35, Mac-2 antigen, IgE-binding protein, L-29, L-34)	many cell types	monomer with one CRD (oligomer formation in solution and on surfaces); Pro-, Tyr-, and Gly-rich repeats in N-terminal section (27–36 kDa); chimera type
galectin-4	colon, small intestine, stomach, oral epithelium, esophagus; lung, testis, breast, liver, and placenta by RT-PCR	monomer with two partially homologous but distinct CRDs connected by a link peptide (36 kDa); proteolysis generates truncated proto-type-like products: tandem-repeat type
galectin-5	reticulocytes, erythrocytes (rat)	monomer with one CRD (17 kDa); proto type
galectin-6	small intestine, colon	tandem-repeat arrangement of two CRDs (33 kDa)
galectin-7	keratinocytes, stratified epithelia, carcinoma cells	homodimer; one CRD per subunit (15 kDa); proto type
galectin-8	several tissues; frequently present in tumor cell lines (link peptide extension possible)	homologous to galectins-4 and -6 (tandem-repeat arrangement of two CRDs with unique link peptide; 34 kDa)
galectin-9	small intestine, liver, lung, kidney, thymus (rat/mouse; small intestinal isoform with 31/32 amino acid extension of link peptide); lymphatic tissue and B cells, T cells and macrophages, pancreas, colon carcinoma cells (human)	homologous to galectins-4, -6, and -8 (tandem-repeat arrangement of two CRDs with unique link peptide; 36 kDa)
Charcot–Leyden crystal protein (galectin-10)	major autocrystallizing constituent of eosinophils and basophils	one CRD-like structure with specificity for D-Man (16.5 kDa)
galectin-11 (ovgal-11)	sheep gastrointestinal tract, induced upon nematode infection	one CRD, resembles proto-type galectins (14 kDa)
galectin-12	several tissues (upregulation in cells synchronized at the G1 phase or G1/S boundary of the cell cycle), adipocytes identical to placental protein 13 (pp13); also expressed in the spleen, kidney, bladder, and in tumor cells	homologous to galectins-4, -6, -8, and -9 (tandem-repeat arrangement of two CRDs with unique link peptide; 35.3 kDa)
galectin-13	sheep gastrointestinal tract, induced upon nematode infection	homodimer; one CRD per subunit (16.1 kDa); close similarity to galectin-7 and the Charcot–Leyden crystal protein
galectin-14	ovine eosinophils, secreted into bronchoalveolar lavage fluid	one CRD resembling proto-type galectins (18.2 kDa)

[a] Taken from ref. [27c], extended, and modified. Please note that the presence of the galectins in humans has not been confirmed in all cases (e.g. rat galectin-5).

glycan is a means of conferring the signal to its carrier for a rather long period of circulation.^[57] Addition of a bisecting GlcNAc residue to the biantennary *N*-glycan considerably increases uptake of the neoglycoprotein into the liver and spleen, which is relevant for clinical imaging. Neither core fucosylation nor use of the glycan free of substitution can achieve the same effect.^[57] As the cited reports describe in further detail, glycan modification by substitution can also bring about notable changes in the affinity of the molecule for soluble lectins, an effect emerging from the presence of distinct substitutions with biological/clinical relevance.^[57] Another route towards optimization of glycosylation for clinical use involves glycoengineering. In this approach, new *N*-glycosylation sequences (the sequon Asn-X-Ser/Thr, where X is any amino acid except Pro) are introduced into protein therapeutics such as recombinant human erythropoietin by site-directed muta-

genesis. The duration of serum presence and the activity of the engineered glycoproteins in vivo have been increased in this way.^[58] The combined use of these chemoenzymatic strategies, which render *N*- and *O*-glycans of choice available,^[57,59] and molecular biological engineering is bound to bring about the rational design of compounds with prolonged bioavailability or refined capacity for specific delivery.

Fixed topological presentation of binding sites, as discussed above, is also a prerequisite for blocking access to bacterial/viral lectins, a new concept for interfering with the adhesion step of infections and the binding of AB₅-toxins.^[30g,k] The fivefold symmetry of the presentation of the binding sites in these toxins provides the potential for extremely tight binding by a suitably designed pentavalent ligand. This configuration is evocative of a starfish and such compounds are 10⁷-fold more potent in inhibition assays than their monomers.^[30k] Since soluble lectins also display binding sites in distinct arrangements, the therapeutic concept may be extended beyond infections toward attenuating lymphocyte accumulation or metastatic spread. Glycodyndrimers have indeed been shown to impair binding of galectins both in solid-phase

assays with selectivity for the glycoprotein ligand and type of galectin, and in cell-binding studies.^[60] The recently delineated involvement of galectins in tissue invasion during glioblastoma progression or within the metastatic cascade (e.g. in colon, breast, or prostate carcinoma)^[16b,61] is a potential area of interest for testing these ideas in applications. In addition to the effects of the spatial presentation of ligands on synthetic scaffolds such as wedge-like glycodendrimers,^[60c,d] the fine specificity differences between these homologous endogenous lectins are being delineated to enhance probe selectivity, another challenge for chemical biology.^[62] A theory is forming that the structure of the ligand and the spatial mode of its presentation modulate binding avidity in markedly different ways for individual lectins of a family. The detection of these differences lends credit to the assumption that intrafamily diversification is accompanied by quantitative alterations of the ligand profile,

Table 8. Functions of animal lectins.^[a]

Activity	Example of lectin
ligand-selective molecular chaperones in endoplasmic reticulum	calnexin, calreticulin
intracellular routing of glycoproteins and vesicles	ERGIC-53 and VIP-36 (probably also ERGL and VIPL), P-type lectins, comitin
intracellular transport and extracellular assembly	nonintegrin 67-kDa elastin/laminin-binding protein
inducer of membrane superimposition and zippering (formation of Birbeck granules)	langerin (CD207)
cell-type-specific endocytosis	hepatic and macrophage asialoglycoprotein receptors, dendritic cell and macrophage C-type lectins (mannose receptor family members of the tandem-repeat type and single CRD lectins such as langerin/CD207), cysteine-rich domain of the dimeric form of the mannose receptor for GalNAc-4-SO ₄ -bearing glycoprotein hormones in hepatic endothelial cells, P-type lectins
recognition of foreign glycans (β1,3-glucans, LPS)	CR3 (CD11b/CD18), dectin-1, <i>Limulus</i> coagulation factors C and G, earthworm CCF
recognition of foreign or aberrant glycosignatures on cells (including endocytosis or initiation of opsonization or complement activation)	collectins, L-ficolin, C-type macrophage and dendritic cell receptors, α/θ-defensins, pentraxins (CRP, limulin), tachylectins
targeting of enzymatic activity in multimodular proteins	acrosin, laforin, <i>Limulus</i> coagulation factor C
intra- and intermolecular modulation of enzyme activities in vitro	porcine pancreatic α-amylase, galectin-1/α2-6-sialyltransferase
bridging of molecules	homodimeric and tandem-repeat-type galectins, cytokines (e.g. IL-2:IL-2R and CD3 of T-cell receptors), cerebellar soluble lectin
induction or suppression of effector release (H ₂ O ₂ , cytokines, etc.)	galectins, selectins, and other C-type lectins such as CD23, BDCA-2, and dectin-1
cell growth control and induction of apoptosis/apoptosis	galectins, C-type lectins, amphoterin-like protein, hyaluronic-acid-binding proteins, cerebellar soluble lectin
cell migration and routing	selectins and other C-type lectins, I-type lectins, galectins, hyaluronic-acid-binding proteins (RHAMM, CD44, hyalectans/lecticans)
cell-cell interactions	selectins and other C-type lectins (e.g. DC-SIGN), galectins, I-type lectins (e.g. siglecs, N-CAM, P _v , or L1)
cell-matrix interactions	galectins, heparin- and hyaluronic-acid-binding lectins such as hyalectans/lecticans, calreticulin
matrix network assembly	proteoglycan core proteins (C-type CRD and G1 domain of hyalectans/lecticans), galectins (e.g. galectin-3/hensin), nonintegrin 67-kDa elastin/laminin-binding protein

[a] Taken from ref. [4c], extended, and modified.

as mentioned at the start of this section. Systematic chemical mapping with ligand derivatives and screening of arrays/libraries to discover potent ligand mimetics are likely eventually to allow molecules to be devised that fit hand-in-glove into a particular galectin (or any other lectin of clinical interest).^[13d,16d,32] As in the case of plant lectins, these reagents will be instrumental to the detection of lectin activities and to their cyto- and histochemical localization, which is relevant to histopathology.^[27] This approach (i.e. tracking down carbohydrate-binding proteins by using synthetic probes) has been termed "reverse lectin histochemistry" to distinguish it from the routine lectin applications listed in Table 5.^[63]

The deployment of mammalian lectins as laboratory tools has lagged behind application of agglutinins from plants. The reasons for this lack of application are definitely the limited availability of, and access to the reagents. The easy-to-follow protocols provided by recombinant technology to solve this

problem have paved the way for endogenous proteins to become tools too. As research tools (see Table 5), the endogenous lectins offer the added advantage of being able to act as potential therapeutic agents that exploit natural substances and signal pathways, for example, to limit tumor proliferation or T-cell-dependent immune disorders.^[28f,41c,64] Since lectins naturally select binding partners for an in situ function, it is a sure bet that assays involving the participation of endogenous lectins will increase in number. In terms of functional considerations, the development of assays with endogenous lectins (instead of plant surrogates) can be considered a quantum leap. Since the fine sugar specificities of plant and mammalian lectins often differ, results obtained with plant lectins suffer from the inevitable drawback that they cannot be reliably extrapolated to in situ functionality.

We have compiled the documented functions of animal lectins for review in Table 8. Evidently, carbohydrates serve as versatile ligands. It is thus logical to ask a fundamental question on the nature of oligosaccharides: "How can flexible molecules act as signals?"^[65] This concern was put into words in a recent review in which the

author states that, "on several occasions I have heard structural biologist colleagues state that the glycan units in a glycoprotein, for instance, cannot be important because they are too flexible to be seen in an X-ray crystal structure or by NMR. In other words, if they do not have a structure, how can they have a function? That this conclusion is gratuitous"^[2] can be seen by turning to the next section.

6. The Third Dimension of the Sugar Code

It is in principle correct to point critically at the inherent flexibility of oligosaccharides. Rapid intramolecular movements can explain the frustrating futility of attempts to obtain crystals from viscous solutions produced by synthetic carbohydrate chemistry. In glycoproteins, the glycan antennae can even behave as nearly separate entities, a noteworthy factor that allows the proteomic complexity to be increased through dis-

tinct template-independent posttranslational modifications, without altering the genomic coding capacity.^[15a] Close inspection of this flexibility by molecular modeling (molecular mechanics and dynamics simulations) and NMR spectroscopy^[65,66] has revealed that "certain glycans have highly favored conformations."^[2] Figures 2 and 3 focus on ligands for the galactose-specific lectins (galectins and mistletoe lectin) introduced above and illustrate that the conformational space of such a free disaccharide is energetically structured like a topographical map is arranged with respect to altitude. The molecules

populate low-energy (valley) positions in the molecular dynamics simulations, and this result is experimentally verified by the detection of time-averaged interresidual resonance transfer between water-insensitive C–H protons (Figures 2 and 3).^[67] Such disaccharides thus have access to more than one position in the Φ, Ψ, E plot characterizing the distinct sets of energetically favored conformations (Figure 2, Figure 3). Since "the carbohydrate moves in solution through a bunch of shapes each of which may be selected by a receptor," Hardy has likened such a carbohydrate ligand to a "bunch of keys,"^[68] with explicit ref-

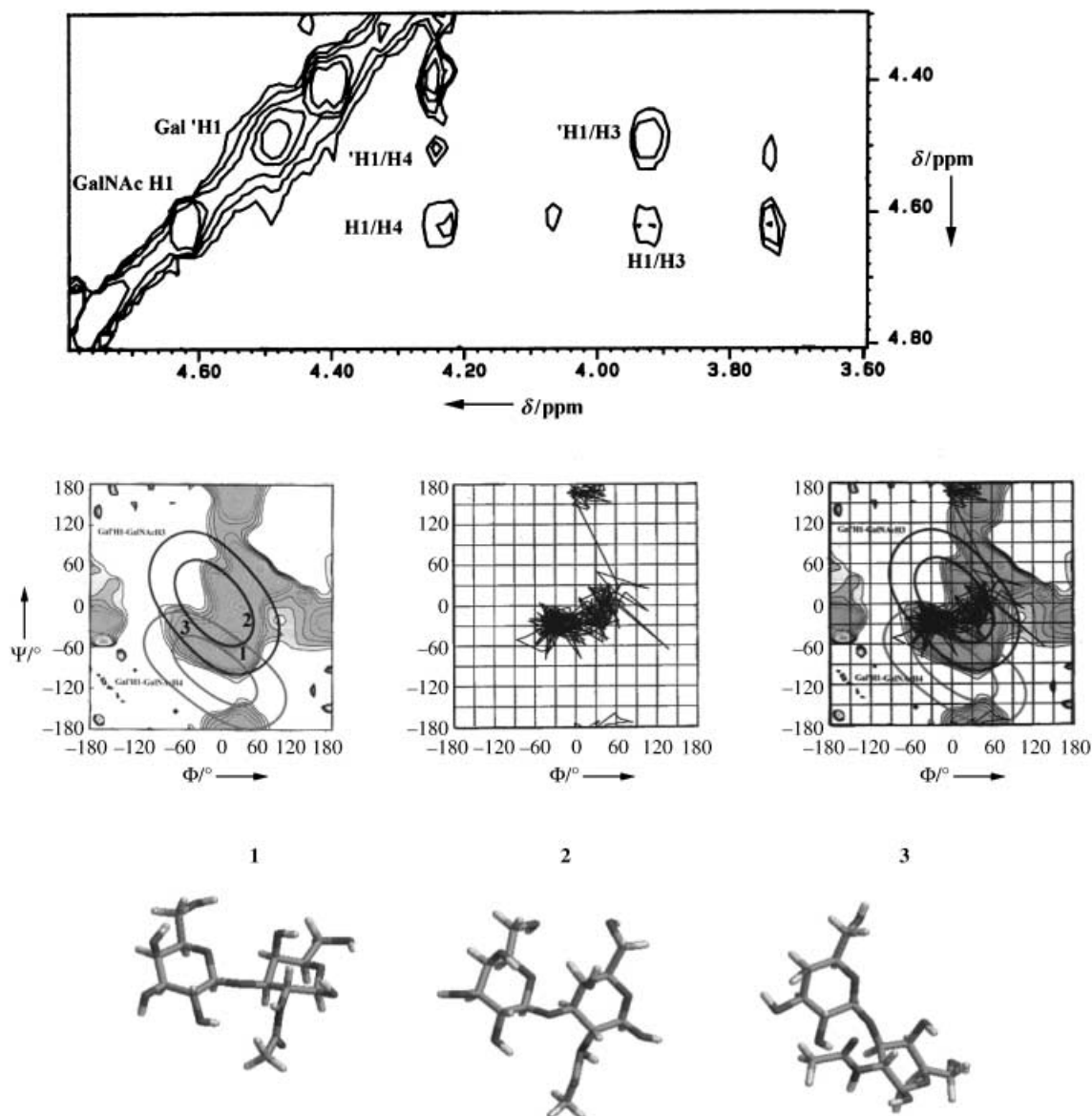


Figure 2. Illustration of conformational aspects of the disaccharide Gal β 1-3GalNAc α / β . This epitope (the α -anomer is the Thomsen–Friedenreich tumor antigen) is a ligand for galectins (for further information on this family of animal lectins, see Tables 7 and 8), as shown by the occurrence of two interresidual trNOE contact signals in the 2D trNOESY spectrum of a mixture (molar ratio 10:1) of the disaccharide with chicken liver galectin (CG-16), recorded at 500 MHz and 298 K with a mixing time of 100 ms (top). Introduction of this information as two pairs of contour lines into the conformational energy map (Φ, Ψ, E plot) derived from molecular mechanics calculations ($\epsilon = 4$) limits the conformational space of the bound ligand. It is also limited in this way when the experimental information is introduced into the molecular dynamics profile (300 K, 1000 ps) derived from calculations that explicitly include water molecules and start from the Φ, Ψ coordinates at $0/180^\circ$ outside the central low-energy valley. These calculations reveal a high population density within this central valley (middle), as described previously.^[67b] Three individual low-energy conformations from the central area, marked 1, 2, and 3 in the energy map, were drawn by using these sets of Φ, Ψ angle combinations to visualize the structural impact of Φ, Ψ angle changes (bottom).

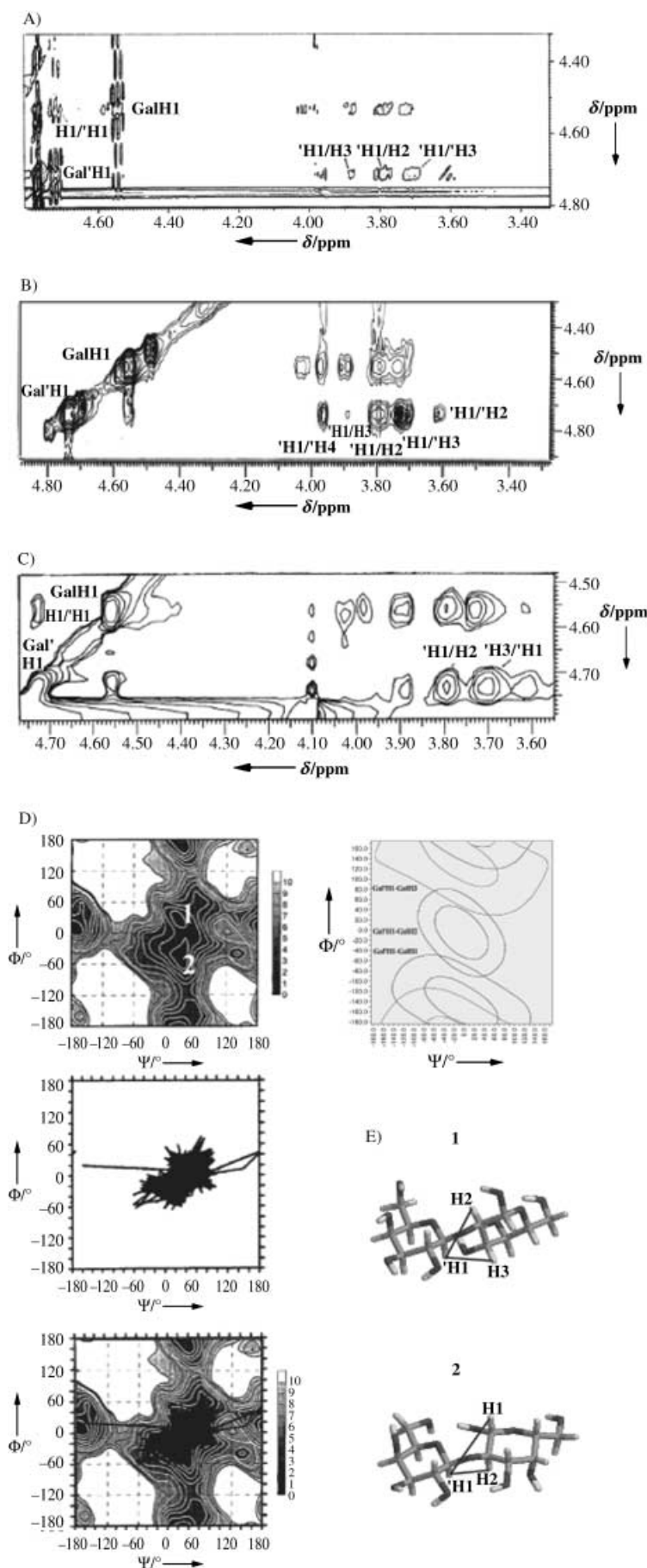


Figure 3. Illustration of the structural aspects of differential conformer selection of a digalactoside by a plant and an animal lectin. Relevant parts of 2D ROESY/trNOESY spectra (recorded at 500 MHz, 298 K with a mixing time of 100 ms) of the free disaccharide Gal β 1-2Gal (A) and of this ligand at a 10:1 molar ratio with (B) the galactoside-specific mistletoe lectin (*Viscum album* L. agglutinin, VAA; see Table 4 for further information) and (C) the chicken liver galectin (CG-16), respectively. The spectra show three interresidual cross-peaks for the free ligand and two such signals for the lectin–ligand complexes, as described previously.^[66a, 67a] The interresidual H1'/H2 cross-peak is shared by the three spectra, whereas only one of the interresidual H1'/H1 and H1'/H3 cross-peaks is present in each of the trNOESY spectra of the ligand with the plant and animal lectins. Molecular mechanics ($\epsilon = 4$) and molecular dynamics calculations ($\epsilon = 80$, CVFF, 300 K, 1000 ps), combined with the NMR-spectroscopy-based contour line pairs (see refs. [66a, 67a] for details), revealed that only one of the two conformers present in solution (labeled as 1 and 2 in the Φ/Ψ , E plot) was bound by each of these two lectins (D). The plant agglutinin and the animal lectin select different conformers of the digalactoside. The structures of the conformers are shown in (E).

reference to the “lock-and-key” paradigm introduced by Fischer in 1894 (see Section 2).^[5] Taken literally, each individual conformer (“key”) is endowed with the potential to interact with a certain complementary receptor site (“lock”).

In other words, a lectin might perform conformer selection, which provides a starting point for hypothesis-driven work. Several agglutinins that share sequence specificity for a disaccharide might subject the ligand population to differential conformer selection. In this sense, recognition is primarily a shape problem (see the passage quoted above), a statement with substantial implications for the design of therapeutic glycomimetics. As illustrated in Figures 2 and 3, experimental data on interresidual proton distances for the ligand in complex with the lectin in solution are obtained by transferred nuclear Overhauser effect (trNOE) spectroscopy, where the signal intensity serves as a molecular ruler.^[66, 69] Whereas the definition of the bound conformation is not unambiguous for the example given in Figure 2 and requires further experimental input or a docking analysis (see below for further discussion and also Figure 4), the information presented in Figure 3 clearly demonstrates the principle of differential conformer selection by lectins.^[66a, 67] In this instance, a single disaccharide (Gal β 1-2Gal) forms two rapidly interconverting shapes. Each specifically interacts with only one of the two different lectins, that is, either with a galectin or a plant lectin. The same ligand can form a bioactive and a bioinert conformation when viewed from the perspective of the galectin tested. As Roseman commented, “it is this interplay between proteins and different conformers that likely allows a single carbohydrate structure (...) to be used in many different ways.”^[2] In terms of methodology, it is the interplay of carbohydrate chemistry, molecular modeling, NMR

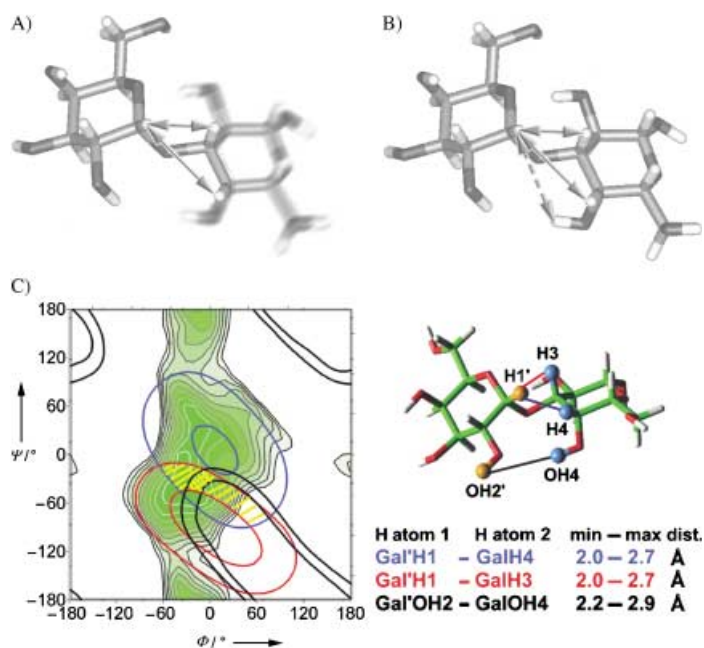


Figure 4. Illustration of the substantial gain of information about the actual conformation of a carbohydrate ligand provided by access to NOE data for contacts involving water-exchangeable protons, and an experimental example to verify the validity of this concept (discussed in detail previously).^[76a] The blurring in (A) demonstrates that even the presence of two interresidual contacts (here H1' to H3 and H4 of Gal' α 1-3Gal β 1-R) does not allow accurate definition of the conformation of the disaccharide (see Figure 2 (middle) and Figure 3D for the size of the area shared by two pairs of contour lines in the E plot). Although the results of molecular mechanics calculations intimate that the bound-state conformations are at low-energy sites in the Φ , Ψ , E plots, further experimental evidence to support this assumption is essential. This verification is symbolized by the substitution of the blurred image by a clear structure (B) after inclusion in calculations of a signal indicating a third contact. In fact, detection of the new water-sensitive contact by analysis of protein–ligand complexes in an aprotic solvent improves the precision of the conformational description by allowing a third pair of contour lines to be added to the E plot. This pair of lines delimits the area of overlap of the two pairs of lines drawn based on water-insensitive contacts (C). Remarkably, this area representing the ligand's bound-state conformation, which is accommodated by a natural immunoglobulin G fraction from human serum, lies within/close to the central low-energy valley.^[76a]

spectroscopy, and biochemical preparation of the receptors that allows the validity of this concept to be convincingly proven.

The power of this integrated approach is again made evident in Figure 5, which shows the bound-state conformers of a glycomimetic. The tested C-glycoside offers the pharmacodynamic advantage of resistance to hydrolytic cleavage. However, an increased degree of flexibility relative to that of the O-glycoside results from the introduction of a methylene bridge in place of the oxygen atom (for further information, see the legend of Figure 5).^[70] By using exclusive interresidual contacts as fingerprint-like characteristics for a certain bound-state topology, differential conformer selection was established and the conformers selected by galectin-1 (*syn*- Φ , Ψ), the B chain of ricin (*anti*- Ψ), and an enzymatically inactive mutant of the bacterial β -galactosidase (*anti*- Φ) were tracked down.^[71] One may wonder whether this result applies only to small ligand structures or also to naturally occurring extended saccharide chains.

A recent example is provided by a combined NMR spectroscopy and molecular modeling study that defined the bound-state topology of a cell-surface-exposed oligosaccharide chain, the pentasaccharide of ganglioside GM₁. The obtained data add further strong support to the concept that a certain low-energy conformer is favored for binding. The carbohydrate chain of the ganglioside is the target for both cross-linking by galectin-1 to induce inhibition of the growth of human SK-N-MC neuroblastoma cells and for the AB₅ toxin of *Vibrio cholerae*.^[30k,64a] This ability of one molecule to act as a ligand for two structurally unrelated receptors prompts questions about the topological aspects of these two recognition processes. As shown in Figure 6, in which the two bound-state conformations are compared, there is indeed a difference at the branch point of the carbohydrate chain.^[72] The dihedral angles of the Neu5Ac α 2-3Gal linkage in the bound ligand are either Φ , $\Psi = 70^\circ/15^\circ$ in the case of galectin-1 (in solution) or about $-172^\circ/-26^\circ$ for cholera toxin (in crystals). The conformations selected for binding represent two of the three lowest-energy conformations of the free ligand. Binding causes no distortion of the topology of the selected "key". This result makes it tempting to suggest that ligand derivatives with the same carbohydrate sequence but conformational restriction at the linkage of the internal branch point could no longer interact with both receptor proteins. After all, it would be clinically desirable to block the action of the AB₅ toxin with an inhibitor while lowering the affinity of the inhibitor to the endogenous lectin to avoid undesired side reactions. This challenge at the interface of synthetic carbohydrate chemistry and chemical biology can be tackled rationally given precise topological information. Beyond selectivity, the binding of a deliberately preformed conformer might also help reduce the entropic penalty in the thermodynamic balance sheet of the overall association and accommodation process.^[13ce,65] When we analyzed the binding of the pentasaccharide to galectin-1 by modeling, we were able to obtain information on the major contact sites and the resulting interaction energy terms, data that provide more input for the design of glycomimetics.^[72]

An intriguing example of the intimate relationship between carbohydrate flexibility and molecular recognition is given by iduronic acid in heparin/heparan sulfates, as outlined in Section 2 (for the position of L-iduronic acid in the anticoagulant heparin pentasaccharide that binds to antithrombin III, see Scheme 4). When latched into the recognition site of the plasma protein antithrombin III, 2-O-sulfated L-iduronic acid is driven toward its skewed ²S₀ conformation. In contrast, the local-kink-forming ¹C₄ conformation is preferred by fibroblast growth factors because it maximizes contact between the target determinant in the glycosaminoglycan and these homologous proteins.^[10,73] This amazing role as a versatile hinge that allows the crucial regions of the glycosaminoglycan to adopt the most favorable spatial topology makes it clear that the development of the epimerase reaction that produces L-iduronic acid was not a fortuitous event but a wise investment. The given examples teach this lesson: the more we learn about the intricacies underlying the virtues of carbohydrates as ligands, the more refined the ideas on the drawing-board for devising

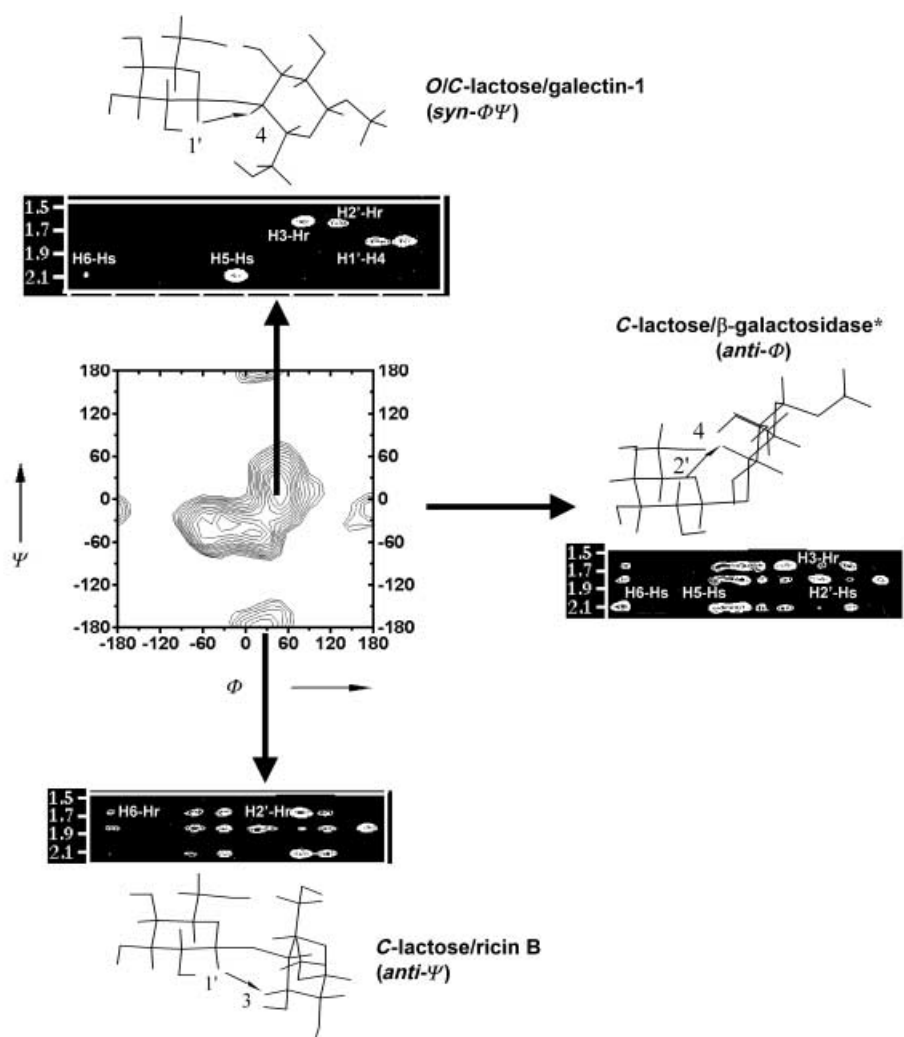


Figure 5. Illustration of the structural aspects of differential conformer selection of C-lactoside by an animal lectin (galectin-1), a plant lectin (the B chain of ricin; see Table 4 for further information), and a catalytically inactive form of *E. coli* β -galactosidase (the asterisk denotes the E537Q mutant). The glycomimetic, which cannot be hydrolyzed, accesses 23% of the conformational space in the Φ , Ψ , χ plot, while 12% is accessed by the O-lactose.^[71b] The increased flexibility of C-lactoside compared to O-lactoside is accompanied by a shift of population density from the *syn* conformation (Φ , Ψ : 55°, 20°) to the *anti-Ψ* conformation (Φ , Ψ : 40°, 180°) to give a 32/54% ratio of the conformers.^[71b] The three conformations of C-lactoside at relative energy minima (*syn*, *anti-Ψ*, and *anti-Φ*) are characterized by the occurrence of distinct interresidual resonance transfer processes, each of which is possible for only one topological constellation and thus establishes an exclusive contact. Each arrow in the figure originates from the respective position in the Φ , Ψ , χ plot and points to the relevant part of a spectrum, shown together with a molecular model in which the pair of protons establishing the exclusive contact is indicated: GalH1/GlcH4 (*syn*), GalH1/GlcH3 (*anti-Ψ*), and GalH2/GlcH4 (*anti-Φ*).^[69b,71d] Detection of cross-peaks arising from any of these exclusive contacts in the 2D trNOESY spectra of the three types of lactoside-binding proteins allows the bound-state conformation of the lactoside to be defined. The animal lectin, the plant agglutinin, and the enzymatically inactive bacterial β -galactosidase select different conformers of the ligand.

ligands with optimal fit and specificity will become. Consideration of the shape of the molecule and its control will play a major role in this process. Rational synthesis and manipulation of the structural details of the molecule, such as the sulfation pattern, as well as screening of oligosaccharide libraries provide routes to augment the affinity of ligands for certain targets and to obtain substances with special biological properties, such as dissociating anticoagulant and antiangiogenic activities.^[9,74] The guidelines for the synthesis clearly rely on the

precision of the information on the bound-state topology of the ligand. As shown in Figures 2, 3, and 5, only C–H protons have been exploited as reporter groups so far. Recruiting hydroxy protons to contribute to the fingerprint of 2D trNOESY/ROESY cross-peaks would improve the quality of our view of bound-state ligand topology, as graphically depicted in Figure 4A and B. To preclude loss of the information from water-exchangeable hydroxy protons, one option was to build on pioneering work with glucosides. Sharp signals were detected for these study objects when they were dissolved in an aprotic solvent (dimethyl sulfoxide).^[75] The concern that the activity of carbohydrate-binding proteins might be harmed substantially by the solvent change was addressed by performing systematic binding assays. These assays revealed that the activity of proteins with a well-structured folding pattern, for example, the jelly roll of galectins, the double β trefoil of the mistletoe lectin, and the Ig fold of immunoglobulin G fractions, is not harmed by such solvent change.^[76] These data square well with encouraging experiences with enzymes in organic solvents.^[77] The results of such experiments also intimate that the folding pattern, at least around the binding site, is not markedly changed by the solvent. Indeed, the accuracy of this assumption has been ascertained experimentally. Formation of dimers of the homo-

dimeric galectin-1, instead of any indication of unfolding, was observed by small angle neutron scattering.^[78] The experimental approach of turning to aprotic solvents for trNOE spectroscopy thus affords the possibility of detecting signals from ligand protons other than those originating from resonance transfer between C–H protons. Figure 4 illustrates results from a proof-of-principle example. The results shown prompted consideration of how the range of applicability of this approach could be extended.

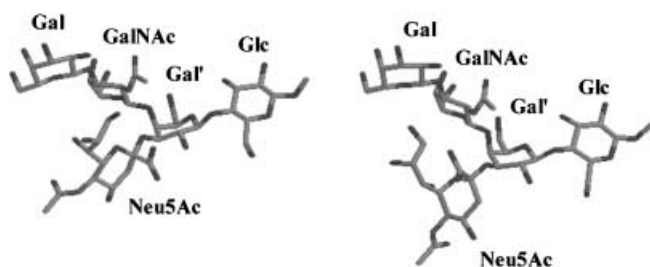


Figure 6. Illustration of the structural aspects of differential conformer selection of the complex carbohydrate chain in ganglioside GM₁ by human galectin-1 (left) and cholera toxin (right). The structures are based on analysis of the solution structure of the galectin-bound pentasaccharide^[72] and data from the Brookhaven Protein Bank (code no. 2CHB/3CHB) in the case of the cholera toxin. The illustrated difference in the Φ, Ψ angle combination for the glycosidic linkage at the branch point connecting the central galactose unit (Gal') with the α -2-3-linked sialic acid residue reflects the differential selection of two conformers from the three relative energy minima of the free state (Φ, Ψ : 70°, 15° for galectin-1; Φ, Ψ : -172°/-26° for cholera toxin).^[72]

A recent study demonstrated that addition of measured amounts of water to an aprotic solvent does not prevent measurement of sharp signals from the hydroxy protons of the ligand, even at temperatures well above 0 °C.^[79] We thus suggest that the use of binary solvent/water mixtures has the potential to enter the panel of strategies for collecting very detailed information on bound-state topology. The aim of these techniques is to enable the complete structure of the complex, including all details of the receptor, to be revealed by analysis in solution. From this information, the way in which ligand binding affects the conformation of the receptor, including its sites for protein–protein interactions (measured for galectin-1 in solution by small angle neutron scattering),^[78] could be discerned. This is a demanding task to accomplish, both for the biochemist, who has to supply (isotope-labeled) material in sufficient quantity and with sufficient solubility for analysis, and for the NMR expert, who is responsible for turning spectra into a structure. This problem has already been solved for a synthetic Thomsen–Friedenreich antigen-binding 15-mer peptide, hevein-domain-containing plant lectins or lectin domains such as the 43 amino acid hevein and GlcNAc oligomers (see Figure 7), the 11-kDa cyanovirin-N from the cyanobacterium (blue-green alga) *Nostoc ellipsosporum* and Man α 1-2Man α , as well as the 198 amino acid adhesin domain of P-pili from uropathogenic *E. coli* (PapGII) and galabiose (Gal α 1-4Gal β).^[80]

In answer to the question that has guided this section, namely how flexible compounds can act as ligands, it has become clear that the conformational space of carbohydrates is structured into several areas. These areas are distinguishable by their relative energy levels. Only a limited set of conformations ("bunch of keys")^[68] is attributed to low-energy valleys, and the accommodation of such conformers is evidently not associated with an insurmountable entropic barrier. Although the molecular details of the overall thermodynamics of the generally enthalpically driven binding reaction are yet to be understood,^[13c,e] the merging of synthetic excellence with *in silico* and *in vitro* techniques guarantees progress toward resolving this issue eventually.

7. Conclusions and Perspectives

The multifarious intermolecular recognition and regulation processes that underlie the efficient and smooth functioning of cell sociology have hitherto been assigned exclusively to nucleic acids and proteins in the central dogma of molecular biology. Despite a fashionable tendency to write off anything beyond genomics, the problem of how the limited panel of primary gene products is increased to serve all purposes properly and even to allow rapid and reversible regulation has engendered a surge in interest in mechanisms of posttranslational modification. Glycan chains have all the properties required for high-density information storage and are therefore qualified to make a mark in this respect. Their finely tuned synthesis even allows for dynamic modulations in response to external signals, and the ensuing interplay with endogenous lectins furnishes cells with an efficient communication system. This transition in the way we look at glycans, which means that the focus is no longer merely on the role of these molecules as biochemical fuel or protective cell wall constituents, has not passed unnoticed. As a consequence, cellular glycoconjugates and lectins are receiving increasing attention and respect. The entry at the bottom of Table 2 concerning the years 2001/2 attests this development. Stepwise refinements in instrumental capacity for structural analysis of carbohydrate oligo- and polymers have made it possible to consider deciphering the sequence of a glycan no longer deterrent.^[15a,81] The same holds true for conformational analysis. The realization of the enormous talents of glycans occurred in a gradual process rather than by a quantum leap.^[82] Fittingly, progress in lectinology also followed this pattern, as the historic survey in Table 2 recounts and the steady increase of publications dealing with lectins reflects.^[16d] The instrumental role of leguminous and eel lectins in the definition of the structure of AB0 histo-blood group epitopes about 50 years ago (see Section 3) sets a precedent for, and shows the enormous potential of merging these lines of research in the glycosciences branch of chemical biology. The design of optimal ligands to block disease-causing lectin activities (e.g. in bacterial infection or tumor invasion) or of lectin-mimetic peptides to elicit clinically beneficial lectin activities (e.g. removal of activated T-cells in autoimmune diseases or destruction of tumor cells by mimicking the capacity of galectin-1 to induce apoptosis/anoikis) are aims for this research. As summarized by Sharon recently, "breaking the glycode and identifying the receptors are of prime importance not only for theoretical reasons, but also to facilitate the development of novel treatments for the many diseases in which carbohydrate recognition plays a key role."^[83]

Acknowledgements

The excellent manuscript processing by R. Ohl and the constructive and exceptionally helpful comments of both reviewers are greatly appreciated, as is the support from the Verein zur Förderung des biologisch-technischen Fortschritts in der Medizin e.V. A sincere apology is directed to colleagues whose original work

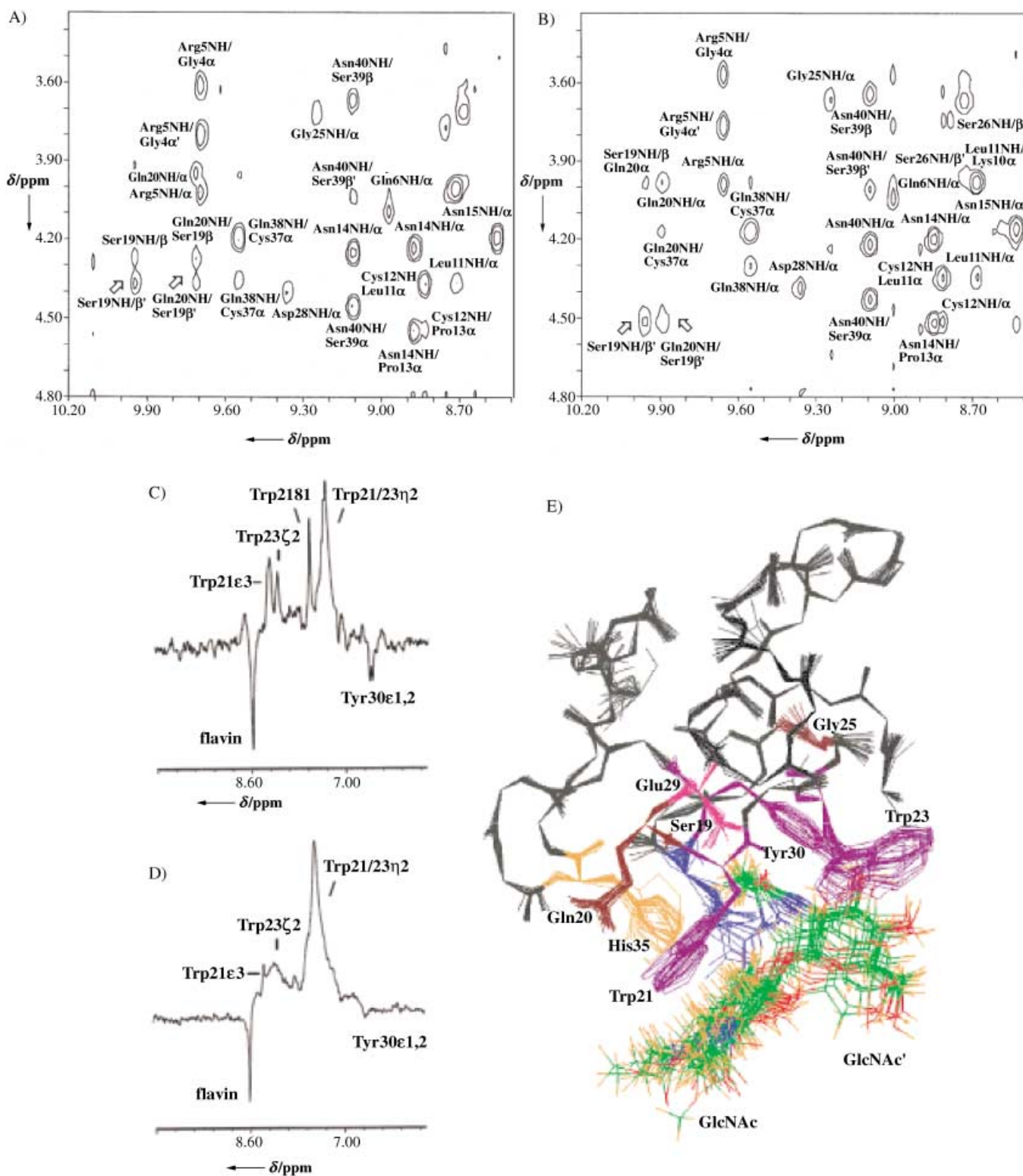


Figure 7. Relevant sections of the 2D NOESY spectra (recorded at 360 MHz, 300 K and with a mixing time of 200 ms) of the 43 amino acid plant lectin hevein (for further information on this lectin, see Table 4) in the absence (A) and in the presence (B) of N,N'-diacetylchitobiose. Characteristic alterations in the Ser19-dependent signals caused by the presence of a ligand are indicated by arrows. Involvement of the aromatic amino acids Trp21, Trp23, and Tyr30 in ligand binding is delineated by laser photo CIDNP difference spectra (aromatic section) of 1 mM hevein in the absence (C) and in the presence (D) of 1 mM N,N'-diacetylchitobiose at pD 4.^[84] The spatial proximity of Ser19 and the three aromatic amino acid side chains to the ligand is depicted by the superposition of twenty snapshots (E) of the lectin–ligand complex taken in the course of a molecular dynamics simulation with explicit inclusion of water molecules, as presented in detail previously.^[76a]

could not be completely included, discussed, and cited because of space limitations and the scope of this review. With regret regarding this aspect of the paper, we set out to produce a primer on the concept of the sugar code as we see it, illustrated by selected proof-of-principle examples to convey a flavor of the field.

Keywords: adhesion · bioinformatics · drug design · glycosylation · lectins · molecular modeling · NMR spectroscopy

- [1] D. A. Rees, *Biochem. J.* **1972**, *126*, 257–273.
- [2] S. Roseman, *J. Biol. Chem.* **2001**, *276*, 41527–41542.
- [3] P. J. Winterburn, C. F. Phelps, *Nature* **1972**, *236*, 147–151.
- [4] a) H.-J. Gabius, *Naturwissenschaften* **2000**, *87*, 108–121; b) J. Hirabayashi, K.-i. Kasai, *Trends Glycosci. Glycotechnol.* **2000**, *12*, 1–5; c) H.-J. Gabius, S. André, H. Kaltner, H.-C. Siebert, *Biochim. Biophys. Acta* **2002**, *1572*, 165–177.
- [5] E. Fischer, *Ber. Dt. Chem. Ges.* **1894**, *27*, 2985–2993.
- [6] R. A. Laine in *Glycosciences: Status and Perspectives* (Eds.: H.-J. Gabius, S. Gabius), Chapman and Hall, London, **1997**, pp. 1–14.
- [7] a) W. D. Comper, *J. Theor. Biol.* **1990**, *145*, 497–509; b) H. Kresse in *Glycosciences: Status and Perspectives* (Eds.: H.-J. Gabius, S. Gabius), Chapman and Hall, London, **1997**, pp. 201–222; c) J. Turnbull, A. Powell, S. Guimond, *Trends Cell Biol.* **2001**, *11*, 75–82.
- [8] a) S. Yamada, K. Sugahara, *Trends Glycosci. Glycotechnol.* **1998**, *10*, 95–123; b) R. Sasisekharan, G. Venkataraman, *Curr. Opin. Chem. Biol.* **2000**, *4*, 626–631.
- [9] a) S. Alban in *Carbohydrates in Drug Design* (Eds.: Z. J. Witzczak, K. A. Nieforth), M. Dekker, New York, **1997**, pp. 209–276; b) R. J. Linhardt, T. Toida in *Carbohydrates in Drug Design* (Eds.: Z. J. Witzczak, K. A. Nieforth), M. Dekker, New York, **1997**, pp. 277–341; c) B. Casu, U. Lindahl, *Adv. Carbohydr. Chem. Biochem.* **2001**, *57*, 159–206; d) I. Capila, R. J. Linhardt, *Angew. Chem.* **2002**, *114*, 428–451; *Angew. Chem. Int. Ed.* **2002**, *41*, 390–412; e) A. S. Gallus, D. W. Coghlan, *Curr. Opin. Hematol.* **2002**, *9*, 422–429; f) M. Sundaram, Y. Qi, Z. Shriver, D. Liu, G. Zhao, G. Venkataraman, R. Langer, R. Sasisekharan, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 651–656.
- [10] B. Casu, M. Petitou, M. Provasoli, P. Sinay, *Trends Biochem. Sci.* **1988**, *13*, 221–225.
- [11] a) N. Perrimon, M. Bernfield, *Nature* **2000**, *404*, 725–728; b) S. B. Selleck, *Trends Genet.* **2000**, *16*, 206–212; c) M. Princivalle, A. De Agostini, *Int. J. Dev. Biol.* **2002**, *46*, 267–278; d) J. Turnbull, K. Drummond, Z. Huang, M. Ford-Perriss, M. Murphy, S. Guimond, *Biochem. Soc. Transact.* **2003**, *31*, 343–348.
- [12] a) L. V. Hooper, S. M. Manzella, J. U. Baenziger in *Glycosciences: Status and Perspectives* (Eds.: H.-J. Gabius, S. Gabius), Chapman and Hall, Weinheim-London, **1997**, pp. 261–276; b) M. Fukuda, N. Hiraoka, T. O. Akama, M. N. Fukuda, *J. Biol. Chem.* **2001**, *276*, 47747–47750; c) J. R. Grunwell, C. R. Bertozzi, *Biochemistry* **2002**, *41*, 13117–13126; d) K. Honke, N. Taniguchi, *Med. Res. Rev.* **2002**, *22*, 637–654; e) J. U. Baenziger, *Biochem. Soc. Transact.* **2003**, *31*, 326–330; f) I. Brockhausen, *Biochem. Soc. Transact.* **2003**, *31*, 318–325.
- [13] a) F. A. Quijcho, *Pure Appl. Chem.* **1989**, *61*, 1293–1306; b) R. U. Lemieux, *Acc. Chem. Res.* **1996**, *29*, 373–380; c) H.-J. Gabius, *Pharmaceut. Res.* **1998**, *15*, 23–30; d) D. Solís, J. Jiménez-Barbero, H. Kaltner, A. Romero, H.-C. Siebert, C.-W. von der Lieth, H.-J. Gabius, *Cells Tissues Organs* **2001**, *168*, 5–23; e) T. K. Dam, C. F. Brewer, *Chem. Rev.* **2002**, *102*, 387–429.
- [14] a) H. Uedeira, H. Uedeira, *J. Sol. Chem.* **1985**, *14*, 27–34; b) J. Hirabayashi, *Quart. Rev. Biol.* **1996**, *71*, 365–380; c) A. M. Striegel, *J. Am. Chem. Soc.* **2003**, *125*, 4146–4148.
- [15] a) G. Reuter, H.-J. Gabius, *Cell. Mol. Life Sci.* **1999**, *55*, 368–422; b) T. Hennet, *Cell. Mol. Life Sci.* **2002**, *59*, 1081–1095; c) R. G. Spiro, *Glycobiology* **2002**, *12*, 43R–56R; d) P. M. Coutinho, E. Deleury, G. J. Davies, B. Henrissat, *J. Mol. Biol.* **2003**, *328*, 307–317; e) D. J. Becker, J. B. Lowe, *Glycobiology* **2003**, *13*, 41R–53R; f) K. G. Ten Hagen, T. A. Fritz, L. A. Tabak, *Glycobiology* **2003**, *13*, 1R–16R.
- [16] a) S. H. Barondes, *Annu. Rev. Biochem.* **1981**, *50*, 207–231; b) H. Kaltner, B. Stierstorfer, *Acta Anat.* **1998**, *161*, 162–179; c) A. Villalobo, H.-J. Gabius, *Acta Anat.* **1998**, *161*, 110–129; d) H. Rüdiger, H.-C. Siebert, D. Solís, J. Jiménez-Barbero, A. Romero, C.-W. von der Lieth, T. Díaz-Mauriño, H.-J. Gabius, *Curr. Med. Chem.* **2000**, *7*, 389–416; e) N. M. Dahms, M. K. Hancock, *Biochim. Biophys. Acta* **2002**, *1572*, 317–340; f) S.-i. Kawabata, R. Tsuda, *Biochim. Biophys. Acta* **2002**, *1572*, 414–421; g) J. Lu, C. Teh, U. Kishore, K. B. M. Reid, *Biochim. Biophys. Acta* **2002**, *1572*, 387–400; h) P. H. Weigel, J. H. N. Yik, *Biochim. Biophys. Acta* **2002**, *1572*, 341–363.
- [17] S. W. Mitchell, *Smithsonian Contrib. Knowledge* **1860**, *XII*, 89–90.
- [18] S. Flexner, H. Noguchi, *J. Exp. Med.* **1902**, *6*, 277–301.
- [19] a) J. Bordet, F. P. Gay, *Ann. Inst. Pasteur* **1906**, *20*, 467–498; b) J. Bordet, O. Streng, *Zbl. Bakteriol. Parasitenkd. Infektionskrankh. Hyg. Abt. I Orig.* **1909**, *49*, 260–276; c) S. Hirani, J. D. Lambris, H. J. Müller-Eberhard, *J. Immunol.* **1985**, *134*, 1105–1109; d) H.-J. Gabius, *Int. J. Biochem.* **1994**, *26*, 469–477; e) G. R. Vasta, M. Quesenberry, H. Ahmed, N. O'Leary, *Dev. Comp. Immunol.* **1999**, *23*, 401–420; f) D. C. Kilpatrick, *Biochim. Biophys. Acta* **2002**, *1572*, 401–413.
- [20] H. Stillmark, *Über Ricin, ein giftiges Ferment aus den Samen von Ricinus comm. L. und einigen anderen Euphorbiaceen* Inaugural Dissertation, Schnakenburg's Buchdruckerei, Dorpat, **1888**.
- [21] J. B. Sumner, *J. Biol. Chem.* **1919**, *37*, 137–142.
- [22] J. B. Sumner, S. F. Howell, *J. Bacteriol.* **1936**, *32*, 227–237.
- [23] a) A. Creite, *Z. Rat. Med.* **1869**, *36*, 90–108; b) K. Landsteiner, *Zbl. Bakteriol. Parasitenkd. Infektionskrankh. Hyg. Abt. I Orig.* **1900**, *27*, 357–362; c) K. Landsteiner, *Wiener Klin. Wschr.* **1901**, *42*, 1020–1024; d) K. Landsteiner, J. van der Scheer, *J. Exp. Med.* **1931**, *54*, 295–305; e) K. O. Renkonen, *Ann. Med. Exp. Biol. Fenn.* **1948**, *26*, 66–72; f) W. C. Boyd, R. M. Reguera, *J. Immunol.* **1944**, *62*, 333–339; g) W. C. Boyd, *Vox Sang.* **1963**, *8*, 1–32; h) N. C. Hughes-Jones, B. Gardner, *Br. J. Haematol.* **2002**, *119*, 889–893.
- [24] W. C. Boyd in *The Proteins* (Eds.: H. Neurath, K. Bailey), Academic Press, New York, **1954**, Vol. 2, Part 2, pp. 756–844.
- [25] a) W. M. Watkins, W. T. J. Morgan, *Nature* **1952**, *169*, 825–826; b) W. T. J. Morgan, W. M. Watkins, *Br. J. Exp. Pathol.* **1953**, *34*, 94–103; c) W. J. Judd, *CRC Crit. Rev. Clin. Lab. Sci.* **1980**, *12*, 172–214; d) G. W. G. Bird, *Transfusion Med. Rev.* **1989**, *3*, 55–62; e) W. M. Watkins, *Trends Glycosci. Glycotechnol.* **1999**, *11*, 391–411; f) H. P. Schwarz, F. Dorner, *Br. J. Haematol.* **2003**, *121*, 556–565.
- [26] a) I. J. Goldstein, R. C. Hughes, M. Monsigny, T. Osawa, N. Sharon, *Nature* **1980**, *285*, 66; b) J. Kocourek, V. Hořejší, *Nature* **1981**, *290*, 188; c) M. B. F. Dixon, *Nature* **1981**, *292*, 192; d) J. Kocourek, V. Hořejší in *Lectins. Biology, Biochemistry, Clinical Biochemistry* (Eds.: T. C. Bøg-Hansen, G. A. Spengler), W. de Gruyter, Berlin, **1983**, Vol. 3, pp. 3–6; e) S. H. Barondes, *Trends Biochem. Sci.* **1988**, *13*, 480–482; f) H.-J. Gabius, *Biochim. Biophys. Acta* **1991**, *1071*, 1–18; g) H. Rüdiger, H.-J. Gabius, *Glycoconjugate J.* **2001**, *18*, 589–613.
- [27] a) H.-J. Gabius, S. André, A. Danguy, K. Kayser, S. Gabius, *Methods Enzymol.* **1994**, *242*, 37–46; b) H.-J. Gabius, C. Unverzagt, K. Kayser, *Biotech. Histochem.* **1998**, *73*, 263–277; c) H.-J. Gabius, *Anat. Histol. Embryol.* **2001**, *30*, 3–31.
- [28] a) H.-J. Gabius, *Cancer Invest.* **1987**, *5*, 39–46; b) L. D. Powell, A. Varki, *J. Biol. Chem.* **1995**, *270*, 14243–14246; c) H.-J. Gabius, *Eur. J. Biochem.* **1997**, *243*, 543–576; d) T. Angata, E. C. M. Brinkman-Van der Linden, *Biochim. Biophys. Acta* **2002**, *1572*, 294–316; e) D. C. Kilpatrick, *Biochim. Biophys. Acta* **2002**, *1572*, 187–197; f) G. A. Rabinovich, N. Rubinstein, M. A. Toscano, *Biochim. Biophys. Acta* **2002**, *1572*, 274–284.
- [29] H. Rüdiger, *Acta Anat.* **1998**, *161*, 130–152.
- [30] a) C. P. Stowell, Y. C. Lee, *Adv. Carbohydr. Chem. Biochem.* **1980**, *37*, 225–281; b) J. D. Aplin, J. C. Wriston, Jr., *CRC Crit. Rev. Biochem.* **1981**, *10*, 259–306; c) H.-J. Gabius, *Angew. Chem.* **1988**, *100*, 1321–1330; *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 1267–1276; d) *Neoglycoconjugates. Preparation and Applications* (Eds.: Y. C. Lee, R. T. Lee), Academic Press, San Diego, **1994**; e) N. V. Bovin, H.-J. Gabius, *Chem. Soc. Rev.* **1995**, *24*, 413–421; f) R. Roy, *Trends Glycosci. Glycotechnol.* **1996**, *8*, 79–99; g) M. Mammen, S.-K. Choi, G. M. Whitesides, *Angew. Chem.* **1998**, *110*, 2908–2953; *Angew. Chem. Int. Ed.* **1998**, *37*, 2754–2794; h) L. L. Kiessling, L. E. Strong, J. E. Gestwicki, *Annu. Rep. Med. Chem.* **2000**, *35*, 321–330; i) N. Yamazaki, S. Kojima, N. V. Bovin, S. André, S. Gabius, H.-J. Gabius, *Adv. Drug Deliv. Rev.* **2000**, *43*, 225–244; j) B. T. Houseman, M. Mrksich, *Top.*

- Curr. Chem.* **2002**, *218*, 1–44; k) C. L. Schengrund, *Biochem. Pharmacol.* **2003**, *65*, 699–707.
- [31] a) W. Straus, *J. Histochem. Cytochem.* **1983**, *31*, 78–84; b) H.-J. Gabius, R. Engelhardt, K. P. Hellmann, T. Hellmann, A. Ochsenfahrt, *Anal. Biochem.* **1987**, *165*, 349–355; c) S. Gabius, K. P. Hellmann, T. Hellmann, U. Brinck, H.-J. Gabius, *Anal. Biochem.* **1989**, *182*, 447–451.
- [32] a) A. Barkley, P. Arya, *Chem. Eur. J.* **2001**, *7*, 555–563; b) S.-I. Nishimura, *Curr. Opin. Chem. Biol.* **2001**, *5*, 325–335; c) K. R. Love, P. H. Seeberger, *Angew. Chem.* **2002**, *114*, 3733–3736; *Angew. Chem. Int. Ed.* **2002**, *41*, 3583–3586; d) L. A. Marcaurrelle, P. H. Seeberger, *Curr. Opin. Chem. Biol.* **2002**, *6*, 289–296; e) C. O. Mellet, J. M. G. Fernández, *ChemBioChem* **2002**, *3*, 819–822; f) O. Ramström, T. Bunyapaiboonsri, S. Lohmann, J.-M. Lehn, *Biochim. Biophys. Acta* **2002**, *1572*, 178–186.
- [33] N. Sharon, *Protein Sci.* **1998**, *7*, 2042–2048.
- [34] a) B. B. L. Agrawal, I. J. Goldstein, *Biochem. J.* **1965**, *26*, 23c; b) J. H. Pazur, *Adv. Carbohydr. Chem. Biochem.* **1981**, *39*, 405–447; c) H.-J. Gabius, *Anal. Biochem.* **1990**, *189*, 91–94; d) H.-J. Gabius in *Protein Liquid Chromatography* (Ed.: M. Kastner), Elsevier, Amsterdam, **2000**, pp. 619–638.
- [35] a) T. Freier, G. Fleischmann, H. Rüdiger, *Biol. Chem. Hoppe-Seyler* **1985**, *366*, 1023–1028; b) H. Rüdiger in *Lectins and Glycobiology* (Eds.: H.-J. Gabius, S. Gabius), Springer, Heidelberg, **1993**, pp. 31–46.
- [36] G. Fleischmann, I. Mauder, W. Illert, H. Rüdiger, *Biol. Chem. Hoppe-Seyler* **1985**, *366*, 1029–1032.
- [37] a) H. Lis, N. Sharon, *Annu. Rev. Biochem.* **1986**, *55*, 35–67; b) I. Damjanov, *Lab. Invest.* **1987**, *57*, 5–20; c) T. Osawa, T. Tsuji, *Annu. Rev. Biochem.* **1987**, *56*, 21–42; d) A. Danguy, F. Akif, B. Pajak, H.-J. Gabius, *Histol. Histopathol.* **1994**, *9*, 155–171; e) J. F. Kennedy, P. M. G. Palva, M. T. S. Corella, M. S. M. Cavalcanti, L. C. B. B. Coelho, *Carbohydr. Polym.* **1995**, *26*, 219–230; f) R. D. Cummings in *Glycosciences: Status and Perspectives* (Eds.: H.-J. Gabius, S. Gabius), Chapman and Hall, London, **1997**, pp. 191–199; g) W. J. Peumans, E. J. M. van Damme, *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 209–259.
- [38] a) V. Kinzel, D. Kübler, J. Richards, M. Stöhr in *Concanavalin A as a Tool* (Eds.: H. Bittiger, H. P. Schnebli), Wiley, London, **1976**, pp. 467–478; b) V. Kinzel, D. Kübler, J. Richards, M. Stöhr, *Science* **1976**, *192*, 487–489.
- [39] a) T. Hajto, K. Hostanska, H.-J. Gabius, *Cancer Res.* **1989**, *49*, 4803–4808; b) T. Hajto, K. Hostanska, K. Frei, C. Rordorf, H.-J. Gabius, *Cancer Res.* **1990**, *50*, 3322–3326.
- [40] a) E. Kunze, H. Schulz, M. Adamek, H.-J. Gabius, *J. Cancer Res. Clin. Oncol.* **2000**, *126*, 125–138; b) H.-J. Gabius, F. Darro, M. Rimmelink, S. André, J. Kopitz, A. Danguy, S. Gabius, I. Salmon, R. Kiss, *Cancer Invest.* **2001**, *19*, 114–126; c) A. V. Timoshenko, Y. Lan, H.-J. Gabius, P. K. Lala, *Eur. J. Cancer* **2001**, *37*, 1910–1920; d) S. Gabius, H.-J. Gabius, *Dtsch. med. Wschr.* **2002**, *127*, 457–459.
- [41] a) S. Chouaib, C. Asselin-Paturel, F. Mami-Chouaib, A. Caignard, J. Y. Blay, *Immunol. Today* **1997**, *18*, 493–497; b) J. A. Sogn, *Immunity* **1998**, *9*, 757–763; c) H.-J. Gabius, *Biochimie* **2001**, *83*, 659–666; d) L. M. Cousens, Z. Werb, *Nature* **2002**, *420*, 860–867; e) J. L. Yu, J. W. Rak, *Breast Cancer Res.* **2003**, *5*, 83–88.
- [42] a) H. Heimpel, *Dtsch. med. Wschr.* **1995**, *16*, 205; b) W. Hagenah, I. Dörge, E. Gafumbegete, T. Wagner, *Dtsch. med. Wschr.* **1998**, *123*, 1001–1004; c) A. M. M. Eggermont, U. R. Kleeborg, D. J. Ruiters, S. Suci in *ASCO Educational Book* (Ed.: M. C. Perry), American Society of Clinical Oncology, Alexandria, VA, USA, **2001**, pp. 88–93; d) E. Ernst, K. Schmidt, M. K. Steuer-Vogt, *Int. J. Cancer* **2003**, *107*, 262–267.
- [43] a) L. M. Brill, C. J. Evans, A. M. Hirsch, *Plant J.* **2001**, *25*, 453–461; b) R. Esteban, B. Dopico, F. J. Muñoz, S. Romo, E. Labrador, *Physiol. Plant.* **2002**, *114*, 619–626; c) W.-d. Yong, Y.-y. Xu, W.-z. Xu, X. Wang, N. Li, J.-s. Wu, T.-b. Liang, K. Chong, Z.-h. Xu, K.-h. Tan, Z.-q. Zhu, *Planta* **2003**, *217*, 261–270.
- [44] a) J.-C. Promé, *Curr. Opin. Struct. Biol.* **1996**, *6*, 671–678; b) A. M. Hirsch, *Curr. Opin. Plant Biol.* **1999**, *2*, 320–326; c) P. Potin, K. Bouarab, F. Küpper, B. Kloareg, *Curr. Opin. Microbiol.* **1999**, *2*, 276–283; d) T. Yamaguchi, Y. Ito, N. Shibuya, *Trends Glycosci. Glycotechnol.* **2000**, *12*, 113–120; e) P. P. G. van der Holst, H. R. M. Schlaman, H. P. Spaink, *Curr. Opin. Struct. Biol.* **2001**, *11*, 608–616.
- [45] a) P. Tomme, R. A. J. Warren, N. R. Gilkes, *Adv. Microbiol. Physiol.* **1995**, *37*, 1–81; b) C. Khosia, P. B. Harbury, *Nature*, **2001**, *409*, 247–252; c) B. W. McLean, A. B. Boraston, D. Brouwer, N. Sanaie, C. A. Fyfe, R. A. J. Warren, D. G. Kilburn, C. A. Haynes, *J. Biol. Chem.* **2002**, *277*, 50245–50254; d) S. Thobhani, B. Ember, A. Siriwardena, G.-J. Boons, *J. Am. Chem. Soc.* **2003**, *125*, 7154–7155.
- [46] a) J. Wang, J. A. Stuckey, M. J. Wishart, J. E. Dixon, *J. Biol. Chem.* **2002**, *277*, 2377–2380; b) S. Ganesh, N. Tsurutani, T. Suzuki, Y. Hishii, T. Ishihara, A. V. Delgado-Escueta, K. Yamakawa, *Biochem. Biophys. Res. Commun.* **2004**, *313*, 1101–1109.
- [47] a) Y. S. Kim, J. H. Lee, G. M. Yoon, H. S. Cho, S.-W. Park, M. C. Suh, D. Choi, H. J. Ha, J. R. Liu, H.-S. Pai, *Plant Physiol.* **2000**, *123*, 905–915; b) A. Barre, C. Hervé, B. Lescure, P. Rougé, *Crit. Rev. Plant Sci.* **2002**, *21*, 379–399; c) M. Nishiguchi, K. Yoshida, T. Sumizono, K. Tazaki, *Mol. Genet. Genomics* **2002**, *267*, 506–514.
- [48] a) P. A. Gleeson, *Curr. Top. Microbiol. Immunol.* **1988**, *139*, 1–34; b) H. Ueda, H. Ogawa, *Trends Glycosci. Glycotechnol.* **1999**, *11*, 413–428; c) S. Chen, A. M. Spence, H. Schachter, *Trends Glycosci. Glycotechnol.* **2001**, *13*, 447–462; d) I. B. H. Wilson, *Curr. Opin. Struct. Biol.* **2002**, *12*, 569–577.
- [49] S. Roth, *Quart. Rev. Biol.* **1973**, *48*, 541–563.
- [50] P. Weiss, *Yale J. Biol. Med.* **1947**, *19*, 235–278.
- [51] a) R. L. Hudgin, W. E. Pricer, Jr., G. Ashwell, R. J. Stockert, A. G. Morell, *J. Biol. Chem.* **1974**, *249*, 5536–5543; b) V. I. Teichberg, I. Silman, D. D. Beitsch, G. Resheff, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 1383–1387; c) T. K. Gartner, K. Stocker, D. C. Williams, *FEBS Lett.* **1980**, *117*, 13–16.
- [52] a) H. Lis, N. Sharon, *Chem. Rev.* **1998**, *98*, 637–674; b) W. J. Peumans, A. Barre, Q. Hao, P. Rougé, E. J. M. van Damme, *Trends Glycosci. Glycotechnol.* **2000**, *12*, 83–101; c) R. B. Dodd, K. Drickamer, *Glycobiology* **2001**, *11*, 71R–79R; d) R. Loris, *Biochim. Biophys. Acta* **2002**, *1572*, 198–208.
- [53] D. N. W. Cooper, *Biochim. Biophys. Acta* **2002**, *1572*, 209–231.
- [54] The *C. elegans* Sequencing Consortium, *Science* **1998**, *282*, 2012–2018.
- [55] R. O. Hynes, Q. Zhao, *J. Cell Biol.* **2000**, *150*, F89–F95.
- [56] a) J. C. Rogers, S. Kornfeld, *Biochem. Biophys. Res. Commun.* **1971**, *45*, 622–629; b) Y. C. Lee, *FASEB J.* **1992**, *6*, 3193–3200; c) L. Fiume, C. Busi, G. Di Stefano, A. Mattioli, *Adv. Drug Deliv. Rev.* **1994**, *14*, 51–65; d) D. K. F. Meijer, G. Molema, *Sem. Liver Dis.* **1995**, *15*, 202–256; e) H.-J. Gabius, *Cancer Investig.* **1997**, *15*, 454–464; f) K. G. Rice in *Glycosciences: Status and Perspectives* (Eds.: H.-J. Gabius, S. Gabius), Chapman and Hall, London–Weinheim, **1997**, pp. 471–483; g) B. G. Davis, M. A. Robinson, *Curr. Opin. Drug Discov. Develop.* **2002**, *5*, 279–288.
- [57] a) S. André, C. Unverzagt, S. Kojima, X. Dong, C. Fink, K. Kayser, H.-J. Gabius, *Bioconjugate Chem.* **1997**, *8*, 845–855; b) C. Unverzagt, S. André, J. Seifert, S. Kojima, C. Fink, G. Srikrishna, H. Freeze, K. Kayser, H.-J. Gabius, *J. Med. Chem.* **2002**, *45*, 478–491; c) S. André, C. Unverzagt, S. Kojima, M. Frank, J. Seifert, C. Fink, K. Kayser, C.-W. von der Lieth, H.-J. Gabius, *Eur. J. Biochem.* **2004**, *271*, 118–134.
- [58] S. Elliott, T. Lorenzini, S. Asher, K. Aoki, D. Brankow, L. Buck, L. Busse, D. Chang, J. Fuller, J. Grant, N. Hernday, M. Hokum, S. Hu, A. Knudten, N. Levin, R. Komorowski, F. Martin, R. Navarro, T. Osslund, G. Rogers, N. Rogers, G. Trail, J. Egrie, *Nat. Biotechnol.* **2003**, *21*, 414–421.
- [59] a) O. Seitz, *ChemBioChem* **2000**, *1*, 214–246; b) M. Mizuno, *Trends Glycosci. Glycotechnol.* **2001**, *13*, 11–30; c) M. J. Grogan, M. R. Pratt, L. A. Marcaurrelle, C. R. Bertozzi, *Annu. Rev. Biochem.* **2002**, *71*, 593–634.
- [60] a) S. André, P. J. Cejas Ortega, M. Alamino Perez, R. Roy, H.-J. Gabius, *Glycobiology* **1999**, *9*, 1253–1261; b) S. André, B. Frisch, H. Kaltner, D. L. Desouza, F. Schuber, H.-J. Gabius, *Pharmaceut. Res.* **2000**, *17*, 985–990; c) S. André, R. J. Pieters, I. Vrasidas, H. Kaltner, I. Kuwabara, F.-T. Liu, R. M. J. Liskamp, H.-J. Gabius, *ChemBioChem* **2001**, *2*, 822–830; d) I. Vrasidas, S. André, P. Valentini, C. Böck, M. Lensch, H. Kaltner, R. M. J. Liskamp, H.-J. Gabius, R. J. Pieters, *Org. Biomol. Chem.* **2003**, *1*, 803–810; e) S. André, B. Liu, H.-J. Gabius, R. Roy, *Org. Biomol. Chem.* **2003**, *1*, 3909–3916; f) S. André, H. Kaltner, T. Furuike, S.-I. Nishimura, H.-J. Gabius, *Bioconjugate Chem.* **2004**, *15*, 87–98.
- [61] a) D. W. Ohannesian, R. Lotan in *Glycosciences: Status and Perspectives* (Eds.: H.-J. Gabius, S. Gabius), Chapman and Hall, London, **1997**, pp. 459–469; b) S. André, S. Kojima, N. Yamazaki, C. Fink, H. Kaltner, K. Kayser, H.-J. Gabius, *J. Cancer Res. Clin. Oncol.* **1999**, *125*, 461–474; c) H. Lahm, S. André, A. Höflich, J. R. Fischer, B. Sordat, H. Kaltner, E. Wolf, H.-J. Gabius, *J. Cancer Res. Clin. Oncol.* **2001**, *127*, 375–386; d) A. Danguy, I. Camby, R. Kiss, *Biochim. Biophys. Acta* **2002**, *1572*, 285–293; e) I. Camby, N. Belot, F. Lefranc, N. Sadeghi, Y. de Launoit, H. Kaltner, S. Musette, F. Darro, A. Danguy, I. Salmon, H.-J. Gabius, R. Kiss, *J. Neuropathol. Exp. Neurol.* **2002**, *61*, 585–596; f) P. Nangia-Makker, J. Konkin, V. Hogan, A. Raz, *Trends Mol. Med.* **2002**, *8*, 187–192; g) N. Nagy, H. Le-

- gendre, O. Engels, S. André, H. Kaltner, K. Wasano, Y. Zick, J.-C. Pector, C. Decaestecker, H.-J. Gabius, I. Salmon, R. Kiss, *Cancer* **2003**, *97*, 1849–1858.
- [62] a) N. Ahmad, H.-J. Gabius, H. Kaltner, S. André, I. Kuwabara, F.-T. Liu, S. Oscarson, T. Norberg, C. F. Brewer, *Can. J. Chem.* **2002**, *80*, 1096–1104; b) J. Hirabayashi, T. Hashidate, Y. Arata, N. Nishi, T. Nakamura, M. Hirashima, T. Urashima, T. Oka, M. Futai, W. E. G. Müller, F. Yagi, K.-i. Kasai, *Biochim. Biophys. Acta* **2002**, *1572*, 232–235; c) A. M. Wu, J. H. Wu, M.-S. Tsai, J.-H. Liu, S. André, K. Wasano, H. Kaltner, H.-J. Gabius, *Biochem. J.* **2002**, *367*, 653–664.
- [63] H.-J. Gabius, S. Gabius, T. V. Zemlyanukhina, N. V. Bovin, U. Brinck, A. Danguy, S. S. Joshi, K. Kayser, J. Schottelius, F. Sinowatz, L. F. Tietze, F. Vidal-Vanaclocha, J.-P. Zanetta, *Histol. Histopathol.* **1993**, *8*, 369–383.
- [64] a) J. Kopitz, C. von Reitzenstein, S. André, H. Kaltner, J. Uhl, V. Ehemann, M. Cantz, H.-J. Gabius, *J. Biol. Chem.* **2001**, *276*, 35917–35923; b) G. Rapp, H. Abken, J. M. Muche, W. Sterry, M. Tilgen, S. André, H. Kaltner, S. Ugurel, H.-J. Gabius, U. Reinhold, *Leukemia* **2002**, *16*, 840–845; c) L. Santucci, S. Fiorucci, N. Rubinstein, A. Mencarelli, B. Palazzetti, B. Federici, G. A. Rabinovich, A. Morelli, *Gastroenterology* **2003**, *124*, 1381–1394; d) J. Kopitz, S. André, C. von Reitzenstein, K. Versluis, H. Kaltner, R. J. Pieters, K. Wasano, I. Kuwabara, F.-T. Liu, M. Cantz, A. J. R. Heck, H.-J. Gabius, *Oncogene* **2003**, *22*, 6277–6288.
- [65] J. P. Carver, *Pure Appl. Chem.* **1993**, *65*, 763–770.
- [66] a) C.-W. von der Lieth, H.-C. Siebert, T. Kozár, M. Burchert, M. Frank, M. Gilleron, H. Kaltner, G. Kayser, E. Tajkhorshid, N. V. Bovin, J. F. G. Vliegthart, H.-J. Gabius, *Acta Anat.* **1998**, *161*, 91–109; b) R. J. Woods, *Glycoconjugate J.* **1998**, *15*, 209–216; c) J. Jiménez-Barbero, J. L. Asensio, F. J. Cañada, A. Poveda, *Curr. Opin. Struct. Biol.* **1999**, *9*, 549–555; d) A. Imberty, S. Pérez, *Chem. Rev.* **2000**, *100*, 4567–4588; e) M. R. Wormald, A. J. Petrescu, Y.-L. Pao, A. Glithero, T. Elliott, R. A. Dwek, *Chem. Rev.* **2002**, *102*, 371–386; f) T. Weimar, R. J. Woods in *NMR Spectroscopy of Glycoconjugates* (Eds.: J. Jiménez-Barbero, T. Peters), Wiley-VCH, Weinheim, **2003**, pp. 111–144.
- [67] a) H.-C. Siebert, M. Gilleron, H. Kaltner, C.-W. von der Lieth, T. Kozár, N. V. Bovin, E. Y. Korchagina, J. F. G. Vliegthart, H.-J. Gabius, *Biochem. Biophys. Res. Commun.* **1996**, *219*, 205–212; b) M. Gilleron, H.-C. Siebert, H. Kaltner, C.-W. von der Lieth, T. Kozár, K. M. Halkes, E. Y. Korchagina, N. V. Bovin, H.-J. Gabius, J. F. G. Vliegthart, *Eur. J. Biochem.* **1998**, *252*, 416–427.
- [68] B. J. Hardy, *J. Mol. Struct.* **1997**, *395–396*, 187–200.
- [69] a) B. Meyer, T. Peters, *Angew. Chem.* **2003**, *115*, 890–918; *Angew. Chem. Int. Ed.* **2003**, *42*, 864–890; b) H.-C. Siebert, J. Jiménez-Barbero, S. André, H. Kaltner, H.-J. Gabius, *Methods Enzymol.* **2003**, *362*, 417–434.
- [70] a) J. Jiménez-Barbero, J. F. Espinosa, J. L. Asensio, F. J. Cañada, A. Poveda, *Adv. Carbohydr. Chem. Biochem.* **2001**, *56*, 235–284; b) H. Yuasa, H. Hashimoto, *Trends Glycosci. Glycotechnol.* **2001**, *13*, 31–55.
- [71] a) J. F. Espinosa, F. J. Cañada, J. L. Asensio, H. Dietrich, M. Martín-Lomas, R. R. Schmidt, J. Jiménez-Barbero, *Angew. Chem.* **1996**, *108*, 323–326, *Angew. Chem. Int. Ed.* **1996**, *35*, 303–306; b) J. F. Espinosa, F. J. Cañada, J. L. Asensio, M. Martín-Pastor, H. Dietrich, M. Martín-Lomas, R. R. Schmidt, J. Jiménez-Barbero, *J. Am. Chem. Soc.* **1996**, *118*, 10862–10871; c) J. F. Espinosa, E. Montero, A. Vián, J. L. García, H. Dietrich, R. R. Schmidt, M. Martín-Lomas, A. Imberty, F. J. Cañada, J. Jiménez-Barbero, *J. Am. Chem. Soc.* **1998**, *120*, 1309–1318; d) J. L. Asensio, J. F. Espinosa, H. Dietrich, F. J. Cañada, R. R. Schmidt, M. Martín-Lomas, S. André, H.-J. Gabius, J. Jiménez-Barbero, *J. Am. Chem. Soc.* **1999**, *121*, 8995–9000; e) J. M. Alonso-Plaza, M. A. Canales, M. Jiménez, J. L. Roldán, A. García-Herrero, L. Iturrino, J. L. Asensio, F. J. Cañada, A. Romero, H.-C. Siebert, S. André, D. Solís, H.-J. Gabius, J. Jiménez-Barbero, *Biochim. Biophys. Acta* **2001**, *1568*, 225–236.
- [72] H.-C. Siebert, S. André, S.-Y. Lü, M. Frank, H. Kaltner, J. A. van Kuik, E. Y. Korchagina, N. V. Bovin, E. Tajkhorshid, R. Kaptein, J. F. G. Vliegthart, C.-W. von der Lieth, J. Jiménez-Barbero, J. Kopitz, H.-J. Gabius, *Biochemistry* **2003**, *42*, 14762–14773.
- [73] a) S. K. Das, J.-M. Mallet, J. Esnault, P.-A. Driguez, P. Duchaussoy, P. Sizun, J.-P. Héroult, J.-M. Herbert, M. Petitou, P. Sinay, *Angew. Chem.* **2001**, *113*, 1723–1726; *Angew. Chem. Int. Ed.* **2001**, *40*, 1670–1673; b) M. Hricovini, M. Guerrini, A. Bisio, G. Torri, A. Naggi, B. Casu, *Semin. Thromb. Hemost.* **2002**, *28*, 325–334; c) R. Raman, G. Venkataraman, S. Ernst, V. Sasisekharan, R. Sasisekharan, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2357–2362.
- [74] a) B. Casu, A. Naggi, G. Torri, *Semin. Thromb. Hemost.* **2002**, *28*, 335–342; b) P. Jemth, J. Kreuger, M. Kusche-Gullberg, L. Sturiale, G. Giménez-Gallego, U. Lindahl, *J. Biol. Chem.* **2002**, *277*, 30567–30573; c) R. Ojeda, J. Angulo, P. M. Nieto, M. Martín-Lomas, *Can. J. Chem.* **2002**, *80*, 917–936.
- [75] B. Casu, M. Reggiani, G. G. Gallo, A. Vigevani, *Tetrahedron* **1966**, *22*, 3061–3083.
- [76] a) H.-C. Siebert, S. André, J. L. Asensio, F. J. Cañada, X. Dong, J.-F. Espinosa, M. Frank, M. Gilleron, H. Kaltner, T. Kozár, N. V. Bovin, C.-W. von der Lieth, J. F. G. Vliegthart, J. Jiménez-Barbero, H.-J. Gabius, *ChemBioChem* **2000**, *1*, 181–195; b) H.-C. Siebert, M. Frank, C.-W. von der Lieth, J. Jiménez-Barbero, H.-J. Gabius in *NMR Spectroscopy of Glycoconjugates* (Eds.: J. Jiménez-Barbero, T. Peters), Wiley-VCH, Weinheim, **2003**, pp. 39–57.
- [77] a) A. M. Klibanov, *Nature* **2001**, *409*, 241–246; b) C. Mattos, D. Ringe, *Curr. Opin. Struct. Biol.* **2001**, *11*, 761–764.
- [78] L. He, S. André, H.-C. Siebert, H. Helmholtz, B. Niemeyer, H.-J. Gabius, *Bioophys. J.* **2003**, *85*, 511–524.
- [79] H.-C. Siebert, S. André, J. F. G. Vliegthart, H.-J. Gabius, M. J. Minch, *J. Biomol. NMR* **2003**, *25*, 197–215.
- [80] a) J. L. Asensio, F. J. Cañada, M. Bruix, A. Rodríguez-Romero, J. Jiménez-Barbero, *Eur. J. Biochem.* **1995**, *230*, 621–633; b) J. L. Asensio, F. J. Cañada, M. Bruix, C. González, N. Khiar, A. Rodríguez-Romero, J. Jiménez-Barbero, *Glycobiology* **1998**, *8*, 569–577; c) J. L. Asensio, H.-C. Siebert, C.-W. von der Lieth, J. Laynes, M. Bruix, U. M. Soedjanaatmadja, J. J. Beintema, F. J. Cañada, H.-J. Gabius, J. Jiménez-Barbero, *Proteins* **2000**, *40*, 218–236; e) J. F. Espinosa, J. L. Asensio, J. L. García, J. Laynez, M. Bruix, C. Wright, H.-C. Siebert, H.-J. Gabius, F. J. Cañada, J. Jiménez-Barbero, *Eur. J. Biochem.* **2000**, *267*, 3965–3978; f) C. A. Bewley, *Structure* **2001**, *9*, 931–940; g) M.-s. Sung, K. Fleming, H. A. Cheng, S. Matthews, *EMBO Rep.* **2001**, *2*, 621–627; h) H.-C. Siebert, S.-Y. Lü, M. Frank, J. Kramer, R. Wechselberger, J. Joosten, S. André, K. Rittenhouse-Olson, R. Roy, C.-W. von der Lieth, R. Kaptein, J. F. G. Vliegthart, A. J. R. Heck, H.-J. Gabius, *Biochemistry* **2002**, *41*, 9707–9717.
- [81] a) E. F. Hounsell in *Glycosciences: Status and Perspectives* (Eds.: H.-J. Gabius, S. Gabius), Chapman and Hall, London, **1997**, pp. 15–29; b) H. Geyer, *Acta Anat.* **1998**, *161*, 18–35.
- [82] J. Montreuil in *Glycoproteins* (Eds.: J. Montreuil, J. F. G. Vliegthart, H. Schachter), Elsevier, Amsterdam, **1995**, pp. 1–12.
- [83] N. Sharon, *Acta Anat.* **1998**, *161*, 7–17.
- [84] H.-C. Siebert, C.-W. von der Lieth, R. Kaptein, J. J. Beintema, K. Dijkstra, N. van Nuland, U. M. S. Soedjanaatmadja, A. Rice, J. F. G. Vliegthart, C. S. Wright, H.-J. Gabius, *Proteins* **1997**, *28*, 268–284.

Received: August 25, 2003