Lectins: tools for the molecular understanding of the glycocode

Moira Ambrosi,[a] Neil R. Cameron*[a] and Benjamin G. Davis*[b]

[a] *Department of Chemistry, University of Durham, South Road, Durham, UK DH1 3LE*
[b] *Department of Chemistry, University of Oxford, Mansfield Road, Oxford, UK OX1 3TA*

OBC
www.rsc.org/obc

Recent progress in glycobiology has revealed that cell surface oligosaccharides play an essential role in recognition events. More precisely, these saccharides may be complexed by lectins, carbohydrate-binding proteins other than enzymes and antibodies, able to recognise sugars in a highly specific manner. The ubiquity of lectin–carbohydrate interactions opens enormous potential for their exploitation in medicine. Therefore, extraordinary effort is made into the identification of new lectins as well as into the achievement of a deep understanding of their functions and of the precise mechanism of their association with specific ligands. In this review, a summary of the main features of lectins, particularly those found in legumes, will be presented with a focus on the mechanism of carbohydrate-binding. An overview of

Moira Ambrosi was born in Florence (Italy) in 1971. She graduated in Chemistry at the University of Florence in 1998 with 1st class honours. She then moved to Durham (UK) to undertake a PhD with Dr Neil Cameron, working on glycopolymers, from October 1999 and graduated in September 2002. She currently has a post-doctoral position at the Chemistry Department of the University of Florence, working with Professor Piero Baglioni on derivatives of vitamin C with biomedical applications. She is also collaborating with Procter & Gamble on liquid detergents containing bleaching agents, and with Covercolor (Italy) on the improvement of mechanical properties of ceramic materials. She has recently become "Cultore della Materia" of Physical Chemistry.

Neil Cameron was born in Ayr, Scotland, in 1969 and studied chemistry at the University of Strathclyde in Glasgow, graduating with 1st class honours in July 1991. He then remained at Strathclyde to undertake a PhD with Professor David Sherrington, working on highly porous polymer materials (PolyHIPEs), from October 1991 and graduated in 1995. His first post-doctoral position, on nitroxide-mediated polymerisations, was in Eindhoven University of Technology with Professor Anton German, from April 1995 until March 1996, after which he returned to the UK as a post-doctoral associate in the laboratory of Professor Iain Cowie at Heriot-Watt University in Edinburgh, investigating ageing of polymer blends. In October 1997 he joined the academic staff in the Chemistry Department at the University of Durham, where he is currently a Reader (from October 2005). His research interests are in synthetic macromolecular chemistry, with a particular emphasis on materials for use in biological and medical applications. His work to date has led to 50 publications and he has significant current funding from a variety of sources (UK research councils, E.U., industry and charities). In 2003 he became the Associate Director (for Durham) of the Interdisciplinary Research Centre in Polymer Science and Technology (Polymer IRC) and he was awarded the 2003 Young Researchers' Award by the Macro Group UK, a joint subject group of the Royal Society of Chemistry and the Society for Chemical Industry.

Ben Davis got his BA (1993) and DPhil (1996) from the University of Oxford. During this time he learnt the beauty of carbohydrate chemistry under the supervision of Professor George Fleet. He then spent two years as a postdoctoral fellow in the laboratory of Professor Bryan Jones at the University of Toronto exploring protein chemistry and biocatalysis. In 1998 he returned to the UK to take up a lectureship at the University of Durham; in autumn 2001 he moved to the Dyson Perrins Laboratory, University of Oxford and a fellowship at Pembroke College, Oxford. His group's research centres on chemical biology with an emphasis on carbohydrates and proteins. In particular, their interests encompass synthesis and methodology, inhibitor design, biocatalysis, enzyme mechanism, protein engineering, drug delivery, molecular modelling, molecular biology and glycoscience. This work has been published in 13 patents, 6 books and book chapters and >50 papers and has been recognised by the 1999 RSC Meldola Prize & Medal, a DTI Smart Innovation Award (1999), a Mitzutani Foundation for Glycoscience Award (2000), an AstraZeneca Strategic Research Fund Award (2000), the 2001 RSC Carbohydrate Chemistry Award, a 2002 Phillip Leverhulme Prize for Biochemistry and Molecular Biology and selection for the 2003 MIT TR100.

Moira Ambrosi   **Neil R. Cameron**   **Benjamin G. Davis**

lectin–carbohydrate interactions will also be given, together with an insight into their energetics. In addition, therapeutic applications of lectins will be discussed.

1 Lectins

Lectins (from *lectus*, the past participle of *legere*, to select or choose)[1] are defined as carbohydrate binding proteins other than enzymes or antibodies[2] and exist in most living organisms, ranging from viruses and bacteria to plants and animals. Their involvement in diverse biological processes in many species,[3] such as clearance of glycoproteins from the circulatory system,[4,5] adhesion of infectious agents to host cells,[6] recruitment of leukocytes to inflammatory sites,[7] cell interactions in the immune system, in malignancy and metastasis,[8] has been shown. A doctoral thesis from 1888 on the agglutination of red blood cells by extracts of castor beans (the active component being a protein named ricin) is often cited as the beginning of lectinology,[9] however the agglutination of erythrocytes by rattlesnake venom, observed around 1860,[10] has in fact been suggested as the first demonstrated example of lectin activity. The first pure lectin, concanavalin A (Con A, from jack beans), was isolated in 1919 by Sumner,[11] who also demonstrated its sugar specificity.[12] Subsequently, lectins played a crucial role in elucidating the molecular basis for blood group specificity.[13] Recent advances in carbohydrate chemistry and the study of protein–ligand interactions mean that lectins are currently the focus of intense research. In particular, the investigation of their role in cell recognition, as well as their employment as invaluable tools for the study of complex carbohydrates in solution and on cell surfaces, is contributing markedly to advances in glycobiology.

Lectins interact with carbohydrates non-covalently in a manner that is usually reversible and highly specific.[14] Classical lectins contain two or more carbohydrate-binding sites; therefore, their interaction with sugars on the surface of erythrocytes results in the cross-linking of several blood cells and their subsequent precipitation. This phenomenon, known as cell agglutination, is a major attribute of the activity of lectins and has been used classically and routinely for their detection and characterisation (however it is now recognised that cell agglutination is not a defining feature of lectins[2]). Both the agglutination and precipitation processes are inhibited by the carbohydrate for which the lectin is specific.

According to the monosaccharide ligand toward which they exhibit the highest affinity, Sharon classified lectins into five groups: mannose, galactose/*N*-acetylgalactosamine, *N*-acetylglucosamine, fucose and *N*-acetylneuraminic acid[15] (how-

ever it should be noted that this classification ignores certain important monosaccharides such as mannose-6-phosphate and *N*-acetylgalactose-4-sulfate). With the exception of fucose, all these sugars are in the D-configuration. Despite the weak intrinsic protein–carbohydrate affinity, with typical dissociation constants in the millimolar range for monosaccharides, lectin specificity is usually high.[16] Therefore, apart from a few exceptions, neither do lectins specific for galactose bind glucose, nor do those specific for mannose bind galactose. Interestingly, certain variations at the C-2 position of the pyranose ring may be tolerated, so that most lectins that bind galactose also interact with *N*-acetylgalactosamine. Certain lectins combine preferentially with either the *α*- or the *β*-anomer, whereas others lack anomeric specificity. The properties of the aglycon can influence markedly the lectin recognition activity. For instance, aromatic glycosides bind to Con A much more strongly than aliphatic ones, indicating the presence of a hydrophobic region in the proximity of the carbohydrate-binding site.[17] Despite being familiar and practically useful, this sugar-specificity based classification is nowadays becoming obsolete. First, there are marked differences in the fine specificities of lectins within a single category. Secondly, an increasing number of lectins that do not show high affinity toward simple saccharides have been identified.

Many lectins can be grouped in families characterised by similar sequences and structural organisation (Table 1). Sequence similarity with known lectins provides a valuable guideline for the identification of new ones. The determination of the amino acid sequences of several hundreds of lectins, together with the elucidation of about thirty 3D-structures, has allowed the replacement of more traditional divisions according to the lectin origin or ligand by a classification based on common structural features.[15] Most lectins belong to three classes: (1) simple, (2) mosaic (or multidomain) and (3) macromolecular assemblies. Simple lectins consist of a small number of subunits, not necessarily identical, each of molecular weight usually below 40 kDa. Each monomeric unit contains a carbohydrate-binding site. This class comprises practically all known plant lectins[18] and most members of the galectin family (formerly known as S-lectins), a group of *β*-galactoside specific animal lectins.[19] Mosaic lectins are composite molecules consisting of several kinds of protein domains, only one of which possesses a carbohydrate-binding site. This class includes diverse proteins from different sources: viral hemagglutinins[20] and animal lectins of C-, P- and I-types.[21] Macromolecular assemblies are common in bacteria. They are filamentous organelles consisting of helically arranged subunits (pilins) assembled in a well-defined order.[22] Within

Table 1 Examples of lectins, the families to which they belong and their glycan ligand specificities[a]

Lectin name	Family	Glycan ligands
Plant lectins[b]		
Concanavalin A (Con A; jack bean)	Leguminosae	Man/Glc
Wheat germ agglutinin (WGA; wheat)	Gramineae	(GlcNAc)$_{1-3}$, Neu5Ac
Ricin (castor bean)	Euphorbiaceae	Gal
Phaseolus vulgaris (PHA; French bean)	Leguminosae	None known
Peanut agglutinin (PNA; peanut)	Leguminosae	Gal, Galβ3GalNAcα (T-antigen)
Soybean agglutinin (SBA; soybean)	Leguminosae	Gal/GalNAc
Pisum sativum (PSA; pea)	Leguminosae	Man/Glc
Lens culinaris (LCA; lentil)	Leguminosae	Man/Glc
Galanthus nivalus (GNA; snowdrop)	Amaryllidaceae	Man
Dolichos bifloris (DBA; horse gram)	Leguminosae	GalNAcα3GalNAc, GalNAc
Solanum tuberosum (STA; potato)	Solanaceae	(GlcNAc)$_n$
Animal lectins		
Asialoglycoprotein receptor (ASGPR) H1	C-type	Gal
Galectin-3	Galectins	Gal
Sialoadhesin[132]	I-type	Neu5Ac
Cation-dependent mannose-6-phosphate receptor (CD-MPR)[143]	P-type	Man6P
C-reactive protein (CRP)[144]	Pentraxins	Gal, Gal6P, galacturonic acid

[a] Selected from ref. 25 and ref. 10, unless alternative source given. [b] Lectin name is in regular type, the scientific name of the plant from which it is obtained is in italics.

each class, proteins can be grouped into families, with similar sequences and structural properties. Nowadays, as the genome analysis of various organisms, including humans, is completed, lectins can also be valuably classified through the concept of protein (gene) families, whose members show evolutionary kinship.[23]

The intention of this review is not to provide a comprehensive survey of the literature on lectins, but rather to serve as an introduction to the topic for the reader unfamiliar with the area. The reader is directed to excellent reviews on plant[24] and animal[10] lectins for further information, and a detailed description of the history of lectinology can be found in the recent article by Gabius *et al.*[25] In our article, the main features of lectins will be summarised, with an emphasis particularly on legume lectins. The abundance of these proteins in plant seeds, their solubilities and their wide range of saccharide specificities make them good model systems, tools for elucidating protein–carbohydrate interactions as well as for biomedical and biotechnological applications.

1.1 Legume lectins

Legume lectins represent the largest and most thoroughly studied family of simple lectins. Around 100 members have been characterised, almost all isolated from the seeds of the plants in which they are present.[18,26] Concanavalin A (Con A), the lectin from the jack bean, is the prototype member of the family. The relative abundance of this protein in jack bean, the ease of its preparation and the large number of saccharides with which it can interact, have led to numerous studies on Con A, markedly accelerated by the discovery in 1969 that cells transformed by DNA tumour viruses or carcinogens were agglutinated by the lectin more readily than normal cells.[27] About 85% of the binding sites for Con A, which are in a cryptic form on normal cells, were found to be exposed on the surface membrane of transformed cells. It was therefore hypothesised that the change in structure of the surface membrane, resulting in the exposure of the sites, could produce the change in cellular regulatory mechanisms associated with transformation.

1.1.1 Structural features. Typically, legume lectins consist of two or four identical or near-identical subunits (protomers) of 25–30 kDa each, which are commonly single polypeptide chains of about 250 amino acids presenting one or two *N*-linked oligosaccharides. Each protomer typically contains a carbohydrate-combining site, a tightly bound Ca^{2+} and a transition metal ion, usually Mn^{2+}. Approximately 20% of the amino acid residues are invariant in all legume lectins and another 20% are similar. The conserved amino acids include several of those involved in the interaction with the saccharide and almost all the residues that coordinate the metal ions. The resolution of 3D-structures of about ten legume lectins has shown that each subunit is constituted largely—nearly 60%—of β-strands mutually connected by loops. For all legume lectins known so far, the tertiary structure is made up of two anti-parallel β-sheets, a six-stranded flat "back" and a seven-stranded curved "front", connected by a five-stranded β-sheet, giving the well known "jellyroll" motif, also referred to as the "lectin fold"[28] (Fig. 1a).

The subunit structures of different legume lectins can be nearly superimposed, irrespective of the protein specificity. Despite their similarities at the primary, secondary and tertiary structural monomeric level, legume lectins exhibit considerable variation in their quaternary structure: small differences in the amino acid sequences at the monomer–monomer interfaces and the presence/absence of glycosylation affect the monomers' association modes. In the case of lectins with "canonical" quaternary structure, such as Con A, pea lectin, favin and *L. ochrus*, dimerisation involves anti-parallel side-by-side alignment of the flat six-stranded β-sheets of the two monomers, resulting in the formation of a continuous 12-stranded sheet that extends across the dimer interface. A considerable portion
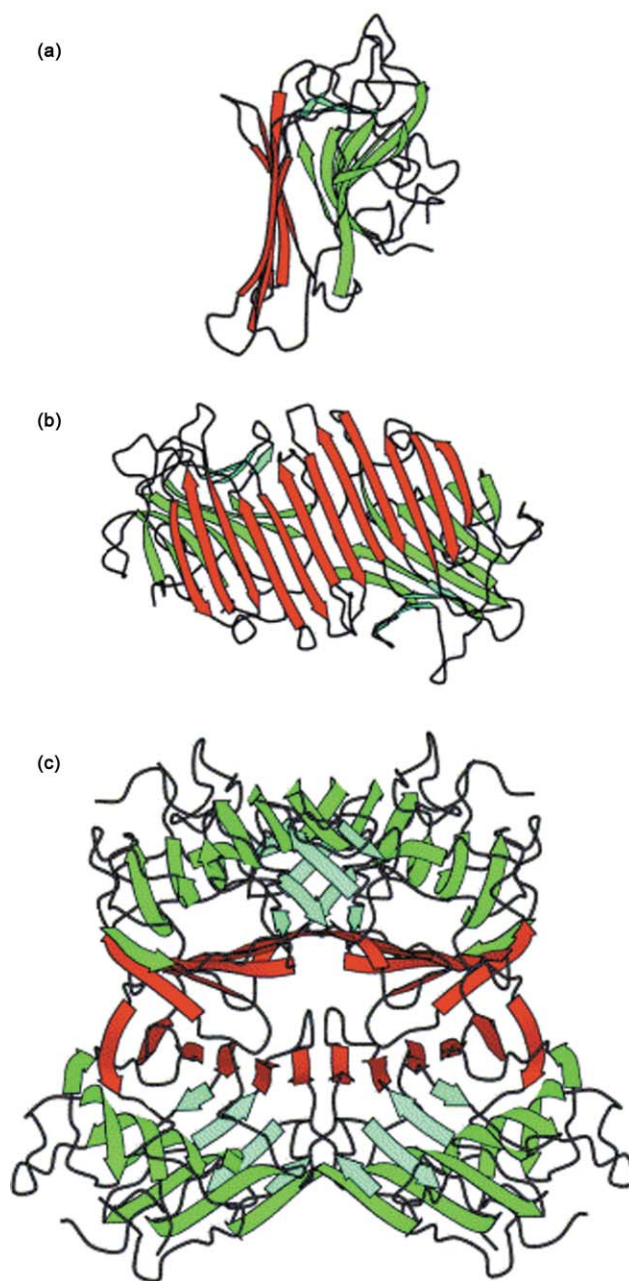

Fig. 1 (a) Representative tertiary structure of a legume lectin monomer; (b) dimerisation in Con A; (c) tetramerisation in Con A. Reprinted with permission from Elsevier.[29]

of the surface area is buried in the process: *ca.* 1000 Å² per monomer (Fig. 1b). Further association of two dimers gives the tetrameric assembly of Con A observed in physiological conditions[29] (Fig. 1c). Peanut agglutinin (PNA), the lectin from *Arachis hypogaea*, shows an unusual "open" quaternary structure, where the homotetramer possesses neither 222 (D_2) nor 4-fold (C_4) symmetry.[30] This structure is a refinement of earlier low resolution X-ray crystallographic work[31] suggesting D_2 symmetry, however the conclusion that the PNA tetramer is a dimer of a dimer, with a dimeric association similar to that observed in lectin IV of *Griffonia simplicifolia* (GS4), was confirmed.[30] The open quaternary association is stabilised mainly by hydrophobic, hydrogen-bonded and water-mediated interactions. While the dimerisation process results in the burial of 1920 Å² surface area of which 71% is non-polar, the percentage of hydrophobic surface area buried during the further formation of the tetramer is relatively low. The dimers interact in such a way that monomers belonging to two different GS4-like dimers associate in a canonical fashion except that the sheets

do not interact directly, but by means of six water bridges. As a result, the formed interface is intrinsically less stable than the analogous interface in Con A.[29] As with most plant lectins, the quaternary structure of PNA depends on the pH. The lectin is a tetramer at physiological pH,[32] which dissociates reversibly into dimers at pH below 5.1. Below pH 3.4 PNA is totally dimeric.[33] Interestingly, it has been found that a partially unfolded intermediate of PNA retains carbohydrate binding ability with affinities that are 75–85% of those of native PNA.[34]

1.1.2 Carbohydrate-binding site.

During the past 15 years there has been significant progress in elucidating the features of lectins involved in carbohydrate binding. X-ray crystallography of the proteins complexed with their ligands, site-directed mutagenesis experiments and molecular modelling have allowed the identification of the chemical groups belonging to both interacting species involved in the binding and of the types of bond formed. Studies of lectin–oligosaccharide complexes are especially interesting, providing the basis for the understanding of the proteins' interaction with natural ligands. Generally, lectins show exquisite specificity for di-, tri- and tetrasaccharides, with association constants significantly higher than those for the corresponding monosaccharides.

Carbohydrate-binding sites are often shallow depressions on the surface of the protein. In all cases the combining site appears to be preformed,[35] since few conformational changes occur upon binding. In all legume lectins, irrespective of their specificity, four invariant amino acid residues participate in the ligand binding: an aspartic acid, an asparagine, a glycine (conserved in all the lectins of the family apart from Con A) and an aromatic amino acid[36] or leucine:[37] Asp83, Gly104, Asn127 and Tyr125 for PNA.[38] Replacement of the aspartic acid or asparagine by site-directed mutagenesis results, in several cases, in the loss of the lectin sugar-binding ability.[39] However, despite the conservation of key amino acids involved in the binding of the carbohydrate, different legume lectins can show different specificity. For instance, while Con A binds mannose and glucose, PNA, ECorL and SBA bind galactose. Therefore, while the constellation of highly conserved amino acids provides the framework required for binding, specificity apparently arises from the variability of amino acid residues in other regions of the combining pocket. The sugar-combining site is made up by amino acid residues residing in four loops, designated A, B, C and D.[40] The invariant aspartic acid and glycine belong to A and B, respectively, whereas the asparagine and the hydrophobic residue are in loop C. Additional interactions are provided by amino acids in loop D, suggesting a correlation between this loop and the lectin specificity. In fact, loop D is highly variable in terms of length, sequence and conformation. Thus, for instance, the size of this loop is identical in all mannose-specific lectins.[41]

The Ca^{2+} and Mn^{2+} (or other transition metal) are situated around 4 Å apart and in close proximity to the sugar-combining pocket. Although not always directly involved in the carbohydrate binding, the cations help the positioning of the amino acid residues interacting with the glycoside. The two invariant aspartic acid and asparagine residues also participate in coordinating Ca^{2+}. A rare *cisoid*-peptide bond between the critical asparagine and the preceding amino acid, usually alanine, confers on the asparagine residue the proper orientation.[15]

Lectins bind carbohydrates through a network of hydrogen bonds and hydrophobic interactions. This is highlighted by the schematic representation of the T-antigenic disaccharide (the carbohydrate structure for which peanut agglutinin shows the highest affinity), Galβ(1→3)GalNAc (Tant), in the binding site of PNA shown in Fig. 2.

The four invariant hydrogen bonds are Asp83 Oδ1-Gal O3, Asp83 Oδ2-Gal O4, Gly104 N-Gal O3 and Asn127 Nδ2-Gal O3.[42–45] There is a key stacking interaction between the aromatic ring of Tyr125 and the hydrophobic α-face patch of
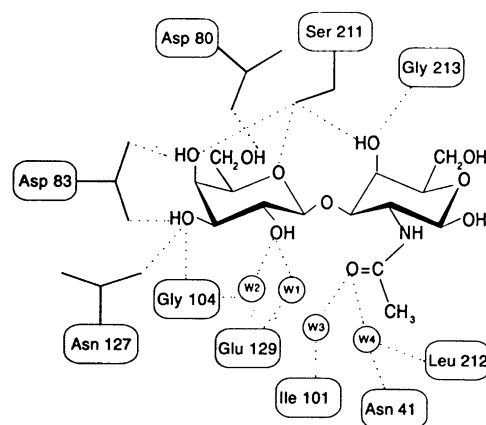

Fig. 2 Schematic representation of protein–carbohydrate interactions in the PNA–Tant complex (reprinted with permission from.[15] Copyright (1998) American Chemical Society).

the galactose unit of the disaccharide. Further hydrogen bonds are Asp80 Oδ2-Gal O6 and that involving Ser211 Oγ.

Van der Waals forces, although rather weak (usually a fraction of 4.2 kJ mol^{-1} for each pair of atoms), are frequently numerous, contributing significantly to the overall binding.[15] The steric disposition of hydroxyl groups in carbohydrates creates hydrophobic patches[46] on the sugar surface that can interact with hydrophobic regions of the protein.[47] For example in the PNA–Tant complex (Fig. 2), about 60 van der Waals contacts are formed between the disaccharide and amino acid residues within 4 Å of the carbohydrate.[44]

Contacts between the ligand and the protein are often mediated by water molecules. Water acts as a molecular "mortar";[48] its small size and its ability to behave as both hydrogen donor and acceptor make it near-ideal for this function. Tightly bound water molecules can in effect be considered as structural, *i.e.* an extension of the protein surface. Thus, water plays a significant role in carbohydrate recognition, imparting in some cases exquisite specificity. Comparison of crystal structures of PNA complexed with different ligands (T-antigen, Tant,[43] methyl-β-galactoside, MeGal,[44] N-acetyllactosamine, LacNAc[44] and lactose, Lac[45]) has shown that water bridges involving water molecules W1 and W2 (Fig. 2) occur in all four complexes. Moreover, major additional interactions between the T-antigen and PNA, compared to the other disaccharides, occur through two water molecules involving the carbonyl O atom of the acetamido group: W3 connects the acetamido O atom to Ile101 O, while W4 connects it to Asn41 Nδ2 and to Leu212 N. These two water molecules also exist in the other complexes, but in these cases they interact only with the protein, due to the lack of sugar atoms in their immediate vicinity. Therefore, the additionally high specificity of PNA for the T-antigenic disaccharide appears, critically, to be generated by W3 and W4.

A detailed study of water molecules in the combining site of PNA has been performed.[44] Interestingly, as already reported for other carbohydrate-binding proteins,[49] ordered water molecules can be found in the unligated form at positions corresponding to hydroxyl groups in the ligated form (for instance, a water molecule occupies the position of the acetamido carbonyl O in the GalNAc moiety of the T-antigen). In general, water molecules in the carbohydrate-binding region mimic the ligand to a substantial extent not only at the primary site, but also in the regions adjacent to it. Molecular dynamics (MD) simulations carried out for the PNA–Tant complex revealed that there is a constant exchange of water molecules between the binding site and the bulk, especially in the proximity of Asn41, Asn127 and Glu129.[50] As evidenced by the short mean residence times and by the trajectories of water molecules during the formation of water bridges, these molecules do not stay in the sugar-combining pocket for long; rather they are in constant motion. MD and

crystallographic results were comparable. It was found that the number of both direct and water-mediated hydrogen bonds calculated by MD were higher than those determined by crystallography. Due to the large number of water molecules that have access to the site, the simulation data revealed 23 water-mediated interactions, including those present in the crystallographic data. The same authors showed that most intramolecular, water-mediated interactions involving the T-antigen galactose moiety are diminished or even lost upon complexation. By contrast, a water bridge with a long mean lifetime is established within the complex between Gal O6 and GalNAc O6. It was hypothesised that this very stable water bridge, occurring frequently during the simulation, could have a role in maintaining the stretched conformation of the saccharide. Intramolecular GalNAc O1-W-GalNAc C=O and GalNAc O4-W-GalNAc C=O, pre-existent to complex formation, were still present in the active site with an increased persistence. As was expected, the hydration number of the T-antigen was significantly reduced upon complexation, going from about 33 to 13. Pratap and co-workers[45] carried out comparative molecular dynamics simulations for PNA–Tant and PNA–Lac complexes. The results showed that the number of alternative binding modes is higher for the former than for the latter, resulting in the "breaking" of the enthalpy–entropy compensation that is characteristic of the binding of monovalent sugars to lectins. Such molecular dynamics simulations provide important contributions that complement the static view of crystallography, and provide a better sense of molecular interactions in the *in vivo* state.

1.1.3 Physiological functions. Despite their long history, the true physiological role of legume lectins is still not well understood. Many hypotheses have been formulated in the course of the years but, at present, no physiological function for any legume lectin has been established with certainty. The difficulty in assigning a precise role arises from several of their features. The defining characteristic of all lectins is their carbohydrate-binding ability. This activity has been preserved during evolution, suggesting that it is essential for the exploitation of their function. Nevertheless, some lectins possess additional activities that may be non-secondary in determining their physiological role. Several legume lectins, for instance, present adenine-combining sites able to bind active forms of the cytokinins, a major class of plant hormones.[51] The specificity of adenine-binding sites has been maintained during evolution despite variations in carbohydrate specificity and tissue distribution. Therefore, it may be possible that in such cases proteins have a primary role unrelated to carbohydrates, *e.g.* hormone-binding, and that the carbohydrate-combining activity plays a regulatory or transducing function.[52]

In general, a single legume plant can contain a variety of lectins that may have evolved by gene duplication and become specialised for different roles in the plant.[53] Furthermore, despite being concentrated in plant seeds, lectins are also present in different tissues, where their function probably requires lower concentrations. One of the most credited theories concerning legume lectins' physiological role, extensible to all plant lectins, considers them as defence agents against predators.[54] An essential feature of any active defence agent is the ability to recognise specifically the pathogen. Based on their carbohydrate-binding specificity and also considering their abundance in plant seeds and bark, lectins seem to possess all the necessary characteristics to exploit this function. Early investigators noted the similarities of lectins to antibodies and hypothesised that lectins might function as plant antibodies.[55] One of the early events in the defence response of legumes is the production of phytoalexins,[56] whose synthesis is stimulated by the release of elicitors, many of which are oligosaccharides that arise from the breakdown of plant or pathogen cell wall components.[57] Therefore, lectins could participate in this step of the defence process either acting as receptors for elicitors or organising the elicitors in structures

required by their receptors. It has been shown that the levels of *Dolichos biflorus* stem and leaf lectin, DB58, increase upon plant wounding.[52]

Again, based on their carbohydrate-binding ability, lectins have also been thought to be involved in the establishment of symbiosis between nitrogen-fixing bacteria and plants.[3,58] Legumes are able to associate specifically and form symbioses with soil bacteria of the rhizobia family, a phenomenon that makes them independent from soil nitrogen supplies. The idea that lectins are responsible for this association was advanced over 20 years ago and it has been the subject of much controversy. The nitrogen-fixing symbiosis is a multistep process that requires the formation of the root nodule, followed by the adhesion of the bacteria to the roots and, finally, the internalisation of the bacteria into the nodule.[59] Initiation of nodulation depends on the production of Nod factor signals by the rhizobia. Nod factors are lipochitooligosaccharides that are modified in different rhizobial strains, so generating the basis for host–strain specificity. The finding that a legume binds to a specific rhizobial species and not to bacteria that are symbionts for other legumes suggests that the interaction occurs between lectins in the roots of the plant and carbohydrates on the bacteria surface. DB46, the *Dolichos biflorus* root lectin present on the surface of root hairs, has been found to bind with several Nod factors.[52] Despite the fact that molecular genetics experiments support this hypothesised role of lectins as receptors for oligosaccharides produced during the symbiosis processes, several inconsistencies can be pointed out. First of all, for most host–symbiont systems there is no proof of the presence of lectins and of the respective ligands on the two interacting species. Secondly, the correlation between the sugar specificity of legume lectins and their ability to recognise bacteria appears not to be particularly strong. Furthermore, and most strikingly, mutants of soybean lacking the lectin are still nodulated by the rhizobial symbiont.[15]

2   Lectin–carbohydrate interactions: towards therapy

The idea that lectins and carbohydrates are excellent as cell recognition markers originates from the findings that both classes of compound are commonly present on the cell surface and that sugars possess tremendous coding capacity. The ability of lectins to distinguish between subtle variations of oligosaccharide structure makes them perfectly suitable as decoders for such carbohydrate-encoded information. In other words, whilst sugars are able to carry the biological information, lectins are capable of deciphering this "glycocode".

2.1   Multivalency

The most striking features of lectin–monosaccharide interactions are that they are relatively weak, with dissociation constants usually in the millimolar range for monosaccharides,[16,60] and that they may show relaxed specificity, when compared to the strict nature of enzyme–substrate associations. The reason for this weakness lies in the solvent-exposed nature of the lectin binding-sites, which are shallow pockets making few direct contacts with the ligands.[61] There is, in fact, a significant difference in affinity between these shallow sites and deep sites, as is well illustrated by the influenza haemagglutinin lectin, which binds sialic acids with an approximately 1000-fold lower affinity than a neuraminidase found in the same virus.[62] Nevertheless, lectins exhibit both high affinity and exquisite specificity for oligosaccharide structures of glycoproteins and glycolipids on the cell surface. If this was not the case, lectins could not act as recognition molecules in biological processes. It has therefore been suggested that multiple protein–carbohydrate interactions are involved in the recognition event, giving the required high affinity and specificity.[60] Thus, polyvalent associations occur throughout biology, showing a number of characteristics that

monovalent interactions do not exhibit.[63] In most cases, biological systems seem to use polyvalent interactions rather than a very strong single one. Therefore, multivalency is employed in Nature not only to achieve the necessary high affinity, but also to ensure the correct functioning of the cells through high specificity. For example, asialoglycoprotein receptor-mediated clearance of erythrocytes occurs only when the density of galactose moieties reaches a critical level.[64] Moreover, multivalent lectin–carbohydrate interactions can also lead to the formation of surprisingly homogeneous complexes and lattices. These have been suggested to be involved in certain physiological processes such as T cell death, the signal for which is induced by glycoprotein clustering in the presence of galectin-1.[65] Furthermore, protein–carbohydrate interactions may imply significant conformational changes that can represent biological signals. Decavalent binding of IgM to the bacterial surface initiates the complement cascade leading to the death of the microorganism.[63]

All classes of antibodies have multiple equivalent receptor sites. Multivalency leads to high affinity binding to surfaces that show repeated epitopes, such as almost all invading pathogens. Mannose residues on the tail (the F_c portion) of the antibody interact with mannose receptors (the F_c receptors) on the surface of a macrophage. The interaction of a single F_c portion with its receptor seems to be too weak to induce a response by the macrophage, whereas multiple antibodies bound to the surface of the bacterium can cross-link the F_c receptors, triggering an internal signal in the macrophage to ingest the infective agent.[66]

2.2 Energetics

A full understanding of the mechanism of protein–carbohydrate association at a molecular level requires the elucidation of both structural and energetic aspects of the process. As briefly discussed in the previous sections, a large amount of structural data is now available. However, the link between structure and energetic properties remains too obscure to allow a description of the energetics of the interaction from structural information.[48] Historically, the strength of protein–ligand complexation has been determined by hemagglutination and precipitation inhibition assays, although several spectroscopic techniques, affinity chromatography and equilibrium dialysis have also been utilised. The so obtained values of the binding constant, K, and of the free energy of binding, ΔG, give little information about the forces involved in the complex formation. Moreover, the enthalpy and entropy of association are derived only indirectly from van't Hoff analyses of the temperature dependence of the free energy:

$$\frac{d \ln K}{dT} = \frac{\Delta H}{RT^2} \qquad (1)$$

The employment of this equation requires the assumption that the enthalpy of binding, ΔH, is not a function of the temperature, i.e. $\Delta C_P = 0$, since $\left(\frac{\partial H}{\partial T}\right)_P = \Delta C_P$. However, the heat-capacity change, ΔC_P, for ligand binding in aqueous media is typically non-zero, making the accuracy of the calculated values doubtful.

The accessibility of commercial high-resolution microcalorimeters has provided a direct method of evaluation of accurate thermodynamic data that can complement the structural information. Isothermal titration microcalorimetry, ITC, is the only technique that allows direct determination of K and ΔH in a single experiment. Values of ΔC_P are also obtained performing titrations at different temperatures. ΔC_P is a powerful indicator of solvent behaviour during binding[67] and provides invaluable information about the driving forces for association. Accurate values of ΔC_P are available only by microcalorimetry.

2.2.1 Monovalent ligands. Calorimetrically determined thermodynamic constants for lectin–carbohydrate interactions are listed in Table 2.

In almost all cases the enthalpy of binding is more negative than the free energy. It is noteworthy that this pattern is generally found for association in aqueous solution and is not a special feature of lectin–ligand complexation.[68,69] It has already been pointed out that water molecules are strongly involved in protein–sugar interactions. Watson et al.[70] have reported that in the case of complexation of glucose with glycogen phosphorylase, water-mediated hydrogen bonds are as strong as the direct ones, reinforcing the concept of structural water as an extension of the protein surface. However, the displacement of a large number of water molecules during binding can also be observed. The effect of this release of water has long been debated. Numerous models have been formulated in order to explain the thermodynamic properties of apolar organic molecules in water, particularly for the surprising cases where the most significant term opposing dissolution is enthalpic rather than entropic.[48] It has been proposed that although the strength of hydrogen bonds in the hydration shell of the protein is higher than in bulk solvent, the fraction of broken bonds is also higher, due to geometric constraints imposed on water by the solute. Thus, the return of water of solvation to bulk water might be enthalpically driven, providing a favourable contribution to the enthalpy of complexation.[71] Monte Carlo simulations, carried out to investigate the structure of water near the surface of the lectin binding site, have shown highly disordered water in the proximity of the protein surface. The phenomenon has been explained by assuming that the protein surface could not be complementary to any low-energy structure that the water could adopt.[72] Chervenak and co-workers[67] have reported a calorimetric evaluation of the thermodynamic binding parameters for several systems in light and heavy water. In all cases desolvation was found to contribute to a significant fraction of the binding enthalpy: ca. 25% for lectin–sugar association. To date, the role of protein–ligand hydrogen bonding vs. solvent reorganisation as contributors to the enthalpy of binding remains the most polemic area of discussion.

Another common feature of lectin–carbohydrate interactions is the strong linear enthalpy–entropy compensatory behaviour. This offset has been interpreted both in terms of changes in the degrees of freedom of the ligand upon binding[73,74] and in terms of solvent reorganisation.[75] Chervenak et al.[76] have determined the configurational and solvation entropies for the formation of a series of complexes with Con A and DGL (Table 3).

The total entropy of complexation can be separated into terms accounting for changes in solvation and losses of configurational, rotational and translational degrees of freedom.

$$\Delta S = \Delta S_{solv} + \Delta S_{config} + \Delta S_{rot} + \Delta S_{trans} \qquad (2)$$

The sum of the first two terms has been referred to as the unitary contribution, while the sum of rotational and translational entropies has been defined as the cratic contribution.[77] Knowing that:

$$\left(\frac{\partial S}{\partial T}\right)_P = \frac{\Delta C_P}{T} \qquad (3)$$

ΔS_{solv} can be expressed as follows:

$$\Delta S_{solv} = \Delta S_{solv}^* + \Delta C_p \ln\left(\frac{T}{T^*}\right) \qquad (4)$$

where T^* is the temperature at which there is no solvent contribution and is equal to 385.15 K. ΔS_{solv}^* contains protonation and electrostatic contributions and for lectin–ligand interactions can be considered equal to zero. Murphy et al.[78] have proposed that for the formation of a 1 : 1 complex with 1 M as the standard state, the entropic cratic contribution, $\Delta S_{rot} + \Delta S_{trans}$, is equal to −33.5 J mol^{-1} K^{-1}. Thus, accurate measurements of ΔC_P provide ΔS_{solv} and, consequently, ΔS_{config}. For the systems

Table 2 Calorimetrically derived thermodynamic parameters of lectin–carbohydrate interactions

Lectin	Carbohydrate	$10^{-3} \times K/\mathrm{M}^{-1}$	$\Delta G/\mathrm{kJ\,mol}^{-1}$	$\Delta H/\mathrm{kJ\,mol}^{-1}$	$\Delta S/\mathrm{J\,mol}^{-1}\,\mathrm{K}^{-1}$	$\Delta C_\mathrm{P}/\mathrm{J\,mol}^{-1}\,\mathrm{K}^{-1}$	Reference
WBAI[a]	Gal	1.2	−17.5	−24.3	−23	—	145
	GalNAc	7.2	−22.0	−28.0	−20	—	
	MeαGal	6.6	−21.8	−23.5	−6	—	
	MeβGal	1.0	−17.1	−19.7	−9	—	
Con A	MeαMan	—	−22.2	−27.6	−18.3	−251	68
	Manα1,6(Manα1,3)Man[b]	—	−30.1	−41.0	−36.5	−460	
Con A	MeαMan	11.9	−22.9	−29.2	−22	—	74
	Man	2.1	−18.6	−23.9	−18	—	
	MeαGlc	2.7	−19.2	−18.1	+4	—	
	Glc	0.6	−15.4	−17.1	−6	—	
Pea	MeαMan	1.9	−18.3	−27.3	−31	—	
	Man	0.9	−16.5	−24.8	−29	—	
	MeαGlc	0.6	−15.5	−13.1	+8	—	
Con A	MeαMan	8.2	—	−34.3	−40.7	—	146
	Manα1,6(Manα1,3)Man	49.0	—	−60.3	−92.7	—	
DGL[c]	MeαMan	—	−20.5	−32.6	−40.7	−234	76
	Manα1,3Man	—	−26.4	−47.7	−71.6	−167	
	Manα1,6Man	—	−21.0	−36.0	−50.5	−92	
	Manα1,6(Manα1,3)Man	—	−34.3	−54.4	−67.4	−402	
Con A	MeαMan	—	−22.2	−28.5	−21.0	−209	
	Manα1,3Man	—	−25.2	−31.0	−19.6	−460	
	Manα1,6Man	—	−22.2	−28.9	−22.5	−184	
	Manα1,6(Manα1,3)Man	—	−31.0	−42.7	−39.3	−389	
SBA[d]	MeβGal	0.5	—	−44.4	−96.2	−393	
	GalNAc	9	—	−39.7	−57.2	−418	
	MeβGalNAc	22	—	−58.1	−111.6	—	
	LacNAc	0.7	—	−34.3	−60.0	—	
ECorL[e]	MeβGal	0.4	—	−18.4	−11.2	—	86
	GalNAc	1.2	—	−29.7	−40.4	—	
	MeβGalNAc	1.3	—	−28.5	−34.9	—	
	LacNAc	4.2	—	−45.6	−82.3	+393	
Gal-1[f]	LacNAc	6.2	—	−27.6	−19.5	−377	
C2S-Gal-1[g]	LacNAc	2.9	—	−11.7	+27.1	—	
N-Gal-1[h]	LacNAc	8.7	—	−2.5	+66.5	—	
	Gal	1.6	−18.2	−13.7	+15.3	—	
	MeαGal	1.4	−18.1	−21.6	−12.0	—	
ECorL	MeβGal	0.7	−16.3	−18.2	−6.6	—	80
	GalNAc	1.3	−17.9	−23.0	−17.1	—	
	Lactose	1.9	−18.8	−41.2	−75.4	—	
	LacNAc	9.7	−22.7	−47.1	−83.2	—	
	2′-FL[i]	3.7	−20.3	−18.0	+7.7	—	
	MeαDNSGalN[j]	351.5	−31.7	−23.1	+30.0	—	
	Fucose	0.5	−15.2	−4.7	+35.2	—	
	Galβ1,4Glc	6.4	−21.7	−20.9	0	—	
	Galβ1,4Fruc	8.2	−22.4	−34.4	−40	—	
Gal-1[k]	Galβ1,4Man	11.3	−23.0	−35.2	−45	—	88
	Galβ1,4Ara	5.4	−21.3	−36.8	−52	—	
	Galβ1,4GlcNAc	22.2	−24.7	−35.9	−38	—	
	Galβ1S1βGal	11.6	−23.1	−46.4	−78	—	
PNA[l]	Galβ1,3GalNAc	20.6	−24.7	−59.0	−115.5	—	45

[a] Basic lectin from winged bean. [b] 3,6-Di-*O*-(*a*-D-mannopyranosyl)-*a*-D-mannopyranoside. [c] Lectin from *Dioclea grandiflora*. [d] Soybean agglutinin. [e] Lectin from *Erythrina corallodendron*. [f] From Chinese hamster ovary cells. [g] Cys to Ser mutant of Gal-1. [h] Monomeric mutant of Gal-1. [i] 2′-Fucosyllactose. [j] Methyl-*a*-*N*-dansylgalactosaminide. [k] From bovine spleen. [l] Peanut agglutinin.

Table 3 Entropic contributions to the binding of mannose oligosaccharides to Con A and DGL

System	$\Delta C_\mathrm{P}/\mathrm{J\,mol}^{-1}\,\mathrm{K}^{-1}$	$\Delta S/\mathrm{J\,mol}^{-1}\,\mathrm{K}^{-1}$	$\Delta S_\mathrm{solv}/\mathrm{J\,mol}^{-1}\,\mathrm{K}^{-1}$	$\Delta S_\mathrm{config}/\mathrm{J\,mol}^{-1}\,\mathrm{K}^{-1}$
Con A				
MeαMan	−209	−20.5	+53.6	−40.6
Manα1,6Man-OMe	−184	−3.8	+47.3	−17.6
Manα1,3Man	−460	−20.1	+118.0	−104.6
Manα1,6(Manα1,3)Man	−389	−51.9	+118.0	−136.4
DGL				
MeαMan	−234	−41.1	+60.2	−67.8
Manα1,6Man	−92	−50.3	+23.4	−40.2
Manα1,3Man	−167	−72.0	+42.7	−81.2
Manα1,6(Manα1,3)Man	−402	−110.5	+102.9	−179.9

reported in Table 3, both solvation and configurational entropies roughly scale with the size of the ligand, as a result of both the number of water molecules released and the number of degrees of freedom restricted during binding. Moreover, unfavourable configurational entropies are responsible for the observed enthalpy–entropy compensation (shown in Table 2). The more favourable entropy of association of the 1→3 dimannoside to Con A than to DGL has been attributed to a greatly reduced solvation contribution in the latter case.

It is now clear that the heat-capacity change contains important information about the mechanism of protein–carbohydrate complexation. The factors affecting ΔC_P have been discussed extensively and, at present, the term is usually thought to reflect solvent reorganisation effects and, to a lesser extent, changes in the protein vibrational modes.[79] For lectin–sugar interactions, ΔC_P values are usually small ($\leq$400 J mol^{-1} K^{-1}) and negative. The significantly larger heat-capacity changes accompanying antibody–saccharide associations (compared to lectin–sugar complexations) can be interpreted in terms of structural differences between the binding sites of the two classes of biological polymers.[80] The antibody combining site contains several aromatic amino acids and the burial of these hydrophobic residues represents the driving force for protein–antigen binding.[81] However, the entropy of binding is still negative, offsetting the enthalpic gain. Mutagenesis studies have revealed that part of the entropic loss may be associated with the reduced mobility of amino acid side chains in the antibody combining site.[82] In this case, therefore, the enthalpy–entropy compensation is also a function of protein structure, albeit with a different underlying molecular mechanism. Considering again the data reported in Table 1 and 2,[76] considerably different mechanisms of association can be observed for the binding of the same oligosaccharide to two different lectins. Small but significant differences in the primary sequences of two proteins may affect the flexibility of the loops forming binding sites. Although lectin binding sites are greatly preorganised, several amino acid side chains move considerably to achieve the optimal hydrogen bonding and van der Waals contacts with the ligand. Thus, an enhanced flexibility would favour hydrogen bonding, but, at the same time, a more severe entropic loss would be observed. However, since the values of ΔC_P are typically small, little change in overall protein conformation is usually involved, as confirmed by X-ray crystallographic determinations.

The larger $-\Delta H$ values obtained for the binding of 3,6-di-O-(a-D-mannopyranosyl)-a-D-mannopyranoside to Con A and DGL compared to methyl-a-D-mannopyranoside indicate the presence of extended binding sites. The trimannoside is, in fact, the minimum carbohydrate epitope that completely fills the combining pocket, giving rise to the maximum number of hydrogen bonds and van der Waals interactions.[83] This oligosaccharide represents a core branching structure in all N-linked glycoproteins.[76]

Despite the general feature of being enthalpically driven with an unfavourable entropic contribution, a few lectin–carbohydrate systems show positive entropy changes. As reported in Table 2, most of the carbohydrates that bind to ECorL show negative ΔS, but positive entropic changes are observed for galactose, fucose, methyl-a-N-dansylgalactosaminide and 2′-fucosyllactose.[80] Favourable entropies are associated with non-polar contacts, for example between the aromatic dansyl moiety at C-2 of MeaDNSGalN and Trp 135 in the binding site of the lectin. From X-ray crystallographic studies,[49] the increase of entropy associated with the binding of galactose is explained by the release of tightly bound water molecules, not still fully compensated by opposing entropically disadvantaged factors.

Differences in the thermodynamics of binding for homologous lectins, for which most of the amino acid residues in the binding sites are conserved, might indicate the indirect but critical role of non-conserved residues away from the combining pocket.[84] Effects of single-site mutations on the conformation of

lectins have been investigated by Siebert and co-workers.[85] The mutation of sites far removed from the carbohydrate binding site in galectin-1 (Gal-1) from CHO cells strongly affects its thermodynamics of complexation with LacNAc.[86] As shown in Table 2, the two Gal-1 mutants C2S-Gal-1 and monomeric N-Gal-1 exhibit significantly lower $-\Delta H$ values than Gal-1, whereas the entropic contributions become favourable.

Importantly, studies carried out with monodeoxy analogues of 3,6-di-O-(a-D-mannopyranosyl)-a-D-mannopyranoside have shown that K and $-\Delta H$ for the analogues are lower than for Man$a$1,6(Man$a$1,3)Man.[87] The sum of the $\Delta\Delta H$ and $\Delta\Delta G$ values for the hydroxyl groups of the trimannoside, obtained from the monodeoxy analogues, do not correspond to the measured ΔH and ΔG for this ligand. This means that the magnitude of $\Delta\Delta H$ and $\Delta\Delta G$ represents not only the loss of the hydrogen bonds involved, but also differences in the solvent and protein contributions to the binding of Man$a$1,6(Man$a$1,3)Man and of the deoxy analogues.

No correlation was found between the calorimetrically derived thermodynamic parameters of a series of ligands for the bovine spleen galectin-1 and the number of protein–sugar close contacts.[88] Interestingly, binding enthalpies calculated from changes in the solvent-accessible areas of the galectin-1 binding site upon complexation showed poor agreement with the calorimetric values, again reflecting the importance of factors other than the burial of hydrophobic surface area.

Everything mentioned above shows the complex balance of forces before and after the binding event. It is therefore clear why both detailed structural and energetic information is essential in order to design ligand mimics. Attempts to overcome certain thermodynamic barriers through the planning of the glycoconjugate structure may not always result in the expected high-affinity association, due to the offset of different contributions.[89]

2.2.2 Multivalent ligands. The "cluster glycoside effect", defined as "the enhancement in the activity of a multivalent ligand beyond what would be expected due to the increase in sugar local concentration (statistical effect) alone"[90] is nowadays generally accepted. Despite the numerous observations of the phenomenon, a molecular interpretation of the effect is difficult and, at present, its physical origin is still not well understood. Efforts to provide a molecular basis for the cluster glycoside effect have been complicated by two issues: 1) polyvalent ligands are often polydisperse and structurally ill-defined and 2) methodologies used to evaluate protein–carbohydrate binding measure several phenomena, including protein–carbohydrate association. Moreover, there may exist a relationship between the measured magnitude of the cluster glycoside effect and the assay utilised for the determination.[91] In a first approximation, a trend of increased enhancement with increasing valency can be drawn.

In principle, at least two models of association can be described: an intramolecular, or chelate, binding and an intermolecular aggregative process. In the former a multivalent ligand spans a number of binding sites on a single protein, while in the latter the spanned binding sites belong to different receptor molecules, resulting in aggregates that may or may not precipitate.

Considering, for simplicity, a bivalent ligand, the thermodynamic parameters describing a chelate complexation can be expressed as follows:[92]

$$\Delta J_{bi} = 2\Delta J_{mono} + \Delta J_i \qquad (5)$$

where: ΔJ_{bi} = change of any thermodynamic parameter for bivalent complexation, ΔJ_{mono} = corresponding term for monovalent association, ΔJ_i = interaction term. Interaction energies have been traditionally considered in entropic terms, where translational and rotational savings and conformational penalties have to be taken into account. As the translational and rotational

entropies of a molecule are, respectively, directly and indirectly proportional to the logarithm of the molecular weight, the binding of two (or more) ligands produces a multivalent ligand with translational and rotational entropy roughly equivalent to that of the monovalent ligand. Therefore, the binding of a bivalent ligand is characterised by a favourable contribution to the entropic term equal to the translational and rotational entropy of the corresponding monovalent ligand. This term presumably accounts for a large part of the interaction free energy. On the contrary, the loss of conformational degrees of freedom of the ligand upon binding results in an unfavourable contribution to the overall entropy. Enthalpy also plays an important role in determining the overall affinity of a multivalent ligand. Alteration of the ligand position within the binding site would result in an unfavourable contribution, while the contribution of the linker may be favourable due to its favourable interactions with the protein surface (at the periphery of the binding site or over the space separating the spanned binding sites) or unfavourable as a result of disadvantaged steric inter-action. This latter aspect can result in a significant unfavourable contribution, due to the lock of linker dihedral angles into gauche or eclipsed orientations, which can, in turn, preclude the complexation by an intramolecular mode. Considering what has just been said and combining it with the fact that distances between binding sites on the receptor molecule are generally of the order of tens of Angstroms (65 Å in the canonical dimer of concanavalin A), it appears clear that the achievement of chelate complexation is challenging.

Most of the multivalent ligands reported so far reach the observed affinity enhancement by means of an aggregative pro-cess, where the formed cross-linked complexes may be stabilised by a range of forces, including protein–protein interactions. Alternatively, a diminished solubility of the aggregate and its precipitation would lead to an apparently enhanced affinity through a coupled equilibrium.[93] In addition, surface plasmon resonance (SPR) studies have indicated that certain multivalent interactions have higher affinity constants due to reduced dissociation (k_{diss}) rather than increased association (k_{ass}) rate constants.[94] A redox-switchable aggregation in a snowdrop lectin mutant has recently allowed the separation of direct binding from aggregation in an effort to understand further these two processes.[95]

Few calorimetric studies of multivalent carbohydrate–protein interactions (indeed few for any multivalent ligand) have been reported so far.[93,96–101] Moreover, ITC has been limited to small or dendritic glycoconjugates possessing a maximum number of carbohydrate residues equal to six. Only very recently ITC has been used to determine the enthalpy of coaggregation between two oral bacterial pairs.[102] After each injection, the number of free and bound streptococci was evaluated microscopically by means of a Petroff–Hausser chamber. Experiments were performed both with a coaggregating and a non-coaggregating species, the latter used as a control. Comparison of the heat released upon coaggregation with the enthalpy of binding cor-responding to lectin–carbohydrate interactions indicates that the number of binding sites involved in the formation of a bacterial coaggregate is relatively huge. It is worthwhile to note that no model was assumed for the determination of the coaggregation constant. Simply, the heat flow measured directly was divided by the number of bound streptococci determined microscopically assuming that all streptococci injected reached the ampoule. Brewer et al.[98] have recently reported the determination of the thermodynamics of binding of small multivalent ligands (with structural valency up to 4) to Con A and DGL. The structures of those that showed an affinity enhancement relative to the corresponding monovalent ligands are reported in Fig. 3. The thermodynamic parameters determined using Con A as receptor protein are reported in Table 4 (italic), together with those reported by Toone et al.[97] (regular) for the interaction of the two series of ligands (dendritic compounds containing
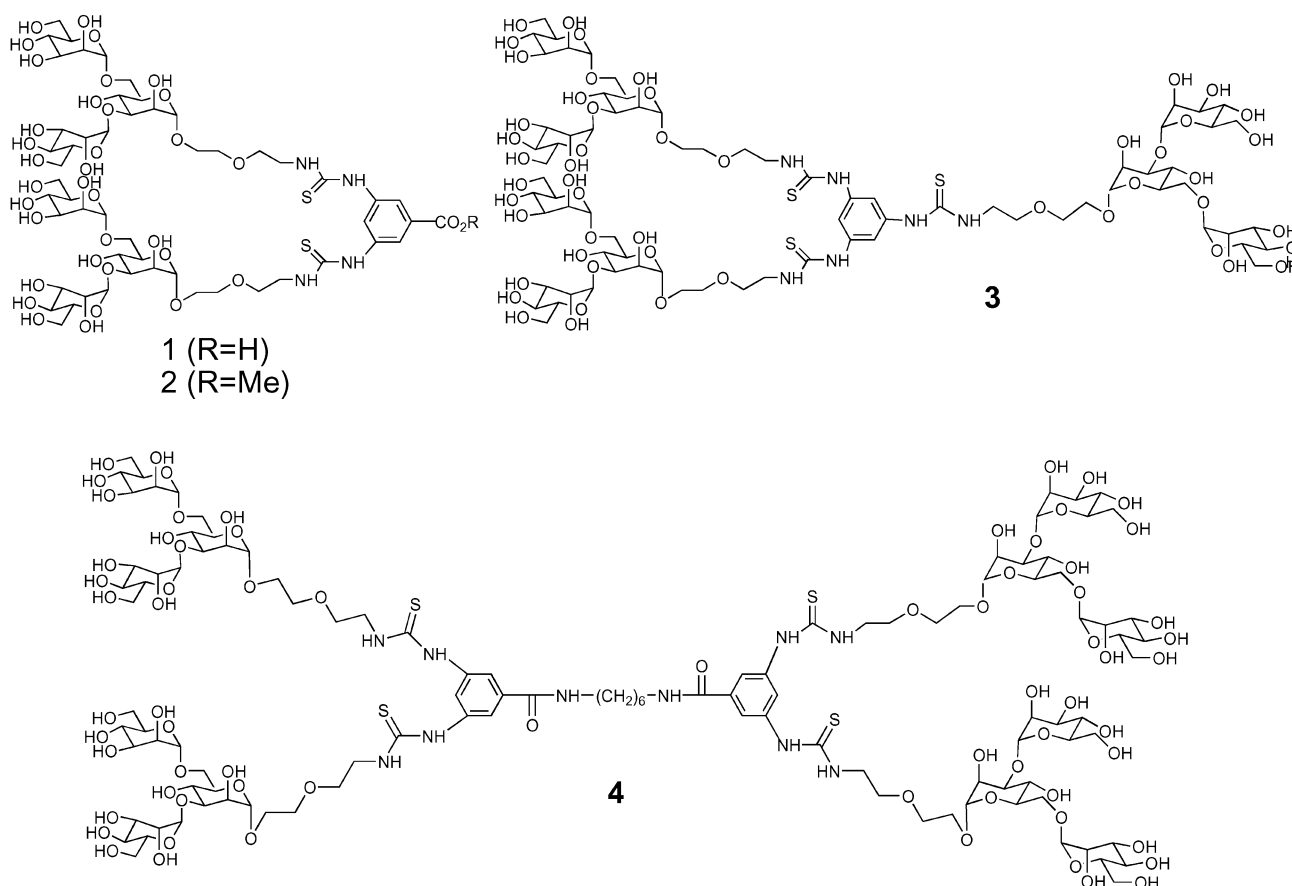

Fig. 3 Multivalent ligands used by Brewer et al.[101] The corresponding thermodynamic binding parameters are reported in italics in Table 4.

Table 4 Thermodynamic binding parameters of Con A with multivalent ligands

Compound[a]	$10^{-4} \times K/M^{-1}$	ΔG/kJ mol^{-1}	ΔH/kJ mol^{-1}	$T\Delta S$/kJ mol^{-1}	n^b
Manα1,6(Manα1,3)Man	*39*	*−31.8*	*−61.5*	*−29.7*	*1.0*
1	*286*	*−36.8*	*−96.6*	*−59.8*	*0.53*
2	*250*	*−36.4*	*−109.6*	*−73.2*	*0.53*
3	*420*	*−37.6*	*−121.3*	*−83.7*	*0.51*
4	*1350*	*−40.6*	*−221.8*	*−181.2*	*0.26*
MeαMan	0.76	−22.2	−27.6	−5.4	1
5	0.92	−22.6	−31.0	−8.4	1
6	0.80	−22.2	−31.4	−9.2	1
7	0.79	−22.1	−32.6	−10.5	1
8	0.75	−22.2	−17.6	+4.6	1
9	ND	—	—	—	—
10	0.80	−22.2	−26.8	−4.6	1
11	0.86	−22.6	−32.2	−9.6	1
12	0.47	−20.9	−29.7	−8.8	1
13	6.2	−27.6	−9.6	+18.0	1
14	150.0	−35.5	−5.4	+30.1	1

[a] Italics: [Ligand] = molar concentration of the overall ligand molecule; regular: [Ligand] = valency-corrected concentration. [b] n = binding stoichiometry, carbohydrate : protein.

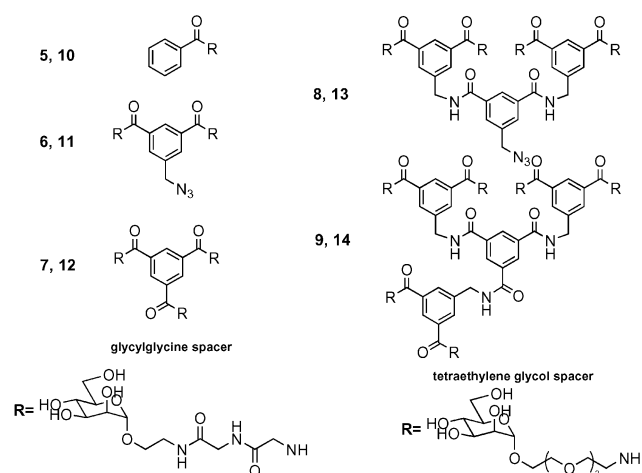up to six binding epitopes) shown in Fig. 4 with the same lectin.

Fig. 4 Dendritic ligands used by Toone *et al.*[97] The corresponding thermodynamic binding parameters are reported in regular type in Table 4.

As previously discussed, the interaction of Con A with monovalent carbohydrates was found to be enthalpically driven, with an unfavourable entropic contribution.[68] Despite the different behaviour of the thermodynamic parameters going from the mono- to the multivalent ligands in each series, the affinity enhancement observed for ligands **1**–**4** and **13**–**14** was, in both cases, explained by means of an intermolecular aggregative mechanism. Assuming non-cooperativity of the different epitopes in the multivalent carbohydrates, Toone *et al.* explained the thermodynamics of lectin–ligand association (ligands **13**–**14**), hypothesising an initial exothermic protein–carbohydrate interaction, which proceeds with thermodynamic parameters equivalent to those of the corresponding monovalent saccharide, coupled with or followed by an endothermic, entropically driven, aggregation process. The diminution of $-\Delta H$ was therefore considered the thermodynamic signature for the intermolecular aggregation process.[91,97]

Studies previously carried out by Toone and co-workers employing another series of dendritic multivalent ligands,[96] where crystals of both bi- and trivalent ligands were grown in the presence of succinylated Con A, showed that the carbohydrate orientation within the binding site was identical to that of the corresponding monovalent ligand. Interestingly, ligands with valencies greater than three showed affinity enhancements in agglutination assay but not by ITC. On the other hand, tetra- and hexavalent carbohydrates showed substantially diminished enthalpy (in absolute values) and visible cloudiness. Therefore, the IC$_{50}$ values were found to correlate with the calorimetrically derived entropies, but not with the free energies of ligand binding. The authors deduced that the agglutination assay did not evaluate the strength of protein–carbohydrate interaction, but rather the ability of multivalent ligands to drive aggregation processes. Based on these results, the overall thermodynamic parameters could be expressed as the sum of two contributions, the actual protein–sugar interaction-characterised by thermodynamic values equal to those determined for the corresponding monovalent ligand—and the aggregation process. Enthalpies of aggregation were thus determined.

Despite the common explanation for the observed enhancement of the binding affinity through an intermolecular aggregative mechanism, the results reported by Brewer *et al.* are rather different from those of Toone and co-workers. In the former case, an increase of $-\Delta H$ was observed going from the monovalent to the multivalent ligand (**1**–**4**). With the exception of compound **3** (which was found to behave as a divalent ligand despite its structural valency being equal to three[98,100]), ΔH scaled proportionally to the number of epitopes in the multivalent ligand, whereas ΔS did not. The entropy variation was instead more negative than the sum of ΔS of individual epitopes. According to the authors, this behaviour of ΔS indicated the binding of the multivalent ligands to separate receptor molecules (aggregation). Moreover, the non linearity of Scatchard and Hill plots was explained by a negative cooperativity for the binding of multivalent ligands,[99] which was experimentally demonstrated by measuring the microscopic thermodynamic parameters of each carbohydrate residue in the ligand by means of reverse isothermal titration microcalorimetry.[100] In this case, solutions of the mono-, bi- and trivalent ligands were placed into the cell and Con A used as the titrant. For instance, the binding constant, K_1, of the first epitope of the bivalent ligand was found to be one order of magnitude greater than those of the second epitope and of the corresponding monovalent carbohydrate. The average of the two microscopic binding constants determined agreed with the macroscopic binding constant determined using the sugars as injectants. The enthalpic contribution of the two epitopes was essentially the same and very close to that for the monovalent ligand. Therefore, the enhancement in affinity observed was associated with a favourable entropy of binding of the first epitope relative to the second one. However, while the sum of the two enthalpic contributions was very close to the enthalpy determined by "normal" microcalorimetry, the sum of the two microscopic entropies was much lower, in absolute value,

than the one reported when the ligand was injected into the lectin solution. Hemagglutination inhibition measurements correlated perfectly with the results obtained by ITC. The authors provided some hypotheses to explain the differences between their results and those reported by Toone and co-workers. For instance, they underlined the difference in the method adopted to express the concentration of the ligands. While Toone *et al.* considered the number of epitope equivalents, Brewer *et al.* referred to the molar concentration of the ligand molecule. Their arguments against the use of valency-corrected concentrations were the observed negative cooperativity of binding and the determined difference between the structural and functional valency.[99]

Most examples of intramolecular/chelate binding involve either bacterial toxins or polymeric ligands. Studies have shown that the enhancement in binding affinity is much higher when different carbohydrate residues of the ligand bind to sites on the same receptor molecule. In 1983, Lee *et al.*[103] reported that a synthetic tetraantennary undecasaccharide showed an inhibition constant for the hepatic Gal/GalNAc receptor 10^6 times greater than an equivalent monoantennary trisaccharide, despite the only 4-fold statistical increase in absolute galactose concentration. It was therefore pointed out that the number of Gal residues per cluster and their branching mode were major determinants of the binding affinity of a ligand to the mammalian hepatic lectin. A striking example of the affinity enhancement achievable through multivalency (10 orders of magnitude compared to monovalent ligand), for a non-carbohydrate system, was given by Rao *et al.*[104]

Most of the currently synthesised polymeric and dendritic multivalent ligands show a "random" multivalency, which rarely allows an affinity enhancement higher than 1,000-fold. Moreover, as already pointed out, the reasons underlying this activity gain are poorly understood.[63] In a different approach, structural information about the spatial arrangement of the target binding sites are taken into account for the design of ligands ideal for maximising the interaction with the receptor.

The modular synthesis of multivalent ligands of the heat-labile enterotoxin (LT) from *E. coli* was recently reported by Fan and co-workers.[105] The five B subunits of the toxin present a 5-fold symmetric configuration. In this study, pentavalent ligands constituted by a semirigid "core" which can adopt a conformation close to 5-fold symmetry, flexible "linkers" that project toward the receptor binding sites and "fingers" represented by 1-β-amidated D-galactose, were synthesised. Different lengths of the linker were tested and the receptor–ligand interaction analysed by an enzyme-linked immunosorbent assay, ELISA. All pentavalent ligands led to significant affinity gains compared to the corresponding monovalent ones. The best one showed an IC_{50} 10^5-fold lower than galactose, approaching the affinity of the oligosaccharide portion of the natural ganglioside GM1. The pentavalent ligand also showed a 10^4-fold affinity gain compared to the corresponding monovalent ligand, or 2000-fold on a valency-corrected basis. Dynamic light scattering ruled out the possibility of an aggregative process, supporting the expected formation of a 1 : 1 toxin–ligand complex. The analysis of the effect of linker-length on the affinity showed that the greatest enhancement was detected for the ligand whose linker effective length[106] best matched the distance between nonadjacent binding sites.

One of the most striking examples of affinity enhancement was recently reported by Kitov *et al.*[107] Shiga-like toxins SLT-I and SLT-II were inhibited by a decavalent ligand, designated STARFISH, whose structure was complementary to that of the receptor. The SLTs are AB toxins constituted by an enzymatic (A) component and a cell-binding (B) part. The A-subunit is situated on one face of the B-component, which is a pentamer of identical subunits. In the absence of the A-subunits, the B-subunits still form pentamers that are functionally equivalent to the overall toxin in their attachment to the host cell. The *in vivo* cytotoxicity of SLTs has been correlated with their

binding affinity toward the glycolipid globotriaosylceramide (Gb_3). The crystal structure of SLT-I B-pentamer complexed with a Gb_3 analogue revealed three Gb_3-binding sites per B-subunit[108] (Fig. 5).
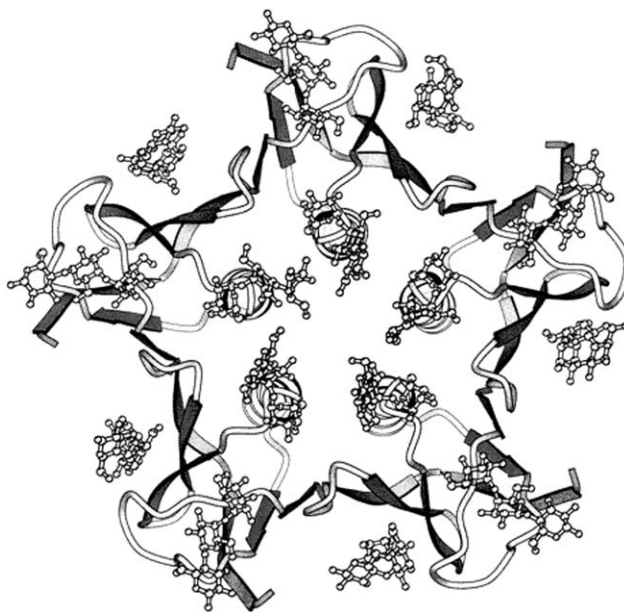

Fig. 5 View along the 5-fold axis of the SLT-I B-pentamer bound to the Pk-MCO trisaccharides. The surface toward the viewer is the sugar binding surface (reprinted with permission from ref. 108, copyright (1998) American Chemical Society).

As determined by means of an ELISA protocol, STARFISH exhibited more than 10^6-fold increase in inhibition over the Pk trisaccharide. STARFISH was designed to achieve high affinity through the simultaneous binding of all five peripheral bridged Pk dimers to the ten sites 1 and 2 of the B-pentamer (with each trisaccharide dimer bridging sites 1 and 2 of the monomer). Crystallographic studies of the formed complex revealed a different mode of binding (Fig. 6). One STARFISH molecule, in fact, bound two B-subunit monomers from separate toxin molecules, with the saccharide residues occupying only site 2. Each trisaccharide interaction was identical to that seen for the corresponding univalent ligand. A higher affinity of binding of site 2 compared to 1 had been already highlighted by previous studies carried out on toxin mutants. Therefore, the observed mixture of aggregative and chelate mechanisms through which STARFISH exerted its activity could be explained by the stronger interactions of the carbohydrate within site 2, possibly coupled with unfavourable entropic and enthalpic contributions of the bridging.

Recent mutation studies have shown that site 2 of the SLT 1B subunit is the most important site for binding of free Pk trisaccharides, however the STARFISH ligand also requires an intact site 3. This suggests that site 3 is specifically involved in binding of the pentavalent ligand, the individual trisaccharides of which bind at sites 1 and 2.[109]

A calorimetric study of the interaction of bivalent glycopeptides with SLT-I was carried out by Toone and co-workers.[93] The two ligands used differed in the nature of the peptidic linker. Both ligands showed affinity gain, but the mechanism through which this was achieved depended on the nature of the linker domain: while the compound characterised by a hydrophobic linker seemed to work through a chelate mechanism, the hydrophilic one bound by means of an intermolecular aggregative process. Again the binding enthalpy of the latter appeared to be significantly lower (in absolute value) than that of the monovalent compound, a trend that this group already recognised as the signature for an intermolecular binding process. The affinity enhancement reached through the intramolecular mechanism
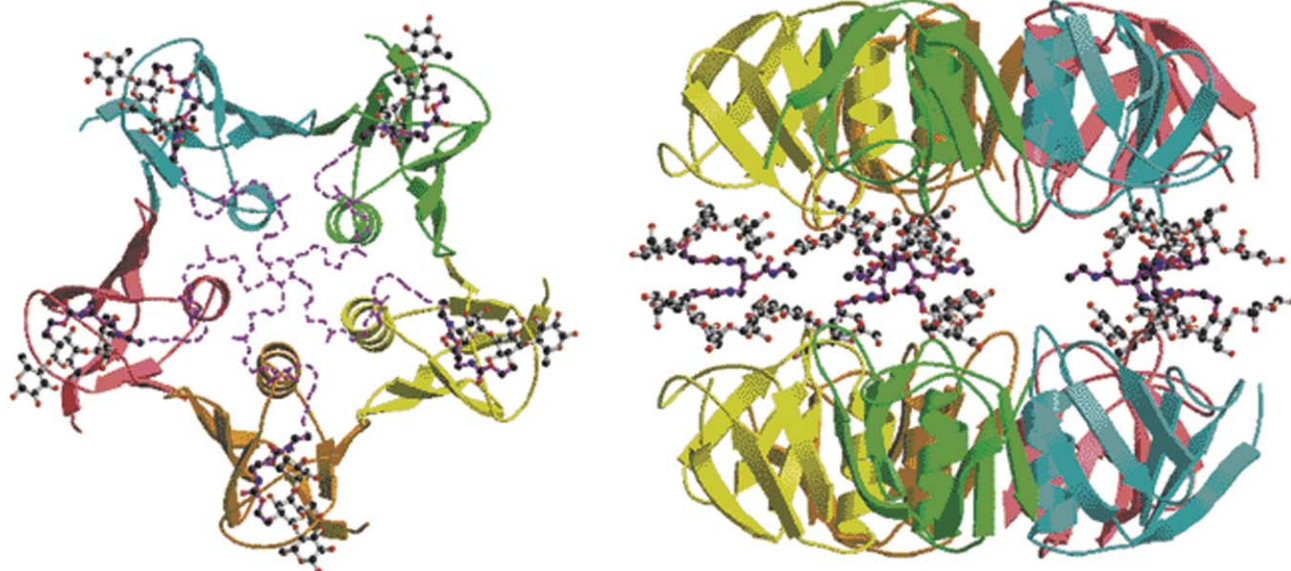
Fig. 6 Diagrams of STARFISH ligand bound to the SLT. Reprinted with permission from Nature Publishing Group.[107]

was slightly higher than that of the hydrophilic compound. Peptidic spacers were chosen in the search for possible favourable additional interactions of the ligand with the protein surface. These interactions were given as an explanation for the higher affinity of the ligand with the hydrophobic linker, and for its intramolecular mechanism. For the hydrophilic linker, either there were no favourable contacts or there were in fact repulsive interactions of the linker with hydrophobic regions on the protein surface. Therefore, the affinity was lower and the binding proceeded *via* an intermolecular mechanism.

The greatest increases in activity on a valency-corrected basis are reported with polymeric ligands.[91] Nevertheless, the only polymeric ligands synthesised so far which seem to bind through an intramolecular mechanism are those prepared by Kanai *et al.*[110] (Fig. 7).
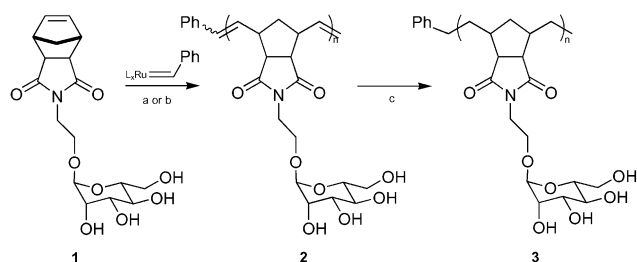

Fig. 7 Scheme for the synthesis of neoglycopolymers by ROMP: (a) MeOH–H₂O–DCM (6 : 1 : 5) then add H₂O, room temperature; (b) H₂O–DCM (2 : 1), DTAB, room temperature; (c) TsNHNH₂, H₂O, 100 °C, 56% (*n* = 10), 91% (*n* = 25), 58% (*n* = 52), 100% (*n* = 143).[110]

Several ligands with different backbone lengths were synthesised and tested as inhibitors for Con A–carbohydrate interaction. The maximum inhibitory activity seemed to coincide with an average length sufficient to span two lectin binding sites belonging to the same protein molecule, supporting the hypothesis of a chelate mechanism of binding. The enhancement in affinity determined for polymers too short to bridge two binding sites was attributed to a high local concentration of sugar moieties (statistical effect), which would perturb the rate of dissociation of the formed complex. The authors concluded that the observed dependence of the inhibition activity on the polymer length was largely due to a combination of statistical and chelation effects. Moreover, despite the differences in backbone flexibility, the most potent ligands of each series, **2** and **3**, had approximately the same efficacy. This latter observation points out an interesting aspect of multivalent ligand–protein binding. It is believed that rigid linkers would favour the interaction, due to a reduction of the entropic penalties accompanying the binding.[105,107,111] However, the great enhancements achieved with flexible linkers and the negligible gain in activity observed with dendritic ligands characterised by rigid arms, have recently led Toone and co-workers to suggest that entropic concerns might be less severe than was previously thought and probably less severe than those related to unfavourable enthalpic contributions.[91]

2.3 Diagnostic and therapeutic applications

There is considerable evidence that lectins are involved in many physiological events (several examples are given in Table 5). The enormous developments that the possibility to inhibit, activate or exploit protein–carbohydrate interactions could

Table 5 Examples of the functions of animal lectins[a]

Function	Examples
Intracellular routing of glycoconjugates	P-type lectins, ERGIC-53, VIP-36
Molecular chaperones during glycoprotein synthesis	Calnexin, calreticulin
Mediation of endocytosis	Asialoglycoprotein receptors, macrophage mannose receptor
Cellular growth regulation	Galectins, sarcolectin, cytokines
Extracellular molecular bridging	*Geodia cydonium* galectin, other galectins, interleukin-2
Cell–cell interactions for homing and trafficking	Selectins, CD22, CD31, CD44
Cell–matrix interactions	Galectins, heparin- and hyaluronic acid binding lectins
Scavenging of cellular debris; anti-inflammatory action	Galectin-9

[a] Selected from ref. 10 and ref. 25.

bring, especially in medicinal chemistry, push the study of lectins and carbohydrates to the forefront of research. For example, the cellular protein glycosylation pattern is influenced by several physiological changes, such as the occurrence of disease. Thus, the altered glycoform population of a given glycoprotein may be diagnostic of the disease responsible for the alteration itself. Abnormal glycosylation has been detected in cancer development.[112] Both quantitative and qualitative lectin-binding differences were observed for cytosolic glycoproteins in benign and malignant thyroid neoplasms: in the majority of carcinomas lectin-binding was weaker in comparison with adenomas and non-neoplastic specimens.[113] Changes in prostate-specific antigen (PSA) glycosylation during malignant transformation may be used for the diagnosis of prostate cancer at an early stage. The quantitative precipitation method of Con A–carbohydrate interaction[114] was used by Basu and co-workers[115] for differentiation between prostate cancer (PC) and benign prostatic hyperplasia (BPH). The carbohydrate content in the precipitate after binding of Con A with serum PSA of PC was significantly lower than that of BPH. It was concluded that a serum value $<3.0\,\mu g\,ml^{-1}$ of the carbohydrate content of Con A–PSA precipitate indicates strong suspicion for prostate cancer, reducing the rate of unnecessary biopsies in men with total PSA value between 4.0 and $10.0\,\mu g\,ml^{-1}$. More recently, wheat germ agglutinin (WGA) has been shown to induce rapid apoptosis in malignant cells *via* a novel mechanism.[116] Studies indicated that GlcNAc-containing glycoconjugates were also involved in WGA-mediated cell death, and that the mechanism probably involved the mitochondrial pathway.

Such valuable diagnostic and potential therapeutic applications highlight the need to gain further insights into lectin-specific ligand interactions and into the physiological processes that these associations initiate and/or regulate. Novel lectins are therefore continuously isolated, such as, in 2003, the *N*-acetylglucosamine-specific lectin from fresh sclerotia of the edible mushroom *Pleurotus tuberregium*,[117] and the two mannan- and L-fucose-binding lectins from the green alga *Enteromorpha prolifera*.[118] At the same time, a deeper understanding of the functioning of known systems is achieved. Loris and co-workers[119] have recently reported the crystal structure of the *Pseudomonas aeruginosa* fucose-specific lectin (LecB) in its metal-bound and metal-free state as well as complexed with fucose, mannose and fructopyranose. The higher protein affinity for fucose was ascribed to favourable interactions involving C6 and O1 of the saccharide, absent for the other ligands, due to the partial shielding of a carboxylate group and one or two hydroxyl groups. A recent crystallographic study upon the lung surfactant protein D (SP-D), a protein capable of interacting directly with carbohydrate residues on pulmonary pathogens and allergens, of stimulating immune cells and of intervening directly in the lungs' immune response, has revealed that a previously unreported calcium ion is essential for biological activity, ensuring the right structural requirements for the established effector mechanism.[120] The lectin from *Vatairea macrocarpa* seeds (VML) was used to induce neutrophil migration in rats in order to explore the pathways by which lectins cause an inflammatory response.[121] The results obtained suggested that the lectin-induced effects do not involve the usual mediators, such as PAF mediator, rather it acts through an indirect mechanism.

One of the most well-known physiological processes initiated by protein–carbohydrate interactions is the clearance of old erythrocytes from the blood.[5] With age the red blood cells become progressively desialylated. When the density of galactose moieties so exposed on the surface is high enough, the cell is removed from the circulation and destroyed through the recognition of the galactosyl residues by hepatocyte receptors (so-called asialoglycoprotein receptors). Clues to the role of galactose in this process were first obtained in 1968 from experiments with rabbits that showed that desialylated ceruloplasmin

(which also has terminal galactose residues) was cleared rapidly from the serum by the liver.[122] The carbohydrate recognition domain (CRD) of the mammalian asialoglycoprotein receptor of hepatocytes is the prototype for the C-lectins,[5,123] a large family of Ca^{2+}-dependent animal proteins. Members of this group are important components of the rapidly-responding innate immune system, including the mannose-binding proteins, MBPs. The role of MBPs in host defence against viral pathogens has been demonstrated by several studies.[124] Like all other collectins—a group of soluble C-type lectins—MBPs have an N-terminal collagenous domain and a globular C-terminal portion, containing the CRD.[125] Thus, while the C-terminal domain mediates the binding to various microorganisms, the N-terminal one interacts with the cell, triggering the carbohydrate-dependent complement activation, which then results either in the killing of the pathogen[126] or in the deposition of complement components on the surface of the microorganism (opsonisation) which is then cleared by phagocytic cells like macrophages and neutrophils. MBPs have been shown to prevent human immunodeficiency virus (HIV) infection of H9 cells, probably by binding to oligomannose units of the envelope gp120 of the virus.[127] A deficiency in MBPs has been claimed to account for a significant number of immunodeficiency cases in children, where the immune system is still not fully capable of mounting an efficient adaptive response.[128] Serum MBP levels are largely affected by polymorphisms of the MBP gene. The presence of the minority allele can result in either a significant reduction of the serum protein concentration or in its almost complete absence. These conditions have been suggested to be associated with recurrent infections.[129] Recent studies have also shown that individuals with the minority alleles of the MBP gene are at risk from autoimmune diseases; serum MBP may also bind to apoptotic cells and so contribute to the clearance of potential autoantigens.[130]

Siglecs (sialic acid-binding immunoglobulin-like lectins), an important subset of the Ig superfamily, act as endogenous receptors of sialic acid residues. Recent advances in the understanding of the function of known siglecs and the identification of new members of the family have contributed significantly to the elucidation of their physiological roles. Thus, a participation in inflammation processes, indicated by the expression of the lectin on inflammatory macrophages found in rheumatoid arthritis, was suggested for sialoadhesin (Siglec-1).[131] CD22 (Siglec-2), a well-characterised B cell inhibitory receptor, has been found to modulate leukocyte activation; in particular, it has been shown that lectin *trans*-interactions with *a*-2,6-linked sialic acids in *N*-glycans can dampen B cell activation.[132] All new family members recently identified are highly related to CD33 (Siglec-3), a marker of myeloid cells. Their expansion in the innate immune system suggests an important role of these proteins in host defence. Similar to the role proposed for CD22, CD33-related siglecs may regulate the autoreactivity of myeloid cells.[132]

Selectins, a subfamily of C-type lectins so called because of their ability to mediate selective contacts between cells, play a key role in the recruitment of leukocytes to inflammatory sites.[10] Selectins represent the first group of mammalian lectins shown to mediate cell–cell contact, in particular they provide rapid and reversible cell adhesion under hydrodynamic flow.[133] The dynamics of the selectin–sialyl Lewis X (sLeX) interaction have been probed by single molecule dynamic force spectroscopy.[134] Unbinding was determined to involve at least two activation barriers; that which determines the response of the complex under high shear is determined by Ca^{2+}-mediated binding of the fucose residue of sLeX to the lectin. During the inflammatory response, damage in tissues surrounding a blood vessel causes the influx of signalling molecules. Selectins are thus rapidly expressed on the inner surface of blood vessels (E-selectins) and on platelets (P-selectins). Highly specific binding of these proteins to sugars on white blood cell surfaces causes the leukocytes to adhere to the vessel walls, which, at this stage, express

L-selectins. By interaction with L-selectins, leukocytes can now roll toward the site of damage, where they pass through to the surrounding tissue.[135] Whilst this is a well-controlled process in healthy individuals, in excess it is a cause of arthritis, asthma and myocardial infarction. Due to their ability to mediate cell–cell adhesion, selectins play a critical role in tumour metastasis.[136] For instance, it has been suggested that carcinoma cells can interact with platelets, leukocytes and endothelia through the selectins. Thus, the blood platelets may form complexes around the tumour cells and facilitate metastasis by masking them from the host defence.[137] Due to their implication in a large number of diseases, suitable selectin inhibitors represent a class of efficient therapeutic agents. P-selectin glycoprotein ligand (PSGL-1) has shown good results in phase II clinical trials to prevent ischemia reperfusion injury and it seems to be a promising candidate as a cancer therapeutic.[137]

Galectins are a growing family of β-galactoside-specific proteins of which 15 members have been identified in mammals.[138] Many tumours, such as colon, thyroid and breast carcinomas, express both galectin-1 and -3, and it has been suggested that galectin-3 is a metastasis marker.[139] Among the degenerative processes that they can mediate, galectins may favour metastasis by binding glycoconjugates of the extracellular matrix and subsequent intravasation of the tumour cell into the blood vessels. A modified citrus pectin (MCP), a complex highly branched polysaccharide rich in galactoside residues, has shown very encouraging results as an anti-cancer agent, significantly reducing lung colonisation of B16-F1 melanoma cells.[140] More recently, 3′-aromatic amide derivatives of 3′-amino-N-acetyllactosamine have been found to be potent inhibitors of galectin-3 (IC_{50} down to 4.4 µM).[141]

Conclusions

Multivalent protein–carbohydrate interactions regulate a myriad of vital cellular events. Lectin–carbohydrate associations are usually weak when compared to others in Nature, with K_D values of the order of mM for monosaccharides. However, when multivalent saccharides of the right type are clustered together with the right geometry, the interaction becomes strong and highly specific. This increase is more than what would be expected on the basis of the increased sugar local concentration. The reasons of this phenomenon, termed "cluster glycoside effect", have not been rigorously determined yet, but its implications are extraordinary.

It has been pointed out repeatedly that a thorough understanding of the lectin–carbohydrate interaction is essential in order for it to be exploited for biomedical applications.[61,63,142] A deep knowledge of both structural and energetic aspects of the process is thus necessary. Indeed, the failure of a number of carbohydrate-based drugs may be attributed to a poor understanding of their supposed mechanism of action rather than to any inherent deficiency of carbohydrate therapeutics.[61] When good understanding is achieved, the results are impressive.[105,107]

References

1 W. C. Boyd and E. Shapleigh, *Science*, 1954, **119**, 419.
2 S. H. Barondes, *Trends Biochem. Sci.*, 1988, **13**, 480.
3 N. Sharon and H. Lis, *Science*, 1989, **246**, 227.
4 G. Ashwell and A. G. Morell, *Adv. Enzymol.*, 1974, **41**, 99.
5 G. Ashwell and J. Harford, *Annu. Rev. Biochem.*, 1982, **51**, 531.
6 J. C. Paulson in *The Receptors*, ed. P. M. Conn, Academic Press, New York, 1985, p. 131; D. C. Wiley and J. J. Skehel, *Annu. Rev. Biochem.*, 1987, **56**, 365; N. Sharon, *FEBS Lett.*, 1987, **217**, 145.
7 J. Melrose, N. Tsurushita, G. Liu and E. L. Berg, *J. Immunol.*, 1998, **161**, 2457; N. Watanabe, H. Kawashima, Y.-F. Li and M. Miyasaka, *J. Biochem.*, 1999, **125**, 826.
8 H. J. Gabius, *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 1267; M. Monsigny, A. C. Roche, C. Kieda, P. Midoux and A. Obrenovitch, *Biochimie*, 1988, **70**, 1633.
9 N. Sharon and H. Lis, *Trends Biochem. Sci.*, 1987, **12**, 488.
10 D. C. Kilpatrick, *Biochim. Biophys. Acta*, 2002, **1572**, 187.
11 J. B. Sumner, *J. Biol. Chem.*, 1919, **37**, 137.
12 J. B. Sumner and S. F. Howell, *J. Bacteriol.*, 1936, **32**, 227.
13 W. M. Watkins and W. T. J. Morgan, *Nature*, 1952, **169**, 825; W. T. J. Morgan and W. M. Watkins, *Br. J. Exp. Pathol.*, 1953, **34**, 94.
14 I. E. Liener, N. Sharon, and I. J. Goldstein, *The Lectins: Properties, Functions and Applications in Biology and Medicine*, Academic Press, Orlando, 1986; N. Sharon and H. Lis, *Lectins*, Chapman and Hall, London, 1989.
15 H. Lis and N. Sharon, *Chem. Rev.*, 1998, **98**, 637.
16 I. J. Goldstein and R. D. Poretz, in *The Lectins: Properties, Functions and Applications in Biology and Medicine*, ed. I. E. Liener, N. Sharon, and I. J. Goldstein, Academic Press, Orlando, 1986, p. 33.
17 R. D. Poretz and I. J. Goldstein, *Biochem. Pharmacol.*, 1971, **20**, 2727.
18 A. Pusztai, *Plant Lectins*, Cambridge University Press, Cambridge, 1991.
19 Y. D. Lobsanov, M. A. Gitt, H. Leffler, S. H. Barondes and J. M. Rini, *J. Biol. Chem.*, 1993, **268**, 27034.
20 N. K. Sauter, J. E. Hanson, G. D. Glick, J. H. Brown, R. L. Crowther, S.-J. Park, J. J. Skehel and D. C. Wiley, *Biochemistry*, 1992, **31**, 9609.
21 K. Drickamer, *J. Biol. Chem.*, 1988, **263**, 9557.
22 I. Ofek and R. J. Doyle, *Bacterial Adhesion to Cells and Tissues*, Chapman and Hall, London, 1994.
23 J. Hirabayashi, *Trends Glycosci. Glycotechnol.*, 1997, **9**, 45.
24 R. Loris, T. Hamelryck, J. Bouckaert and L. Wyns, *Biochim. Biophys. Acta*, 1998, **1383**, 9.
25 H. J. Gabius, H. C. Siebert, S. Andre, J. Jimenez-Barbero and H. Rudiger, *ChemBioChem*, 2004, **5**, 741.
26 Y. Konami, K. Yamamoto, T. Osawa and T. Irimura, *Glycoconjugate J.*, 1995, **12**, 128.
27 M. Inbar and L. Sachs, *Nature*, 1969, **223**, 710.
28 N. Srinivasan, S. D. Rufino, M. B. Pepys, S. Wood and T. L. Blundell, *Chemtracts: Biochem. Mol. Biol.*, 1996, **6**, 149.
29 V. R. Srinivas, G. B. Reddy, N. Ahmad, C. P. Swaminathan, N. Mitra and A. Surolia, *Biochim. Biophys. Acta*, 2001, **1527**, 102.
30 R. Banerjee, S. C. Mande, V. Ganesh, K. Das, V. Dhanaraj, S. K. Mahanta, K. Suguna, A. Surolia and M. Vijayan, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 227; R. Banerjee, K. Das, R. Ravishankar, K. Suguna, A. Surolia and M. Vijayan, *J. Mol. Biol.*, 1996, **259**, 281.
31 D. M. Salunke, M. J. Swamy, M. Islam Khan, S. C. Mande, A. Surolia and M. Vijayan, *J. Biol. Chem.*, 1985, **260**, 13576.
32 W. W. Fish, L. M. Hamlin and R. L. Miller, *Arch. Biochem. Biophys.*, 1978, **190**, 693.
33 M. Decastel, H. De Boeck, Y. Goussault, C. K. De Bruyne, F. G. Loontiens and J.-P. Frénoy, *Arch. Biochem. Biophys.*, 1985, **240**, 811.
34 G. B. Reddy, V. R. Srinivas, N. Ahmad and A. Surolia, *J. Biol. Chem.*, 1999, **274**, 4500.
35 W. I. Weis and K. Drickamer, *Annu. Rev. Biochem.*, 1996, **65**, 441.
36 N. Sharon, *Trends Biochem. Sci.*, 1993, **18**, 221; R. Adar and N. Sharon, *Eur. J. Biochem.*, 1996, **239**, 668.
37 A. Imberty, F. Casset, C. V. Gegg, M. E. Etzler and S. Pèrez, *Glycoconjugate J.*, 1994, **11**, 400.
38 R. Banerjee, K. Das, R. Ravishankar, K. Suguna, A. Surolia and M. Vijayan, *J. Mol. Biol.*, 1996, **259**, 281.
39 P. Adhikari, K. Bachhawat-Sikder, C. J. Thomas, R. Ravishankar, A. A. Jeyaprakash, V. Sharma, M. Vijayan and A. Surolia, *J. Biol. Chem.*, 2001, **276**, 40734.
40 N. M. Young and R. P. Oomen, *J. Mol. Biol.*, 1992, **228**, 924.
41 V. Sharma and A. Surolia, *J. Mol. Biol.*, 1997, **267**, 433.
42 D. M. Salunke, M. Islam Khan, A. Surolia and M. Vijayan, *J. Mol. Biol.*, 1982, **154**, 177; D. M. Salunke, M. Islam Khan, A. Surolia and M. Vijayan, *FEBS Lett.*, 1983, **156**, 127.
43 R. Ravishankar, M. Ravindran, K. Suguna, A. Surolia and M. Vijayan, *Curr. Sci.*, 1997, **72**, 855; R. Ravishankar, M. Ravindran, K. Suguna, A. Surolia and M. Vijayan, *Curr. Sci.*, 1999, **76**, 1393.
44 R. Ravishankar, K. Suguna, A. Surolia and M. Vijayan, *Acta Crystallogr., Sect. D*, 1999, **55**, 1375.
45 J. V. Pratap, G. M. Bradbrook, G. B. Reddy, A. Surolia, J. Raftery, J. R. Helliwell and M. Vijayan, *Acta Crystallogr., Sect. D*, 2001, **57**, 1584.
46 B. G. Davis, *J. Chem. Soc., Perkin Trans. 1*, 1999, 3125.
47 J. M. Rini, *Annu. Rev. Biophys. Biomol. Struct.*, 1995, **24**, 551.
48 E. J. Toone, *Curr. Opin. Struct. Biol.*, 1994, **4**, 719.
49 S. Elgavish and B. Shaanan, *J. Mol. Biol.*, 1998, **277**, 917.

50 E. R. Caffarena, J. R. Grigera and P. M. Bisch, *J. Mol. Graphics*, 2002, **21**, 227.

51 C. V. Gegg, D. D. Roberts, I. H. Segel and M. E. Etzler, *Biochemistry*, 1992, **31**, 6938.

52 M. E. Etzler, *Trends Glycosci. Glycotechnol.*, 1998, **10**, 247.

53 M. E. Etzler, in *Glycoconjugates. Composition, Structure and Functions*, ed. H. J. Allen and E. C. Kisailus, Marcel Dekker Inc., New York, 1992, p. 521.

54 W. J. Peumans and E. J. M. Van Damme, in *Lectins. Biomedical Perspectives*, ed. A. Pusztai and S. Bardocz, Taylor & Francis, London, 1995, p. 1; J. W. Kijne, *Chemtracts Biochem. Mol. Biol.*, 1996, **6**, 180.

55 G. W. G. Bird, *Vox Sang.*, 1959, **4**, 307; W. C. Boyd, *Vox Sang.*, 1963, **8**, 1.

56 R. A. Dixon and N. L. Paiva, *Plant Cell*, 1995, **7**, 1085.

57 F. Cote and M. G. Hahn, *Plant Mol. Biol.*, 1994, **26**, 1379.

58 M. E. Etzler, in *The Lectins: Properties, Functions and Applications in Biology and Medicine*, ed. I. E. Liener, N. Sharon, and I. J. Goldstein, Academic Press, Orlando, 1986, p. 371.

59 P. Mylona, D. Pawlowski and T. Bisseling, *Plant Cell*, 1995, **7**, 869.

60 Y. C. Lee and R. T. Lee, *Acc. Chem. Res.*, 1995, **28**, 321.

61 B. G. Davis, *Chem. Rev.*, 2002, **102**, 579.

62 M. N. Janakiraman, C. L. White, W. G. Laver, G. M. Air and M. Luo, *Biochemistry*, 1994, **33**, 8172.

63 M. Mammen, S.-K. Choi and G. M. Whitesides, *Angew. Chem., Int. Ed. Engl.*, 1998, **37**, 2754.

64 M. Yehuda and M. Moshe, *Glycoconjugate J.*, 2002, **19**, 1.

65 C. F. Brewer, M. C. Miceli and L. G. Baum, *Curr. Opin. Struct. Biol.*, 2002, **12**, 616.

66 Z. K. Indik, J. G. Park, S. Hunter and A. D. Schreiber, *Blood*, 1995, **86**, 4389.

67 M. C. Chervenak and E. J. Toone, *J. Am. Chem. Soc.*, 1994, **116**, 10533.

68 B. A. Williams, M. C. Chervenak and E. J. Toone, *J. Biol. Chem.*, 1992, **267**, 22907.

69 M. Kaul and D. S. Pilch, *Biochemistry*, 2002, **41**, 7695.

70 K. A. Watson, E. P. Mitchell, L. N. Johnson, J. C. Son, C. J. F. Bichard, M. G. Orchard, G. W. J. Fleet, N. G. Oikonomakos, D. D. Leonidas, M. Kontou and A. Papageorgiou, *Biochemistry*, 1994, **33**, 5745.

71 N. Muller, *Acc. Chem. Res.*, 1990, **23**, 23.

72 H. Beierbeck, L. T. J. Delbaere, M. Vandonselaar and R. U. Lemieux, *Can. J. Chem.*, 1994, **72**, 463.

73 J. E. Leffler and E. Grunwald, *Rates and Equilibria of Organic Reactions*, John Wiley & Sons, New York, 1963.

74 F. P. Schwarz, K. D. Puri, R. G. Bhat and A. Surolia, *J. Biol. Chem.*, 1993, **268**, 7668.

75 G. I. Makhatadze and P. L. Privalov, *J. Mol. Biol.*, 1993, **232**, 639; P. L. Privalov and G. I. Makhatadze, *J. Mol. Biol.*, 1993, **232**, 660.

76 M. C. Chervenak and E. J. Toone, *Biochemistry*, 1995, **34**, 5685.

77 W. Kauzmann, *Adv. Protein Chem.*, 1959, **14**, 1.

78 K. P. Murphy, D. Xie, K. S. Thompson, L. M. Amzel and E. Freire, *Proteins: Struct., Funct., Genet.*, 1993, **18**, 63.

79 J. M. Sturtevant, *Proc. Natl. Acad. Sci. USA*, 1977, **74**, 2236.

80 A. Surolia, N. Sharon and F. P. Schwarz, *J. Biol. Chem.*, 1996, **271**, 17697.

81 R. F. Kelley and M. P. O'Connell, *Biochemistry*, 1993, **32**, 6828.

82 D. A. Brummell, V. P. Sharma, N. N. Anand, D. Bilous, G. Dubuc, J. Michniewicz, C. R. MacKenzie, J. Sadowska, B. W. Sigurskjold, B. Sinnott, N. M. Young, D. R. Bundle and S. A. Narang, *Biochemistry*, 1993, **32**, 1180.

83 C. F. Brewer, L. Bhattacharyya, R. D. I. Brown and S. H. Koenig, *Biochem. Biophys. Res. Commun.*, 1985, **127**, 1066.

84 T. K. Dam, B. S. Cavada, T. B. Grangeiro, C. F. Santos, V. M. Ceccato, F. A. M. de Susa, S. Oscarson and C. F. Brewer, *J. Biol. Chem.*, 2000, **275**, 16119.

85 H. C. Siebert, R. Adar, R. Arango, M. Burchert, H. Kaltner, G. Kayser, E. Tajkhorshid, C. W. von der Lieth, R. Kaptein, N. Sharon, J. F. Vliegenthart and H.-J. Gabius, *Eur. J. Biochem.*, 1997, **249**, 27.

86 D. Gupta, M. Cho, M. D. Cummings and C. F. Brewer, *Biochemistry*, 1996, **35**, 15236.

87 D. Gupta, T. K. Dam, S. Oscarson and C. F. Brewer, *J. Biol. Chem.*, 1997, **272**, 6388.

88 F. P. Schwarz, H. Ahmed, M. A. Bianchet, L. M. Amzel and G. R. Vasta, *Biochemistry*, 1998, **37**, 5867.

89 N. Navarre, N. Amiot, A. van Oijen, A. Imberty, A. Poveda, J. Jimenez-Barbero, A. Cooper, M. A. Nutley and G.-J. Boons, *Chem. Eur. J.*, 1999, **5**, 2281.

90 Y. C. Lee, in *Neoglycoconjugates: Preparation and Applications*, ed. Y. C. Lee and R. T. Lee, Academic Press, London, 1994, p. 3.

91 J. J. Lundquist and E. J. Toone, *Chem. Rev.*, 2002, **102**, 555.

92 W. P. Jencks, *Proc. Natl. Acad. Sci. USA*, 1981, **78**, 4046.

93 J. J. Lundquist, S. D. Debenham and E. J. Toone, *J. Org. Chem.*, 2000, **65**, 8245.

94 E. Duverger, N. Frison, A. C. Roche and M. Monsigny, *Biochimie*, 2003, **85**, 167.

95 R. E. McDonald, D. Hughes and B. G. Davis, *Angew. Chem., Int. Ed.*, 2004, **43**, 3025.

96 S. M. Dimick, S. C. Powell, S. A. McMahon, D. N. Moothoo, J. H. Naismith and E. J. Toone, *J. Am. Chem. Soc.*, 1999, **121**, 10286.

97 J. B. Corbell, J. J. Lundquist and E. J. Toone, *Tetrahedron: Asymmetry*, 2000, **11**, 95.

98 T. K. Dam, R. Roy, S. K. Das, S. Oscarson and C. F. Brewer, *J. Biol. Chem.*, 2000, **275**, 14223.

99 T. K. Dam, R. Roy, D. Pagé and C. F. Brewer, *Biochemistry*, 2002, **41**, 1351.

100 T. K. Dam, R. Roy, D. Pagé and C. F. Brewer, *Biochemistry*, 2002, **41**, 1359.

101 T. K. Dam and C. F. Brewer, *Chem. Rev.*, 2002, **102**, 387.

102 F. Postollec, W. Norde, H. C. van der Mei and H. J. Busscher, *J. Microbiol. Methods*, 2003, **55**, 241.

103 Y. C. Lee, R. R. Townsend, M. R. Hardy, J. Lönngren, J. Arnarp, M. Haraldsson and H. Lönn, *J. Biol. Chem.*, 1983, **258**, 199.

104 J. Rao, J. Lahiri, L. Isaacs, R. M. Weis and G. M. Whitesides, *Science*, 1998, **280**, 708.

105 E. Fan, Z. Zhang, W. E. Minke, Z. Hou, C. L. M. J. Verlinde and W. G. J. Hol, *J. Am. Chem. Soc.*, 2000, **122**, 2663.

106 R. H. Kramer and J. W. Karpen, *Nature*, 1998, **395**, 710.

107 P. I. Kitov, J. M. Sadowska, G. Mulvey, G. D. Armstrong, H. Ling, N. S. Pannus, R. J. Read and D. R. Bundle, *Nature*, 2000, **403**, 669.

108 H. Ling, A. Boodhoo, B. Hazes, M. D. Cummings, G. D. Armstrong, J. L. Brunton and R. J. Read, *Biochemistry*, 1998, **37**, 1777.

109 A. M. Soltyk, C. R. MacKenzie, V. M. Wolski, T. Hirama, P. I. Kitov, D. R. Bundle and J. L. Brunton, *J. Biol. Chem.*, 2002, **277**, 5351.

110 M. Kanai, K. H. Mortell and L. L. Kiessling, *J. Am. Chem. Soc.*, 1997, **119**, 9931.

111 H. S. Park, Q. Lin and A. D. Hamilton, *J. Am. Chem. Soc.*, 1999, **121**, 8.

112 J. W. Dennis, S. Laferte, C. Waghorne, M. L. Breitman and R. S. Kerbel, *Science*, 1987, **236**, 582; J. Roth, *Chem. Rev.*, 2002, **102**, 285.

113 A. Krzeslak, L. Pomorski, Z. Gaj and A. Lipinska, *Cancer Lett.*, 2003, **196**, 101.

114 L. L. So and J. Goldstein, *J. Biol. Chem.*, 1967, **242**, 1617.

115 P. S. Basu, R. Majhi and S. K. Batabyal, *Clin. Biochem.*, 2003, **36**, 373.

116 B. Gastman, K. Wang, J. Han, Z. Y. Zhu, X. J. Huang, G. Q. Wang, H. Rabinowich and E. Gorelik, *Biochem. Biophys. Res. Commun.*, 2004, **316**, 263.

117 H. Wang and T. B. Ng, *Protein Expression Purif.*, 2003, **29**, 156.

118 A. L. Ambrosio, L. Sanz, E. I. Sánchez, C. Wolfenstein-Todel and J. J. Calvete, *Arch. Biochem. Biophys.*, 2003, **415**, 245.

119 R. Loris, D. Tielker, K.-E. Jaeger and L. Wyns, *J. Mol. Biol.*, 2003, **331**, 861.

120 A. K. Shrive, H. A. Tharia, P. Strong, U. Kishore, I. Burns, P. J. Rizkallah, K. B. M. Reid and T. J. Greenhough, *J. Mol. Biol.*, 2003, **331**, 509.

121 N. M. N. Alencar, A. M. S. Assreuy, V. B. M. Alencar, S. C. Melo, M. V. Ramos, B. S. Cavada, F. Q. Cunha and R. A. Ribeiro, *Int. J. Biochem. Cell Biol.*, 2003, **35**, 1674.

122 A. G. Morell, R. A. Irvine, I. Sternlieb, I. H. Scheinbe and G. Ashwell, *J. Biol. Chem.*, 1968, **243**, 155.

123 M. Meier, M. D. Bider, V. N. Malashkevich, M. Spiess and P. Burkhard, *J. Mol. Biol.*, 2000, **300**, 857.

124 E. M. Anders, C. A. Hartley and D. C. Jackson, *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 4485; A. J. Tenner, S. L. Robinson and R. A. B. Ezekowitz, *Immunity*, 1995, **3**, 485.

125 W. I. Weis, *Structure*, 1994, **2**, 147.

126 M. Ohta, M. Okada, I. Yamashina and T. Kawasaki, *J. Biol. Chem.*, 1990, **265**, 1980.

127 T. Feizi and M. Larkin, *Glycobiology*, 1990, **1**, 17.

128 M. Sumiya, M. Super, P. Tabona, R. J. Levinsky, T. Arai, M. W. Turner and J. A. Summerfield, *Lancet*, 1991, **337**, 1569.

129 P. Garred, F. Larsen, H. O. Madsen and C. Koch, *Mol. Immunol.*, 2003, **40**, 73.

130 A. Tsutsumi, H. Ikegami, R. Takahashi, H. Murata, D. Goto, I. Matsumoto, T. Fujisawa and T. Sumida, *Hum. Immunol.*, 2003, **64**, 621.

131 T. Angata and E. C. M. Brinkman-Van der Linden, *Biochim. Biophys. Acta*, 2002, **1572**, 294.

132 P. R. Crocker, *Curr. Opin. Struct. Biol.*, 2002, **12**, 609.

133 R. P. McEver, *Curr. Opin. Cell Biol.*, 2002, **14**, 581.
134 X. H. Zhang, D. F. Bogorin and V. T. Moy, *ChemPhysChem*, 2004, **5**, 175.
135 L. A. Lasky, *Annu. Rev. Biochem.*, 1995, **64**, 113.
136 P. M. Sass, *Cancer Invest.*, 1998, **16**, 322; L. Borsig, R. Wong, J. Feramisco, D. R. Nadeau, N. M. Varki and A. Varki, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 3352; I. T. Makagiansar, H. Yusuf-Makagiansar, A. Calcagno, A. Ikesue and T. J. Siahaan, *Curr. Top. Pept. Protein Res.*, 2001, **4**, 69.
137 P. Nangia-Makker, J. Conklin, V. Hogan and A. Raz, *Trends Mol. Med.*, 2002, **8**, 187.
138 C. A. Gray, D. L. Adelson, F. W. Bazer, R. C. Burghardt, E. N. T. Meeusen and T. E. Spencer, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 7982.

139 Y. Takenaka, T. Fukumori and A. Raz, *Glycoconjugate J.*, 2002, **19**, 543.
140 D. Platt and A. Raz, *J. Nat. Cancer Inst.*, 1992, **84**, 438.
141 P. Sorme, Y. Qian, P.-G. Nyholm, H. Leffler and U. J. Nilsson, *ChemBioChem*, 2002, **3**, 183.
142 G. Molema and D. K. F. Meijer, *Adv. Drug Delivery Rev.*, 1994, **14**, 25.
143 N. M. Dahms and M. K. Hancock, *Biochim. Biophys. Acta*, 2002, **1572**, 317.
144 R. T. Lee and Y. C. Lee, *Glycobiology*, 2003, **13**, 11.
145 F. P. Schwarz, K. Puri and A. Surolia, *J. Biol. Chem.*, 1991, **266**, 24344.
146 D. K. Mandal, L. Bhattacharyya, S. H. Koenig, R. D. I. Brown, S. Oscarson and C. F. Brewer, *Biochemistry*, 1994, **33**, 1157.