Genomic analysis of C-type lectins

Kurt Drickamer1 and Andrew J. Fadden

Glycobiology Institute, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Abstract

Many biological effects of complex carbohydrates are mediated by lectins that contain discrete carbohydrate-recognition domains. At least seven structurally distinct families of carbohydrate-recognition domains are found in lectins that are involved in intracellular trafficking, cell adhesion, cell–cell signalling, glycoprotein turnover and innate immunity. Genome-wide analysis of potential carbohydrate-binding domains is now possible. Two classes of intracellular lectins involved in glycoprotein trafficking are present in yeast, model invertebrates and vertebrates, and two other classes are present in vertebrates only. At the cell surface, calcium-dependent (C-type) lectins and galectins are found in model invertebrates and vertebrates, and two other classes are present in vertebrates only. At the cell surface, calcium-dependent (C-type) lectins and galectins are found in model invertebrates and vertebrates, but not in yeast; immunoglobulin superfamily (I-type) lectins are only found in vertebrates. The evolutionary appearance of different classes of sugar-binding protein modules parallels a development towards more complex oligosaccharides that provide increased opportunities for specific recognition phenomena. An overall picture of the lectins present in humans can now be proposed. Based on our knowledge of the structures of several of the C-type carbohydrate-recognition domains, it is possible to suggest ligand-binding activity that may be associated with novel C-type lectin-like domains identified in a systematic screen of the human genome. Further analysis of the sequences of proteins containing these domains can be used as a basis for proposing potential biological functions.

Animal lectin structure and function

Complex glycans on cell surfaces, in the extracellular matrix and on soluble, secreted glycoproteins can carry information that must be decoded by animal lectins or sugar-binding proteins [1]. Recognition events between glycans and lectins mediate glycoprotein trafficking and turnover, as well as cell adhesion

1To whom correspondence should be addressed (e-mail kd@glycob.ox.ac.uk).
and cell–cell communication. Roles for animal lectins in cell–cell interactions in the immune system and in pathogen recognition in the innate immune response have been extensively investigated [2]. Similarly, the importance of sugars in the maturation and delivery of proteins in the luminal compartments of eukaryotic cells has been well established (see Chapter 6 in this volume) [3]. The dramatic effects on development caused by modifying or eliminating various types of glycosylation suggest that similar recognition events take place at various stages in development, although identification of the relevant lectins is not so far advanced [4]. Although it is tempting to speculate on the importance of glycan diversity in brain development, it has proven particularly difficult to identify receptors that might utilize this diversity in the process of brain development. Thus, although good examples of animal lectin function are available, the overall picture of what processes they may mediate remains unclear. The availability of the complete human genome sequence has created the opportunity to gain a broad perspective on carbohydrate-binding proteins.

A key finding in the animal lectin field has been the identification of discrete, modular domains that mediate carbohydrate recognition [5]. These domains fall into a number of different structural categories; the members of the largest and most complex group are designated C-type lectins because they require Ca\(^{2+}\) in order to bind glycan ligands. The carbohydrate-recognition domains (CRDs) within these proteins share a common overall fold, which was first described in the structure of an N-terminal fragment of rat serum mannose-binding protein [6]. As illustrated in Figure 1, the sugar binding takes place at a conserved Ca\(^{2+}\)-binding site and requires that key hydroxyl groups on the sugar ligands from a set of hydrogen bonds with acidic and amide amino-acid side chains on the protein. These amino-acid side chains and the sugar hydroxyl groups also form co-ordination bonds with the Ca\(^{2+}\). The C-type CRDs represent a subset of a larger family of protein modules that are denoted C-type lectin-like domains (CTLDs) [5]. Many of the CTLDs bind proteins rather than sugar ligands, and this binding is often Ca\(^{2+}\)-independent (Figure 2).

**Genomic screening for CTLDs**

The conceptual approach to the identification of possible animal lectins is summarized in Figure 3. The first step is to identify protein modules that have structural features of C-type CRDs. Proteins containing CTLDs are examined further in two general ways. In the first, the CTLDs are separated from the remainder of the protein sequence and aligned with known CRDs. The presence of residues that form sugar-binding sites in *bona fide* CRDs is used to make predictions about possible sugar-binding activity. In the second, the sequences of the entire proteins are analysed to see the context of the putative CRD and to establish likely locations, usually in the plasma membrane, the extracellular matrix or as soluble secreted proteins. Other modules in the protein may also suggest possible biological functions of the molecule.

Some of the steps required to implement this general strategy are indicated in more detail in Figure 4. Our knowledge of CTLDs can be used to devise a systematic approach to the identification of genes encoding CTLDs.
Figure 1 Ligand-binding site of the CRD of the prototypical C-type lectin, mannose-binding protein. Spheres represent carbon (white), oxygen (light grey), nitrogen (black) and Ca\(^{2+}\) (dark grey). Hydrogen bonds are shown as broken arrows and co-ordination bonds as solid lines.

Figure 2 Relationship between CRDs and the larger family of CTLDs.
The original motif used to spot domains of this type was based on a set of invariant and highly conserved residues that were characteristic of the members of the family that were known at the time. This approach has now been supplanted by more sophisticated computational methods involving domain profiles. Such profiles are generated from multiple sequence alignments of all known members of the domain family. Once the sequences are aligned optimally, the likelihood of each possible amino acid appearing at each position is tabulated to create the profile; for example, an invariant disulphide bond would be indicated by two positions at which the probability of cysteine residues occurring is 1.0 and the probability of each of the other 19 amino acids occurring is 0.

Predicted protein sequences are aligned with the profile to achieve the optimal match to the established frequencies of amino acids at each position. Based on alignments with randomized sequences, the statistical significance of the matching scores can be evaluated. The process can be iterated, with new sequences added to the profile, which is then refined to reflect the changed frequencies. Profiles that identify CTLDs have been established in several databases, including the PROSITE database set up at the Swiss Institute of Bioinformatics [7], the Pfam (protein families) database maintained at the Sanger Institute [8], and the InterPro and SMART databases that are run by the European Bioinformatics Institute [9,10].

The *Caenorhabditis elegans* and *Drosophila melanogaster* sequences have been completed to a very high degree of reliability, and annotation of possible genes is, for the most part, accurate [11,12]. The presence of introns is the major
source of potential confusion since incorrectly predicted splicing events can lead to severe corruption of deduced protein sequences. The human genome sequence has yet to be established at the same degree of reliability. Thus, annotation in the Ensembl database at the European Bioinformatics Institute and the Sanger Institute is changing continuously as sequences are edited and corrected. Analysis of this sequence-in-progress has been supplemented by screening the database of cDNA sequences generated from various tissue libraries. No new proteins containing potential CTLDs came to light in the 3 months prior to the writing of this chapter, which suggests that coverage of the genome is nearly complete. Thus, it appears that there is a total of approximately 63 proteins containing CTLDs in the human genome.

Although the profiling methodology is quite sophisticated, sequences can be incorrectly identified as containing potential CTLDs. Examination of all of the sequences in the SwissProt database that are identified as matching one of the CTLD profiles suggests that 5–10% of the sequences that receive this annotation do not contain CTLDs. The absence of CTLDs can be demonstrated by repeating the profile scans, which in these cases fail to show even a statistically dubious match to the profiles. The reason for this discrepancy is not clear. It does seem that the mismatched sequences are often cysteine-rich and, in several
cases, make good matches to other domain profiles. Alignments with these sequences confirm that they do not show the expected patterns of core hydrophobic residues that are required to form the CTLD fold. It is also possible that the profile screen can miss potential CTLDs; however, in terms of identifying potential sugar-binding proteins this is less of a problem. Highly divergent sequences are unlikely to form binding sites that are analogous to those in the known C-type lectins.

Identification of potential CRDs

Using our knowledge of how C-type CRDs bind to Ca\(^{2+}\) and glycans, it is possible to examine novel human CTLDs to see which ones are likely to bind these ligands (Figure 4). There are five amino-acid side chains that contribute ligands for the primary Ca\(^{2+}\)-binding site in most of the known C-type CRDs. These side chains (aspartate, asparagine, glutamate and glutamine) are polar and have the potential to form multiple hydrogen bonds. In a few cases, one of these ligands is missing and serine, threonine or a water molecule serves as a Ca\(^{2+}\) ligand. Predictions about the presence of this Ca\(^{2+}\) site can thus be made with some confidence.

Because of the way that sugar ligands interact with the conserved Ca\(^{2+}\) in the C-type CRDs, only those CTLDs that contain this site are considered as likely to display sugar-binding activity. In crystal structures of C-type CRDs complexed with saccharide ligands, the hydrogen-bonding scheme requires the presence of acid and amide side chains that form co-operative hydrogen bonds with the 3- and 4-hydroxyl groups of the sugar ligand, although the geometrical configurations of the bound sugars can differ. However, given that some of the known lectins have serine or threonine residues in at least some of these positions, it seems prudent to suggest that any CTLD with a probable conserved Ca\(^{2+}\)-binding site must be considered a potential sugar-binding protein. In cases in which the pattern of liganding side chains exactly matches a pattern associated with binding of a specific class of saccharide ligand, a reasonably strong prediction about probable selectivity of binding can be made.

The full selectivity and affinity of glycan binding by the C-type CRDs is achieved by extension and combination of the primary monosaccharide-binding sites [13–15]. Although a number of types of extended binding sites are known, it is difficult to make firm predictions from the relatively limited database. Similarly, oligomerization can often be predicted from the nature of accessory domains adjacent to the CTLDs, but the implications of such oligomers for ligand-binding activity are difficult to discern; for example, helical coiled-coil domains cause the association of three CRDs in both mannose-binding protein and the asialoglycoprotein receptor, yet the former binds clusters of widely spaced monosaccharides on cell surfaces and the latter binds closely spaced terminal sugars in a single N-linked glycan [16,17].
Groups of human proteins containing CTLDs

CTLD-containing proteins can be classified by defining the overall domain organization of the proteins in which they are found. The profile-screening approach is used to identify what other types of domains are found in addition to the CTLDs in these proteins. The proteins can also be analysed using simpler sequence motifs to identify hydrophobic signal sequences and transmembrane domains, as well as repeated patterns that are characteristic of the coiled coils of α-helices and collagenous domains.

When these approaches are applied to the CTLD-containing proteins, a relatively small number of different protein organizations are revealed. The 14 groups of CTLD-containing proteins defined in this way are summarized in Figure 5. The genomic screen has revealed almost no novel domain organizations, as all the groups except XIII and XIV were known previously. The numbers of proteins that fall into each group are indicated in Table 1.

![Figure 5 Domain organization of proteins that contain CTLDs.](image)

CCP, complement control protein; CUB, complement components C1r/C1s, Uegf and bone morphogenetic protein-1; EGF, epidermal growth factor; Fn, fibronectin; LDL-R, low-density lipoprotein receptor; LLR, leucine-rich repeat; SS, disulphide bond; VWF-C, von Willebrand factor.
Evolution of human CTLD sequences

One approach for the analysis of CTLD evolution is to analyze the degree of sequence similarity between CTLDs. In order to make good comparisons, it is necessary to identify the CTLD-coding sequences precisely. The end points of CRDs have been well defined using protease-resistance and expression studies [18]. The loop-out topology of the C-type CRDs brings the N- and C-terminal ends of the domain close to each other as a pair of β-strands that end in register at the domain boundaries [6]. The limits of the CRD also correspond to exon boundaries in many of the genes that encode these proteins [19]. Thus, it is possible to excise the regions of protein sequences that correspond to CTLDs and then to use multiple sequence alignment and cluster analysis to make comparisons that reflect only the evolution of the CTLDs and not their assortment with other protein modules. Dendrograms describing the relative similarities of the different domains reveal that many of the CTLDs fall into well-defined groups of relatively closely related domains. Remarkably, CTLDs in each of these groups always derive from proteins in the same structural groups defined by the modular organization of the parent proteins [19]. Thus, for the most part, it can be concluded that there was a single precursor for members of each group. A few CTLDs are too divergent to allow a robust demonstration that they are particularly closely related to any other members of the family. Most of the proteins containing such divergent CTLDs have unique domain organization.

Of the 63 proteins that contain CTLDs, approximately half are likely to function as lectins. The sugar-binding domains can be classified into two broad groups, based on the relative orientations of the 3- and 4-hydroxyl groups in the monosaccharides that interact with the primary ligand-binding site. Approximately two-thirds of the lectins are expected to bind mannose (Man) and N-acetylglucosamine, while the others bind galactose (Gal) and N-acetyl-galactosamine; these two broad-specificity classes are referred to as Man-type

---

**Table 1 Summary of CTLDs in the human genome**

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of members</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (proteoglycans)</td>
<td>4</td>
</tr>
<tr>
<td>II (type 2 receptors)</td>
<td>11</td>
</tr>
<tr>
<td>III (collectins)</td>
<td>6</td>
</tr>
<tr>
<td>IV (selectins)</td>
<td>3</td>
</tr>
<tr>
<td>V (natural killer cell receptors)</td>
<td>19</td>
</tr>
<tr>
<td>VI (mannose receptor family)</td>
<td>4</td>
</tr>
<tr>
<td>VII (free CTLDs)</td>
<td>5</td>
</tr>
<tr>
<td>VIII (layilin)</td>
<td>2</td>
</tr>
<tr>
<td>IX (tetranectin)</td>
<td>3</td>
</tr>
<tr>
<td>X (polycystin)</td>
<td>1</td>
</tr>
<tr>
<td>XI (attractin)</td>
<td>2</td>
</tr>
<tr>
<td>XII (myelin basic protein)</td>
<td>3</td>
</tr>
<tr>
<td>XIII (IDD)</td>
<td>1</td>
</tr>
<tr>
<td>XIV (endosialin)</td>
<td>1</td>
</tr>
</tbody>
</table>

---

K. Drickamer and A.J. Fadden

Table 1 Summary of CTLDs in the human genome

Copyright 2002 Biochemical Society
and Gal-type respectively. In some cases, all members of a structural group show similar ligand-binding characteristics; for example, all of the collectins (Group III) bind Man-type sugars. In several cases, however, some members of a group bind sugars while other members of the group do not. Of particular interest are the members of Group II, which are clearly closely related to each other in terms of sequence, although some bind Man-type sugars while others bind Gal-type sugars.

Comparisons with model organisms

Because many developmental processes are much easier to analyse in relatively simple model organisms, the study of sugar-recognition molecules in these organisms could contribute significantly to our understanding of how glycans serve as recognition markers during development. For this reason, extensive studies of the CTLDs in *C. elegans* and *Drosophila* have been undertaken [20,21]. Unfortunately, the results suggest that CTLDs in these invertebrates have evolved along rather different lines compared with the mammals.

The differences between the CTLDs in different parts of the animal kingdom can be seen in two ways. Sequence comparisons between the CTLDs reveal that all of the *Drosophila* CTLDs are more related to each other than they are to any of the *C. elegans* or human CTLDs. Similarly, the *C. elegans* CTLDs form an evolutionary cluster that is separate from those in the other organisms. Comparison of the domain architecture of proteins containing CTLDs reveals that the only type of domain organization common to both mammals and either of the invertebrates is that seen in Group VII: CTLDs in isolation with no accessory domains. The organization of the CTLD-containing proteins in *C. elegans* is completely different to that in *Drosophila*. Therefore, it is not possible to identify clear orthologues of any of the human CTLD-containing proteins in the model organisms.

These results indicate that CTLDs have radiated independently in the lineages leading to humans, fruit flies and nematode worms. It is therefore interesting that there are some CTLDs in each of the invertebrate model organisms that are known lectins or are predicted from sequence analysis to have sugar-binding sites similar to those in human C-type CRDs. This finding indicates that sugar-binding activity must have arisen independently in each of the lineages of CTLDs, as indicated in Figure 6. This ‘convergent’ evolution of a sugar-binding site in the CTLD framework on multiple occasions probably reflects its relative simplicity, since only a few critical amino acids need to be present to form a monosaccharide-binding site. Another striking feature of the CTLDs in *C. elegans* and *Drosophila* is the relatively large number of these domains relative to genome size (Table 2). Moreover, the number of CTLDs that is likely to bind sugar is not correlated with the overall developmental complexity of the organisms, as the number of CTLDs predicted to bind sugars is substantially larger in *C. elegans* than in *Drosophila*.

Predictions about sugar-binding activity are, necessarily, rather tentative, so the likely presence of the conserved Ca$^{2+}$-binding site warrants investigation of the possible sugar-binding activity. The small pool of potential sugar-bind-
ing CTLDs in Drosophila makes it an attractive target for a functional genomics pilot project, in which the CTLDs are systematically expressed and their sugar-binding activity is analysed. A strategy for this type of study is outlined in Figure 7. cDNAs libraries from various developmental stages can be used as templates for the amplification of the CTLD-encoding regions, which can then be expressed in a bacterial system in which most CTLDs have been shown to fold directly after secretion into the periplasm. The activity of the proteins expressed in this way can be examined initially by affinity chromatography on immobilized sugars, which also serves as a convenient way to purify correctly folded proteins that can be used in further tests of ligand-binding activity. Such studies will ultimately be undertaken on the set of human CTLDs as well.

**Table 2** Comparison of lectins in various genomes.

R-type, ricin-type lectin, I-type, immunoglobulin superfamily lectin.

<table>
<thead>
<tr>
<th>Genes</th>
<th>C. elegans</th>
<th>Drosophila</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of genes in each organism</td>
<td>19000</td>
<td>13600</td>
<td>( \approx 35000 )</td>
</tr>
<tr>
<td>Calnexin</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>L-type</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>P-type</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>M-type</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>R-type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I</td>
<td>11</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Group II</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Galectins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Group II</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Group III</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Group IV</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C-type</td>
<td>125</td>
<td>32</td>
<td>66</td>
</tr>
<tr>
<td>CTLDs</td>
<td>183</td>
<td>33</td>
<td>96</td>
</tr>
<tr>
<td>CRDs</td>
<td>19</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>I-type</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Total lectin genes</td>
<td>53</td>
<td>29</td>
<td>76</td>
</tr>
</tbody>
</table>
Comparison with other types of animal lectins

Genome-wide studies of C-type lectins should be put in the overall context of genomic analysis of other structural categories of animal lectins [21]. The results of surveys of several genomes are summarized in Table 2. Different patterns of evolution of the intracellular and extracellular lectins are observed. The intracellular lectins involved in trafficking of proteins in between luminal compartments fall into the calnexin, L-type lectin and mannose 6-phosphate receptor families. Each of these families is limited to two or three members in each of the genomes examined.

In contrast with the intracellular lectins, the extracellular galectin and siglec families, like the C-type lectins, have expanded extensively within the vertebrate lineage. Among these families, the immunoglobulin-like siglecs (I-type lectins) are the least divergent in terms of sugar-binding activity, as they all bind glycans that contain sialic acid, and in overall organization, since they all consist of multiple immunoglobulin-type domains (see Chapter 7 in this volume). Siglecs have been identified only in vertebrates. The vertebrate galectins fall into three general structural categories, all of which are also found in *C. elegans*; only one of these groups is represented in *Drosophila*. A fourth group, not found in vertebrates, is found in *C. elegans* and *Drosophila*; for the most part, these proteins bind β-galactosides, although there are more divergent examples that bind other types of sugars including mannose. The radiation of extracellular lectins within the vertebrate lineage suggests that these proteins have diverged to fill novel carbohydrate-recognition roles in vertebrates.

One of the most striking features of the known families of lectins is that sugar-binding activity has evolved in the context of many different protein...
folds. The shallow binding sites and the weak affinities for monosaccharides that characterize these binding sites have apparently evolved on multiple occasions. This observation suggests that there may be additional protein folds that can accommodate sugar-binding activity in human proteins; two examples of such folds have emerged relatively recently.

R-type CRDs are related to the sugar-binding domains of ricin. These CRDs are unusual in that they are found in bacteria as well as in plants and animals. In all cases, they have a targeting function. In ricin, the R-type CRD targets transport of a ribosome-inactivating toxin for transport into the cytoplasm via a retrograde pathway through endosomes, the Golgi apparatus and the endoplasmic reticulum [22]. The R-type CRDs in bacterial proteins direct hydrolytic enzymes to their polysaccharide targets, e.g. cellulose [23]. In humans, R-type CRDs are found in two distinct types of proteins. The larger Group I consists of N-acetylgalactosaminyl transferases that initiate O-linked glycan synthesis [24]. The lectin domains appear to recognize previously glycosylated sites and allow the processive addition of glycans to nearby serine and threonine residues. The other proteins containing modules that resemble R-type CRDs are the members of the mannose receptor family (Group II). In addition to binding mannose and related sugars through C-type CRDs, the mannose receptor binds sulphated N-acetylgalactosamine residues via the R-type CRD [25]. Structural analysis suggests that the homologous domains from other members of the family almost certainly do not have this activity.

More recently, an additional family of intracellular lectins involved in glycoprotein trafficking has been identified. The term M-type lectin has been suggested for these proteins, which resemble mannosidases found in the endoplasmic reticulum, but lack key catalytic residues. These proteins function in the process of endoplasmic-reticulum-associated degradation, in which glycoproteins that have remained incorrectly folded in the endoplasmic reticulum for an extended period are directed to the cytoplasm for proteolytic degradation [26–29]. Like the other intracellular lectin families, the M-type lectin family is modest in size and has not been extensively elaborated in vertebrates compared with invertebrates.

The relatively recent discovery of R- and M-type lectins suggests that families of proteins that bind sugars will continue to be identified. Indeed, other families with such activities, e.g. the interleukins and several types of proteins that bind glycosaminoglycans, have already been described [30,31]. It is clear that sugar-binding activity can arise in the context of many different protein backgrounds, so that the profile-scanning approach would not be useful in detecting such proteins. In addition, sugar-binding proteins are not characterized by simple, contiguous peptide sequence motifs, which reflects the fact that sugar-binding sites arise from the interaction of residues that are close in space, but are not necessarily close in the primary structure of the protein. Thus, for the most part, detection of novel sugar-binding protein families will be an empirical matter.
World-wide web-based resource for animal lectins

The types of bioinformatics and functional genomics discussed in the preceding sections are directed towards an understanding of what genes encode glycan-binding proteins and what their potential glycan ligands may be. As indicated in Figure 3, the domain architecture of proteins containing potential sugar-binding domains may provide clues about their biological functions. However, one of the great advantages of the genomic approach to analysis of sugar-binding proteins is that it will facilitate application of genetic approaches to address issues of function.

An important genetic approach will be analysis of phenotypes associated with mutations in lectin genes. In many cases, identification of such mutations will follow from studies of disease conditions that can be mapped using positional cloning techniques. In order to make use of the information derived from such studies, it is important that the properties of candidate genes will be readily accessible to investigators carrying out mapping studies. For this reason, a database of sugar-binding proteins has been established as a genomic resource for animal lectin studies. The database is still undergoing development, but in its present version it is accessible on the world-wide web (http://ctld.glycob.ox.ac.uk).

There are several aspects to the current database. Part of the information is general background on animal lectin structures and known functions. A second section is devoted to compilation of the proteins containing CTLDs in humans and in model organisms. In its current form, the database is orientated toward structural analysis, but information on function at the level of glycan-binding activities and broader biological functions is also being incorporated.

In addition to providing a link between the work in glycobiology and the world of genomics, the lectin database is intended to assist in the co-ordination of work within the field of carbohydrate recognition. The relatively modest number of lectins that have been identified suggests that, although complex, the universe of protein–carbohydrate interactions is finite. As the identification phase of animal lectin biology comes to a close, the accumulated data on the lectins and their glycan ligands will facilitate the much larger task of understanding their biological functions.

This work was supported by the Wellcome Trust and the Biotechnology and Biological Sciences Research Council.

References